Some Anomalies in the E-layer of the Ionosphere

Thesis submitted by

A.J. Lyon, M.A.,

for the degree of Doctor of Philosophy

University of Edinburgh

September 1956.

Preface.

Some of the work described here has been reported in the four published papers, reprints of which will be found in an envelope attached to the rear inside cover, and a further paper dealing with the statistical work is at present in preparation. Items not previously reported include in the theoretical section the treatment of non-stationary conditions and the discussion of the effects of drift on the shape of a layer, and on the experimental side the special measurements taken at Slough.

The writer is deeply indebted to many colleagues including the joint authors of the above papers and to others, particularly Mr. W.R. Piggott of the Radio Research Station, Slough for much invaluable aid, to the Director of that station for permission to make measurements there, to Mrs. Moorat who did much of the reduction of records, and above all to Principal Sir Edward Appleton, without whose encouragement, inspiration, criticism, instruction and generous assistance in many ways none of this work would have been possible.

CONTENTS.

Index of Figures and Tables

Chapter 1.

Introduction.

The E-layer of the ionosphere, occupying the region from about 100 to 120 km. above the earth's surface, is usually considered to be the simplest, the most stable and the best understood of all the ionospheric layers. In particular, its temporal variations and other characteristics, as revealed by the conventional techniques of radio-sounding, show a remarkable measure of agreement with those predicted many years ago by Chapman (1931), on the assumption that atmospheric ionization is due chiefly to the absorption of solar radiation.

In recent years, however, and particularly since the end of World War II, over seventy ionospheric stations have come into regular operation, and the accuracy and reliability of their measurements have steadily improved. As a consequence of these developments it is now possible, even using routine data alone, to study and assess the behaviour of the E-layer with much greater precision and in much greater detail than ever before.

Such a study has been in progress at Edinburgh for a number of years and has revealed several anomalous features in the behaviour of the E-layer, which show that it is by no means as simple as might at first appear. Briefly the principal anomalies

are: (i) a pre-noon perturbation in the E-layer critical frequency fE and in the minimum equivalent height h'E; (ii) a variation of fE with solar zenith distance χ substantially more rapid than theory predicts; and (iii) the complete absence in most cases, of the increasing asymmetry in the diurnal variation of fE which the known relaxation properties of the layer would lead us to expect.

For the proper study of these anomalies two things are required: first, an accurate knowledge of precisely what variations theory does predict, and of precisely how these are altered by various modifications of simple theory, such as varying temperature or recombination coefficient, or by such perturbing factors as vertical drifts; and second, an adequate supply of accurate values of fE and other E-region parameters.

The first of these problems is largely the mathematical one of finding suitable approximate solutions of the main differential equation, the so-called "continuity equation", which governs the variations of ionization density in the ionosphere. Also involved however is the collation and assessment of any information regarding the properties and processes of the high atmosphere - including rocket data, astrophysical evidence etc. - which may bear on the problem.

For the second requirement - accurate values of fE - the great mass of routine data now available is extremely useful, especially when random variations are smoothed out by appropriate statistical procedures. To supplement this information however two special sequences of intensive E-region measurements were taken at the Radio Research Station, Slough, one in June 1954 and the other in September 1955. A standard ionospheric recorder was used but with a modified technique designed to give more detailed and accurate readings of E-layer characteristics than is normally obtained in routine practice; and records were taken at quarter-hourly, instead of the usual hourly intervals. These measurements have provided useful checks and other additional information on the behaviour of the E-layer.

The anomalies under review will thus be considered along three separate but complementary lines of approach: the theoretical approach aiming to assess their magnitude precisely and to appraise critically any suggested explanations; the statistical approach extending the field of view as widely as possible in space and time (i.e., to cover all seasons, all phases of the solar cycle and all, or most, parts of the globe); and the experimental approach not only seeking more precise information at one place and period but also allowing a critical

examination of the methods by which the ionospheric records are reduced - for it is by no means impossible that the source of some at least of the apparent anomalies lies not in the ionosphere at all, but rather in incorrect analysis of the data.

Chapter 2.

Theoretical Studies.

2.1 Chapman's Theory of Ionospheric Layer-Formation.

The classical theory of ionospheric layer-formation, as given by Chapman (1931), makes, in the first instance, three simplifying assumptions: first, that the ionizing radiation is monochromatic, second, that only one species of molecule is ionized, and third, that the ionosphere may be supposed plane, (i.e. that its sphericity may be neglected). In later papers Chapman (1939, 1931b) discussed the removal of respectively the first and third assumptions, but if for the moment the three assumptions are admitted, it is readily shown that q the number of ions produced per unit volume per second is given by

$$q = n A S_\infty \exp\left\{- n A H \sec\chi\right\} \qquad (1)$$

where S_∞ = the number of photons/cm^2/sec. incident on the outer atmosphere,

χ = the solar zenith distance,

n = the number of molecules of the absorbing gas per unit volume,

A = its absorption cross-section, and

H = its "scale-height" at height h, as defined by the equation

$$\int_\chi^\infty n \, dh = n H \qquad (2).$$

Provided that $\underline{n}$, the density of the absorbing constituent remains proportional to that of the atmosphere as a whole then H as defined by (2) is equal to H_1, the atmospheric scale-height given by

$$H_1 = kT/mg \qquad (3).$$

It is to be noted however that equations (1) and (2) make no assumptions whatever about the variation of $\underline{n}$ with height, which so far may be considered arbitrary.

The condition defining the level of maximum ion-production, is found from (1) to be

$$dn/dh = - n^2 A \sec \chi \qquad (4)$$

or, using (2),

$$n_m = \frac{\cos \chi \ (1 + dH/dh)}{AH} \qquad (5).$$

These equations are again valid for any arbitrary function n(h), but it is now useful to introduce the assumption that the scale-height varies linearly with height according to the equation

$$H = H_0 + \mu h \qquad (6)$$

where μ is a constant. This of course includes the simple case of constant scale-height, i.e. $\mu = 0$, but since recent evidence (e.g. Jackson, 1956) suggests a substantial temperature gradient in the E-region,

equation (6) is probably a more realistic assumption to make, and the linear approximation is likely to be sufficient over the relatively small height-range of E-layer variations.

Introducing a new height parameter z defined by the equation

$$dz = dh/H \qquad (7),$$

we find that

$$nH = n_0 H_0 \, e^{-z} \qquad (8),$$

and if we further choose the level of n_0 so that

$$n_0 \, A \, H_0 = 1 + \mu \qquad (9)$$

then (5) gives the height of maximum ion-production $z(q_m)$ as

$$z(q_m) = \ln \sec \chi \qquad (10),$$

whilst (1) becomes

$$q = q_0 \exp\left\{(1 + \mu)(1 - z - e^{-z} \sec \chi)\right\} \qquad (11)$$

where $$q_0 = \frac{(1 + \mu) \, S_\infty}{H_0 \exp(1 + \mu)} \qquad (12).$$

The parameter z is related to the height h by the equation

$$z = \frac{1}{\mu} \ln(1 + \mu h/H_0) \qquad (13)$$

so that, in general, it measures height on a

logarithmic scale; but for a constant scale-height (or approximately if $\mu h/H_0$ is small) we have the linear relation

$$z = h/H \qquad (13a).$$

As (8) shows, n in that case decreases exponentially with height - the well-known result for an isothermal atmosphere.

If we measure heights from the level of maximum q, i.e. putting

$$z = \ln \sec \chi + y \qquad (14)$$

then (11) becomes

$$q = q_0(\cos \chi)^{1+\mu} \exp\left\{(1+\mu)(1 - y - e^{-y})\right\} \qquad (15)$$

and hence the maximum value of q is

$$q_m = q_0(\cos \chi)^{1+\mu} \qquad (16).$$

From these equations the conclusions of Chapman's theory as applied to a plane-stratified atmosphere with a linear gradient of scale-height may be summarized as follows:-

(i) that the level of maximum ion production for zero χ, determined by (9), depends on H_0, μ and the absorption coefficient A but is independent of the intensity of the radiation;

(ii) that the departure from this level of the height of maximum q for other values of χ depends

only on χ, being given by the simple relation (10);

(iii) that q_0, the maximum q for zero χ, depends on H_0 and the intensity of the radiation S_∞, and that q_m for other values of χ varies according to the simple law expressed by (16); and

(iv) that the ratio q/q_m at a given distance from the q_m level is independent of H_0, S_∞ and χ, or in other words, for given scale-height gradient μ, a single curve will suffice to express the shape of the height distribution curve for all times of day and regardless of variations of the solar radiation.

2.1.1 The Chapman Function for low cos χ.

We must now consider briefly the effects of modifying the three assumptions on which the elementary theory so far presented is based. First if we take into account the spherical stratification of the ionosphere Chapman (1931b) showed that sec χ must be replaced by a new function, $Ch(R,\chi)$, R being the distance of the layer from the centre of the earth, measured in scale-heights. For $\chi \leq 80^\circ$ a useful approximation for this function is

$$Ch(R,\chi) = \sec\chi \left(1 - \frac{\tan^2\chi}{R}\right) \qquad (17)$$

For other values of χ detailed tables of $Ch(R,\chi)$ published by Wilkes (1954) are available.

2.1.2 Absorption over a band of frequencies.

If several separate monochromatic radiations are absorbed, or if radiation is absorbed by different constituents, then since the absorption coefficient will be different in each case so will the height of the corresponding layer-maximum, and in fact a series of separate layers will be produced. These may be quite separate as with the E and F layers; or alternatively there may be merging of several layers in one region.

A special case of this studied by Chapman (1939) is where absorption takes place over a band of frequencies within which there is a continuous variation of the absorption cross-section. He assumes this variation to be of the form

$$A_\lambda = A_0 \exp\left(\frac{\lambda - \lambda_0}{b}\right)^2,$$

and concludes that although the shape of the layer and the height of the maximum will be altered q_m will still be proportional to $\cos\chi$ and the height of the maximum relative to the zero-χ maximum will still vary as $\ln \sec\chi$.

We can easily prove that this is true in the more general case where A_λ is an arbitrary function of λ, and the intensity at wavelength λ over a small range $d\lambda$ is $S_\lambda\, d\lambda$. The total ion production is then

$$q = \int q_\lambda \, d\lambda$$

where

$$q_\lambda = n A_\lambda S_\lambda \exp\left\{- n A_\lambda H \sec\chi\right\} .$$

Thus each range of wavelength $d\lambda$ will produce a small Chapman layer for which equations (11) and (15) will be applicable, except that now we must measure z from an arbitrary fixed level so that z in (11) must be replaced by $z - z_\lambda$ where z_λ is given by

$$z_\lambda = \ln \frac{n_0 A_\lambda H_0}{1 + \mu},$$

the noughts referring to the arbitrary reference-level. Similarly in (15) y must be replaced by $y - z_\lambda$: y then measures height from a point $\ln \sec\chi$ above the zero reference-level.

Hence the height-distribution of electron-production is given by

$$q = (\cos\chi)^{1+\mu} \, F(y) \qquad (18)$$

where

$$F(y) = \int \frac{S_\lambda (1+\mu)}{H_0 \exp(1+\mu)} \exp\left\{(1+\mu)\left[1 - (y - z_\lambda) - e^{-(y - z_\lambda)}\right]\right\} d\lambda \qquad (18a)$$

If $F(y)$ has a maximum F_m at $y = y_m$ then

$$q_m = F_m (\cos \chi)^{1+\mu} \qquad (19)$$

and the height of this maximum level is

$$z(q_m) = y_m + \ln \sec \chi \qquad (20).$$

In these equations F_m and y_m are independent of χ, being determined by the functions S_λ and A_λ only. Hence the $\cos \chi$ law for q_m and the $\ln \sec \chi$ law for the height of maximum are unaffected by absorption covering an arbitrary range of frequencies. It is important to establish this point because one currently favoured theory is that E-layer ionization is due mainly to absorption of X-rays covering a relatively large band of frequencies - perhaps from 50Å to 500Å, according to Rawer and Argence (1954).

One further complication is that the E-layer is formed near, and probably partly within the transition region above which oxygen molecules are mainly monatomic. Such a change in the composition of the atmosphere could affect the mean absorption-coefficient, i.e. could make A_λ a function of height. The effect of this would be to change the effective scale-height which would then have to be defined by the equation

$$\int_h^\infty n A \, dh = n A H', \qquad (21)$$

and in equation (5) μ must then be replaced by $\mu' = dH'/dh$.

2.2 The Continuity Equation and Methods for its Solution.

At any point in space the electron-density must satisfy a continuity equation which in its most general form may be written

$$\frac{\partial N}{\partial t} = q(z,t) - d(N) - \mathrm{div}(Nv), \qquad (22)$$

where d(N) is the rate of disappearance of electrons by way of atomic or molecular reactions, and $\underline{v}$ is the mean velocity of electron transport (possibly including a component of neutral air movement, but excluding thermal motions). If the quantities d(N), $\underline{v}$ and $\underline{q}$ are known functions of $\underline{z}$ and $\underline{t}$ then, in principle at least, equation (22) may be solved for N, and in particular for its diurnal variations. For the simplest case we assume

(i) that $\underline{q}$ is given by (11) with $\mu = 0$, i.e. that it varies in accordance with simple Chapman theory,

(ii) that electron loss occurs by a process at least effectively recombination so that

$$d(N) = \alpha N^2,$$

where the rate coefficient α is assumed constant, and

(iii) that drift is negligible, i.e. that we may put $v = 0$.

The continuity equation then becomes

$$\frac{\partial N}{\partial t} = q_0 \exp(1 - z - e^{-z} \sec \chi) - \alpha N^2 \qquad (23)$$

For any assigned value of z this has the form of a Riccati equation and several methods of solution are available. A numerical method may be necessary in some cases, especially for the periods near sunrise or sunset, but such methods are laborious, and analytical solutions, even if only approximate, are generally preferable.

2.2.1 The quasi-stationary case: simple method.

If the recombination process (whose time-constant is of order $1/\alpha N$) is fast compared to the variation of q, then a quasi-stationary equilibrium tends to be set up with

$$q \approx \alpha N^2 \qquad (24).$$

Under these conditions quite accurate solutions may be obtained by a simple iterative process. Thus from (24), using primes to denote differentiation with respect to t, we have

$$N' = \frac{q'}{2\alpha N} = \frac{q'}{2(\alpha q)^{\frac{1}{2}}},$$

and substituting in (23) yields the next approximation for N. Continuing this process as far as the third-

order approximation, we find

$$N_{(3)} = \left(\frac{q}{\alpha}\right)^{\frac{1}{2}}\left\{1 - \frac{1}{4(q\alpha)^{\frac{1}{2}}}\frac{q'}{q} + \frac{1}{32q\alpha}\frac{4qq'' - 5q'^2}{q^2} - \frac{1}{128(q\alpha)^{3/2}}\frac{8q^2q''' - 36qq'q'' + 30q'^3}{q^3}\right\} \quad (25).$$

It should be noted that the iteration does not yield a convergent series, and should be continued only as far as the successive terms are decreasing. When the terms in q" and higher derivatives are negligible, as is usually the case, the calculation of successive terms is simplified and gives, for example

$$N_{(5)} = \left(\frac{q}{\alpha}\right)^{\frac{1}{2}}\left\{1 - \tfrac{1}{2}\epsilon - \tfrac{5}{8}\epsilon^2 - \tfrac{15}{8}\epsilon^3 - 8.7\,\epsilon^4 - 53\,\epsilon^5\right\} \quad (26)$$

where $$\epsilon = \frac{1}{2(q\alpha)^{\frac{1}{2}}}\frac{q'}{q} = \frac{dN/dt}{\alpha N^2} \quad (27).$$

The quantity $1/2(q\alpha)^{\frac{1}{2}} \approx 1/2\alpha N = \tau$, say, has been called by Appleton (1953) the "relaxation-time" of the layer and occurs frequently in the theory of layer-formation.

Equation (26) may be used to indicate how far the iteration can profitably be carried for any required degree of accuracy. Two or three terms of

the sequence will easily give an accuracy better than 1% in the range $|\epsilon| < 1/10$, which under E-layer conditions covers the greater part of the day.

One disadvantage of this method however is that it does not permit the initial conditions to be taken into account; it is simply assumed that any departure from the quasi-stationary solution (25) has disappeared before the first time for which this solution is used. A somewhat more satisfactory method including the initial conditions has been treated by Rydbeck and Wilhelmsson (1954) and will be considered in detail in the next section.

2.2.2 The quasi-stationery case with given initial conditions: the W.K.B. method.

If we write

$$N = u'/\alpha u, \qquad (28)$$

the continuity equation (23) becomes

$$\frac{d^2u}{dt^2} - \alpha q u = 0, \qquad (29)$$

i.e. it is transformed into a wave-equation, and solutions applicable to the latter are therefore valuable in some cases for solving the continuity equation. In particular the so-called W.K.B. method may be used. Thus writing

$$u = e^{\sqrt{\alpha}\, S}$$

we find from (29) that

$$\frac{1}{\sqrt{\alpha}} S'' + S'^2 = q.$$

This is a similar equation to (23) and the iterative process previously described yields two linearly independent solutions, valid for quasi-stationary conditions, namely,

$$S' = \sqrt{q}\left\{1 - \frac{1}{4(q\alpha)^{\frac{1}{2}}}\frac{q'}{q} + \frac{1}{8q\alpha}\left[\frac{q''}{q} - \frac{5}{4}\left(\frac{q'}{q}\right)^2\right] - \ldots\right\}$$

$$T' = -\sqrt{q}\left\{1 + \frac{1}{4(q\alpha)^{\frac{1}{2}}}\frac{q'}{q} + \frac{1}{8q\alpha}\left[\frac{q''}{q} - \frac{5}{4}\left(\frac{q'}{q}\right)^2\right] + \ldots\right\}$$

Hence the general solution is

$$u = A\, e^{\sqrt{\alpha}S} + B\, e^{\sqrt{\alpha}T}$$

so that

$$N = \frac{1}{\sqrt{\alpha}} \frac{AS'\, e^{\sqrt{\alpha}S} + BT'\, e^{\sqrt{\alpha}T}}{A\, e^{\sqrt{\alpha}S} + B\, e^{\sqrt{\alpha}T}}$$

We shall suppose that the initial conditions are given in the form

$$N = N_1 \quad \text{at} \quad t = t_1.$$

Then

$$N = \frac{S'}{\sqrt{\alpha}} + \frac{\delta\,(S' - T')\, e^{-2W}}{1 - \delta\, e^{-2W}}$$

where $W = \frac{1}{2}\sqrt{\alpha}\,(S - T)$

$$= \int_{t_1}^{t} (q\alpha)^{\frac{1}{2}} \left\{ 1 + \frac{1}{8q\alpha}\left[\frac{q''}{q} - \frac{5}{4}\left(\frac{q'}{q}\right)^2 \right] + \cdots \right\}$$

(30)

and $\delta = -B/A$. Hence

$$N = \left(\frac{q}{\alpha}\right)^{\frac{1}{2}} \left\{ 1 + \frac{2\delta\, e^{-2W}}{1 - \delta e^{-2W}} - \frac{1}{4(q\alpha)^{\frac{1}{2}}} \frac{q'}{q} + \cdots \right\}$$

(31).

At $t = t_1$ $W = 0$ and (31) gives δ in terms of N_1 and q_1. Approximately (if δ is small) we have

$$2\delta = \frac{N_1}{(q_1/\alpha)^{\frac{1}{2}}} - 1 \qquad (32).$$

After an interval Δt

$$N = \sqrt{\frac{q}{\alpha}}\,(1 + 2\delta\, e^{-\Delta t/\tau}) \qquad (33)$$

where, by (30), $1/\tau$ is the mean value of $2(q\alpha)^{\frac{1}{2}}$ over the interval Δt. In other words if at $t = t_1$

N differs from the equilibrium value then it will thereafter approach the equilibrium value in an approximately exponential manner with a time-constant equal to the relaxation-time of the layer.

It should be noted however that for the above solution to be valid it is necessary that during the whole range of time considered, from t_1 to t, quasi-stationary conditions should obtain. It is not therefore valid if the initial conditions are given at or near sunrise.

2.2.3 The non-stationary case: q increasing.

When the rate of change of q is too great to permit the use of a quasi-stationary solution as, for example, during the early morning just after sunrise, the simplest analytical solution is obtained, as shown by Rydbeck and Wilhelmsson (1954), by using a straight-line approximation for q, say

$$q = b(t - t_o) \qquad (34)$$

where b is a constant. The wave-equation (29) then becomes

$$\frac{d^2u}{dt^2} - \alpha b(t - t_o)\, u = 0,$$

a general solution of which is

$$u = A(\alpha b)^{1/6}(t-t_0)^{1/2} I_{1/3}\left\{\tfrac{2}{3}(\alpha b)^{1/2}(t-t_0)^{3/2}\right\} + B(\alpha b)^{1/6}(t-t_0)^{1/2} I_{-1/3}\left\{\tfrac{2}{3}(\alpha b)^{1/2}(t-t_0)^{3/2}\right\} \quad (35)$$

$I_{\frac{1}{3}}$ and $I_{-\frac{1}{3}}$ being modified Bessel functions in the standard notation. Now it is readily shown that

$$\frac{d}{dx}\left[x^{1/2} I_{1/3}\left\{\tfrac{2}{3}x^{3/2}\right\}\right] = x\, I_{-2/3}$$

and

$$\frac{d}{dx}\left[x^{1/2} I_{-1/3}\left\{\tfrac{2}{3}x^{3/2}\right\}\right] = x\, I_{2/3}$$

and hence (28) gives

$$N = \left(\frac{b}{\alpha}\right)^{\frac{1}{2}}(t-t_0)^{\frac{1}{2}} \frac{I_{2/3} + \delta I_{-2/3}}{I_{-1/3} + \delta I_{1/3}} \quad (36)$$

where the argument of the Bessel functions is β given by

$$\beta = \tfrac{2}{3}(\alpha b)^{\frac{1}{2}} |t - t_0|^{\frac{3}{2}} \quad (37).$$

If $N = N_1$ at $t = t_1$

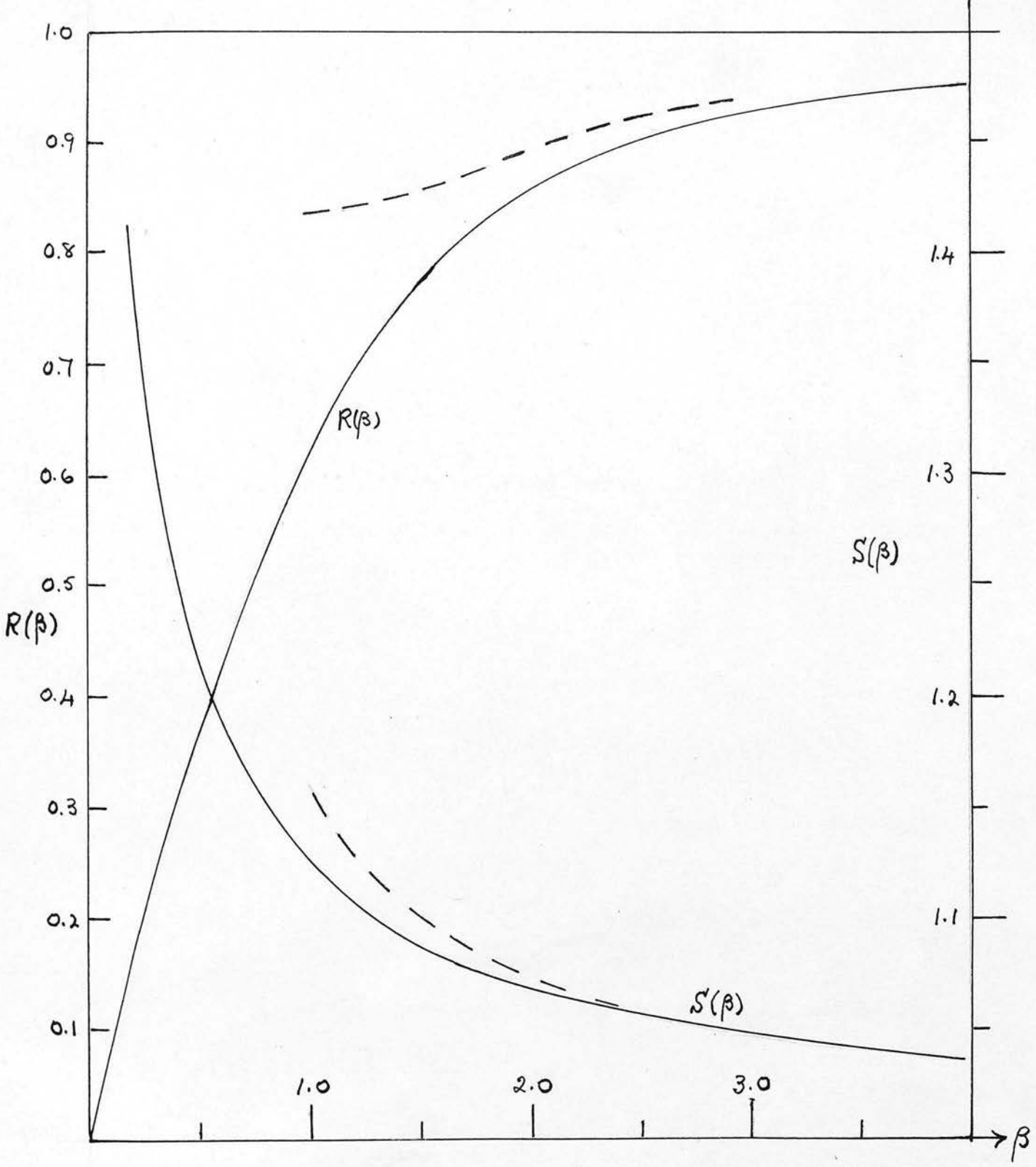


Fig. 1 The functions $R(\beta)$ and $S(\beta)$ compared with the quasi-stationary solutions (dashed lines).

$$\delta = \frac{k\, I_{-1/3}(\beta_1) - I_{1/3}(\beta_1)}{I_{-2/3}(\beta_1) - k\, I_{1/3}(\beta_1)} \qquad (38).$$

Alternatively if $N = N_0$ at $t = t_0$

$$\delta = \frac{N_0}{0.51}\left(\frac{\alpha^2}{3b}\right)^{\frac{1}{3}} \qquad (38a).$$

Finally, if $N = 0$ at $t = t_0$, $\delta = 0$ and

$$N = \left(\frac{b}{\alpha}\right)^{\frac{1}{2}} (t - t_0)^{\frac{1}{2}} \frac{I_{1/3}}{I_{-1/3}} = \sqrt{\frac{q}{\alpha}}\, R(\beta) \qquad (39).$$

The function $R(\beta)$, which gives the departure of $\underline{N}$ from the equilibrium value, is plotted in Fig. 1.

2.2.4 The non-stationary case: q decreasing.

For the case when $\underline{q}$ is decreasing rapidly, as for example near sunset, we may write

$$q = b\,(t_0 - t) \qquad (40)$$

and, $N = N_1 = k\left(\frac{q_1}{\alpha}\right)^{\frac{1}{2}}$ at $t = t_1$

then, as before

$$N = \left(\frac{q}{\alpha}\right)^{\frac{1}{2}} \frac{\delta I_{-2/3} - I_{2/3}}{I_{-1/3} - \delta I_{1/3}} \qquad (41)$$

and

$$= \frac{\kappa I_{-1/3} + I_{2/3}}{\kappa I_{1/3} + I_{-2/3}} \qquad (41a)$$

the argument being β in the first case and β_1 in the second.

For $\beta_1 > 1$ (which will certainly be true if t_1 lies within the quasi-stationary period) δ differs negligibly from unity. In that case we shall write

$$N = \left(\frac{q}{\alpha}\right)^{\frac{1}{2}} S(\beta) \qquad (42).$$

The function $S(\beta)$ is also shown in Fig. 1.

At $t = t_0$ we have

$$N(t_0) = 0.51(3b/\alpha^2)^{\frac{1}{3}} \qquad (43).$$

2.2.5 The Relation between the quasi-stationary and non-stationary solutions.

We have now obtained two main solutions of the continuity equation, one valid for quasi-stationary conditions, that is, for dq/dt small, and the other valid for linear variations of q and therefore particularly useful when dq/dt is large; in other words one solution appropriate to the middle part of the day and the other to the morning and evening. Naturally to

obtain a complete solution valid for the whole day we must find a connecting link between our two solutions; and fortunately this is not difficult.

The morning and afternoon solutions may be written

$$N_{a.m.} = \left(\frac{q}{\alpha}\right)^{\frac{1}{2}} \left\{1 - \tfrac{1}{2}|\epsilon| - \tfrac{5}{8}\epsilon^2 - \ldots\right\} \rightleftharpoons \left(\frac{q}{\alpha}\right)^{\frac{1}{2}} R(\beta) \quad (44)$$

$$N_{p.m.} = \left(\frac{q}{\alpha}\right)^{\frac{1}{2}} \left\{1 + \tfrac{1}{2}|\epsilon| - \tfrac{5}{8}\epsilon^2 + \ldots\right\} \rightleftharpoons \left(\frac{q}{\alpha}\right)^{\frac{1}{2}} S(\beta) \quad (45)$$

where the symbol $\rightleftharpoons$ has been used to indicate the merging of the two solutions. By definition, in the one case

$$\epsilon = \frac{1}{2(q\alpha)^{\frac{1}{2}}} \frac{q'}{q}$$

and in the other case

$$\beta = \tfrac{2}{3}(\alpha b)^{\frac{1}{2}} |t - t_0|^{\frac{3}{2}}$$

and $q = b|t - t_0|$ so that $b = q'$ and the last equation may therefore be written

$$\beta = \tfrac{2}{3}(q\alpha)^{\frac{1}{2}} \frac{q}{q'} = \frac{1}{3\epsilon} \qquad (46).$$

This equation clearly gives us the required link. In Fig. (1) the functions $R(\beta)$ and $S(\beta)$ have been plotted from the Bessel-function tables given by Jahnke-Emde (1945, p. 235) and are shown by full lines:

the quasi-stationary solution for $\epsilon = 1/3\beta$ are shown by dotted lines on the same graphs. The two functions become coincident as $\beta \to \infty$ (or $\epsilon \to 0$): and are very close at $\beta = 4$ (or $\epsilon = 1/12$) in the morning and at $\beta = 3$ (or $\epsilon = 1/9$) in the afternoon but from these points to $\beta = 1$ the error in the quasi-stationary solution becomes increasingly large and for $\beta > 1$ this solution is of course obviously invalid because of the rapid divergence of the series.

So long as linear approximations (or series of linear approximations) are valid for the non-stationary periods, the two solutions can evidently be combined to give a single complete solution covering the whole day.

2.3 The Height and Magnitude of Maximum Electron Density.

We have seen that if the rate of electron-production q is given as a function of time then two main methods of solution of the continuity equation are available. The first, expressed by equation (25), is satisfactory when q is varying slowly (more precisely when $\epsilon \ll 1$); and the second, resulting in equations (36) and (41) is appropriate when q is varying quickly, but is restricted to such a range of t that the variation of q may be considered linear.

So far, however, we have taken no account of the variation of q with height z. The above methods are

of course valid for any constant height; and will yield N as a function of time at any given height, but in the normal practice of ionospheric radio-sounding it is not practicable to measure N at a given height. The quantity measured is the so-called critical frequency of the layer f_c which by way of the well-known result

$$N_m = 1.24 \times 10^4 f_c^2 \qquad (47)$$

yields a value for the spatial maximum of electron density, N_m, that is, the value of N at the level where $\partial N/\partial z = 0$ at any particular time. We shall denote the height of this level by h_m when it is measured in km above the ground, or by z_m when it is measured in scale-heights above the Chapman zero-level defined by (9). h_m for the E-layer is not measured as a matter of routine at the ionospheric stations, but it can be computed from the ionospheric record when a suitable E-layer trace is present and examples of such determinations are given in Chapter 4. The routine measurement is of h'E the minimum recorded height of the layer: and in practice this may to some extent depend on f_{min}, the minimum frequency for which echoes are observed, which in turn depends on the level of absorption.

2.3.1 The Continuity Equation for N_m.

We must now examine whether it is possible to solve our equations for the measurable quantities h_m (or z_m)

and N_m, and for this purpose we shall assume that $\underline{q}$ is given by the expression for a simple Chapman layer, namely,

$$q = q_o \exp (1 - z - e^{-z} \sec \chi).$$

Now if we follow the N_m level

$$\frac{dN_m}{dt} = \left(\frac{\partial N}{\partial t} + \frac{\partial N}{\partial z} \cdot \frac{dz_m}{dt} \right)_{z = z_m}$$

$$= \left(\frac{\partial N}{\partial t} \right)_{z = z_m},$$

and so the continuity equation becomes

$$\frac{dN_m}{dt} = q(z_m) - \alpha N_m^2 \qquad (48)$$

where $q(z_m)$ is the value of $\underline{q}$ at the height z_m , i.e. at the level where N is a maximum. In general this will be different from q_m the maximum value of $\underline{q}$ given by

$$q_m = q_o \cos \chi \qquad (16a)$$

which occurs at the level

$$z(q_m) = \ln \sec \chi \qquad (10).$$

Simple physical considerations suggest that there will be a certain delay in the variation of N compared with the corresponding variation in $\underline{q}$, so that in the morning when the q_m level is coming down we should expect that, at any given time, the N_m level will be

higher than the q_m level; and conversely in the afternoon. To clarify this the relative variations of N and q with height at a given instant are sketched roughly (and in exaggerated form) in Fig. 2.

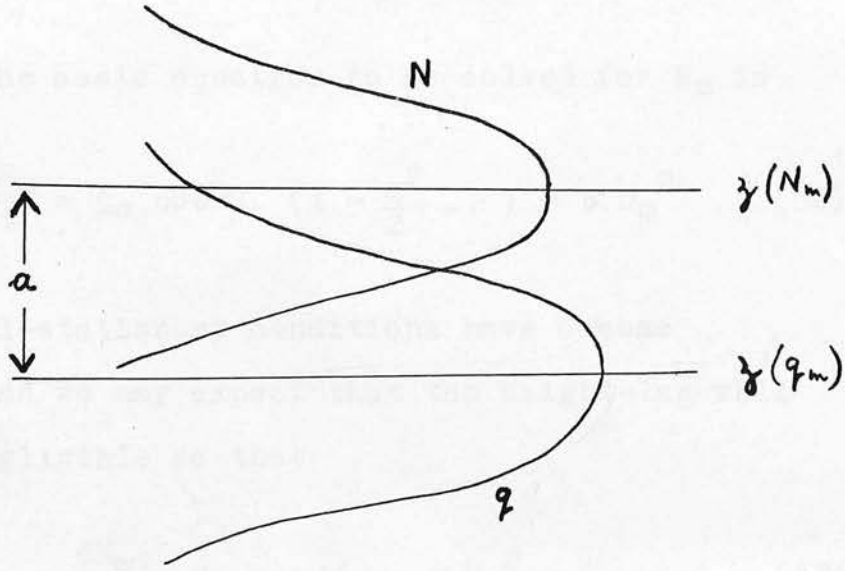


Fig. 2

The difference in height between the two levels, termed by Appleton and Lyon (1953) the "height-lag", we shall denote by a, that is

$$z_m = z(N_m) = z(q_m) - a, \qquad (49)$$

$$= \ln \sec \chi - a \qquad (49a)$$

in the case under consideration.

If a is small then at the N_m level,

$$e^{-z} \sec \chi = e^{a} = 1 + a + \frac{a^2}{2} + \ldots \qquad (50)$$

and

$$q(z_m) = q_o \cos\chi \exp\left(-\frac{a^2}{2} - \frac{a^3}{6} \ldots\right)$$

$$= q_o \cos\chi \left(1 - \frac{a^2}{2} \ldots\right) \qquad (51)$$

so that the basic equation to be solved for N_m is

$$\frac{dN_m}{dt} = q_o \cos\chi \left(1 - \frac{a^2}{2} \ldots\right) - \alpha N_m^2 \qquad (52).$$

Once quasi-stationary conditions have become established we may expect that the height-lag will become negligible so that

$$\frac{dN_m}{dt} = q_o \cos\chi - \alpha N_m^2 \qquad (52a)$$

and the zero-order approximation, valid near noon, is

$$N_m = \left(\frac{q_o}{\alpha}\right)^{\frac{1}{2}} (\cos\chi)^{\frac{1}{2}} \qquad (53).$$

To obtain better approximations we must allow both for the dN_m/dt term and also for the height-lag a and in particular we must determine the value of a.

2.3.2 The height-lag and the value of N_m: quasi-stationary case.

At any given height, under quasi-stationary conditions, equation (25) gives, after some reduction,

$$\frac{\partial N}{\partial \gamma} = \frac{1}{2} a \left(\frac{q}{\alpha}\right)^{\frac{1}{2}} - \frac{1}{4\alpha}(1+a)\tan\chi\,\frac{d\chi}{dt} + \frac{5}{16\alpha^{3/2}q^{1/2}}\left(\tan\chi\,\frac{d\chi}{dt}\right)^2$$

$$= 0 \qquad (54)$$

neglecting terms of the third degree and higher in a or $\tan\chi \; d\chi/dt$. If we neglect also terms of the second degree we obtain the first-order approximation

$$a = \frac{\tan\chi \; d\chi/dt}{2(q\alpha)^{\frac{1}{2}}} = \epsilon \,, \qquad (55)$$

which confirms our expectation that as conditions become more nearly stationary the height-lag tends to zero.

Inserting this value in second-degree terms gives the second-order approximation

$$a = \epsilon - \frac{3}{2}\epsilon^2 \qquad (56)$$

and by (51)

$$q(z_m) = q_0 \cos\chi \left(1 - \frac{\epsilon^2}{2} + \frac{3\epsilon^3}{2} - \dots\right) \qquad (57)$$

Hence by (52) and (22)

$$N_m = \left(\frac{q_0 \cos\chi}{\alpha}\right)^{\frac{1}{2}} \left\{1 - \tfrac{1}{2}\epsilon - \tfrac{7}{8}\epsilon^2 - \tfrac{9}{8}\epsilon^3 - \dots\right\} \qquad (58).$$

If ϵ^2 is negligible, as is usually the case, we have the following useful results

$$\overline{N_m} = \tfrac{1}{2}\left\{N_m(+\chi) + N_m(-\chi)\right\}$$
$$= \left(\frac{q_o \cos\chi}{\alpha}\right)^{\frac{1}{2}} \qquad (59)$$

and

$$\Delta N_m = N_m(+\chi) - N_m(-\chi)$$
$$= \epsilon N_m = \frac{\tan\chi \; d\chi/dt}{2\alpha} \qquad (60).$$

The average $\overline{N_m}$ and the difference ΔN_m of values of N_m at the same χ in the morning and afternoon are, as we shall see, particularly useful quantities for the experimental investigation of the behaviour of an ionospheric layer.

2.3.3 N_m and the height-lag in the non-stationary case.

An alternative expression for the height-lag, not necessarily confined to quasi-stationary conditions, has been given by Appleton and Lyon (1955) on the basis of simple physical considerations. This states that, to a first approximation, the height-lag is given by the product of the relaxation time τ and the speed of descent of the layer, i.e.

$$a = \frac{1}{2\alpha N_m}\frac{d}{dt}(\ln \mathrm{Ch}\,\chi) \qquad (61).$$

This result permits an estimate of the height-lag correction

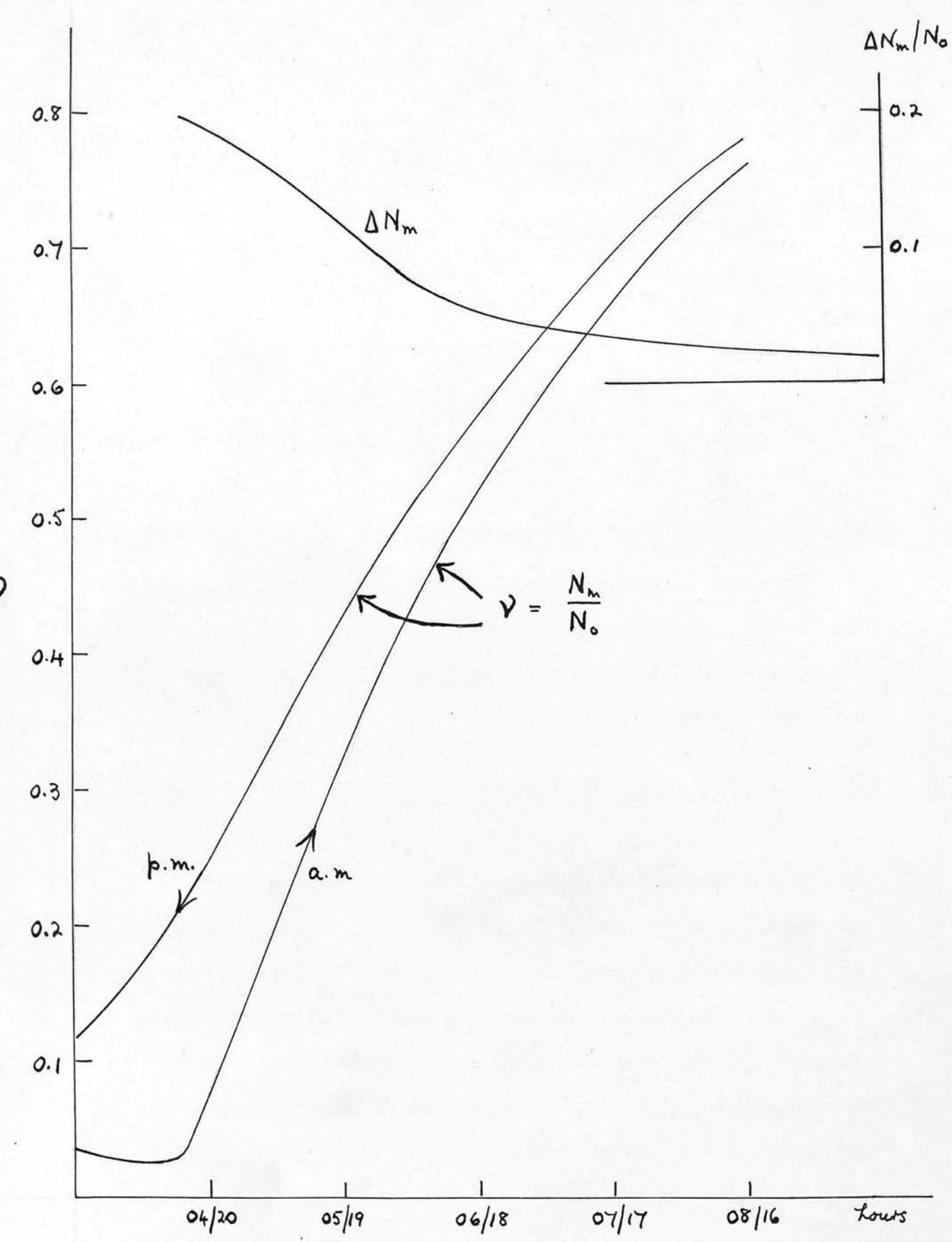


Fig. 3 Theoretical variations of N_m and ΔN_m for a summer day at latitude 50°N.

at times quite close to sunrise and sunset.

To use the functions $R(\beta)$ and $S(\beta)$ to obtain N_m in the sunrise and sunset periods we require a linear approximation to $q(z_m)$ - the rate of ion-production at the layer maximum. As it turns out the effects of the Chapman-function correction and the height-lag correction tend to cancel one another out, with the result that a linear approximation for the hour or so after sunrise or before sunset is a remarkably good one, and is readily deduced from a curve of $1/\mathrm{Ch}\,\chi$ against time for the place and time in question.

Given the curve just mentioned it is now a simple matter to calculate the theoretical variation of N_m throughout the day and night. The result of such a procedure for June at latitude 50°N is shown in Fig. 3. It will be observed that as we approach sunrise or sunset the quantity ΔN_m departs increasingly from the simple expression (60) though it does continue to increase as $\cos\chi$ decreases.

It has been mentioned that our results for non-stationary conditions are based on the work of Rydbeck and Wilhelmsson (1954) relating to variations of N_m during an eclipse. These authors do not mention the height-lag correction, but it will be clear from the present investigation that they are correct to neglect this factor in eclipse computations except in the immediate vicinity of sunrise or sunset. Even at 05 h

in the example we have chosen the correction in question is only of order 1%.

2.4 The Effects of Drift on the Characteristics of a Layer.

If the electrons in a layer are subject to a vertical drift velocity v, which itself varies with height z, the continuity equation then becomes

$$\frac{\partial N}{\partial t} = q - \alpha N^2 - N \frac{\partial v}{\partial z} - v \frac{\partial N}{\partial z} \qquad (62)$$

and at the layer maximum

$$\frac{dN_m}{dt} = q - \alpha N_m^2 - N_m \frac{\partial v}{\partial z} \qquad (63).$$

Assuming that dN_m/dt and the drift term are small compared with the other two terms we obtain for N_m the equation

$$N_m = \sqrt{\frac{q}{\alpha}} \left\{ 1 - \tfrac{1}{2}\epsilon - \frac{1}{2\alpha N_m} \frac{\partial v}{\partial z} \right\} \qquad (64)$$

according to which we see that

(i) the relaxation and drift effects are independent,

(ii) the drift effect depends only on the gradient of drift not on the drift velocity itself, and

(iii) for a given drift gradient the change, δN_m say, due to drift is proportional to the relaxation time τ: in fact

$$\frac{\delta N_m}{N_m} = - \tau \frac{\partial v}{\partial z} \qquad (65).$$

We have assumed in the above remarks that any change, δz say, in the height of the layer-maximum due to drift is a small quantity of the first order so that any effect of this on $\underline{q}$, and hence on N, is of the second order only. We must now consider in detail the magnitude of this quantity δz. Since relaxation and drift effects are, in the main, independent, it will clarify the argument and lead to no serious loss of generality, if we assume for the moment that conditions are stationary, i.e. that $\partial N/\partial t = 0$, at all levels. We therefore have

$$\alpha N^2 = q - N \frac{dv}{dz} - v \frac{dN}{dz} \qquad (66).$$

Differentiating with respect to $\underline{z}$ and then putting $dN/dz = 0$ we obtain for the height z_m of maximum electron density the condition

$$\frac{dq}{dz} = N \frac{d^2v}{dz^2} + v \frac{d^2N}{dz^2} \qquad (67)$$

which can be solved for z_m by successive approximations.

The first approximation, which was given by Appleton and Lyon (1954), is

$$\delta z = \frac{v}{2 \alpha N_m} = \tau v \qquad (68)$$

and it is not difficult to deduce a second approximation, namely,

$$\delta z = \tau v(1 + \tfrac{1}{2}\tau v - 3\tau v') \qquad (69).^{*}$$

Allowing for this small departure from the level of maximum q we now have for N_m the result

$$N_m = \sqrt{\frac{q}{\alpha}}\,(1 - \tau v' - \tfrac{1}{4}\tau^2 v^2) \qquad (70).$$

To obtain an expression for N in the neighbourhood of the layer-maximum we may use a Taylor expansion and this yields the equation

$$N = N_m \left\{1 - \tfrac{1}{4}\right\}^2(1 - \tau v') + \tfrac{1}{12}\Big\}^3 - \ldots\Big\} \qquad (71)$$

where $\}$ is the distance, in scale-heights, from the layer maximum. It will be seen that a positive drift-gradient <u>increases</u> slightly the apparent thickness of the layer near its maximum.

This result, however, is valid only for small $\}$, and to calculate the effect of drift on the shape of the layer as a whole a different procedure is required. If we may still assume that the drift has continued long enough for conditions to become stationary at all levels, the problem reduces to solving the differential equation (66) for N. Putting $\nu = N/N_m$ the latter equation becomes

$$\frac{d\nu}{dz} = \frac{q}{vN_m} - \frac{\alpha N_m}{v}\nu^2 - \frac{v'}{v}\nu \qquad (72)$$

$* \quad v' \equiv \partial v/\partial x$

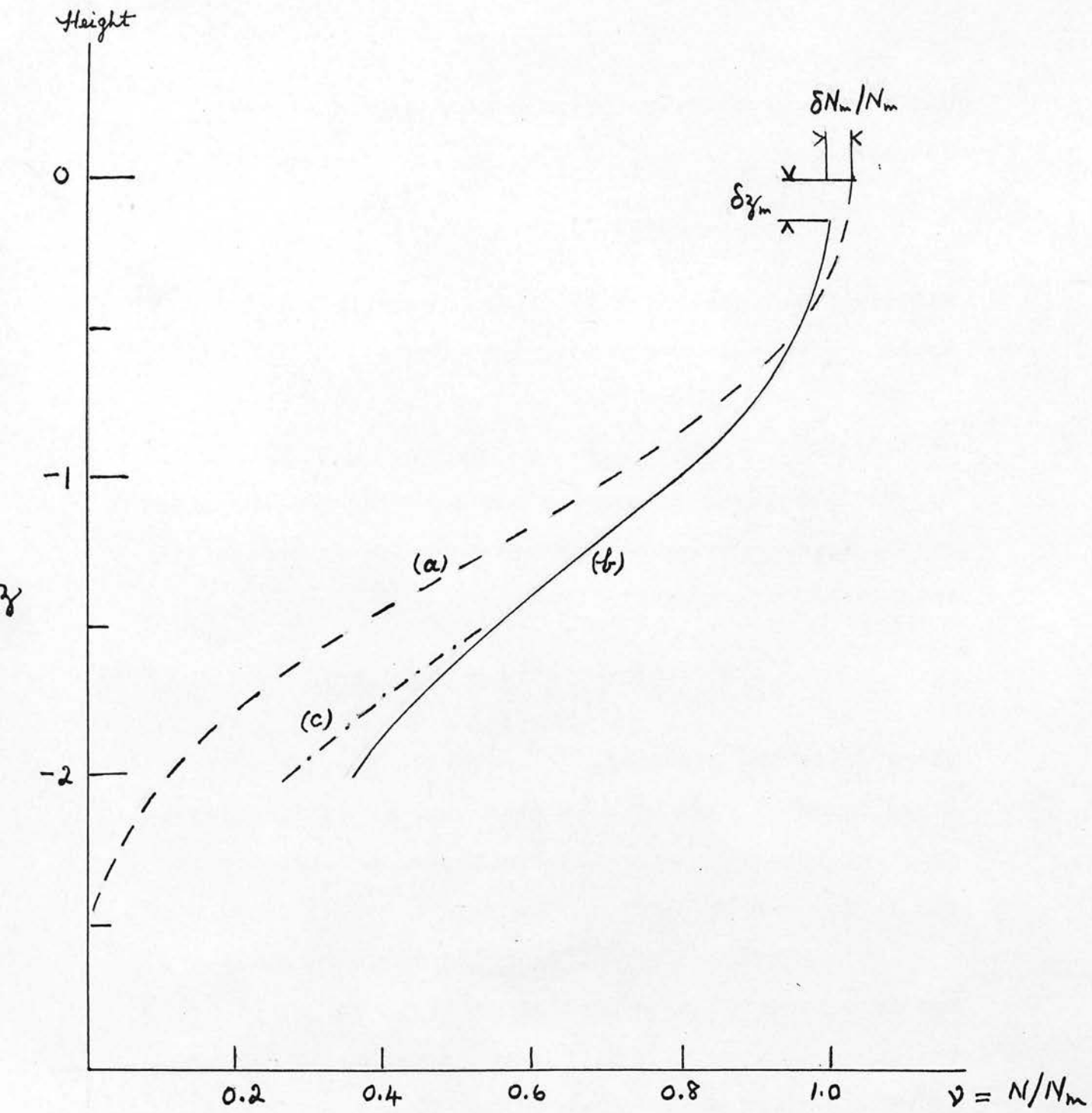


Fig. 4 The effect of drift on layer shape.

(a) Undisturbed simple Chapman layer.

(b) Layer with drift (equilibrium position).

(c) Estimated position after one hour.

where the coefficients may all be functions of z. It is evident that, in general, only numerical solutions of such an equation are possible. For a set of numerical values* which are consistent with the experimental evidence at Slough the result of such a numerical solution is drawn in Fig. 4. This figure shows the small decrease in N_m and in z_m due, respectively, to the positive drift gradient and the downward drift velocity, in accordance with equations (65) and (68), and also shows the considerable increase in ionization-density in the lower part of the layer resulting from the downward drift.

However it must be remembered that the lower down we go in the layer the greater becomes the relaxation-time τ, and hence the longer will the ionization take to reach the equilibrium condition denoted by the full-line curve of Fig. 4. If we assume, as a rough indication, that this condition is approached in an exponential manner with time-constant τ, then the dash-dot continuation would give the position after one hour under average E-layer conditions; and for the practical case of a tidal drift-velocity this may give a more realistic picture than the equilibrium curve at the bottom of the layer.

---

* In fact, $\alpha N_m/v = -3\, e^{-z/5}$.

Chapter 3.

Statistical Investigation of the E-layer.

3.1 Previous statistical studies.

Most statistical investigations of the E-layer up to now have concentrated attention on finding the best value of n in the assumed relationship

$$fE = f_0(\cos \chi)^n \qquad (73)$$

or of n' in the alternative expression

$$(fE)^{n'} = A \cos \chi \qquad (74).$$

Thus Allen (1948) examined this question in relation to a study of sun-spot activity and its effects. Using noon values and the mean of values at 09 h and 15 h throughout the year he calculates n' for each of 14 stations, and obtains an average value of n' = 3.72 (or n = 0.27). This however is a value primarily for the annual variation rather than the diurnal variation of cos χ, and as we shall see the two are in general different.

Tremellen and Cox (1947) in a discussion of ionospheric predictions state that the average value of n for the diurnal variation at moderate latitudes is 0.31, with somewhat larger values in local summer and at sunspot-maximum. They also note (i) an increase of fE at given χ as the equator is approached and (ii) a seasonal effect causing a small reduction of fE in

local summer.

Harnischmacher (1950) on the basis of an analysis of data from 4 stations only, obtains the following empirical formulae for f_0 and $\underline{n}$

$$f_0 = 2.25 + 1.5 \cos \phi + (0.01 - 0.007 \cos \phi)R$$
$$n = 0.21 + 0.12 \cos \phi + 0.002R \qquad (75)$$

where ϕ is the latitude and R the sunspot number. He thus finds that $\underline{n}$ at sunspot minimum decreases with latitude from 0.33 at the equator to 0.21 at the poles; and he finds that $\underline{n}$ increases with sunspot number. Harnischmacher also found a seasonal variation of $\underline{n}$ at Washington with a minimum in winter and a maximum in summer.

Scott (1952) finds for summer months an average value $\underline{n} = 0.33$ for moderate latitudes but notes a sharp drop to between 0.10 and 0.20 in the north auroral zone followed by a rise to 0.25 further north.

Saha (1953) computes values of $\underline{n}$ and f_0 for Calcutta: and states that the two quantities tend to vary together, having an apparent seasonal variation with minima in summer and maxima in winter, and further that afternoon values of $\underline{n}$ (which he computes separately) tend to be smaller than morning values. He finds a wide variation of $\underline{n}$ between 0.10 and 0.50.

Menzel and Wolbach (1954) give what is perhaps the most elaborate analysis to date. To study annual

variation they mass data for all available stations for each month and they find (like Saha) that the resulting monthly values of $\underline{n}$ and f_0 tend to vary together and also (unlike Saha) that they show a correlation with the monthly geo-magnetic index K_p, having maxima at the equinoxes and minima at the solstices. To study latitude variation they mass all data throughout the year for each station and find that both f_0 and $\underline{n}$ decrease with latitude. Thus $\underline{n}$ (which in this case will include effects of both seasonal diurnal variation) varies from around 0.315 at the equator to around 0.26 at 65°N. A third method of analysis they use is to compare the value of fE at a given χ and a given station with the average at this χ for all stations. Averages of this ratio were then computed for different ranges of χ and the variation with latitude examined. The results showed that around noon ($0^\circ < \chi < 27^\circ$) there was a marked increase in the relative value of fE towards the equator but this effect became increasingly less marked for higher ranges of χ.

3.2 The Edinburgh work on the E-layer.

A weakness of the work of Allen and also that of Menzel and Wolbach is the failure to distinguish carefully between the three ways in which $\cos\chi$ may be varied, namely by varying time of day, season and latitude. The $\cos\chi$ relation, as expressed by

equation (73), must be studied separately for each type of variation, and this isolation of the various factors at work has been one of the main objects of the statistical work carried out at Edinburgh.

In one of several methods used for this purpose, the months were grouped by season: northern summer (May to August), equinox (February to May and August to November) and northern winter (November to February). By making use of mass plots of fE against latitude at individual hours (corrected to apparent time), mean values of fE can be derived for a series of latitudes (say at intervals of 10^{o}) for every hour or the day; and from these, smoothed diurnal curves for each season and latitude can be drawn, and any required type of variation studied. This method was used both at sunspot-minimum and sunspot-maximum.

An alternative procedure, also used at Edinburgh, is to average values of fE at each hour of the day, and in each month of the year, over a large number of years for individual stations (e.g. Slough, Washington, Christchurch, Singapore etc.). This is particularly useful for studying annual variations in detail.

We have seen in Section 2.3.2 that for a simple ionospheric layer produced in the manner proposed by Chapman it is possible to make certain theoretical predictions which are readily expressed in terms of fE as follows:-

(i) If we average fE for hours of the same χ in the morning and afternoon the result, $\overline{fE}$ say, should obey a law of the form

$$\overline{fE} = f_0 (\cos \chi)^n \qquad (76)$$

where $n = \frac{1}{4}(1 + \mu)$ and μ is the fractional increase of scale-height (or temperature) per scale-height, and may be expected to be of the order of 0.1, so that the exponent of $\cos \chi$ should be around 0.27.

(ii) f_0 in (76) will vary with sunspot activity but should not vary with latitude, and if corrected for sunspot activity should not vary with season: or in other words $\overline{fE}$ at the same χ should be the same at all seasons and latitudes.

(iii) If we take the difference Δf between values of fE at times of the same χ in the morning and afternoon we should find that Δ fE increases steadily with χ according to the law

$$\Delta fE = \frac{df/dt}{\alpha k f^2} = \frac{n}{\alpha k f} \tan \chi \frac{d\chi}{dt} \qquad (77)$$

in the quasi-stationary period. As Fig. 4 illustrates, at large χ fE will vary rather less rapidly with than this would indicate, but should still increase with χ.

In fact however, as we shall see, none of these predictions is fulfilled in practice.

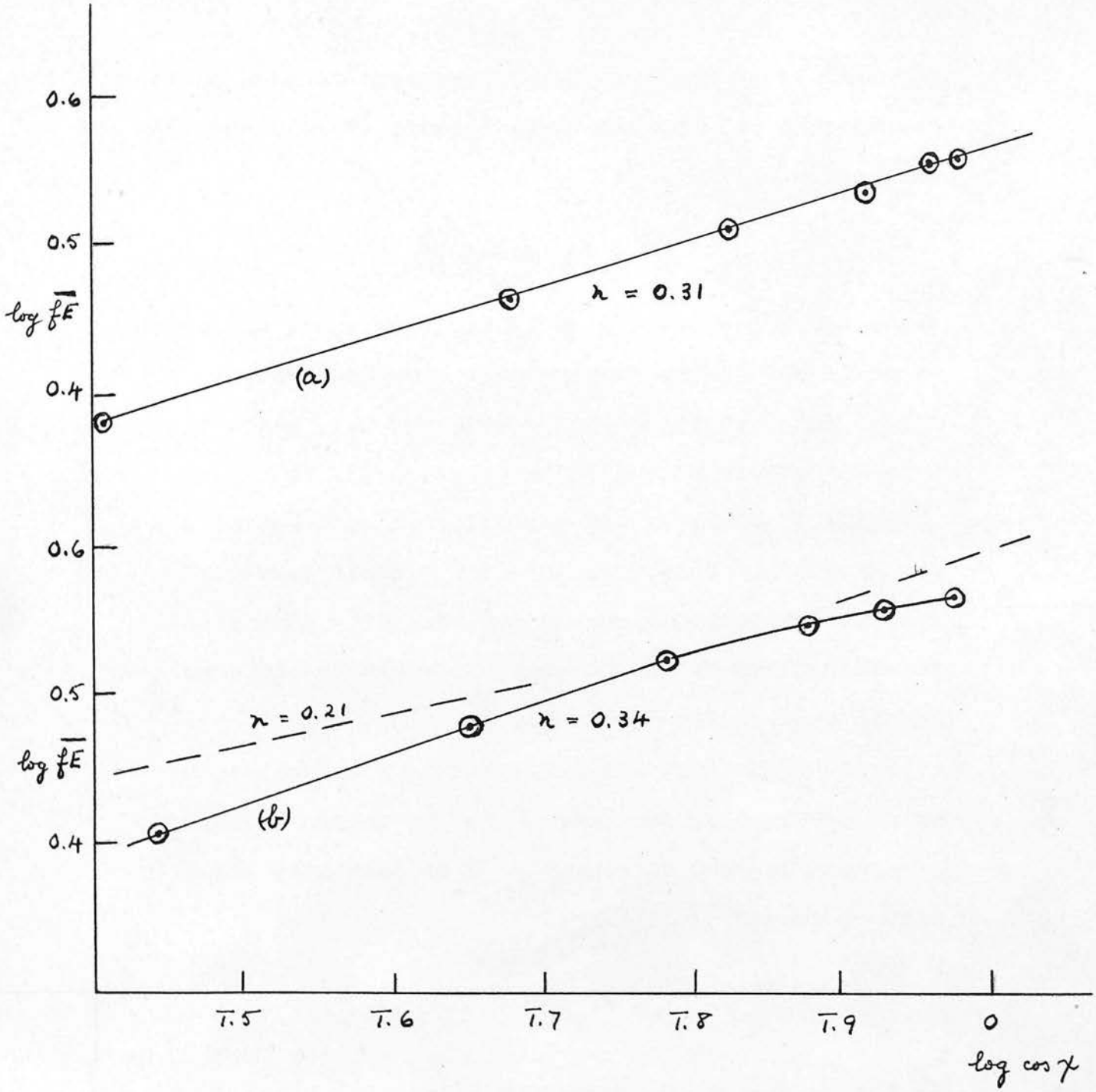


Fig. 5 The variation of $\overline{fE}$ with cos χ

(a) Singapore, July, mean 1950 -53.

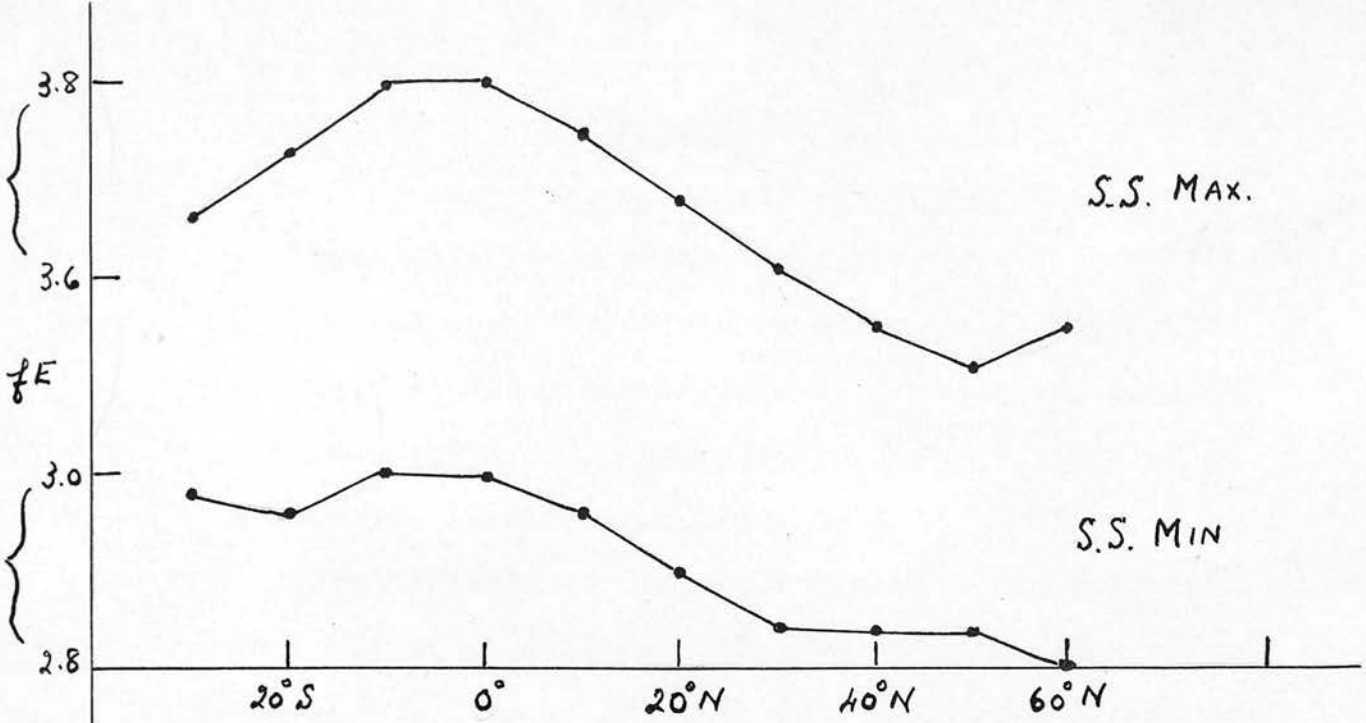
(b) Christchurch, December, mean 1947-53.

3.2.1 The E-layer diurnal cos χ -law.

To obtain the diurnal value of n it is most convenient to plot log $\overline{fE}$ against log cos χ and two typical examples of such plots are shown in Fig. 5. In both cases n has values near 0.33. The curious downward inflection of the first curve at high cos χ is typical of summer and equinox at mid-latitudes, and a small upward inflection at low cos χ is also common. If the central straight portion of the curve is used to determine n in such cases the value of n is found to depart little from 0.33 (remaining for the most part between 0.30 and 0.36 regardless of season or latitude. No very definite significant trends can be observed in the variations of n that do occur, though it does appear to be slightly higher at sunspot-minimum than at sunspot-maximum and possibly tends to rise slightly towards winter. If a best straight line were used in the inflected cases (instead of the slope of the straight portion) this might lead to a slight decrease of n with latitude, as found by Harnischmacher and Menzel and Wolbach. It will be noted that at high cos χ , in the example shown, the slope gives n = 0.21.

3.2.2 $\overline{fE}$ at constant χ: latitude and seasonal variations.

When $\overline{fE}$ for the same χ is plotted against latitude a curve such as that in Fig. 6 is obtained, showing a



$\overline{fE}$ at constant χ. Summer months.

(From smoothed data using all stations. Cf. p.39)

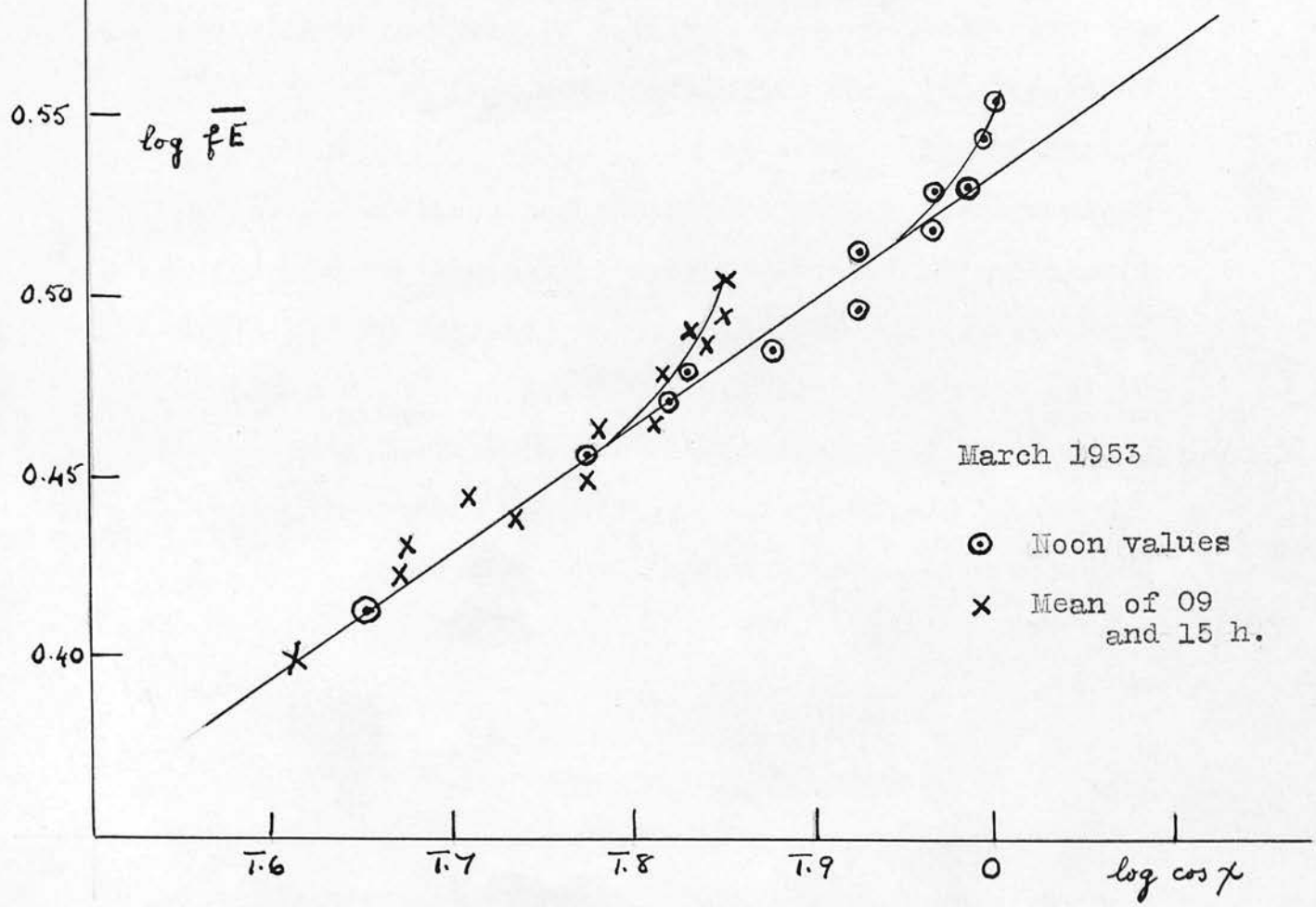


Fig. 6 Latitude variations of fE.

marked rise near the equator: this agrees with the results of Menzel and Wolbach (1954) and also with the earlier observations of Tremellen and Cox (1947) already quoted. What is essentially the same phenomenon can be observed in a plot of log fE against log cos χ, say for noon, for various latitudes: as Fig. 6 shows the result is an upward inflection near the equator.

When in a similar way $\overline{fE}$ for some constant χ is plotted against month of the year, as in Fig. 7, it is found that at mid-latitudes winter values are distinctly higher than summer values, and sometimes there is a secondary maximum in summer. This result is again in agreement with the findings of Tremellen and Cox (1947).

This type of seasonal variation can also be expressed in another way, by determining the cos χ - exponent for seasonal variation of fE at noon. For Slough and Washington the seasonal n turns out to be very close to the theoretical value of $\frac{1}{4}$, as compared to the diurnal value of $\frac{1}{3}$. This is in fact equivalent to the statement that at the same χ fE will be less in summer, as the diagram of Fig. 8 illustrates.

The same figure also shows that in keeping χ constant but varying season, time of day is also necessarily varied: hence the apparent seasonal effect could also be interpreted as implying that fE at the same χ is greater at noon than in the morning or evening. It is thus not possible to tell from the

statistical analysis alone whether a true seasonal variation is present or not.

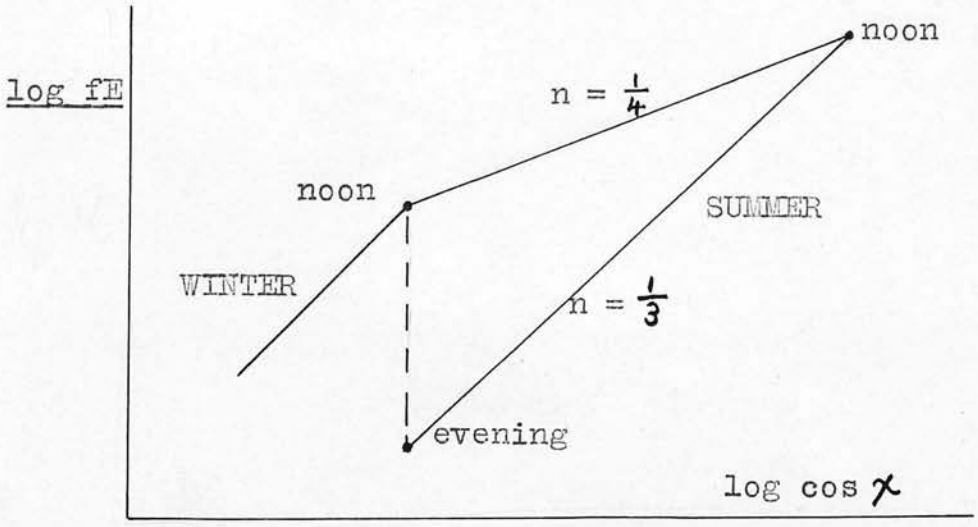


Fig. 8

If Beynon and Brown (1956) are correct the seasonal variation and the value of n for the seasonal cos χ - law are strongly affected by tidal variations.

It may be mentioned that no evidence whatever was found for the type of seasonal variation observed by Menzel and Wolbach (1954), with minima at the solstices: indeed our results show the exact contrary of this.

3.2.3 Anomalies in the diurnal asymmetry ΔfE: the S_q perturbations.

The diurnal asymmetry ΔfE rarely if ever shows the theoretical variation indicated by equation (77) - that is, a steady increase for increasing χ. For times not far from noon ΔfE is too large at high

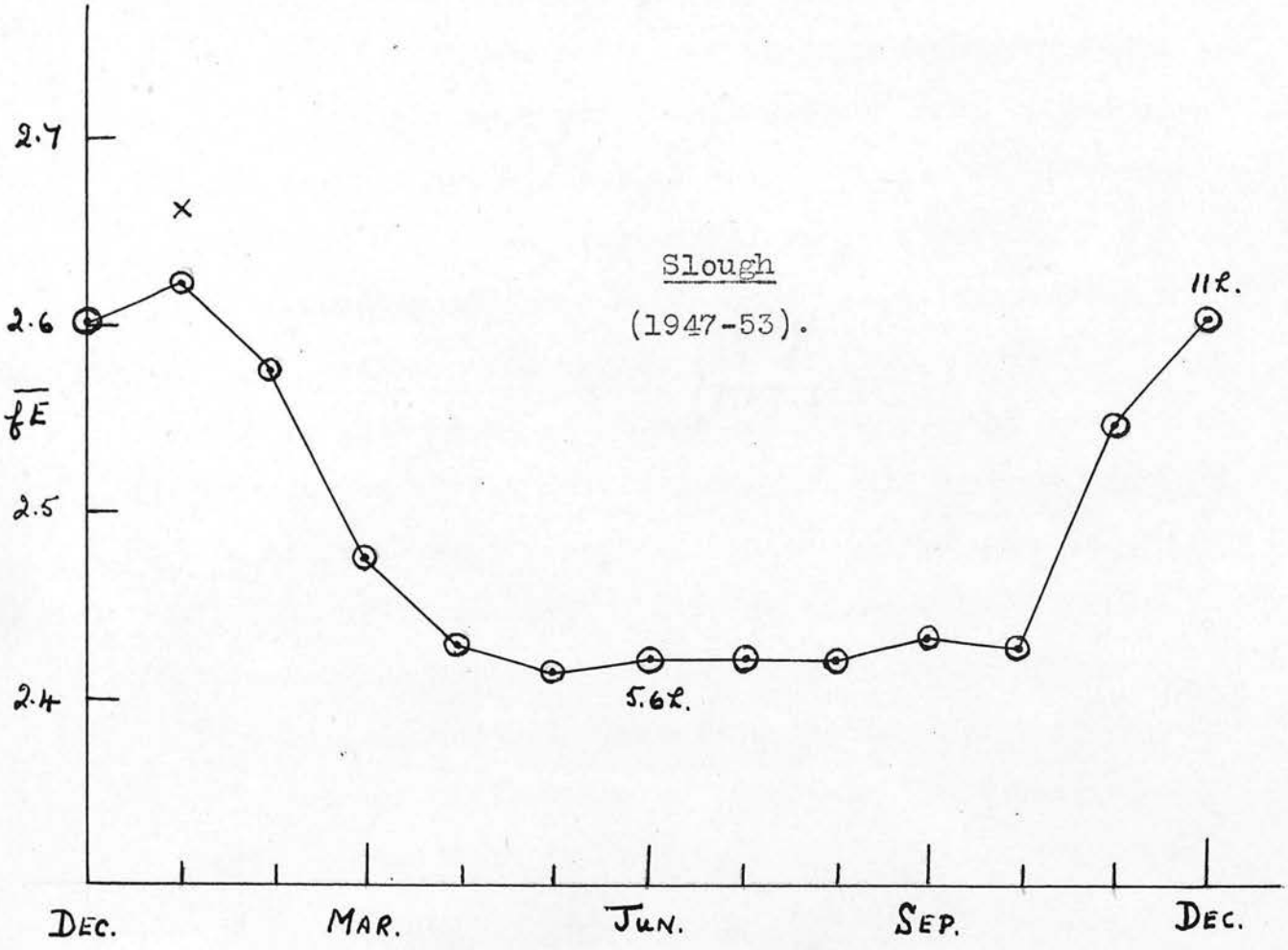


Fig. 7 Seasonal variation of $\overline{fE}$ at constant χ.

X Indicates value before correction for the earth's smaller distance from the sun.

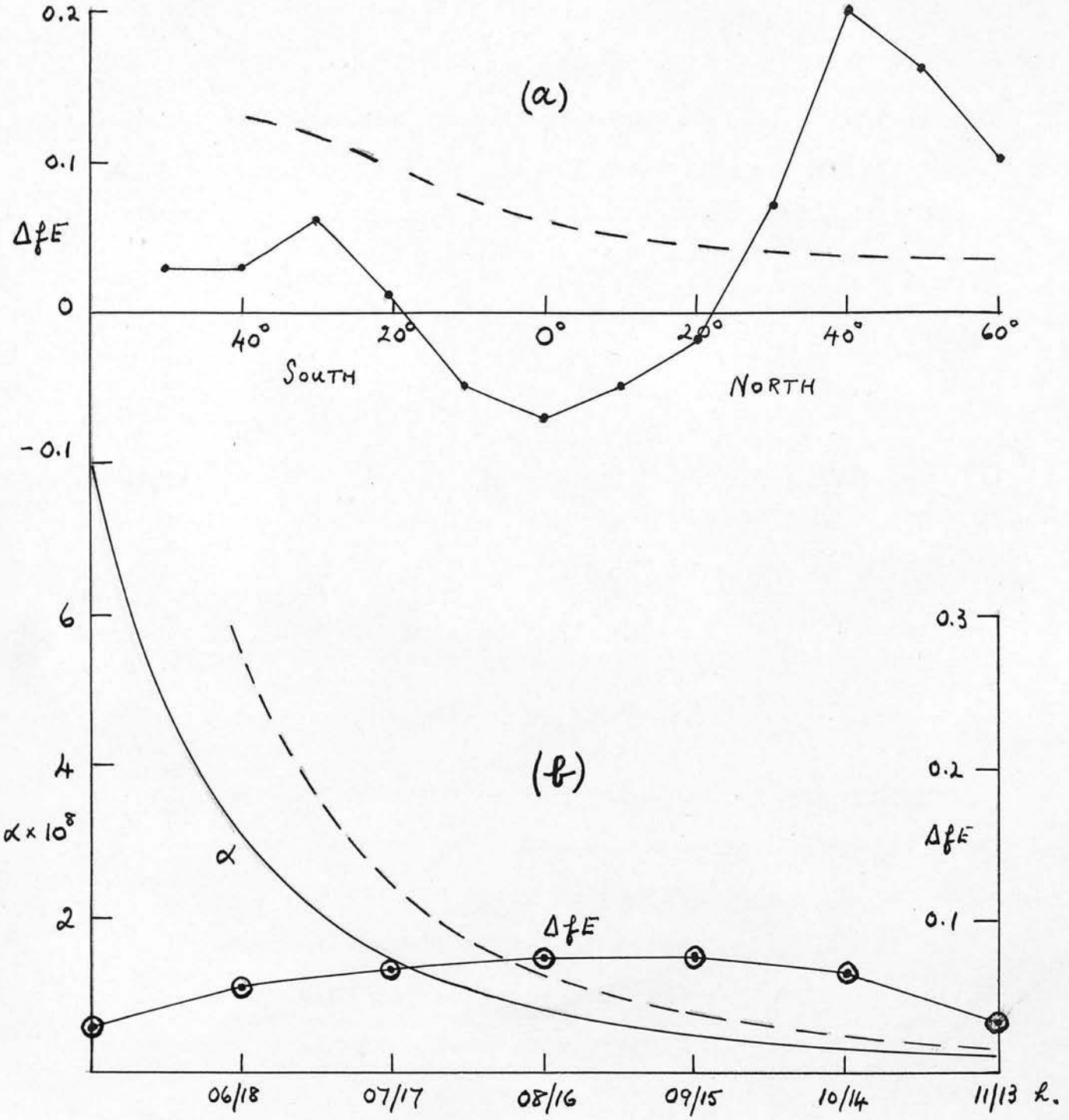


Fig. 9 (a) Latitude variation of ΔfE 15h. - 09h. (Summer months, sunspot maximum.) *

(b) Diurnal variation of ΔfE at Slough, and corresponding apparent variation of α. Summer 1952.

Dashed-line curves indicate theoretical variations of ΔfE.

* 1948 Smoothed data

latitudes and at low latitudes it becomes negative: a plot of ΔfE against latitude is shown in Fig. 9 for 09 h and 15 h, and a diurnal plot of ΔfE typical of summer conditions at mid-latitudes is also shown in the same diagram. The latter curve illustrates a second anomalous feature of the ΔfE variation, namely that ΔfE tends to zero at low $\cos\chi$ instead of continuing to increase as theory predicts. If α is determined from diurnal curves of fE in the manner proposed by Appleton (1937), that is by using equation (77), these low values of ΔfE mean that the apparent value of α increases very rapidly with χ at large χ, as indicated by the dashed line in the figure. This phenomenon was noted by Grace (1951) who considered that it represented a true increase of α with height. However there seem to be insuperable objections to such an explanation: in the first place such a rapid increase of α with height would be very difficult to account for theoretically; and further since q_0/α decreases only slightly with χ it would mean that q_0 also increased enormously at large χ, which is quite unacceptable. We conclude then that there is some extraneous perturbation in fE either in the early morning or the late afternoon; but the nature or cause of this effect is so far unknown.

As to the anomalous values of ΔfE in the midday period however, the fact that the anomaly changes

phase from high to low latitudes provides a clue which enables us to give an explanation, in the manner set out by Appleton, Lyon and Pritchard (1955). As Martyn (1953) has shown, whenever a current with an east-west component flows in an ionospheric region where the earth's field has a horizontal component, the resulting mechanical force will cause vertical drift of neutral ionization (electrons and positive ions) in accordance with the equation

$$v = \frac{j_y H}{(m_i \nu_i + m_e \nu_e)N} \qquad (78)$$

where j_y is the component of current towards geo-magnetic east, H the horizontal component of the earth's field, and m_i, ν_i and m_e, ν_e are the masses and collision frequencies of ions and electrons respectively.

Now it is well-known that the so-called S_q current system, associated with the solar atmospheric tide, flows eastwards in low latitudes and westwards in high latitudes reaching maximum intensity around 10 h local time. There should therefore be an upwards drift at this time in low latitudes and a downwards drift at high latitudes. Further if this drift has a velocity-gradient such that the magnitude of the velocity is decreasing upwards, then the change in N_m in accordance with equation (65) will cause a

pre-noon dip in fE at high latitudes and a pre-noon rise at low latitudes, which is exactly what we require to explain the observed anomalies in ΔfE.

The correctness of this explanation can be confirmed in several ways. In the first place there should be, according to equation (68), corresponding effects in h'E and in h_mE, and these have in fact been observed (see Chapter 4). In the second place the increase in fE at low latitudes will, as the configuration of the S_q current system shows, be continued over the greater part of the midday period and therefore explains the observed tendency already mentioned for fE at given χ to increase towards the equator. Finally the same effect probably also explains why the downward deflection of the diurnal log fE - log cos χ curves at high cos χ (noted in Fig. 5) does not occur at high latitudes.

From this consistency of several independent lines of evidence it seems reasonable to conclude firstly that the S_q currents do indeed flow in the E-region, secondly that they do in fact give rise to vertical drifts in the manner first suggested by Martyn (1947), and finally that these drifts do account satisfactorily for some at least of the observed anomalous features in the behaviour of the E-layer.

Chapter 4.

Experimental Investigation of the E-region.

4.1 Previous experimental studies.

We have seen that according to the statistical evidence a number of anomalies exist in the behaviour of the E-layer, some of them apparently explained in a satisfactory manner by the influence of a tidal perturbation others so far without explanation. Various anomalous, or at least unexplained, features have also been observed in experimental studies of the E-layer.

One of the earliest detailed studies of the E-layer and its diurnal variation was carried out by Appleton and Naismith (1932), who give diurnal curves for f_0E for three individual days, one in summer, one in winter and one at equinox. The main object at that time was to establish (a) that the layer variation was under solar control and (b) that it was produced not by corpuscular streams but by absorption of solar radiation in the manner described by Chapman. It is however remarkable to note that on two of the days analysed there is a marked pre-noon dip around 0930 h identical in character with the dip which we have already noted in the statistical results and ascribed to tidal perturbation (Section 3.2.3).

A weakness of the earliest studies was however the failure to deal appropriately with the conditions

now usually described as E_s. The existence of these conditions was noted by Appleton and Naismith (1935), but for some years there was doubt amongst ionospheric workers as to whether in an ionogram such as the one sketched in Fig. 10 f_0E should be read at a where the transition to the "abnormal" layer occurs or at b (now usually called the blanketing frequency,) where F-region echoes first appear. Appleton, Naismith and Ingram (1936, 1939) eventually showed beyond doubt that the former criterion was the consistent one, and broadly speaking it is the one now in general use.

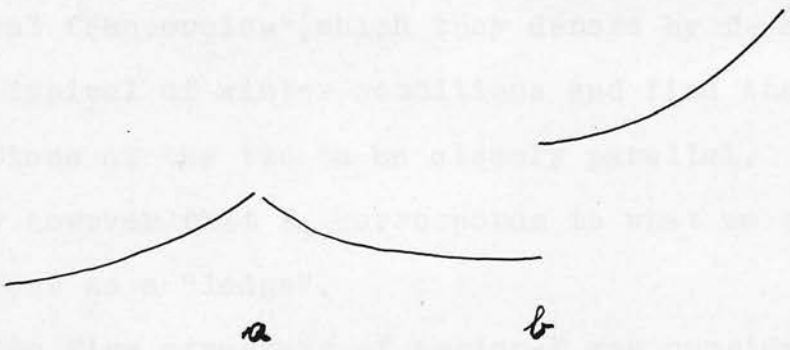


Fig. 10

Halliday (1936) used a differentiating circuit to improve the accuracy of equivalent heights and studied in detail a number of h'-f curves for region-E taken during the autumn months (between June 13 and November 21). An interesting feature which he noted was a discontinuity of height amounting on the average to 5 km and occurring between 1.7 and 2.3 Mc/s. Since

it was usually accompanied by group-retardation, Halliday ascribed the phenomenon to penetration of a low ionized layer. The height of this low layer varied between 90 and 99 km with an average of 95 km.

Another early study of region-E is that of Best, Farmer and Ratcliffe (1938), but their results are somewhat vitiated by their adoption of the b (or blanketing frequency) criterion for the summer period. Their curves do not show any pre-noon dip nor do they appear to show any of the $\cos\chi$ -law or ΔN anomalies under discussion in the present investigation, but it is interesting that they find the occurrence of two critical frequencies (which they denote by f_1 and f_2) to be typical of winter conditions and find the diurnal variations of the two to be closely parallel. It seems likely however that f_1 corresponds to what we should now treat as a "ledge".

The fine structure of region-E was considered in detail by Whale (1950) who noted a number of new and characteristic features of the E-region ionogram. Firstly following Appleton and Naismith (1940) he considers the theoretical h'-f curves for the case when a thin sharply reflecting layer is present at, slightly above or slightly below the normal-E maximum, and he interprets some experimental h'-f curves in terms of such subsidiary layers. Where the apparent height of the subsidiary layer is much greater than that of

normal-E it is treated as a separate E_2-layer: and cases of this kind are said to be common in the early morning and late evening, but in the morning, it is stated, normal-E generally increases faster than the E_2 and so eventually obscures it. Ledges and subsidiary layers are generally supposed by Whale to be transitory phenomena and examples are given of a ledge apparently "dropping off" the F-region, gradually losing height and finally appearing as a sort of E_s-layer, and also of a ledge moving downwards on the E-trace with a quasi-period of about half an hour.

Whale also distinguishs between "rough" echoes showing considerable variation of echo-amplitude with frequency and "smooth" echoes having no such irregularities. Following the work of Ratcliffe (1948) and of Briggs and Phillips (1950) on diffracting screens, Whale ascribes the roughness to irregularities of electron density in a horizontal plane, and describes the reflecting layer as rough or smooth according as these irregularities are present or not.

Bibl (1951) has studied the E-region with particular reference to the precise definition of f_oE for the purposes of routine reduction. He argues that f_oE is to be distinguished from the penetration frequencies of lower, subsidiary layers (a) by the presence of the substantial group retardation to be expected from a layer of the "normal" semi-thickness

of the E-layer (of the order of 20 km) and (b) by a substantial height-discontinuity (~ 15 km at least) in the transition to the next layer. He thus defines f_0E as "the highest critical frequency of a layer of normal thickness, in the E-region, not greater than the first discontinuity of apparent height". Examples of ionograms given by Bibl, and the position of f_0E indicated according to his definition are shown in Fig. 11.

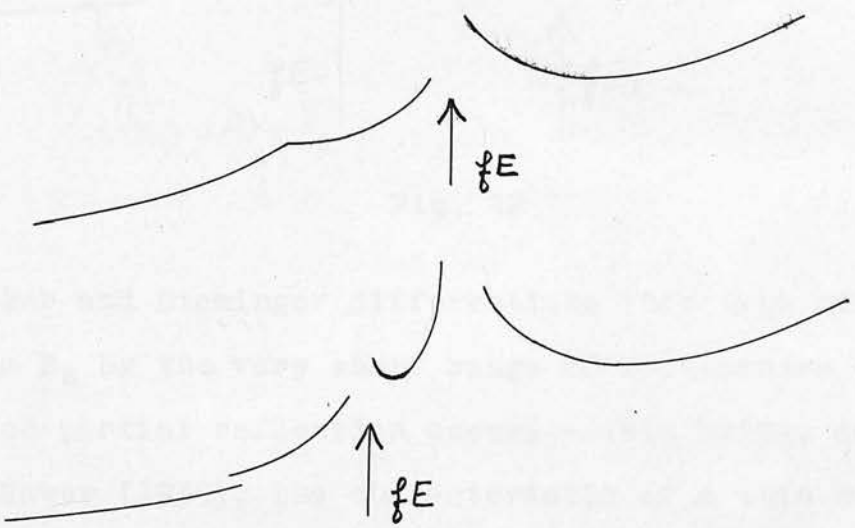


Fig. 11

When extrapolation is necessary due to the presence of E_s Bibl recommends that the value be considered doubtful if the uncertainty exceeds 0,1 Mc/s.

The E_2 layer has been investigated by Becker and Dieminger (1950) but with a somewhat different

interpretation, indicated in Fig. 12, according to which the name E_2 is given to a layer which would be usually described as an E_s layer - and certainly so thin that it shows no group-retardation at the transition to F.

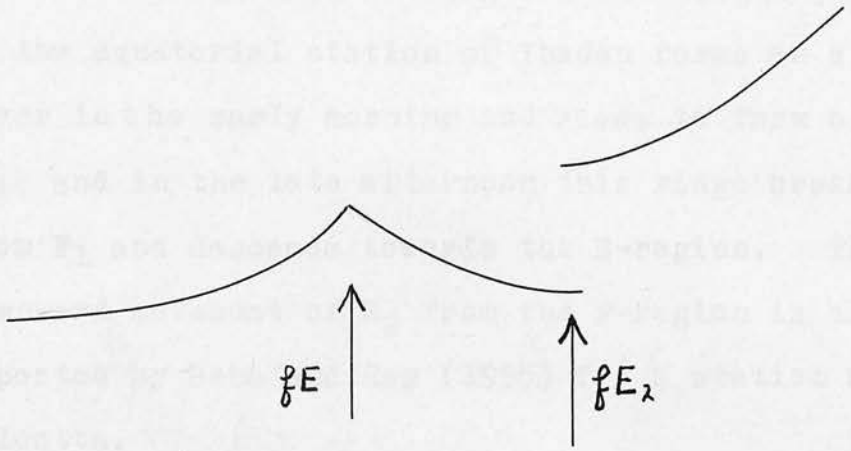


Fig. 12

Becker and Dieminger differentiate this type of layer from E_s by the very short range of frequencies over which partial reflection occurs - this being, according to Rawer (1939), the characteristic of a thin but homogeneous layer. The term E_s is apparently reserved by these authors for the case of reflections, partial or total, extending far above the normal-E critical frequency, and not showing any definite penetration frequency.

A somewhat similar interpretation for fE_2 is adopted by Rastogi (1954), but he then follows the dubious convention of labelling as f_oE what by Bibl's

definition would be f_oE, and of calling f_oE what would normally be considered a subsidiary ledge below E. It is of some interest, however, that he has observed these three frequencies, f_o, f_1 and f_2, with considerable regularity at Ahmedabad.

According to Skinner, Brown and Wright (1954) E_2 at the equatorial station of Ibadan forms as a separate layer in the early morning and rises to form a ridge in F_1; and in the late afternoon this ridge breaks away from F_1 and descends towards the E-region. The downward movement of E_2 from the F-region is also reported by Saha and Ray (1955) for a station near Calcutta.

Instead of the usual P'-f and P'-t records obtained from the ionospheric recorder it is also possible to obtain records of echo-amplitude against time or frequency, and this method has been applied with some ingenuity by Briggs (1951) to the investigation of E_s or abnormal-E, and of subsidiary "ledges" in the E-region. Regarding the abnormal-E phenomena he notes that there have been two theories put forward: the "scattering cloud" theory (e.g. Appleton, Naismith and Ingram, 1939) and the "thin layer" theory (e.g. Rawer, 1939). According to the first theory reflection from the E-region at frequencies above the normal penetration frequency is due to reflections from clouds or patches of higher density embedded in the region, and there is

an overlap between the minimum frequency which can penetrate the main or interstitial electron density and the maximum frequency which can be reflected by the patches. The cases which Briggs finds to be well explained by this theory show a considerable overlap (of the order of 0.5 Mc/s or more), a gradual "trailing" cut-off and the rapid fading to be expected from a patchy or "rough" layer. The alternative theory (Rawer, 1939 and Rydbeck, 1944) invokes the partial reflections which occur, as a full wave treatment shows, when the thickness of the layer becomes comparable with the wavelength used. The cases mentioned above cannot be explained satisfactorily on this basis since the implied layer-thickness would be too small to be maintained against diffusion. Briggs has however observed cases of a small overlap (indicating a semi-thickness of 2 to 3 km), associated with the slow fading of a smooth layer: and these are more readily accounted for in terms of the thin-layer theory.

Briggs also notes the presence of "ledges" in the E-region which show up markedly as absorption bands on the amplitude-frequency records. He finds that the penetration frequency associated with these ledges is always <u>decreasing</u> in frequency at whatever time of the day they may be observed, the average rate of decrease being about 10 kc/s per minute (or 0.6 Mc/s per hour).

The presence of a distinct layer of E_s type below

the normal E-layer has been observed by Naismith (1954) and he calls it the meteoric-E or E_m layer. It occurs at heights between 90 and 97 km, remains at a constant height on any one day, and shows no group-retardation effects. The echoes are intermittent in character either on P'-f or P'-t records and with sufficiently high power, reflections have been obtained at frequencies as high as 73 Mc/s. E_m echoes occur in about 85% of records but the rate of occurrence does show a marked minimum around 22 h and a minor maximum around 04 h; otherwise its diurnal variation is relatively flat. In most cases E_m reflections occur at higher frequencies than any other E_s reflections so that, it is stated, the top E_s frequency is really as a rule fE_m.

A different type of study which may also be relevant to our problem is that of Jones (1955) who examines the results of a series of phase-height measurements at 150 kc/s. The total height-variation throughout the day of the 150 kc/s reflection level is, he finds, much less than would be predicted from Chapman theory, the curve being anomalously flat across the greater part of the day.

4.2 Objects of the Slough measurements.

The primary aims of the measurements to be described here are the confirmation and clarification of the anomalies in fE described in Section 3. All

these anomalies were found using smoothed curves derived from a very large number of raw data, which have of course the advantage of the generality gained by using statistical averages; but it is not as a rule possible to exhibit these anomalies using routine data for single days, or even using the published monthly medians, taking one month at a time. The reason for this may lie partly in the inherent variability of the ionosphere, but undoubtedly the main reason is that the magnitude of the anomalies is near to, or less than, the experimental error in the routine measurements, which is at least 0.1 Mc/s in fE and 10 km in h'E.

It would therefore be of some interest to make more accurate measurements of fE, seeking evidence of the anomalies from data, if not of single days, at least of small groups of days. Not only would this be useful confirmation of the reality of the anomalies; but it would also yield independent, and possibly better, estimates of their magnitude. Moreover in using routine data there is always the possibility that anomalies arise from consistent errors in the reduction. It would therefore be of great value to develop a more accurate and consistent method of reduction which would eliminate errors of this nature.

We have seen that anomalies have also been observed in h'E, and it is therefore also desirable to examine

the variation of this parameter with increased accuracy on a few days; and what might be of still greater value would be to obtain variations of h_m, since this is the quantity whose theoretical behaviour can be predicted most precisely.

It has been customary, in F-layer measurements, to record not only h_m but also y_m the apparent semi-thickness of the layer assuming it to have a parabolic shape in the first approximation, and indeed the determination of h_m is also dependent on this parabolic approximation. The method has not however been generally adopted for the E-layer partly because the thickness of the E-layer is of less practical importance and partly because the requisite measurements are more difficult, though some isolated determinations of y_m have been made on a number of occasions. Thus Appleton (1937) gives the value of 23 km for y_m, but with no corrections made for the effect of the earth's field; and Pfister (1950) finds y_m = 18 km without correction, and 15 km with correction for the earth's field.

No attempt has previously been made however to study the diurnal variation of y_m. This would be of particular interest both in relation to the alleged tidal movements (which may well effect the shape of the layer) and also in relation to suggestions that there is a considerable variation of the scale-height

if established, could, it will be recalled, explain the anomalously high exponent of $\cos\chi$ in the fE-$\cos\chi$ relation.

Since E_s and other forms of ledging can affect the apparent value of fE, a further question to be investigated is whether any of the anomalies in question can be explained in this way.

In addition to these queries relating directly to certain specific anomalies, sets of records in which attention has been specially concentrated on the E-region may give useful information regarding other problems raised by previous experimental work connected with this region. For example it may help to answer the question whether there is a true E-layer and if so how fE_2 is to be measured. Also it should be possible to consider in detail what types of E_s are observed and what regularities, if any, they exhibit.

To sum up then, the object of the measurements is to obtain as accurate and detailed information as possible on the diurnal variation of E-region characteristics at the latitude of Slough, including in particular the variation of the height of maximum electron density h_m and the effective semi-thickness y_m, and with the specific aim of seeking more information on, and clearer understanding of the E-layer anomalies already found in routine data. These aims entail a more accurate technique both in the primary measurements and

also in the subsequent analysis of these data. We shall consider these two problems in the following sections.

4.3 General description of the experiments.

The ionospheric measurements to be described in this chapter were carried out at the Radio Research Station, Slough, the first sequence lasting 4 days at the end of June 1954, and the second lasting 8 days at the end of September 1955.* A standard Mark II D.S.I.R. automatic ionospheric height-recorder was used, but with the time-base sweep-velocity increased considerably above normal so that the record covered only 300 to 400 km of equivalent height instead of the usual 1000 km or so covered in routine records. Examples of records taken in this way are shown in Fig. 13 (from the paper by Lyon and Moorat, 1956). This expansion of the height-scale, which was possible only because the measurements were confined to E-region characteristics, permitted direct readings of apparent virtual height to the nearest km. (The reason for the qualification "apparent" will be explained below.) The heights were measured with the aid of a transparent scale carefully ruled with parallel lines at 10 km intervals.

---

*Actually, 28th, 29th June and 1st, 2nd July, 1954, and 23rd to 30th September (inclusive) 1955.

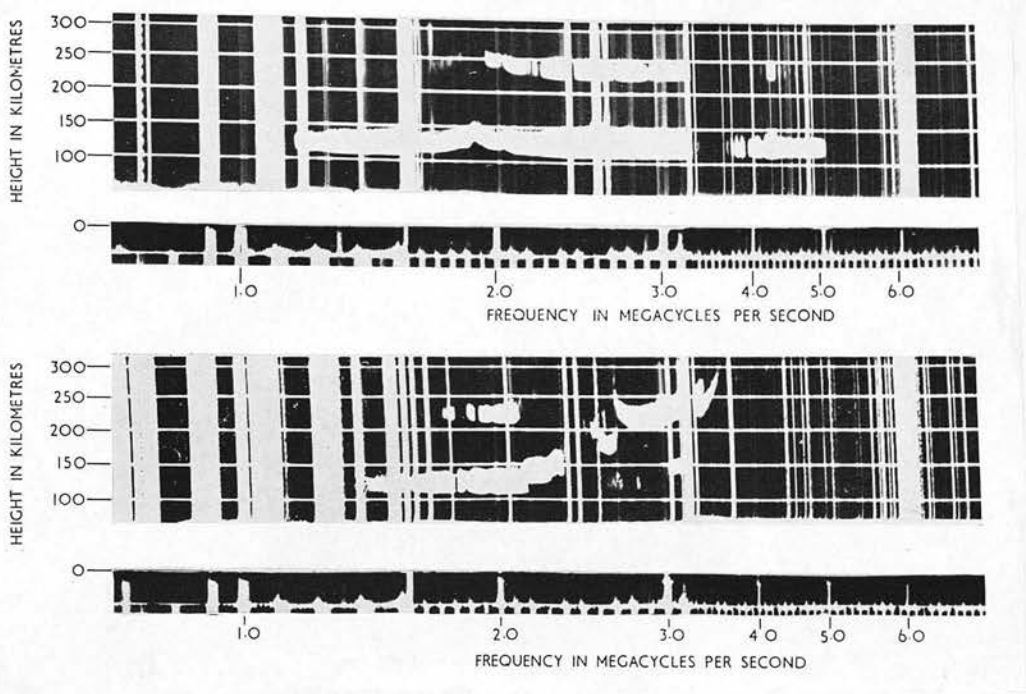


Fig. 13

Records were taken every quarter of an hour during daylight, and considerable care was taken with the timing. The clock time, which is photographed on the record by a flash occurring just before the beginning of the first (Band I) frequency sweep, can be read to the nearest 10 sec., and the various corrections were determined to the same order of accuracy. The recorder clock was checked daily against Greenwich Mean Time and the appropriate correction noted. Since accurate comparison with theory requires that the time used should be true (or in astronomical usage "apparent") solar time,

corrections were also necessary for the "equation of time" and for the longitude of the station. Further, since the time required is that at which the critical frequency fE is recorded, the time elapsing between the clock flash and the recording of the critical frequency must also be allowed for. (This is readily estimated from the fact that on this recorder each frequency band is swept in exactly one minute.)

Although an accuracy of the order of 10 seconds is not in fact required for the present measurements, nevertheless the asymmetries being studied correspond to time differences as small as 5 or 10 minutes, and hence it is clearly essential that times should be correct at least to the nearest minute, and somewhat better if possible.

4.4 Calibration of the Recorder.

In the routine reduction of ionospheric records it is customary to measure the equivalent-height for any frequency at the lower edge of the trace, and this is no doubt sufficient for the level of accuracy usually required. When however greater accuracy is sought allowance must be made for irregularities introduced by random fading.

Whale (1951) considered this question, noting that the bottom of the trace varied with echo-amplitude, and decided that the best procedure was to use the mid-

point of the recorded trace to give at any rate the relative variation of equivalent-height with frequency. Lyon and Moorat (1956) have however re-examined the matter and give a more accurate method of correction which makes it possible to obtain not only the relative variation but also the absolute values of the virtual heights.

In the ideal case of a perfectly square transmitted pulse which remains perfectly square throughout its whole journey to the ionosphere and back again, and through the receiver circuits to the cathode-ray tube of the recording apparatus, there would of course be no variation of apparent height with echo-amplitude. In practice however the pulse at the receiver output (which may be studied by a monitor C.R.T.) has a shape similar to that illustrated in Fig. 14.

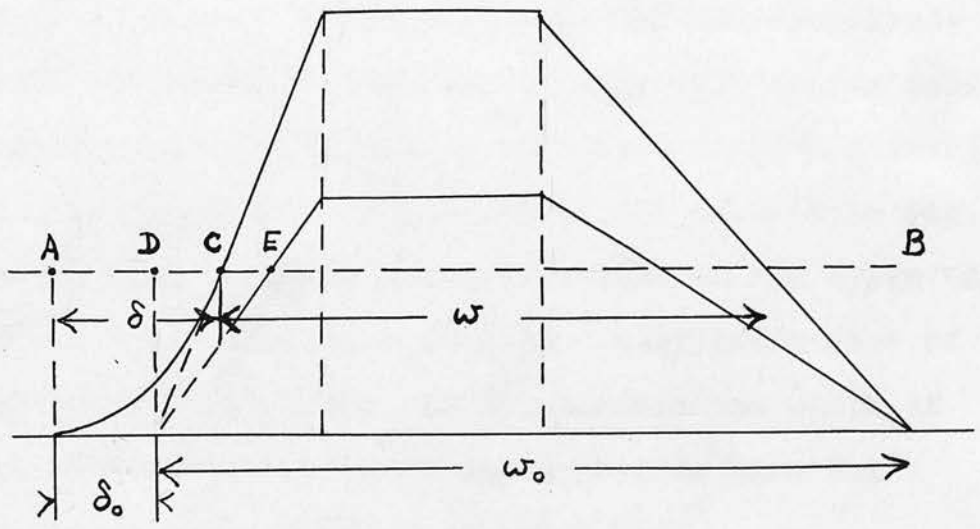


Fig. 14

Moreover in the actual design of an ionospheric recorder there is necessarily a certain threshold level of echo-amplitude (indicated by AB in the figure) below which echoes are not visible on the record. The lower edge of the recorded trace corresponds to the point C, so that there is an error of amount AC, equal to δ, say, in the recorded height. This error consists of two parts, an initial error AD, equal to δ_0, and a part DC depending on the slope of the leading edge of the pulse.

So far as distortion in the receiver is concerned the elementary theory of pulse transmission in nearly loss-free circuits shows that the rise-time and decay-time are constant, the former being given by $1/\Delta f$, where Δf is the "band-pass" of the circuit in cycles/sec. If the condition of constant times of rise and decay is satisfied, clearly when the echo-amplitude falls the slope of the leading edge will become less and the recorded height at the fixed threshold level will be <u>greater</u> - for example at the point E in Fig. 14. At the same time the recorded height of the upper edge of the trace (corresponding to the trailing edge of the pulse) will fall: in other words the width of the echo trace on the record decreases when the amplitude of the echo decreases.

It is not difficult to confirm experimentally

that when a pulse of constant width but variable amplitude is passed through the receiver of the recorder then the width of the recorded trace varies when the amplitude varies: nor is it difficult to determine the magnitude of the variation. In fact this provides us with a means of calibrating the instrument.

The method, described by Lyon and Moorat (1956), simply consists in the attenuation of the transmitter pulse by using a dummy aerial in place of the normal transmitter aerial, so that when picked up by the receiver the pulse simulates an ionospheric echo. The bottom of the trace produced by this pulse should of course coincide with the zero height-marker (which is triggered by the same pulse as sets off the transmitter pulse). In fact however, and as the above argument predicts, the bottom of the trace is always above the zero level because of the combined action of pulse distortion and the existence of a threshold level for a visible echo.

The amplitude of the "dummy echo" can be varied by altering the attenuator controls and again as predicted the magnitude of the error increases as the amplitude decreases. From records obtained in this way it is possible to plot the variation of the error in the recorded height with trace-width, at any rate in so far as this error arises from receiver pulse-

distortion or any asynchronism in the triggering arrangements. A typical calibration record is shown in Fig. 15, from the above-mentioned paper.

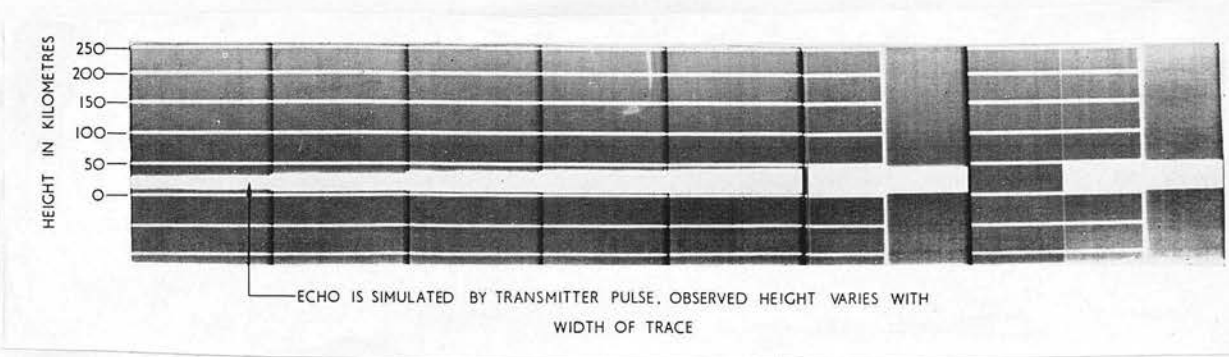


Fig. 15

The results show that the variation is very closely linear - as it obviously should be from the geometry of Fig. 14, if the rise-time and decay-time are constant, or indeed provided that they maintain a constant ratio. The relation between the error δ (which represents a correction to be deducted from the observed height) and the pulse width $\underline{w}$ may thus be written

$$\delta = a - b\,w, \qquad (79)$$

or in terms of Fig. 14

$$\delta = \delta_0 + \frac{w_0 - w}{\beta + 1} \qquad (79a)$$

where β is the ratio of decay-time to rise-time and the meaning of w_0 is indicated in the figure.

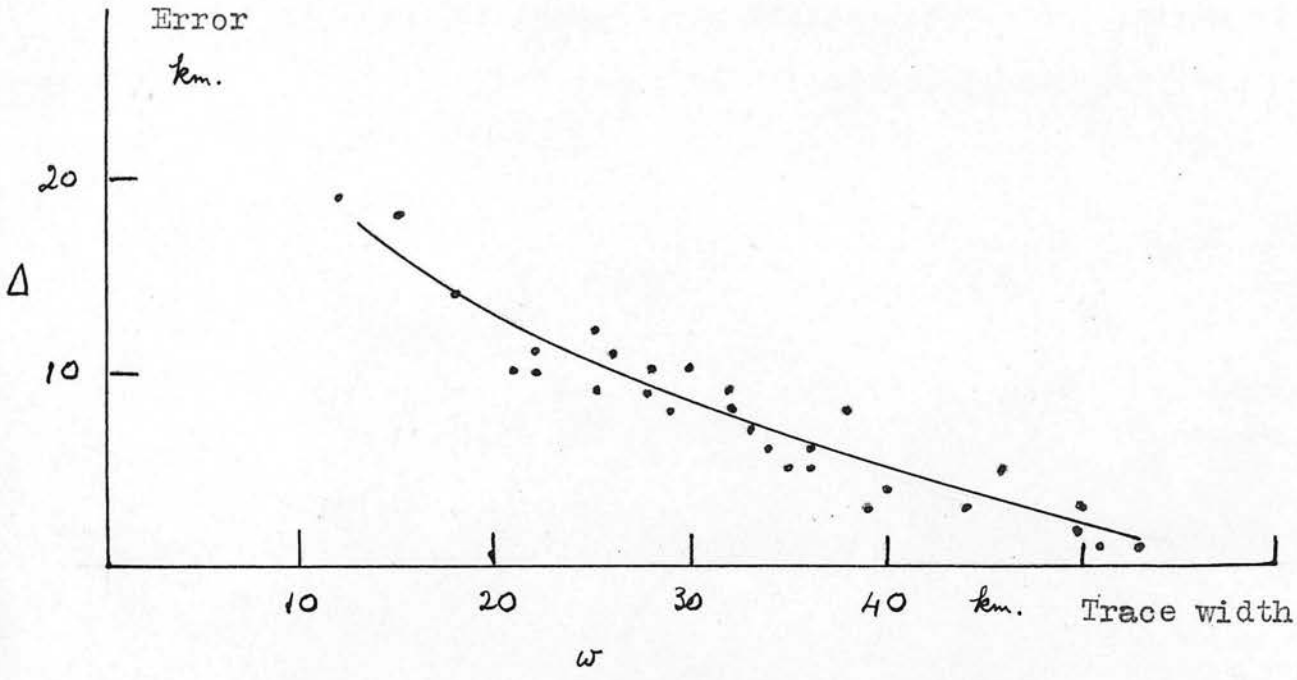


Fig. 16 Calibration curve for ionospheric recorder.

For the D.S.I.R recorder used in the present measurements a was found in the first calibration (June 1954) to be of the order of 15 km and b about $\frac{1}{4}$, so that $\beta = 3$; in other words, the slope of the leading edge is three times that of the trailing edge. For the widths usually encountered this means errors varying between about 4 and 11 km. Such errors are clearly far from negligible in E-region measurements if anything but the lowest level of accuracy is called for.

A slight change in the calibration was found to be necessary for the second set of measurements (September 1955). In this case the curve shown in Fig. 16 gives a better fit to the experimental points than the best-fitting straight line, and accordingly this was used as a calibration curve for the September results.

It remains now to check whether such corrections measured on artificial echoes are applicable to true ionospheric echoes. One method of doing this is to measure the variation of the height of the lower edge of the trace, and of its width, during random fading whilst the true height remains constant (e.g. well below the critical frequency, or on an E_s trace). If these two quantities are plotted (height against width) the slope of the best straight line between the points should equal the slope b in equation (79) - provided that the above-mentioned assumption of constant true

height is valid. A test of this kind has given a value of b very close to that obtained in the calibration measurements described above.

Another check employed by Lyon and Moorat was to compare echo heights obtained with the more accurate technique used in absorption measurements with simultaneous measurements using the standard recorder, but corrected in accordance with the calibration described above. The values agreed in all cases within the accuracy of the measurements, that is, within about 2 km.

Still another test of the calibration makes use of multiple echoes. When the time base is used at its normal setting (to show some 1000 km or so of virtual height) as many as 7 or 8 multiple echoes are sometimes observed. If no corrections were required we should have

$$h_n' = n\, h_1'$$

where h_n' is the measured virtual height of the n^{th} echo. If this equation is applied directly very considerable errors are found, but if corrections are made in accordance with the calibration, these errors disappear, i.e. we have

$$h_n' - (a - bw_n) = n\, h_1' - (a - bw_1)$$

w_n being the width of the n^{th} echo. This equation may

be written

$$\frac{nh_1' - h_n'}{n-1} = a - b\,\frac{nw_1 - w_n}{n-1} \qquad (80)$$

and hence it is possible by plotting $\frac{nh_1' - h_n'}{n-1}$ against $\frac{nw_1 - w_n}{n-1}$ to determine the values of $\underline{a}$ and $\underline{b}$ from multiple echo measurements alone.

Each instrument must of course be calibrated separately at a series of frequencies in the range to be covered in the measurements. There is no need however to use the above confirmatory tests in each individual case. The point of these subsidiary tests is rather to test the applicability of the method of calibration in general. It may be noted that there is one important case, or group of cases, when the calibration is not valid, namely when the echo is complex, i.e. a combination of echoes from different but neighbouring heights, or of echoes of different types (i.e. ordinary and extra-ordinary), or of one main echo and of scattered echoes spread continuously over some range of apparent height. In all these cases increased width of the trace does not indicate increased amplitude but the addition of echoes from greater (or less) effective heights and obviously the corrections considered here are irrelevant. With experience in the interpretation of the records most cases of this kind can be readily diagnosed, though

there probably remains a small residue of cases where it is impossible to tell from the trace that complex echoes are present, and some error in such cases is inevitable. The corrections we have described do however undoubtedly remove the major source of error in the equivalent-height measurements.

4.5 Inversion of the equivalent-height v. frequency curve.

Having seen how we may derive from the records reasonably accurate h'-f curves, we now require a convenient and reliable method of analysing these curves for information regarding the ionization-density and its distribution within the layer.

According to the elementary theory of electromagnetic propagation in an electron plasma the refractive index at any point in the ionosphere is given by

$$\mu = (1 - N/kf^2)^{\frac{1}{2}} \qquad (81)$$

f being the frequency of the incident wave*, N the electron density at height z, and k being given by

$$k = m/4\pi e^2 \qquad (81a)$$

where m and e are the mass and charge of an electron. The group velocity v is then

$$v = c\mu$$

---

* In rad./sec.

where $\underline{c}$ is velocity of light; and the virtual height is

$$h'(f) = \int_0^{\underline{h}} \frac{dz}{\mu} = \int_0^{\underline{h}} \frac{dz}{(1 - N/kf^2)^{\frac{1}{2}}} \qquad (82)$$

where $\underline{h}$ is the maximum true height reached by the wave of frequency $\underline{f}$, being the value of $\underline{z}$ for which $\mu = 0$, i.e. for which

$$N(z) = kf^2.$$

The ionospheric record gives h' as a function of $\underline{f}$ and hence of N. The problem therefore is to invert the integral (82) so that it gives a solution for the true height $\underline{h}$ of a given N, i.e. for the distribution of electron-density with true height. The solution, first given by Appleton (1930), is

$$h = \frac{2}{\pi} \int_0^{\underline{f}} \frac{h'\, dp}{(f^2 - p^2)^{\frac{1}{2}}} + h_0 \qquad (83)$$

h_0 being the true height of the bottom of the layer and $\underline{p}$ being used to denote frequency when it is an integration parameter; or by a simple transformation

$$h = \frac{2}{\pi} \int_0^{\pi/2} h'(f \sin \theta)\, d\theta.$$

As pointed out by Kelso (1952) this yields the very simple numerical quadrature formula

$$h = \frac{1}{n} \sum_{k=1}^{n} h'(f \sin \theta_k) + h_o \qquad (84)$$

where $\theta_k = \frac{2k - 1}{2n}$.

4.5.1 Inclusion of the magnetic field: the Shinn-Kelso method.

The above equations however neglect the presence of the earth's magnetic field. When the effect of this field is included the equation for the refractive index is given by the complete magneto-ionic formula (Appleton, 1932) and a simple inversion of the integral (82) is no longer possible. It has however been shown by Shinn (1956) that at any given place one can use a simple variant of the Kelso formula of the form

$$h = \frac{1}{n} \sum_{k=1}^{n} h'(a_k f) + h_o \qquad (85)$$

where the quantities a_k are functions of the angle of dip and of the ratio of the frequency f to the gyro-frequency f_H given* by

$$f_H = eH/2\pi mc \qquad (86).$$

The values of the constants a_k, computed by Shinn for the magnetic latitude of Slough and for n = 10, are reproduced in Table 1.

---

* In cycles/sec.

Table 1

Coefficients for the Shinn-Kelso Method ($a_k \times 10^3$)

(Ten-point rule)

Station: Slough. Assumed dip: 67^o

Assumed field: 0.48 oersted.

f	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
1.7 Mc/s	072	209	338	454	560	658	750	836	915	983
1.7-2.4 Mc/s	075	218	350	468	574	672	762	844	919	983
2.4-3.4 Mc/s	077	226	361	481	589	686	774	854	924	984

$$h(f) = 1/10 \left\{ h'(a_1 f) + h'(a_2 f) + ... + h'(a_{10} f) \right\}$$

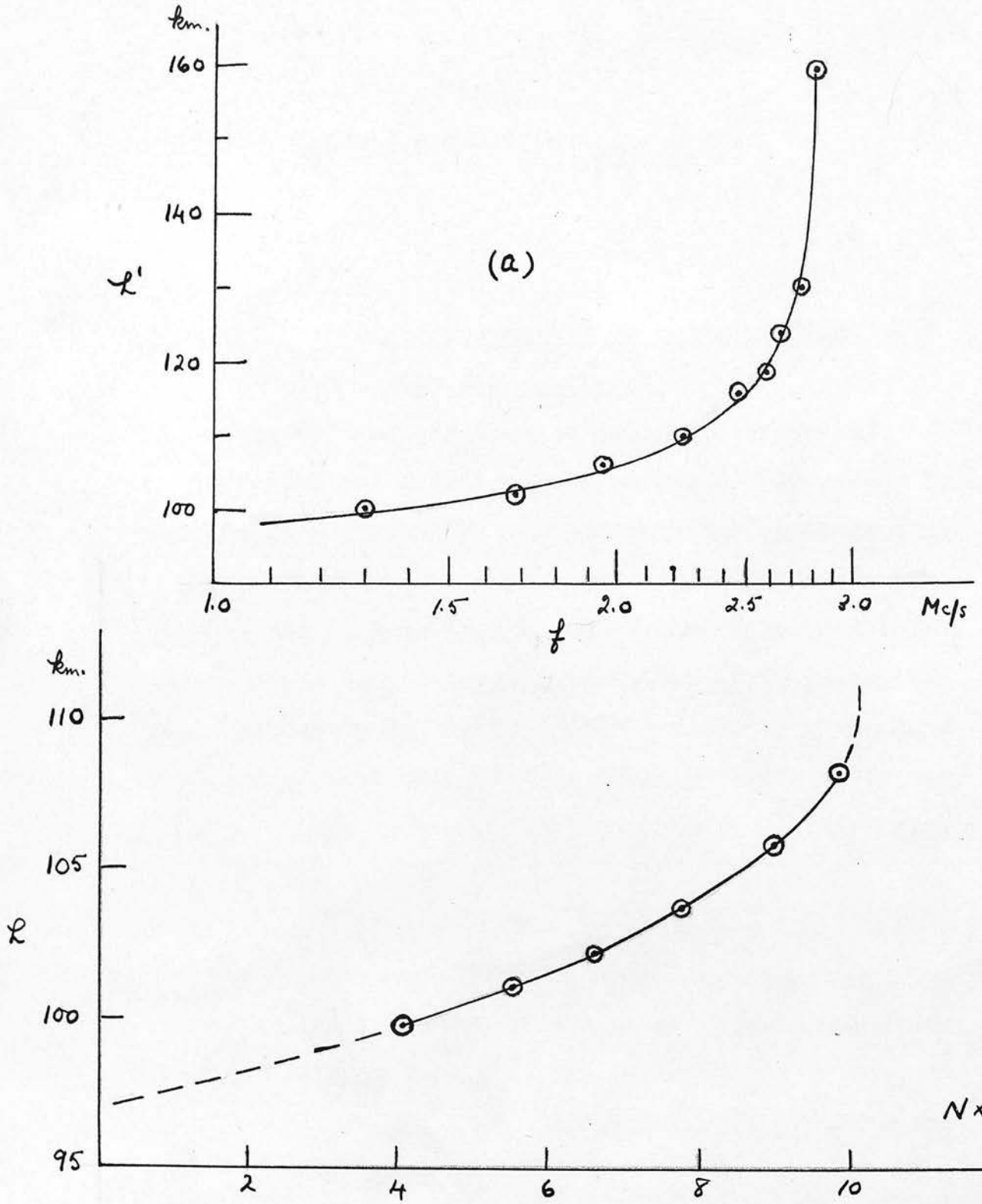


Fig. 17 (a) Points from an experimental h'-f curve, fitted to a theoretical curve for $y_m = 14$.

(b) N-h curve from the same data, derived by the Shinn-Kelso method. The dashed lines are continuations on the parabola indicated by the Shinn-Ratcliffe method.

This method, despite its simplicity and accuracy, has nevertheless two disadvantages in relation to the present experiment. Firstly the method breaks down at the critical frequency itself since the equivalent height becomes, in theory, infinite. This means that it does not give the variation near the maximum and hence to determine h_m (and so an effective y_m) a somewhat precarious extrapolation is required. This is illustrated in Fig. 17 where despite the ideal trace the position of the maximum is doubtful to the extent of a few kilometres. The difficulty becomes all the more serious the more truncated the normal-E trace i.e. the further from the critical frequency the highest visible echo. This is particularly unfortunate for the E-layer where a disappearance of echoes well before the critical frequency is the general rule. The second disadvantage of the method is that it requires knowledge of equivalent heights down to 100 kc/s or less, and in the present measurements these are not of course available, so that extrapolation to low frequencies introduces still another element of uncertainty in the final result.

4.5.2 Use of transparent sliders: the Shinn-Ratcliffe method.

An alternative approach used by Appleton and Beynon (1940), and also by Booker and Seaton (1940), is to assume that as a first approximation the distribution

of N with height obeys a parabolic law of the form

$$N = N_m \left\{ 1 - \frac{h_m - h}{y_m} \right\}^2 \qquad (87).$$

It is then readily shown that

$$h'(f) = h_o + \tfrac{1}{2} y_m \frac{f}{f_c} \ln \frac{1 + f/f_c}{1 - f/f_c}$$

$$= h_m + y_m \phi(f/f_c) \qquad (88)$$

where $\phi(x) = \frac{1}{2} x \ln \frac{1 + x}{1 - x} - 1.$

By plotting h' against $\phi(f/f_c)$ Appleton and Beynon obtained effective values of h_m and y_m.

A variation of the method suggested by Ratcliffe (1951) is to prepare standard graphs of h' against $\underline{f}$ on a transparent sheet for a series of values of y_m. If the frequency is on a logarithmic scale the same standard graphs will be valid for any critical frequency. y_m is determined by choosing the standard curve which best fits the experimental curve, and a fixed line on the transparent slider then gives h_m also.

Again of course this procedure neglects the earth's magnetic field, but Shinn (1953) has shown how it too may be suitably modified to allow for this field. It is now impossible to tabulate equivalent height h' against the ratio f/f_c as in the simple case, for h' is

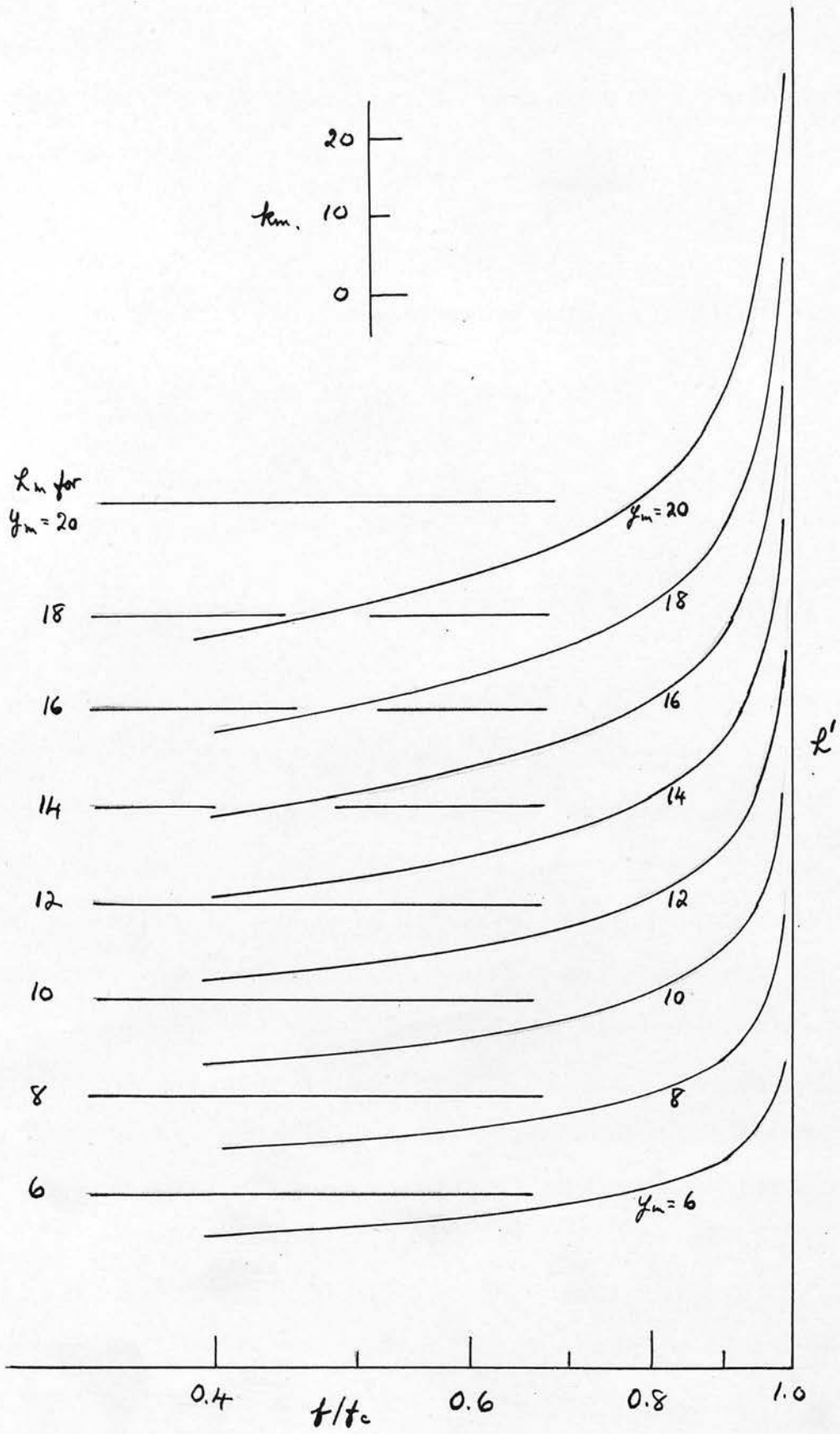


Fig. 18 Theoretical h'-f curves for parabolic layers at the magnetic latitude of Slough (from data by Shinn).

now also a function of the angle of dip δ and of the ratio f_H/f_c where f_H is the gyro-magnetic frequency. For the case of south-east England ($\sec \delta = 1.20$, $f_H = 1.149$ at 300 km) Shinn tabulates (in his Table 4) values of $(h' - h_o)/y_m$ against f/f_c for various values of f_c. This means that the shape of a theoretical h'-f curve varies with the critical frequency (because of its varying ratio to f_H the gyro-magnetic frequency). However for the range of frequencies encountered in the E-region (2.0 to 3.0 Mc/s in the experiments described here) this variation is in practice small, and if over this range we use one set of curves plotted from the last column (i.e. for $f_c = 2.66$ Mc/s), the resulting errors will be entirely negligible. Accordingly such curves (plotted for various values of the semi-thickness y_m) provide a suitable basis of comparison with experimental h'-f curves, provided the parabolic approximation is allowable. A set of these theoretical h'-f curves is shown in Fig. 18.

4.5.3 Comparison of parabolic and simple Chapman layer.

Expansion of the Chapman function gives

$$\exp \tfrac{1}{2}(1 - y - e^{-y}) = (1 - \frac{y^2}{4} + \ldots)$$

where y is measured in scale heights, so that near the layer maximum the apparent y_m is 2H. In consequence

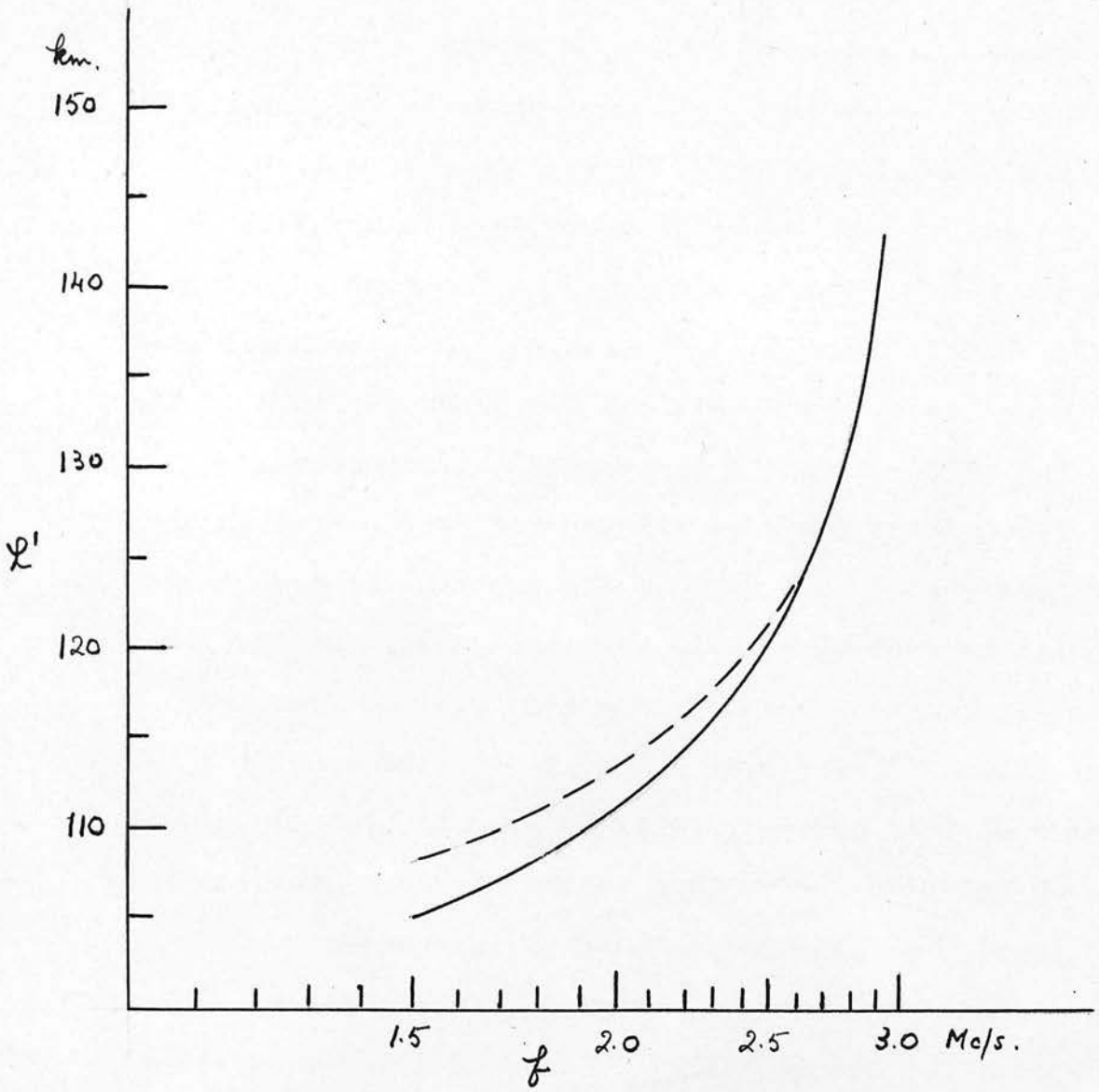


Fig. 19 Comparison of h'-f curves for a parabolic layer with y_m = 20 (full line) and for a simple Chapman layer with H = 10 km (dashed line). No magnetic field.

the value $\frac{1}{2}\ y_m$ is frequently used as giving an indication of the scale-height of the region.

The validity of a parabolic approximation to a simple Chapman layer as a whole has been considered by Pierce (1940) and by Jaeger (1940). Pierce uses an approximate method of calculating the h'-f curve for a Chapman layer with no magnetic field, and from his results concludes that if y_m is the apparent semi-thickness assuming a parabolic distribution (and using the Appleton-Beynon method) then the true scale-height of the Chapman distribution is $H = 0.6\ y_m$, and the true h_m is 0.14 H below the apparent.

Jaeger gives a more exact solution for the equivalent height for reflection in a simple Chapman region, and notes the existence of "large differences" between this and the parabolic case, but does not attempt a numerical comparison. Fig. 19 shows a curve of h' for a Chapman distribution with H = 10 km derived from Jaeger's data, and the corresponding curve for a parabolic distribution with y_m = 20 km. As we should expect the curves are indistinguishable near the maximum (for f/f_c greater than about 0.87) but for lower frequencies they diverge substantially. In fact the greater ionization at low heights in the Chapman distribution considerably increases the group retardation of the lower frequencies and consequently reduces the apparent semi-thickness and raises the

apparent h_m. In the present example comparison with simple no-field curves shows that a parabolic h'-f curve with y_m = 18 km fits as exactly as the eye can judge the whole range of the Chapman curve for which $f/f_c > \frac{1}{2}$. At the same time the apparent h_m has been raised by 3 km or 0.3 H - rather more than the amount given by Pierce.

It follows from this that a parabolic technique will yield an equivalent y_m and h_m for a Chapman distribution - or, we may assume, for any broadly similar variation - but the errors in these quantities may not be negligible. In particular any substantial variations in the amount of ionization in the "tail" of the layer may produce variations in apparent h_m and y_m even when the actual height of the maximum and the thickness of the main part of the layer remain constant.

4.5.4 Discussion of methods of h'-f analysis.

Despite the disadvantage mentioned above, the Shinn-Ratcliffe method - that is, fitting the layer shape to a parabola with the aid of a transparent slider - has a number of advantages.

In the first place it is quick and easy to use and can therefore be applied to a large number of experimental curves; and moreover it yields a set of simple and definite numerical parameters (fE, h_m and y_m). By contrast the Shinn-Kelso method - which in effect

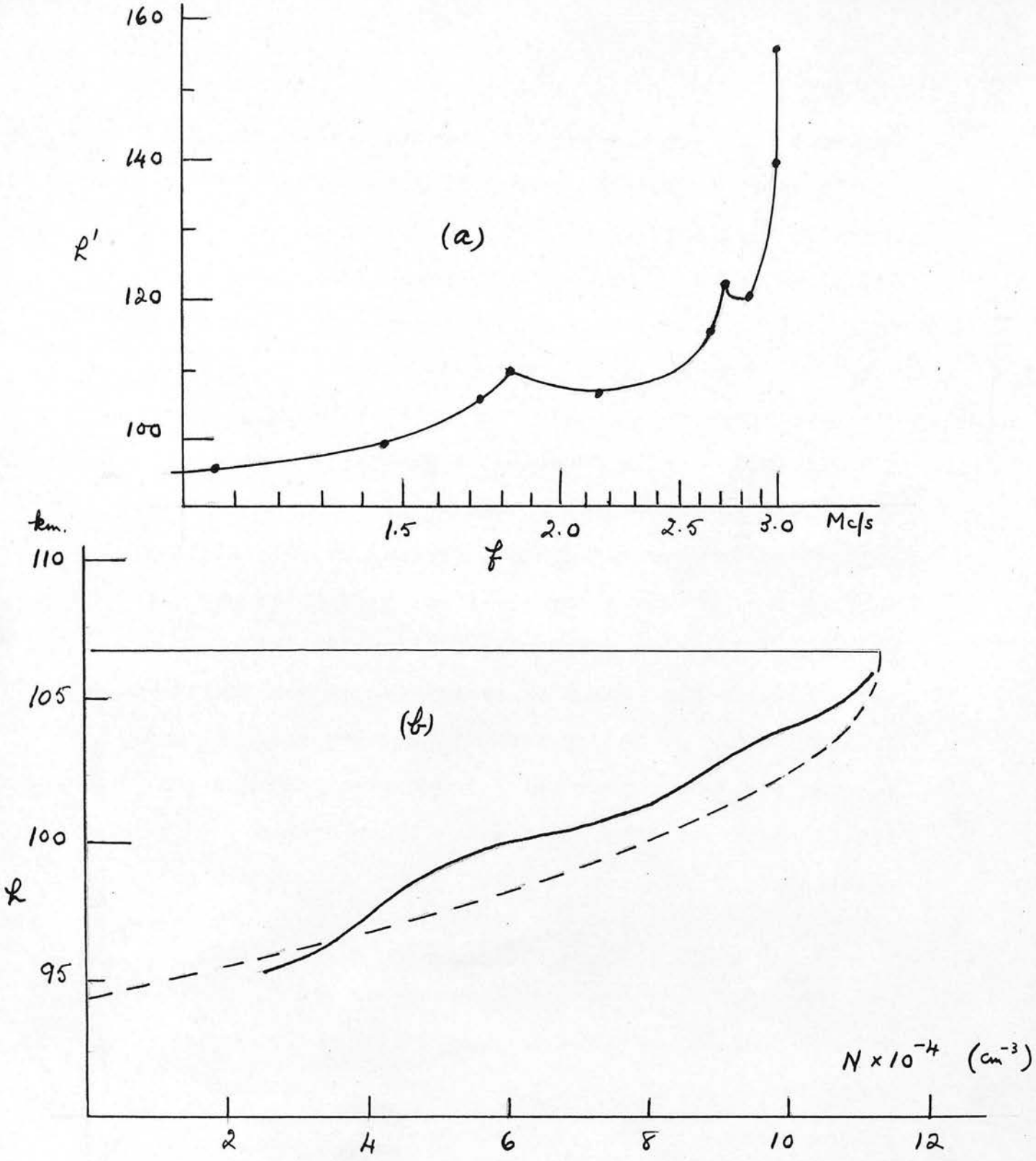


Fig. 20 (a) An experimental h'-f curve showing ledges.

(b) The resulting N-h curves using the Shinn-Kelso method (full line), and using the parabolic, or Shinn-Ratcliffe method (dashed line).

solves the problem by numerical integration - is much more laborious, and in the end yields another curve which still requires further analysis if comparisons are to be made.

In the second place the slider method provides a means of extrapolation to the critical frequency in those cases where extrapolation is necessary. Since theoretically the layer shape should be nearly parabolic at the maximum, the assumption of such a shape seems the most reasonable basis of extrapolation, and permits the reading of a much more accurate and reliable value than a simple visual estimate: indeed the technique increases considerably the number of cases where the reading of a critical frequency is possible.

It may be objected that complex and non-parabolic features are very common in the E-layer, and that therefore the meaning of a forced parabolic fit may be somewhat dubious; and there is perhaps some substance in this objection. Nevertheless if one wishes a first-approximation picture of the layer the parabolic method seems the most reasonable to adopt, and even if incorrect in detail should show up any major changes in the position or thickness of the layer.

Detailed comparison of the two methods in selected cases has confirmed the general correctness of this contention. For example, Fig. 20 shows a case where the h'-f curve has two quite marked "kinks".

Using the Shinn-Kelso method gives the N-h curve shown by the full line whereas fitting the best Shinn-Ratcliffe curve gives the parabola shown by the dashed line. It will be seen that the parabola gives quite a reasonable picture of the position and general distribution of ionization in the layer even in a case where irregularities are known to be present.

4.6 Experimental results.

We have seen that the most promising method for the investigation both of critical frequencies and of layer shape is the determination of the effective h_m and y_m of the best-fitting parabolic distribution using the transparent-slider technique. For this purpose a set of theoretical h'-f curves for various values of semi-thickness y_m (6, 8, 10 ... 20) were plotted on a transparent material (a sheet of "Koda-trace"), using the calculations of Shinn (1953) already mentioned. (The set of curves is shown in Fig. 18.) Since this method requires the experimental h'-f curve to have a logarithmic frequency-scale whereas the recorder used has linear frequency-scales, and since moreover the apparent height of the recorder trace requires correction for echo amplitude (Section 4.4), the first step in the reduction of the records is the careful plotting of h'-f curves on logarithmic graph paper making the appropriate corrections at the same time.

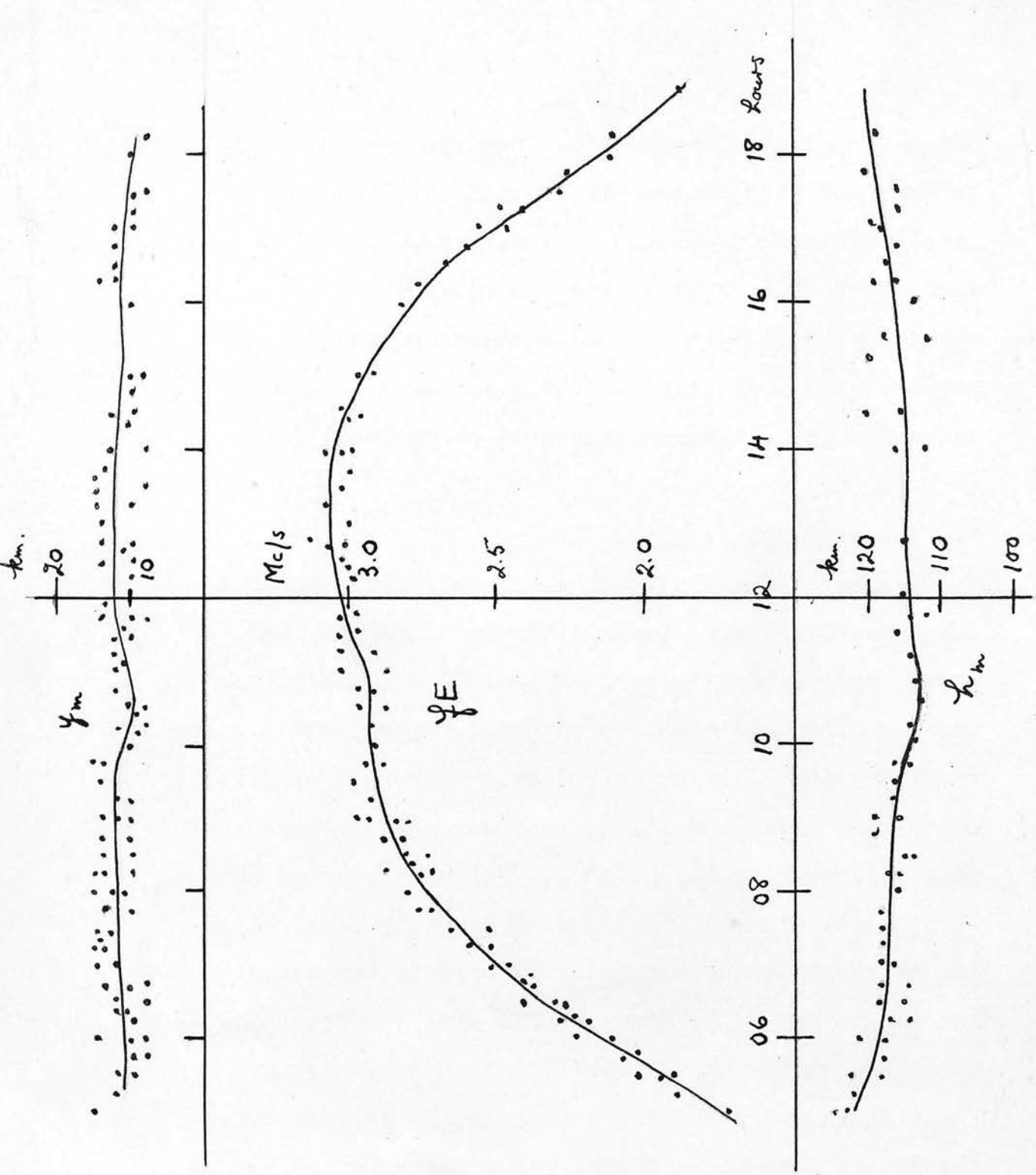


Fig. 21 Plots of fE, y_m and h_m at Slough on four days of June/July 1954.

The next step is to select the best-fitting theoretical curve with the aid of the transparent slider and to read off the corresponding values of y_m, h_m and fE. It was found that a reasonable fit could be obtained in a majority of cases.

4.6.1 The summer measurements.

Mass plots of the three quantities determined in this way are shown for the June 1954 measurements in Fig. 21. The chief conclusions to be drawn from these summer results are as follows:-

(1) There is a marked pre-noon dip in fE, in agreement with the results obtained from massed routine data. The dip lasts about 4 hours and has its maximum at about 1030 h. As far as can be judged by comparison with afternoon values the maximum magnitude of the dip is about 0.15 Mc/s i.e. some 5% of fE. This corresponds to a 10% decrease in electron density.

(2) The apparent semi-thickness y_m varies between 8 and 15 km having an average value of about 12 km - which corresponds roughly to a scale-height of 6 km. There is no indication of any increase in y_m for low $\cos\chi$ (and increased height): indeed if anything y_m tends to be slightly lower in the early morning and late afternoon. These results therefore lend no support to the view that the $\cos\chi$ exponent of about $\frac{1}{3}$ is due to a sharp increase of scale-height with height in the E-region.

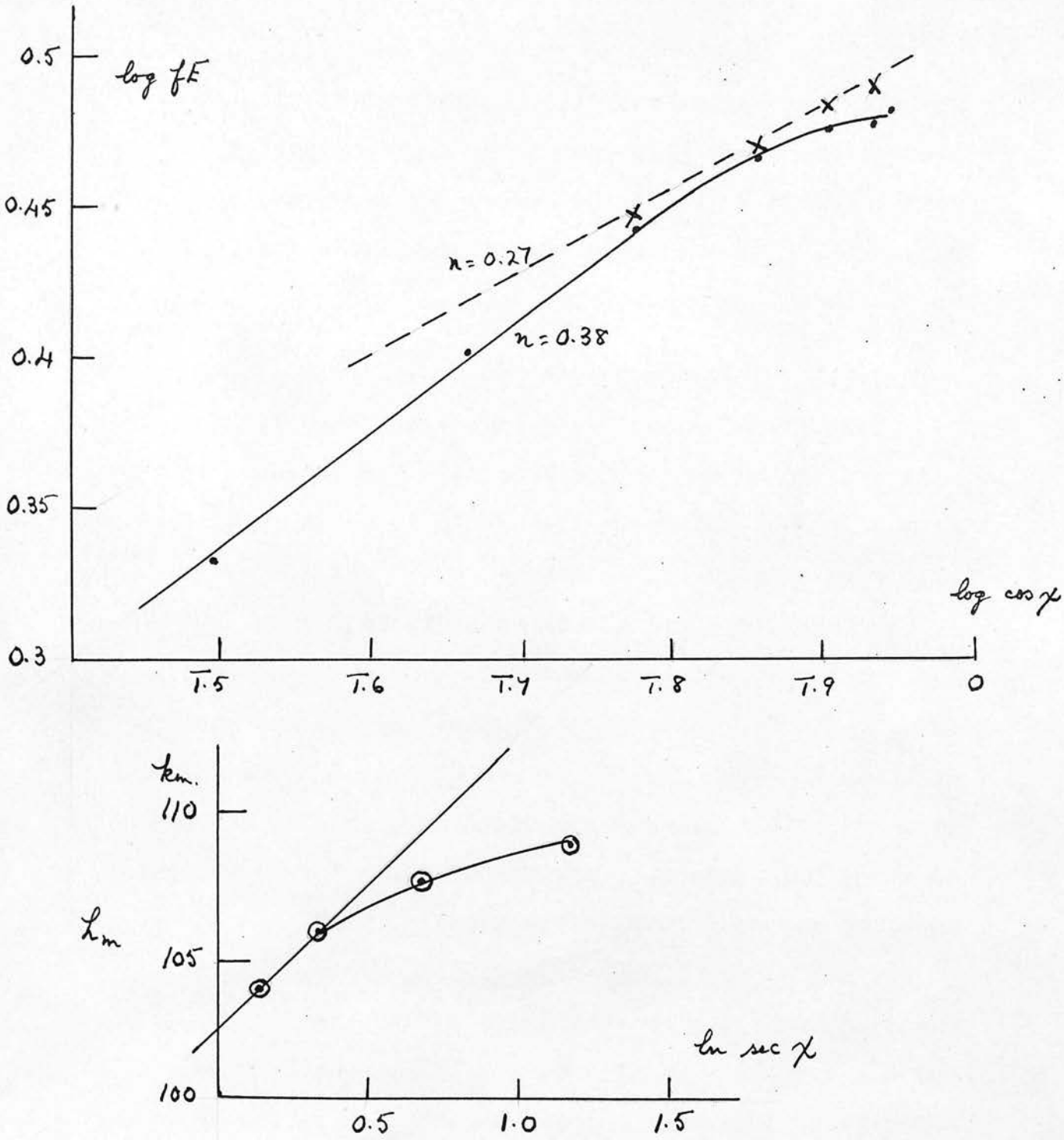
There seems however to be a slight dip in y_m at around 1030 h, coinciding with the dip in fE.

(3) The noon value of h_m is about 104 km - considerably lower than the figure of 120 km or so sometimes quoted. This arises partly from the width corrections and partly from the rather low semi-thickness. Again there is a dip at around 1030 h - of amount about 2 km.

(4) In Fig. 22 log $\overline{fE}$ is plotted against log cos χ, and it will be seen that the curve shows an inflection similar to that observed with routine data, and that the straight portion gives an exponent n = 0.38, a little higher that the values obtained from routine data. It is interesting to note that the values for the period 13 h to 16 h give a slope of about 0.27, very close to the Chapman value.

(5) The second curve in Fig. 22 is a plot of h_m against ln sec χ and shows a marked decrease in apparent H with sec χ. Assuming that h_m increases as H ln sec χ, the value of H between 09 and 12 h is about 10 km, but between 06 and 09 h it is about 4 km. In other words in the morning and evening the rise of h_m with sec χ is much less than Chapman theory would require.

(6) At low cos χ the diurnal curve of fE becomes very closely symmetrical i.e. we have $\Delta fE \rightarrow 0$ at low cos χ, again in agreement with the statistical results.



From the data of Fig. 21.

Fig. 22 (a) Variation of log fE with log cos χ

• represents mean of forenoon and afternoon.
X represents afternoon values only.

(b) Variation of h_m with ln sec χ

(Mean of forenoon and afternoon values.)

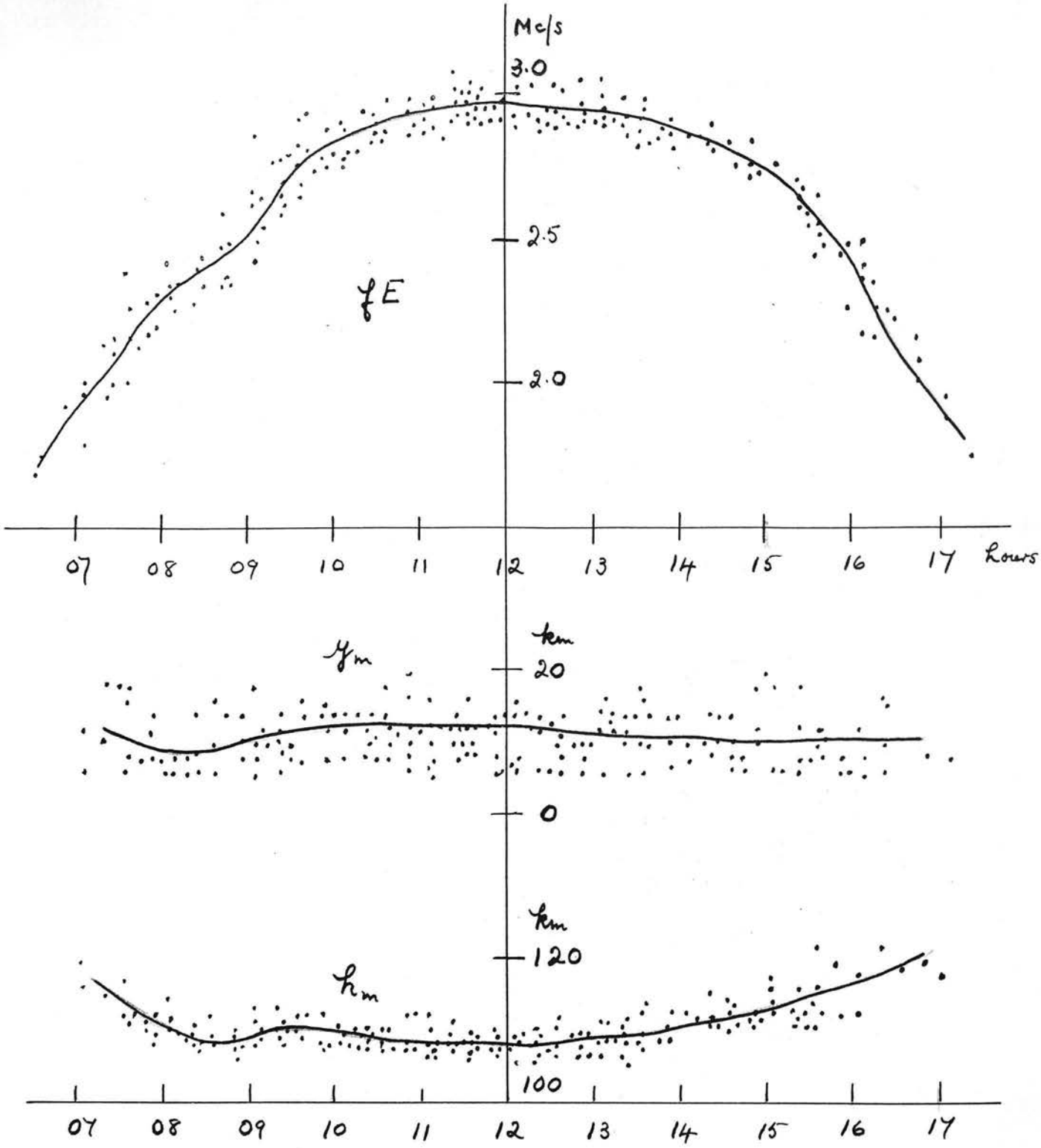


Fig. 23 Mass plots of fE, y_m and h_m for eight days in September, 1955, at Slough.

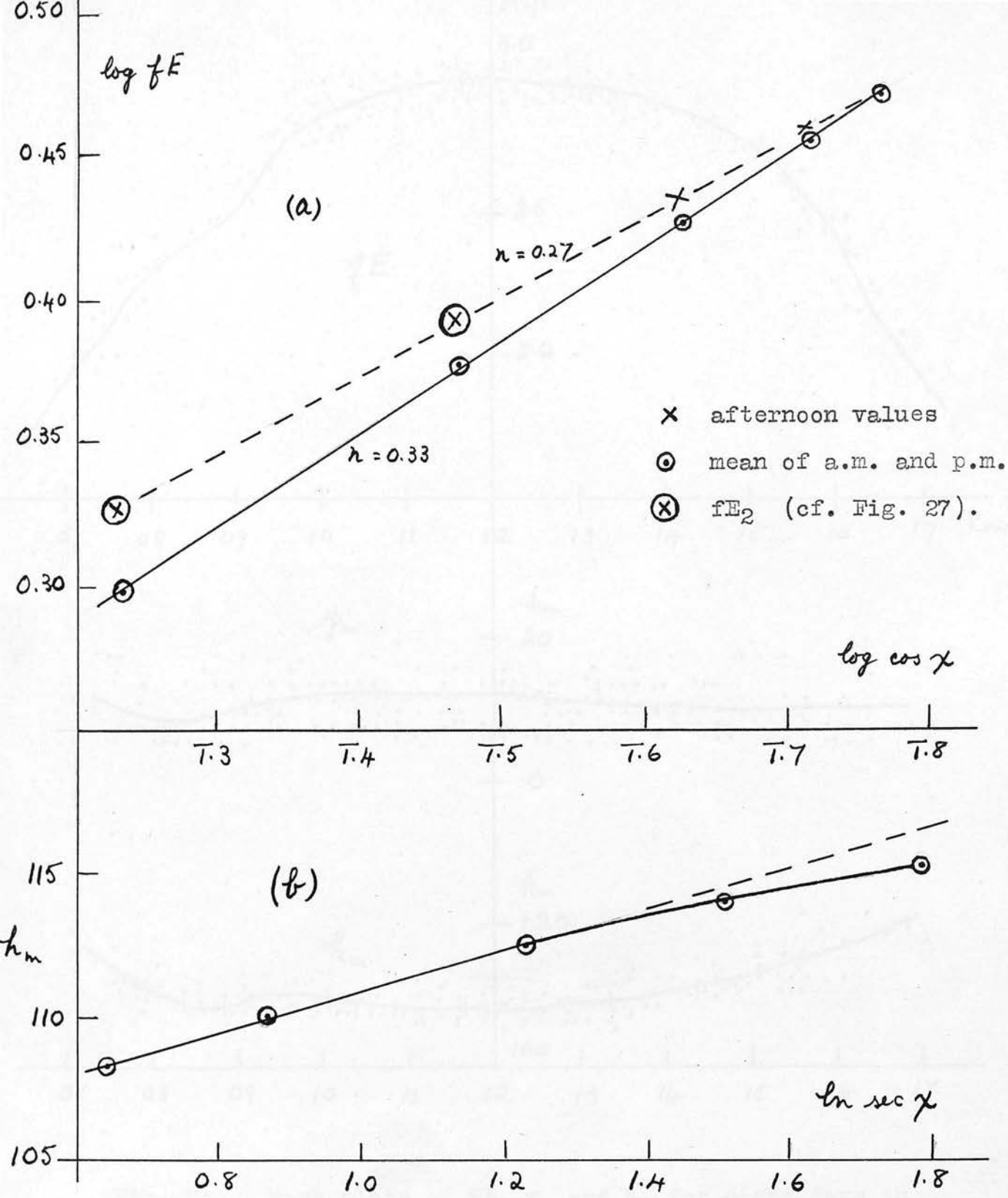


Fig. 24 (a) Variation of log fE with log cos χ

(b) Variation of h_m with ln sec χ

Values from the data of Fig. 23 in both cases.

4.6.2 The equinox measurements.

The results for September 1955 are shown in Fig. 23, and broadly they confirm the June findings:-

(1) There is again a pre-noon dip in fE though in this case its maximum occurs about 09 h - that is, about $1\frac{1}{2}$ hours earlier than in the June case and its magnitude is about 0.07 Mc/s, i.e. about half that in the June case. (This is in agreement with the reduced S_q current densities, and hence the reduced drifts to be expected at equinox.)

(2) The range of variation of y_m is somewhat greater - from 6 to 18 km mainly - and the mean value is 10.5 km. There is still no evidence of y_m increasing at low $\cos\chi$ but there again seems to be a dip of 2-3 km between 08 and 09 h, corresponding to the dip in fE. It may be noted that the best values of y_m (those giving a good fit over a wide range) are usually between 12 and 16 km.

(3) The noon value of h_m is about 108 km and there is a dip of at least 2 km around 0830 h.

(4) The log fE - log $\cos\chi$ plot (Fig. 24) gives on the straight part an exponent of 0.33, and again we find that the points for the period 12 to 15 h lie on a straight line giving n = 0.27, whereas after 15 h there is a sharp change in the slope.

(5) As before there is a decrease in the slope of the h_m-ln sec χ curve after 15 h, though a smaller one than

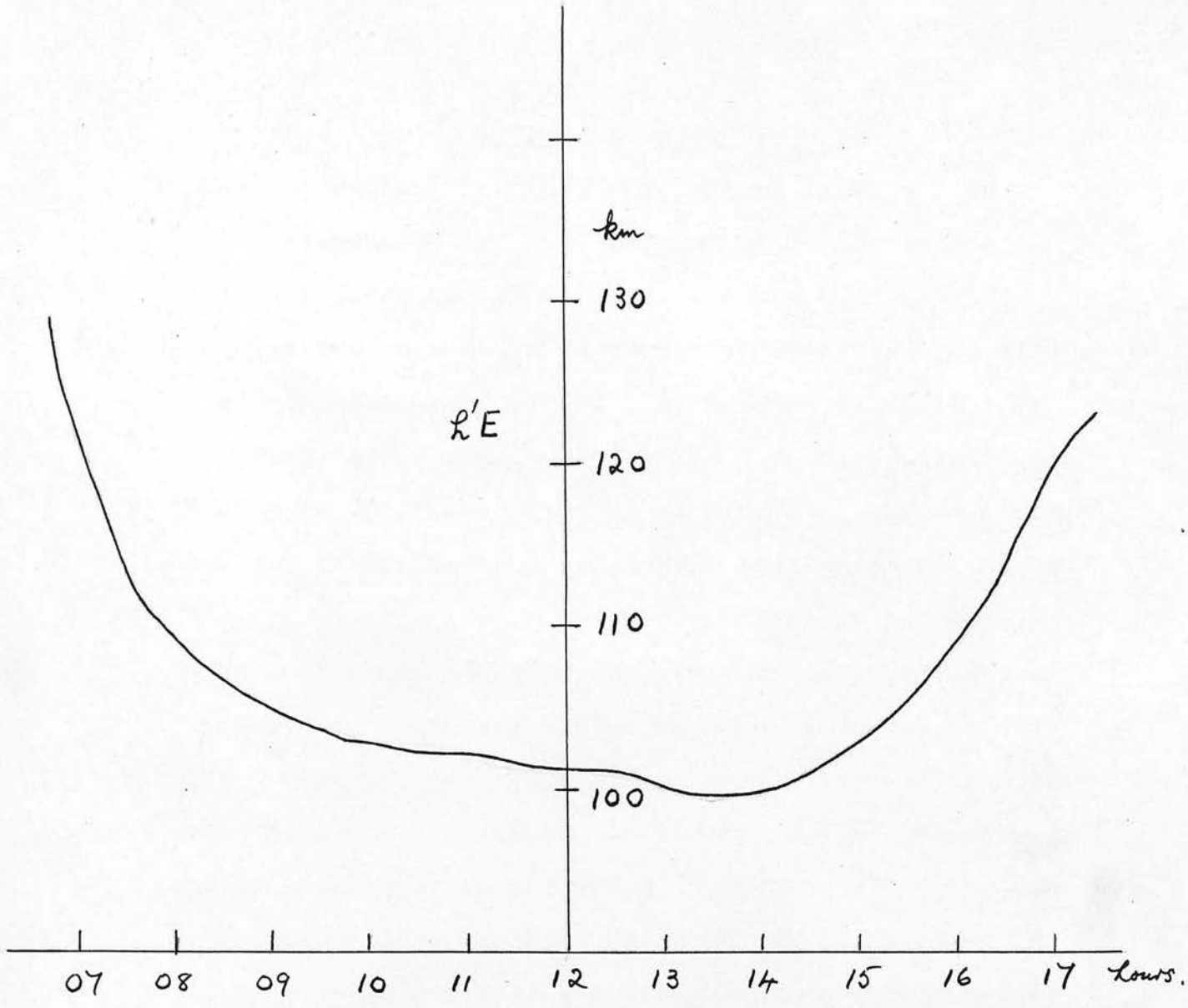


Fig. 25 Mean variation of h'E during the September sequence, 1955.

in the summer case - corresponding to a drop in H from $7\frac{1}{2}$ km to about 6 km.

(6) An additional quantity scaled in the preliminary reduction of the September data and relevant to the present discussion is h'E, measured at 1.6 Mc/s. (This frequency was chosen to allow a maximum number of values to be read.) The values are plotted in Fig. 25 and show the surprising result that between 0830 and 1100 h h'E at 1.6 Mc/s is consistently 2-3 km higher than at the corresponding times in the afternoon.

With routine data from Slough averaged over a number of years it has been shown (Appleton, Lyon and Turnbull, 1955) that there is a clear dip of about 1 km in h'E in the pre-noon period at all seasons; and there is a corresponding rise at the same time at Singapore. Our present result cannot therefore be typical. It may be that we have low afternoon values associated with the curious ledging phenomena discussed in Section 4.6.5.

(7) At low $\cos\chi$ we again have ΔfE tending to zero, or even becoming slightly negative.

In addition to the mass plots of Figs. 21 and 23 separate plots have been made of fE for each individual day, and in these all "ledge" and E_2 frequencies (whether cusps or discontinuities) have been included. Typical examples from the September sequence are shown in Fig. 26. From these more detailed results certain additional conclusions may be drawn:-

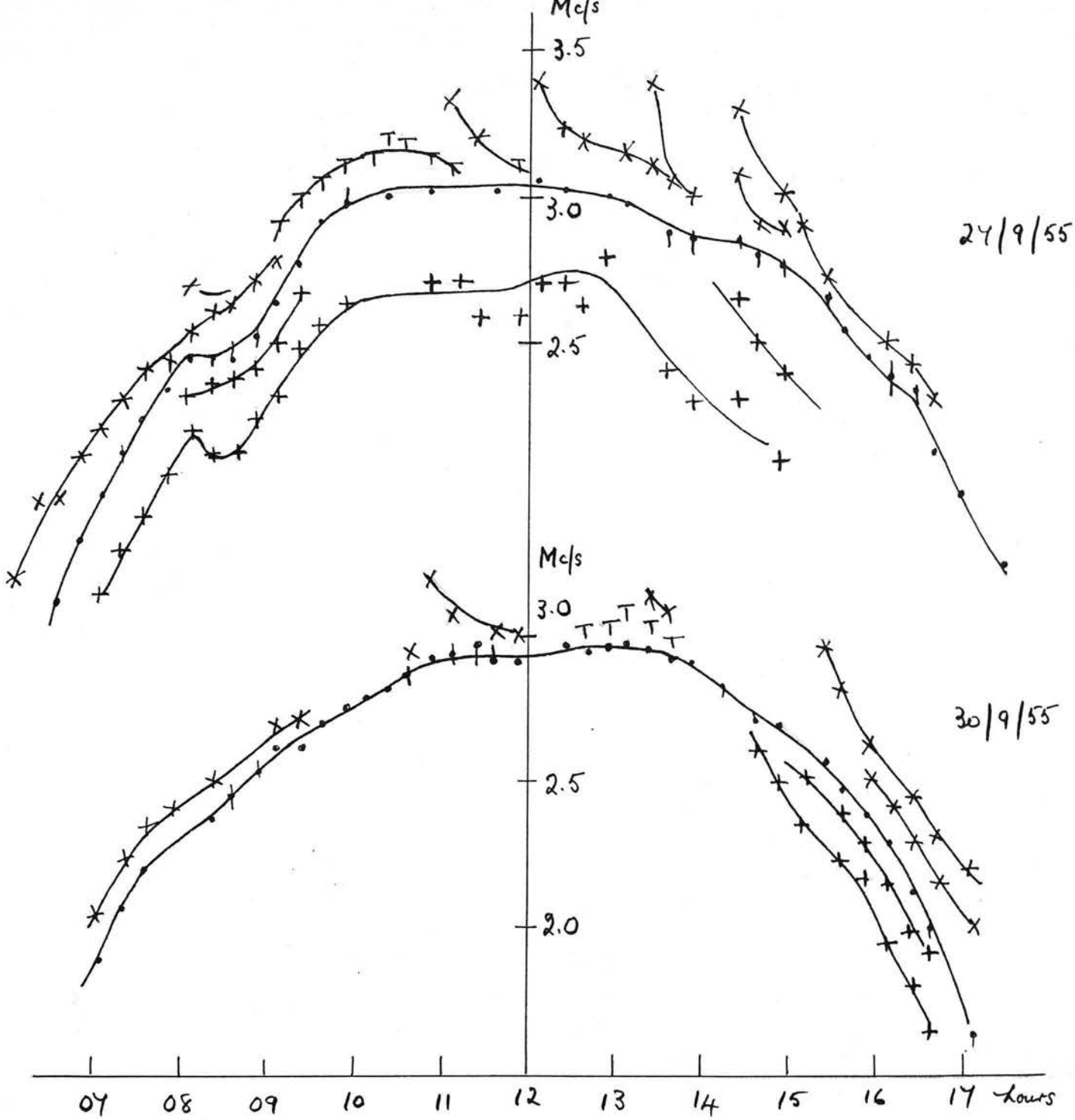


Fig. 26 Variations of fE, fE_2 and ledge frequencies on two days of the September sequence, 1955.

• fE × fE_2

\+ ledge frequency

T abnormal-E (as at b in Fig. 10).

(i) Stratification in one form or another appears to be a typical feature of the E-region: and frequently the series of ledges appears to be remarkably stable. (ii) The pre-noon dip is visible on most days not only in fE but also as a rule in all the attendant ledges. (iii) E_2 of the "dropping" or "sequential" type appears to be very common in the midday period, but in the morning and evening fE_2 tends on the contrary to be very stable, running closely parallel to fE.

4.6.3 The pre-noon perturbation.

We have seen (in Section 3.2.3) that the pre-noon perturbation of fE and h'E observed in routine results can be satisfactorily accounted for in terms of vertical drifts associated with the S_q current system. The special measurements have contributed to this work by confirming, as the foregoing sections have shown, that the perturbations can be observed on individual days, and also by showing that the predicted dip in h_m does in fact occur. The curious dip in y_m however would not be predicted in theory, as Fig. 4 shows, and its significance is not at present clear.

We can now make use of our theoretical results and with the aid of equations (65) and (68) we can determine the values of drift velocity and drift gradient required to produce the observed changes in h_m and N_m. Thus, according to these equations, if δh_m is 2 km $\underline{v}$ is about 4 m/sec.; and if δfE is 0.07 at 2.8 Mc/s $\partial v/\partial z$ is

about 0.1 m/sec./km, α being assumed 10^{-8} in both cases.

We can also make a quite independent estimate on the basis of equation (78) using the magnetic data to give us a value of j_y, the eastward current density. According to Chapman and Bartels (1940) the S_q current density at 50°N at equinox during sunspot minimum is 23,000 amp. per 10° of latitude, or 2×10^{-5} e.m.u./cm. If we assume the ions to be of molecular oxygen (and it will make little difference if ions of molecular nitrogen are also present) $m_i = 5.4 \times 10^{-23}$ gm. The electron collision frequency at the E-layer maximum is in some doubt but following the calculations of Huxley (1955) a value of 1.5×10^4 may not be far wrong. For the ratio ν_i/ν_e we may follow Martyn (1953) who gives it equal to $9.3\sqrt{m_e/m_i}$. This therefore gives ν_i = 5800 and by equation (78) we have

$$v = \frac{120}{T} \text{ m/sec.}$$

where T is the effective thickness of the current sheet. For agreement between the two estimates we should require T = 30 km and this is by no means an unreasonable value.

4.6.4 The exponent of $\cos\chi$.

We have seen that the value of n in the fE-$\cos\chi$ relation has a value considerably higher than would be expected on theoretical grounds; and several explanations

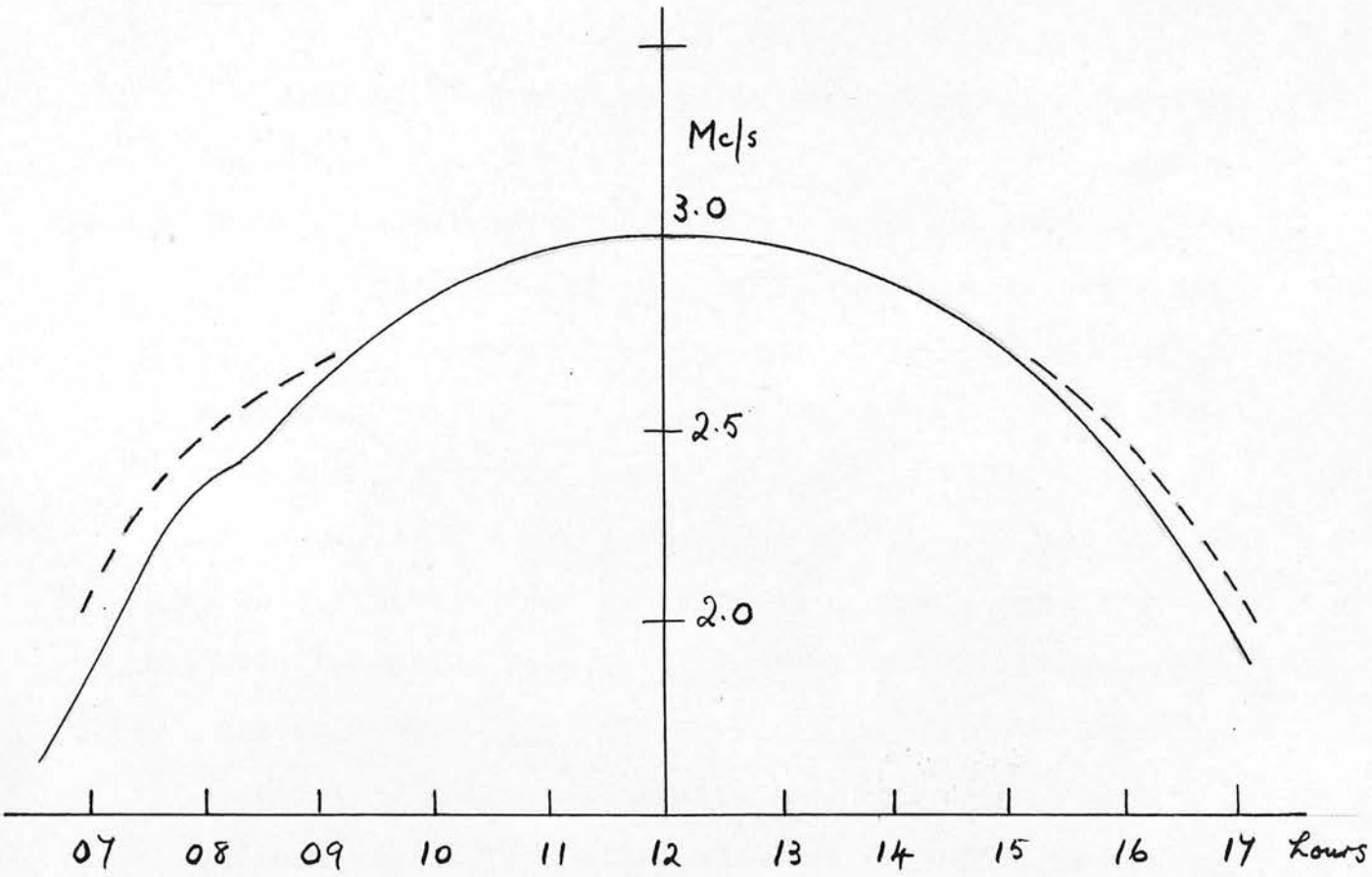


Fig. 27 Mean variations of fE (full-line curve) and of fE_2 in the morning and evening (dashed line curves) during the September sequence, 1955, at Slough.

of this might be proposed. For example if the E-layer is produced by ionization of molecular oxygen (a view supported by Rawer and Argence (1954) and others) then the effect of the transition to mainly atomic oxygen, which is known to occur near the E-region level, might well produce a variation of the kind observed.

The present results suggest however an alternative explanation which may be worthy of consideration. We have seen that E_2 tends to be much more stable in the morning and evening than in the midday period; but may this not mean that we are labelling as E_2 what are in fact two quite different phenomena? If instead we assume that what we call fE_2 in the morning and evening should in fact be called fE, then the variation of fE with $\cos\chi$ would be significantly altered. A comparison of these two interpretations is shown in Fig. 27. Since E_2 of the unstable or sequential type has been disregarded fE is substantially unaltered in the vicinity of noon; and the resulting value of $\underline{n}$ * according to this new interpretation is 0.28; a value very close to what would be predicted in theory.

This hypothesis (and we place it no higher) would also explain the curious change in slope of the log fE - log $\cos\chi$ curve at about 09 or 15 h; indeed it would make the slope of the whole line agree with the slope at midday. Moreover the anomalously low values of h_m in morning and evening would also fit in with this

---

* Cf. Fig. 24.

explanation since they would then refer not to the true E-layer maximum but to a ledge below it.

4.6.5 E_s and other subsidiary ledges.

We now turn finally to some curious ledging phenomena observed in the September sequence which may be relevant to some of our problems. E_s of the "intense", totally reflecting type, extending to high frequencies was observed only rarely in the September records, and "rough" reflections of the patchy-layer type were also uncommon; but "thin-layer" reflections having a definite penetration frequency were extremely common. These echoes show the characteristics predicted by theory (Rawer, 1939) in that there is usually a small overlap at the transition to the higher layer, but no group-retardation; and they also have the characteristics observed by Briggs (1951), that is to say they are "smooth" echoes, with little sign of fading, and they are strongly reflecting, often showing multiple reflections. The "ledges" producing these echoes occur at three different levels: first, near or slightly above the normal-E maximum, at equivalent heights over 110 km; second, in an intermediate position, say between 100 and 115 km; and third, near the bottom of the E-layer, between 90 and 100 km. The first type is the familiar "abnormal-E" which Becker and Dieminger (1950) would describe as E_2; the second type corresponds

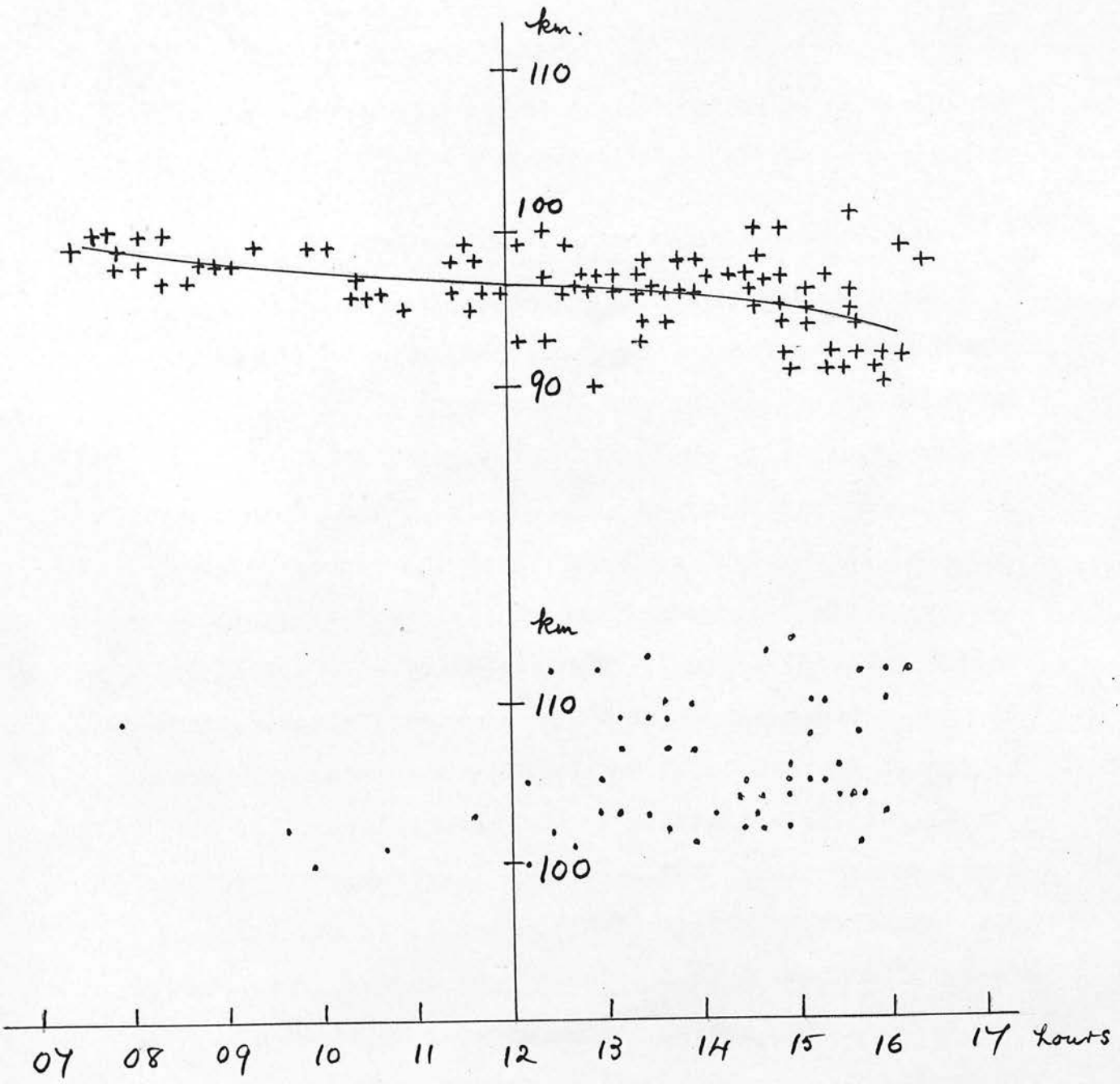


Fig. 28 Heights of ledges of thin-layer type in the E-region.

(a) Low ledges. (b) Intermediate ledges.

From the data of the September sequence, 1955.

to the ledges described by Briggs (1951), but does not seem to have aroused comment so far as conventional-type records are concerned; and the third type corresponds to the low ledge noted by Halliday (1936). A typical instance of the intermediate type is shown in the lower record in Fig. 13. These ledges are not mutually exclusive, two ledges frequently occurring simultaneously at different levels and occasionally three may be present. Moreover there seems to be no sharp distinction between ledges which show group-retardation at the transition and those which do not. To this extent the classification of the upper type as E_2 may be justified, though sometimes a separate E_2 at a higher level is also present.

The heights of the intermediate and low ledges of thin-layer type are shown in Fig. 28, which brings out two points of interest. First both types of ledge are significantly more frequent in the afternoon: indeed the intermediate sharp ledge is almost exclusively an afternoon phenomenon. Second the low ledge shows a small but significant and steady drop in height as the day advances.

Whatever their origin these curious phenomena do establish one thing, at least for equinox conditions, namely that the detailed structure of the E-region is not constant, nor is it symmetrical with respect to noon, but rather it appears to undergo some regular

process of stratification in the later part of the day. If this process affects fE it might have some bearing for example on the ΔfE anomaly at low cos χ , and it might also contribute to the high exponent of cos χ , though for the present such suggestions must remain in the realm of speculation.

Chapter 5.

Conclusions.

If in the course of our investigation some problems have been clarified it is evident from the foregoing discussion that many characteristics of the E-layer remain mysterious, either having several equally plausible explanations or none at all; and there is no dearth of theoretical problems awaiting solution. Let us however summarize briefly the salient features of the present position.

On the theoretical side we can now, with the aid of the formulae developed in Chapter 2, readily compute values of the maximum electron density N_m throughout the day for any assigned variation of electron production q with solar zenith distance χ, assuming a recombination law of electron loss. As a result, particularly with the extension of the calculation to the non-stationary conditions near sunrise and sunset, we may be certain that neither the anomalous variation of ΔfE nor the high exponent of $\cos\chi$ can be put down to inaccuracies of the theory. Hence we have been led to seek their origin elsewhere.

Undoubtedly the main step forward in this respect has been the identification of the pre-noon perturbations in N_m and in the height of the layer as effects

of the vertical drift of neutral ionization associated with the S_q current system. This not only accounts for several previously unexplained anomalies, including increased values of fE near the equator, but is also perhaps of somewhat broader scientific interest; for it provides independent evidence of the existence and the location of these ionospheric currents previously only inferred from the data of the magnetic variations at the ground.

This identification was first made on the basis of a statistical investigation of routine data; but the special measurements made at Slough have provided useful confirmation and have shown that the predicted dip in h_m does in fact occur. Our theoretical investigations of drift have also contributed by clarifying the nature of the drift effects and by enabling us to calculate the magnitudes of the drift velocity and of its gradient.

As to the high exponent of $\cos\chi$ the September results have suggested as a possible explanation that stratification of the layer in the morning and evening may lead to incorrect readings of fE. It seems that if at such times the upper discontinuity is taken to indicate fE instead of the lower one selected in normal practice, then the anomalous variation with $\cos\chi$ would disappear. A decision on this question must however await further evidence.

Finally we have the low values of ΔfE at low $\cos\chi$ leading to an apparent rapid increase of recombination coefficient with height; but so far no satisfactory explanation of this anomaly has been forthcoming, and there is only the somewhat nebulous suggestion that it may be related to the curious "ledging" phenomena which develop in the afternoon period.

Ledging may also be responsible for other unexplained features such as the unexpected pre-noon dip in y_m, and the anomalous variation of h'E in the September data. Indeed it is hardly possible to study a large number of E-region records without becoming convinced that stratification, including all the phenomena variously described as "E_s", is a major problem for future investigation. It may be that, following clarification of the S_q effects, little progress will be made in understanding other peculiarities of the region until the problem of these various subsidiary layers, or "ledges", has been solved.

References. (1) Theoretical.

Appleton, Sir E.V. and Lyon, A.J. (1955) "Report on the Physics of the Ionosphere" p. 20 (The Physical Society).

Appleton, E.V. (1953) J.A.T.P. 3, 282.

Chapman, S. (1931 a,b) Proc. Phys. Soc. 43, 26 and 483.

(1939) Proc. Phys. Soc. 51, 93.

Chapman, S. and Bartels, J. (1940) "Geo-magnetism".

Jackson, F.S. (1956) J. Geophys. Res. 61, 71.

Jahnke-Emde (1945) "Tables of Functions".

Martyn, D.F. (1947) Proc. Roy. Soc. A 189, 241.

(1953) Phil. Trans. 246, 306.

Rawer, K. (1939) Ann. der Phys. 35, 385.

Rawer, K. and Argence, E. (1954) Phys. Rev. 94, 253.

Rydbeck, O.E.H. (1944) Chalmers Tek. Högsk. Handl. No. 34.

Rydbeck, O.E.H. and Wilhelmsson, H. (1954) Chalmers Tek. Högsk. Handl. No. 149.

Wilkes, M.V. (1954) Proc. Phys. Soc. B 67, 304.

Huxley, L.G.H. (1955) T.A.T.P. 8, 118.

References. (2) Statistical Studies.

Allen, C.W. (1948) Terr. Mag. 53, 433.

Appleton, Sir E.V., Lyon, A.J. and Pritchard, Mrs. A.G. (1955) J.A.T.P. 6, 292.

Appleton, Sir E.V., Lyon, A.J., and Turnbull, Mrs. A.G. (1955) Nature 176, 897.

Harnischmacher, E. (1950) Comptes Rendus 230, 1301.

Menzel and Wolbach (1954) Harvard Coll. Observatory Sci. Rep. No. 19.

Saha, A.K. (1953) Ind. J. Phys. 27, 431.

Scott, J.C.W. (1952) J. Geophys. Res. 57, 362.

Tremellen, K.W. and Cox, J.W. (1947) J.I.E.E. 94, Pt. III A, 200.

Beynon, W.J.G. and Brown, G.M. (1956), Nature 177, 583.

References. (3) Experimental.

Appleton, E.V. and Naismith, R. (1932) Proc. Roy. Soc. A 137, 36.
(1935) Proc. Roy. Soc. A 150, 685.
(1940) Proc. Phys. Soc. 52, 402.

Appleton, Naismith and Ingram (1937) Phil. Trans. A 236, 191.
(1939) Proc. Phys. Soc. 51, 81.

Appleton, E.V. (1937) Proc. Roy. Soc. A 162, 451.

Becker and Dieminger (1950) Naturwissenschaften 4, 90.

Best, Farmer and Ratcliffe (1938) Proc. Roy. Soc. A 164, 96.

Bibl, K. (1951) Ann Geophys. 7, 208.

Briggs, B.H. (1951) Proc. Phys. Soc. B 64, 255.

Grace, C.H. (1951) J. Geophys. Res. 56, 452.

Halliday (1936) Proc. Phys. Soc. 48, 421.

Jones, R.E. (1955) J.A.T.P. 6, 1.

Lepechinsky, D. (1956) J.A.T.P. 8, 297.

Naismith, R. (1954) J.A.T.P. 5, 73.

Pfister, W. (1950) Proc. of Conf. on Ionospheric Physics (Pa.) p. 235.

Rastogi, R.G. (1954) Proc. Ind. Acad. Sc. 40, 158.

Saha, A.K. and Ray (1955) J.A.T.P. 7, 107.

Skinner, Brown and Wright (1954) J.A.T.P. 5, 92.

Whale, H.A. (1951) J.A.T.P. 1, 233.

References. (4) Analysis of Ionospheric Data.

Appleton, E.V. (1930) Proc. Phys. Soc. 42, 321.

(1932) Proc. I.E.E. 71, 642.

Appleton, E.V. and Beynon, W.J.G. (1940) Proc. Phys. Soc. 52, 518.

Booker, H.G. and Seaton, S.L. (1940) Phys. Rev. 57, 87.

Jaeger, J.C. (1947) Proc. Phys. Soc. 59, 87.

Kelso, J.M. (1952) J. Geophys. Res. 57, 357.

Lyon, A.J. and Moorat, A.J.G. (1956) J.A.T.P. 8, 309.

Pierce, J.A. (1947) Phys. Rev. 71, 698.

Ratcliffe, J.A. (1948) Nature 162, 9.

(1951) J. Geophys. Res. 56, 463.

Shinn, D.H. (1953) J.A.T.P. 4, 240.

(1956) Private communication.

(*Reprinted from Nature, Vol.* 176, *pp.* 897–899, *Nov.* 12, 1955)

DISTORTION OF THE *E* LAYER OF THE IONOSPHERE BY ELECTRICAL CURRENTS FLOWING IN IT

By Sir EDWARD APPLETON, G.B.E., K.C.B., F.R.S., A. J. LYON and Mrs. A. G. TURNBULL
University of Edinburgh

IT has long been known that the E layer of the ionosphere varies, both diurnally and seasonally, in a fairly regular manner, in that its maximum ionization density, N_m, is closely dependent on the solar zenith distance, χ. In the earliest study of this region, for example, Appleton and Naismith[1] found that, for the seasonal variation of noon maximum ionization in south-east England, N_m varied as $(\cos\chi)^{1/2}$. Because of its substantially regular behaviour, we have selected the E layer as a subject for more detailed study, to ascertain how far its properties conform with those of an ideal ionized layer, as classically expounded by Chapman[2]. We have used, for this purpose, ionospheric data obtained by radio sounding from stations now operating all over the world.

Chapman's well-known theory of ionized layer formation is based on two simple specified assumptions: the earth's atmosphere is considered to be isothermal, and the process by way of which electrons disappear is taken to be that of recombination. A further tacit assumption is that there is no movement of ionization, electrons appearing, and eventually disappearing, at the same point in space.

We began our investigation by attempting to make a careful comparison of the actual diurnal variation of N_m in the E layer with that predicted by Chapman's theory. It can readily be seen physically, and indeed is an essential result of the theory, that, due to the effects of recombination, the temporal variation of ionization density in a given layer lags slightly behind the corresponding variation of the rate of electron production. Thus the day-time maximum value of N_m is expected to occur after noon; while afternoon values of N_m are expected to exceed their corresponding forenoon values. Two well-known theoretical methods of finding the value of the recombination coefficient α are, in fact, based respectively on these two effects. For example, Appleton[3] has shown that the time delay Δt of the maximum of N_m relative to noon is given by $(2\alpha N_m)^{-1}$; so that, if the delay is

measurable, α can be determined, since N_m is known. The other method of determining α is a little more complicated, but depends essentially on the determination of ΔN_m, the difference between afternoon and forenoon values, for local times equally spaced relative to noon. Using the great wealth of ionospheric data available for the Slough station, we have used both the Δt and the ΔN_m methods of finding α, from the diurnal asymmetry of the E layer. The results obtained are very remarkable in indicating extremely large variations of the *apparent* values of α. Our problem has therefore been that of deciding whether, in fact, α really does vary in such a remarkable fashion or whether the basic theory used is inadequate. In the latter case, if there were some additional perturbing factor operative, one might hope that, as often happens in such cases, the magnitude and nature of the discrepancy between theory and experiment would lead to the identification and measurement of the disturbing influence. As we shall see, such a hope has been, to a considerable extent, realized.

By extending our studies of the diurnal variation of N_m to a range of latitudes other than that of Slough, we soon found evidence indicating that the values of *apparent* α obtained by the methods used could not possibly refer to the true recombination coefficient and were in fact being influenced by a perturbing factor. For example, while stations in the higher latitudes behaved like Slough, in that Δt and ΔN_m were both found to be positive there, stations situated in a belt centred on the equator exhibited an inverse phenomenon, in that Δt and ΔN_m were found to be negative. Evidently, then, for stations in low latitudes, the influence of the perturbing factor is so great that it can actually reverse the sign of the diurnal asymmetry and cause afternoon values of N_m to be less than the corresponding forenoon values.

We were therefore led to examine afresh the whole basis of the theory of ionized layer formation, taking into account the possible influence of a likely perturbing factor. Now Chapman's theory is based on the simple continuity equation

$$\frac{dN}{dt} = q_0 \cos\chi - \alpha N^2 \qquad (1)$$

where N is the electron density, q_0 the rate of electron production when the sun is vertical, and t the time. If, however, we wish to take account of the possibility of electron transport phenomena, (1) must be written

$$\frac{dN}{dt} = q_0 \cos\chi - \alpha N^2 - \text{div}\,(Nv) \qquad (2)$$

where the term div (Nv) expresses the electron transport-rate out of the unit volume under consideration. Now, as Martyn[4] has suggested, vertical drift of neutral ionization is a type of electron transport likely to occur in the ionosphere ; and Appleton and Lyon[5] have developed the Chapman theory so as to allow for the effects of such drift under conditions applicable to the E layer. The theory of Appleton and Lyon is, moreover, formulated directly in terms of the value of N_m for a given layer, which is the quantity with which we are directly concerned in practice. From a practical point of view we are concerned with two effects of any vertical drift which may be operative, namely, the *alterations* it provokes in both the atmospheric level of the layer maximum electron density and in the magnitude of the maximum ionization density itself. We shall refer to these two departures from normality as $\delta h(N_m)$ and δN_m, respectively. The appropriate expressions for these quantities, given by Appleton and Lyon, are as follows :

$$\delta h(N_m) = \frac{v}{2\alpha N_m} \qquad (3)$$

and

$$\frac{\delta N_m}{N_m} = \frac{-\dfrac{\partial v}{\partial h}}{2\alpha N_m} - \frac{1}{4H^2}\left(\frac{v}{2\alpha N_m}\right)^2 \qquad (4)$$

where v is the velocity of vertical drift, $\dfrac{\partial v}{\partial h}$ is its vertical gradient (both measured positive upwards) and H is the scale height of the atmospheric ionized component responsible for the E layer. Equations (3) and (4) tell us, then, the corrections to the Chapman theory which are necessary to take vertical drift into account.

Now in looking for the possible intervention of vertical drift effects as a perturbing factor, we are naturally led to consider the S_q overhead current system, which is responsible for the daily solar magnetic variations, as a likely origin, since there is ample evidence that this current system flows at or near the level of maximum ionization in the E layer. We were additionally encouraged to do this since it is known that the S_q currents flow in opposite directions in high and low numerical latitudes, and, as described above, our study of E-layer diurnal abnormalities suggests the intervention of some perturbing agent which has opposite influences in high and low latitudes. Now it is an important characteristic of the S_q system that the overhead currents constituting it attain their maximum value, not at noon, but at about 1000–1100 h. local time.

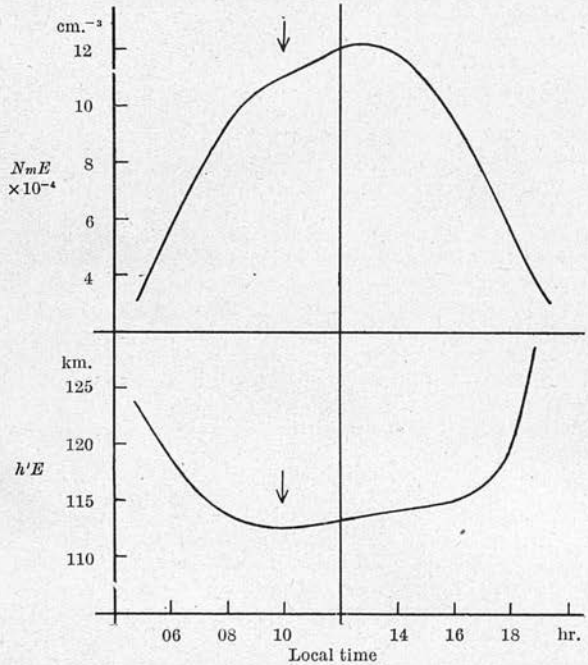


Fig. 1. Showing (upper curve) diurnal variation of Slough E layer maximum ionization density, N_m; and (lower curve) diurnal variation of Slough E layer height, $h'E$. In both cases abnormal variations (depressions) are to be noted in the forenoon when the ionospheric S_q currents attain their maximum intensity

We were therefore led to look carefully for any abnormalities in $h(N_m)$ or N_m at that period of the day, in view of equations (3) and (4). Unfortunately, values of $h(N_m)$ were not available, so we were obliged to use values of $h'E$, the height of the under-boundary of the E layer, instead.

In Fig. 1 are shown the diurnal variations of $h'E$ and of N_m for Slough over the summer months of the year. On the same diagram is marked the approximate local time when the overhead S_q current (and the corresponding value of the horizontal north-seeking magnetic force, X_q) reach their maximum value. It will be at once apparent that, at this particular local time, the value of $h'E$ is abnormally decreased, while the value of N_m is similarly reduced. These results, we therefore submit, indicate that the S_q current system can markedly influence E layer phenomena and encourage us to count such an agency as likely to be one of the major perturbing factors responsible for the divergences we have identified

between E layer behaviour and the Chapman theory.

We may first note a very simple result of electrodynamical theory. Since, at Slough, the S_q current flows from east to west, athwart the northwards horizontal component of the earth's magnetic field, we should expect the velocity of drift to be downwards; and, as equation (4) indicates, we should then expect the height of the layer (as measured by $h(N_m)$ or $h'E$) to fall in sympathy. We note that the abnormal variation of $h'E$, at the time of maximum overhead east-to-west current, has just the sign to be expected. It is, however, a novel result, which could not have been predicted, that δN_m and $\delta(h'E)$ are of the same sense. Since only the first term on the right-hand side of (4) is of numerical significance here, we can deduce, from (3) and (4) together, that v and $\frac{\partial v}{\partial h}$ are of opposite sign. In other words, we find that, when the S_q current flows from east to west, both the velocity of drift and the layer-height variation are downwards; while at the same time the drift velocity falls off with increasing height, leading to a reduction of the layer-ionization maximum N_m. We have now only to assume that all these ionospheric changes are reversed in sense in low latitudes (where, as we know, the S_q current flows from west to east) to be able to explain the E layer anomalies mentioned above. In low latitudes, for example, due to the fact that the S_q current system is not situated symmetrically to the local noon meridian, the E layer is subject to vertical drift effects which are different in the forenoon and afternoon. Such vertical drift causes a marked enhancement of N_m in the forenoon, which so far disturbs the natural diurnal asymmetry of N_m as to cause both Δt and ΔN_m to become negative, instead of positive, as simple theory would require.

Now although the S_q system of overhead currents is not situated quite symmetrically with respect to the noon meridian, its influence is generally exerted in the same sense over practically the whole of the middle of the day. Our study of abnormal diurnal asymmetry in N_m, as described above, shows that a west-to-east S_q current tends to raise the value of N_m, while an east-to-west S_q current tends to reduce it. We can therefore see that the overall effect of such an agency of vertical drift is to influence the general level of N_m in different ways in different latitudes; in low latitudes N_m is increased, at the latitude of the S_q current focus N_m is sensibly unaltered, while in still higher latitudes N_m is decreased. We are therefore encouraged to consider whether

this simple general result, obtained from a study of the abnormal diurnal behaviour of the E layer, also explains any of the other abnormal features of the E layer which have been described in recent years.

Now Appleton[6] has shown that, while the value of N_m in the E layer is closely dependent on the solar zenith distance, it is not uniquely dependent on $\cos\chi$. There is, additionally, a tendency for the value of N_m to be high in low latitudes. For example, at the equinoxes, the contours of the geographical locations of constant N_m are not circles, with the sub-solar point as centre, as would be expected if N_m were uniquely dependent on $\cos\chi$. They are, instead, of elliptical shape, the major axis lying along the equator. Again, in June and December at noon, it has been found that N_m does not attain its maximum value at the sub-solar point but at a latitude displaced therefrom by some 10° towards the equator. It will readily be seen that both these anomalies are also explicable in terms of a perturbation which, *ceteris paribus*, causes N_m to be greater in low latitudes than in higher latitudes. Moreover, a further study shows that these two anomalous features of E-layer morphology, namely, the abnormal diurnal asymmetry and the latitude dependence of N_m, are quantitatively reconcilable.

Finally, it may not be out of place to indicate what we consider to be the significance of these results in the general development of ionospheric studies. Many years ago it was boldly suggested by Balfour Stewart[7] that the daily rhythmic variations of the earth's magnetic field might be due to circulating electrical currents in the upper atmosphere. The geomagneticians then indicated the intensity and configuration of the upper-atmospheric current system which would be required. The discovery of the ionosphere and the elucidation of its properties by radio sounding has shown that the upper atmosphere is, in fact, sufficiently conducting to be the location of such currents; and the work now described has confirmed the presence of ionospheric currents, of the type prescribed, by way of the distortion they engender in the medium in which they flow.

[1] Appleton, E. V., and Naismith, R., *Proc. Roy. Soc.*, A, **150**, 685 (1935).

[2] Chapman, S., *Proc. Phys. Soc.*, **43**, 26 and 483 (1931).

[3] Appleton, E. V., *J. Atmos. Terr. Phys.*, **3**, 282 (1953).

[4] Martyn, D. F., *Proc. Roy. Soc.*, A, **180**, 241; **190**, 273 (1947); **194**, 429 and 445 (1948).

[5] Appleton, E. V., and Lyon, A. J., "Report on the Physics of the Ionosphere", p. 20 (The Physical Society, 1955).

[6] Appleton, E. V., Proc. Fourth Meeting Mixed Commission on the Ionosphere, p. 14 (1955).

[7] Stewart, B., "Terrestrial Magnetism", Encycl. Brit., Ninth Ed. (1882).

Printed in Great Britain by Fisher, Knight & Co., Ltd., St. Albans.

Journal of Atmospheric and Terrestrial Physics, 1956, Vol. 8, pp. 309 to 317. Pergamon Press Ltd., London

Accurate height measurements using an ionospheric recorder

A. J. Lyon* and A. J. G. Moorat†

Official communication from D.S.I.R. Radio Research Station, Slough

(*Received* 2 *November* 1955)

Abstract—A method for the calibration of an ionospheric recorder is described, which corrects errors in height measurement arising from the distortion of the echo-pulse in its passage through the receiver. The amount of this error depends on the echo-amplitude, and is shown to vary in an approximately linear manner with the width of the recorded echo-trace. Several methods of checking the calibration confirm that it is reliable to within ± 2 km. Using a calibrated recorder and an expanded time-base, it is possible to measure E-region equivalent heights to this order of accuracy. The systematic error due to pulse distortion will, in general, cause the heights recorded in routine ionospheric measurements to be from 5 to 15 km too high. Some consequences of this, e.g., for MUF predictions, are mentioned.

1. Introduction

The necessary conditions for accurate height measurement have been well known for many years and are fully exploited in pulse navigational aids. Conventional automatic ionospheric recorders, however, do not conform with these conditions, and some errors are therefore to be expected in their measurements (Naismith, 1949). In this paper we discuss a technique for correcting the error in height measurement due either to the distortion of the pulse during its passage through the receiver, or to any slight asynchronism between the transmitted pulse and the height-calibration markers.

An attempt has recently been made to provide more accurate h'-f curves for the E layer than are required in normal practice. The equipment used was a D.S.I.R. automatic ionospheric height recorder (Clarke and Shearman, 1953) of the type widely used, but with the time-base sweep adjusted to record about 300 km of equivalent height instead of about 1000 km. Examples of records, showing the considerable expansion of E-region traces compared with normal, are given in Fig. 1. By this means the nominal reading accuracy for reflections from layers below about 300 km can be improved from ± 5 km to ± 1 km.

Inspection of these and similar records shows, however, that the bottom of the trace, corresponding to the leading edge of the echo-pulse, fluctuates as the width of the trace varies. Thus, the lower edge of the trace cannot be used without correction even for a relative measure of equivalent height, and it is essential to determine the necessary corrections.

2. Measurement of the Height Errors of an Ionospheric Recorder

The D.S.I.R. automatic ionospheric height recorder, in accordance with internationally accepted practice, is so arranged that the pulse in the modulator of the transmitter is synchronized to one of the height calibration markers. Hence,

* University of Edinburgh: Now at University College, Ibadan, Nigeria.
† Radio Research Station.

under normal operating conditions, the start of the large ground-wave signal coincides with this marker, which forms a natural origin from which to measure the heights of reflection.

The total error in the observed height, measured to this origin, and its variation with echo-amplitude, can be very simply determined by using the transmitted pulse (or "ground-wave") as an artificial echo. To do this, the receiver aerial is replaced by a probe small enough to ensure that the r.f. stages of the receiver are not overloaded. In this way the transmitter pulse input to the receiver is reduced to the amplitude of a normal echo, and by varying the receiver attenuation a sequence of signal amplitudes may be obtained, corresponding to the normal range of echo-amplitudes.

An example of the photographic records obtained by this procedure is shown in Fig. 2. The required total error (whether due to receiver or transmitter) is the distance from the bottom of the zero height-marker to the bottom of the signal trace. Clearly the error decreases steadily as the width of the trace increases —from some 10 km for a weak signal to 2 or 3 km for strong signals.

A series of records similar to the one in Fig. 2 were made for ten frequencies ranging from 0·8 Mc/s to 6·0 Mc/s, using the normal transmitter pulse-width of 150 microseconds. In each case the error Δ, measured as above, was plotted as a function of the width of the trace w. For any single frequency, the error closely follows a linear law of the form

$$\Delta = a - bw. \tag{1}$$

With the equipment in use, it is convenient to divide the frequency band into two parts. For the range 0·8 to 2 Mc/s the best straight line is

$$\Delta = 17{\cdot}5 - 0{\cdot}25w \tag{1a}$$

and for the range 2·5 to 6·0 Mc/s

$$\Delta = 14{\cdot}5 - 0{\cdot}25w \tag{1b}$$

The true equivalent height of an actual echo may now be obtained by subtracting the error as given by these equations from the observed height. With this instrument, as will be seen from Fig. 3, the observed heights will be too high by amounts ranging mostly between 5 and 10 km.

3. The Origin of the Height Errors and Their Variation with Echo-amplitude

As is usual with ionospheric pulse-transmitters, the delay between the modulator pulse and the r.f. pulse radiated from the aerials is found to be very small—certainly less than 2 km (13 μsec). It follows that the major part of the observed errors must arise in the receiver circuits. The character of the distortion is shown diagrammatically in Fig. 4, which gives the equivalent shapes of an originally square pulse, after passage through the receiver, for two particular amplitude levels, assuming that the rise-time and decay-time remain constant. Now, in practice there is a threshold level, OA, below which no signal appears on the record; consequently, the width of the pulse recorded is that of a section across

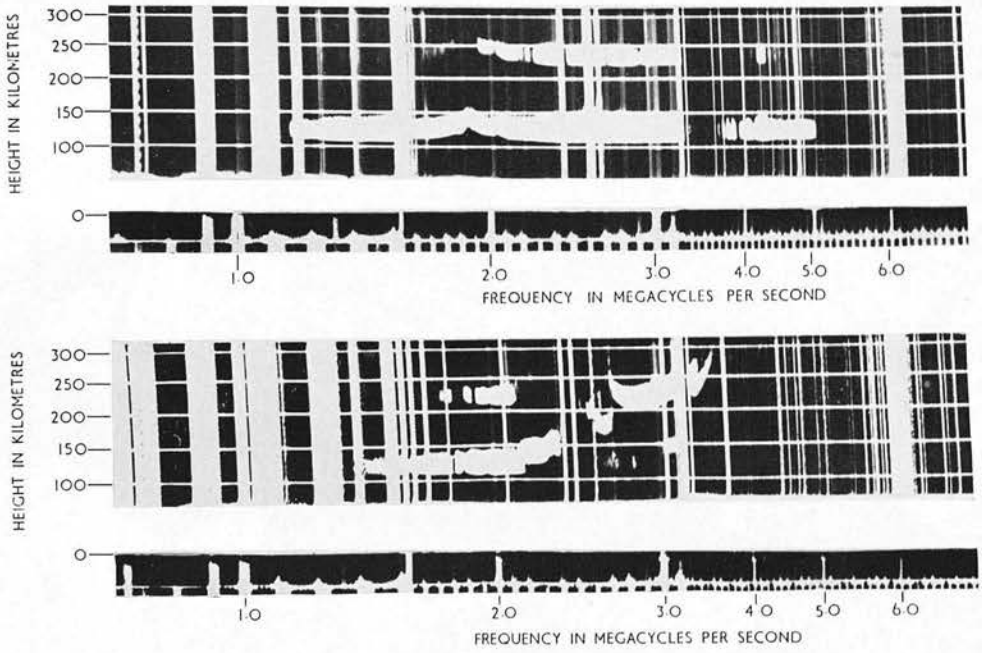


Fig. 1. *h'f* records made using a D.S.I.R. ionospheric height recorder, with the time-base sweep adjusted to 300 km instead of about 1000 km. (The thin horizontal lines are the 50-km height markers; and the irregular vertical streaks are due to broadcast and other interference.)

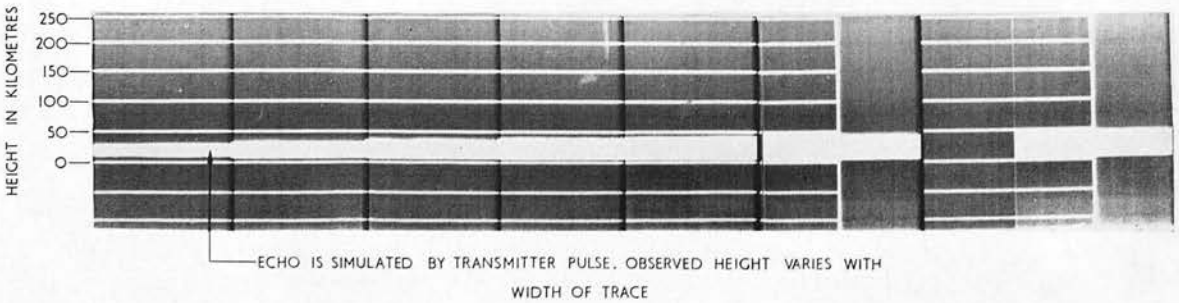


Fig. 2. A calibration record. The observed delay from the zero height-marker varies with the width of the trace, and hence the amplitude of the signal.

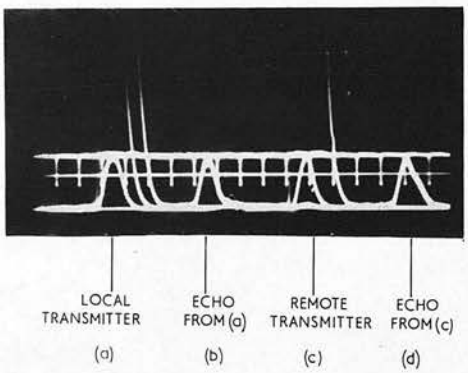


Fig. 6. Ground-wave and E echoes using local and remote transmitters seen on the display of the absorption equipment. The frequency is 1·6 Mc/s, the height markers are at 25 km intervals, and the reading accuracy is $\pm$ 1 km.

the pulse at height OA, i.e., for the smaller pulse the section BC. When the amplitude of the pulse increases, the width of the section BC increases, and the error in the observed height decreases. To give $b = 0{\cdot}25$, the slope of the leading edge is three times that of the trailing edge, so that when the recorded trace-width increases by 4 km, the bottom of the trace will fall by 1 km and the top will rise by 3 km.

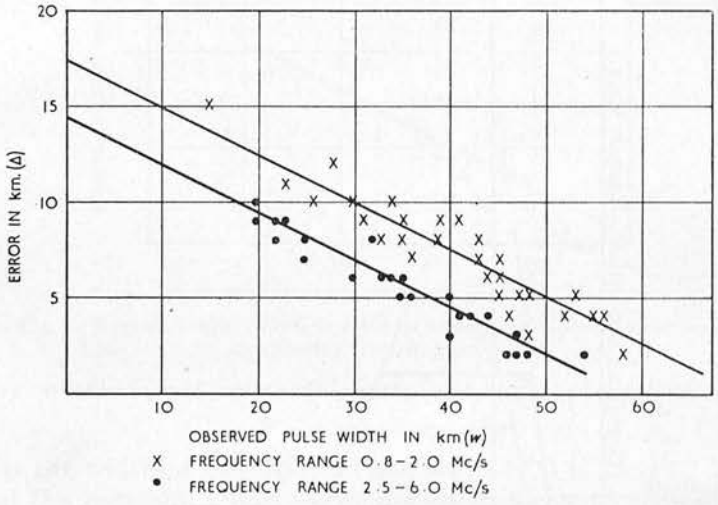


Fig. 3. Recorder height error as a function of observed pulse-width, constructed from calibration records of the type shown on Fig. 2.

Expressing delay-times in kilometres of equivalent height (so that 1 km = 6·6 μsec), the time of rise of the leading edge of the pulse, as measured from Fig. 4, is 7·5 km, which would be the theoretical rise-time for an ideal rectangular band-pass of 20 kc/s. Now, the actual band-pass of the receiver, measured to points 6 db below the peak, is 15 kc/s, and since of course the response curve is not ideal, this may be considered very reasonable agreement. The delay-time measured to the centre of the time of rise is 6 km (40 μsec): but the observed delay may be greater or less than this, depending on the pulse-amplitude.

The quantity a is determined solely by the shape of the pulse after passage through the receiver (it can be measured from the intersection of the lines EK and GL in the figure); hence it does not vary with the threshold level or with the position of the C.R.T. "brightness" control. It will vary, however, with the transmitter pulse-width used, being a little less if the pulse-width is shorter; recalibration is therefore advisable when the pulse-width is changed. The small variation of a with frequency in the instrument under test appears to indicate a slight difference in the receiver characteristics at low and high frequencies: this would not be unreasonable.

4. Checks of the Calibration Procedure

It is desirable to check first that the linear law for the variation of apparent height with trace-width—and particularly the gradient of 0·25—are still true for actual echoes. This can be tested by taking $h' - t$ records at constant frequency, and the echo trace-width can be varied by altering the receiver attenuation

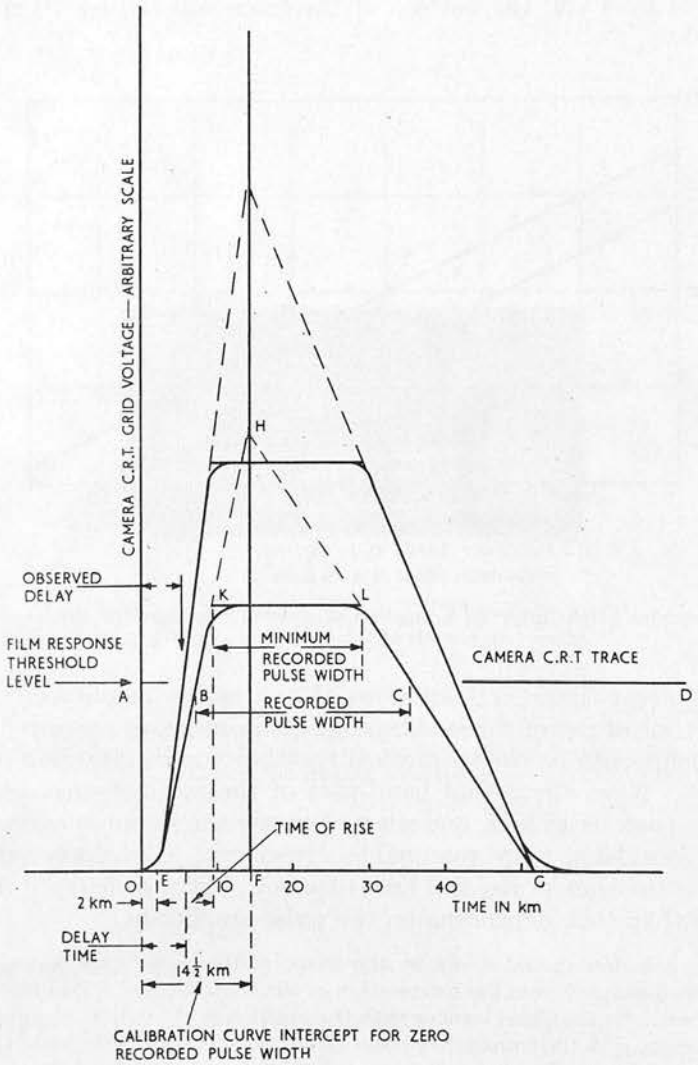


Fig. 4. Equivalent pulse-shape plotted from the position of the recorded trace (Figs. 2 and 3). It is *not* the shape of the receiver output-pulse, but reproduces the essential delay characteristics of that pulse.

or by using natural fading. It will be seen from Fig. 5, plotted from such records, that the gradient $d\Delta/dw$ is again very near to 0·25. This test verifies that our correction equation (1) with $b = 0{\cdot}25$ is satisfactory at least for relative heights.

We may test the accuracy of the correction equations for absolute heights with the aid of multiple echoes. Let the true equivalent height of reflection be

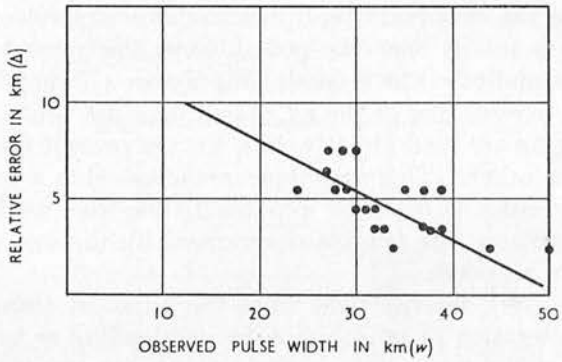


Fig. 5. Recorded relative height error as a function of observed pulse-width constructed from $h' - t$ records.

h'_0, then by equation (1) the observed equivalent height of the rth order echo will be

$$h'_r = rh'_0 + a - bw_r$$

where w_r is the width of the rth echo and b (=.0·25) is known. Hence the true height and the constant a may be expressed in terms of the apparent heights, say of the 1st and rth echoes, by the equations

$$h'_0 = \frac{1}{r-1}\left\{(h'_r + bw_r) - (h'_1 + bw_1)\right\} \tag{2}$$

and

$$a = \frac{1}{r-1}\left\{r(h'_1 + bw_1) - (h'_r + bw_r)\right\} \tag{3}$$

Using equation (3), we have determined a in ten cases where a second echo was present (this being the highest multiple echo that can be recorded with our expanded time base). For eight of these cases (mostly for frequencies greater than 2 Mc/s) the average value of a was 15 km, in good agreement with (1b). The two remaining cases gave much lower values (2 km and 7 km respectively); these anomalous results are probably due to the presence of composite echoes whose lower component is not visible in the second echo.

Overlapping echoes from different apparent heights are not uncommon in E-region reflections. Though not always visible in the records, such composite echoes may be clearly seen on the monitor display, which, unlike the photographic record, shows the variation of signal-amplitude with equivalent height. In such a case, if the lower component is the weaker it may not be visible after a second reflection, so that the observed h'_1 would be too small in relation to the observed h'_2, and by equation (3) the derived value of a would then be too small.

It is worth noting that the crude test of comparing the observed heights from ground-wave to first echo and from first to second echo, without applying any width correction, usually underestimates the real error. When, however, eight or nine multiple echoes are visible, the presence of the error is unmistakable even without width corrections.

A further method of checking the correction formula (1) is to compare the corrected heights with heights measured simultaneously, using the type of manual equipment employed for absorption measurements (Piggott, 1955). With this instrument, the ground-pulse is attenuated at i.f. to a standard amplitude and a height-marker is set to a point on its leading edge. The attenuation is then reduced until the echo reaches the standard amplitude, and the equivalent height is read to the corresponding point on the echo-pulse. Thus the ground-wave and echo signals are distorted equally by the receiver, and no error in height measurement is involved, provided overloading of the r.f. stages does not occur.

Normally the heights are read directly, but, for the present tests, photographs of the tube face were taken. This technique made possible a reading accuracy of about ± 1 km, the same as with the expanded time-base method. In several tests at various frequencies, the heights determined by the two methods agreed to within 2 or 3 km in all cases.

To show that the very intense field from the adjacent transmitter did not cause errors due to alteration of pulse-shape by overloading or leakage, the same technique was used with a transmitter sufficiently remote from the receiver not to overload it. The gain of the absorption receiver was adjusted so as to bring to a standard amplitude in quick succession:

(a) the ground-wave from the local transmitter,
(b) the echo from the local transmitter,
(c) the ground-wave from the distant transmitter, and
(d) the echo from the distant transmitter,

the output at each setting being photographed. A typical example, using a frequency of 1·6 Mc/s, is reproduced in Fig. 6. In this case, the apparent height measured between (a) and (b) is 105 km, and that between (c) and (d) is 106 km; and simultaneous measurements on the D.S.I.R. automatic height recorder gave 105 km when corrected. Other results, using local and remote transmitters, also agreed within the accuracy of the measurements.

These various checks confirm the general soundness of the primary method of calibration described in Section 2, and indicate that expansion of the time-base enables one to achieve, for *E*-region equivalent heights, an accuracy of the order of ± 2 km in cases where the echo observed is simple. The corrections are of course equally applicable, and necessary, when a normal time-base sweep is in use.

Increasing the time-base-sweep velocity is not the only method of improving the reading accuracy. We have found that optical magnification is almost equally effective, though less convenient, even when the records are made on paper.

5. Height Corrections in Routine Measurements

It has been recommended internationally (C.C.I.R., 1948) that routine height measurements should normally be made with an accuracy of ± 10 km, and most stations attempt to reach this accuracy. For many purposes, irregular errors of this magnitude are comparatively unimportant. Systematic deviations of the type discussed above, however, introduce discrepancies which can prove significant both in the prediction of maximum usable frequencies and in the study of the

physics of the ionosphere. For example, a systematic error causing the observed heights to be 10 km too high, which is probable using existing techniques, will make the MUF factors for the E and F layers (respectively) about 6% and 3% too low at medium or long distances. This is an appreciable fraction of the discrepancy known to occur (C.C.I.R., 1953).

Our analysis shows that the correction needed with the type of recorder used at all D.S.I.R. out-stations should vary between about 5 and 10 km. This is confirmed by analysis of a few typical night-time records having a large number of

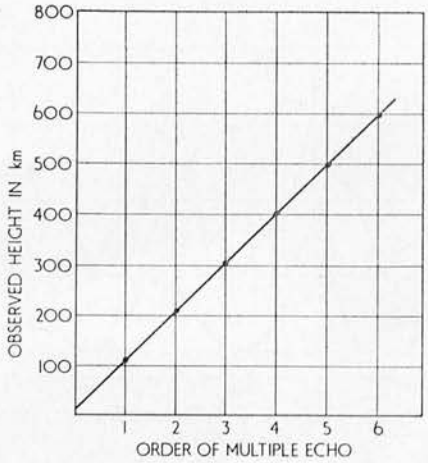


Fig. 7. Heights of multiple echoes from a routine record taken with the prototype recorder at Slough after small corrections (<5 km) have been made for observed pulse-width. The error of the first echo, in this case 13 km, is the intercept on the observed height axis.

multiple echoes, made by the D.S.I.R. recorders at Singapore and Khartoum. These records give values of a, derived by equation (3), of about 13 km, and the corresponding errors in the apparent heights of first-order echoes are between 6 and 8 km. (Since the pulse-width used is shorter than that used in our tests, a smaller value of a is to be expected.) Records from the prototype recorder (Naismith and Bailey, 1951) used at Slough for routine measurements, give an average value for the systematic error of between 10 and 15 km. For example, Fig. 7 shows the heights of six successive multiple echoes from a routine record, after small corrections have been made for pulse-width. In this case the error in the apparent height of the first echo was 13 km; and the average error from seven similar examples was about 12 km.

This result can be checked by comparing the monthly mean values of $h'E$, measured using the same instrument, with the equivalent heights at 2·0 Mc/s obtained during routine absorption measurements at Slough. These two parameters are plotted in Fig. 8 for the years 1951–53. In winter a 2-Mc/s signal is often reflected from heights considerably below the normal E layer, but most of these observations can be removed by omitting heights at or below 105 km in winter months, as shown by the dashed curve. The mean difference between the heights obtained on the two instruments, after allowing for the abnormal winter reflections, is 13 km—in excellent agreement with the figure of 12 km quoted above.

Since the band-widths used in other routine recorders cannot differ greatly from those employed at Slough, and since a considerable part of the delay must be ascribed to the time of rise of the signal, which is determined by the band-width, similar errors are probably present at all stations. It is desirable that these errors should be estimated elsewhere before commencing the intensive observational programme of the International Geophysical Year. For reasons of continuity, however, it would be inadvisable to circulate corrected heights or corrected MUF factors unless the magnitudes of the corrections applied are clearly stated.

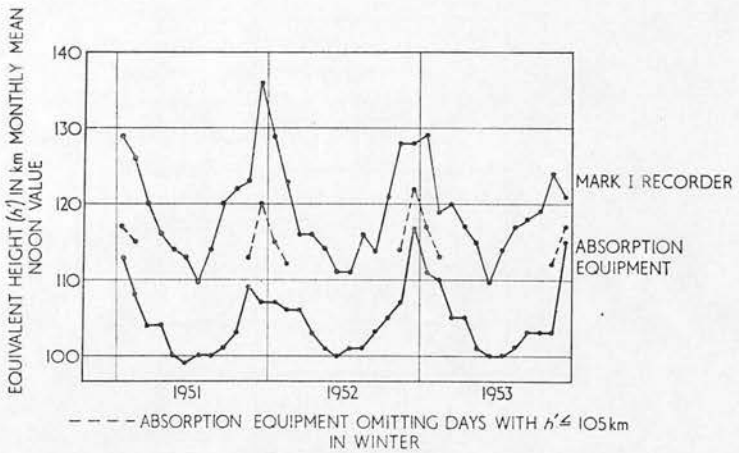


Fig. 8. Comparison of $h'E$ from the prototype recorder records with h' measured at 2·0 Mc/s (from routine data for Slough, England) with the absorption apparatus, showing a mean difference of 13 km between the heights measured on the two equipments (omitting heights <105 km in winter months).

6. Conclusions

We conclude that it is possible to correct the apparent heights obtained from a routine recorder, so that the equivalent height of E reflections can be measured with an accuracy of ± 2 km or better; and, since such corrections are not made in normal practice, that significant systematic errors are probably present in height and MUF measurements from ionospheric stations all over the world. The accuracy demanded by international standards, ± 10 km, may not in fact be reached unless these corrections are applied.

Acknowledgements—The authors wish to thank Messrs. W. R. Piggott and R. W. Meadows of the Radio Research Station, Slough, for many helpful discussions and valuable suggestions, and Mr. A. M. Howatson (University of Edinburgh) for a number of useful comments. In particular, the experimental techniques used were based chiefly on suggestions by Mr. R. W. Meadows. One of us (A. J. L.) is indebted to the Director, Dr. R. L. Smith-Rose, for permission to participate in this work at the Radio Research Station.

This work was carried out as part of the programme of the Radio Research Board, and this paper is published by permission of the Director of Radio Research of the Department of Scientific and Industrial Research.

References

———	1948	C.C.I.R. Recommendation No. 6, C.C.I.R. Fifth Meeting, Stockholm, p. 23.
———	1953	C.C.I.R. Report No. 23, C.C.I.R. Seventh Plenary Assembly, London, **1**, 212.
Clarke C. and Shearman E. D. R.	1953	*Wireless Engr.* **30**, 211.
Naismith R.	1949	*U.R.S.I. Proceedings*, 251.
Naismith R. and Bailey R.	1951	*Proc. Inst. Elect. Engrs. Part III*, **28**, 271.
Piggott W. R.	1955	*Wireless Engr.* **32**, 164

RESEARCH NOTE

The detection of the S_q current system in ionospheric radio sounding

(*Received* 4 *September* 1955)

USING the results now available from the world's ionospheric stations, a critical examination of the behaviour of the E-layer of the ionosphere has been conducted with the object of discovering how far such behaviour is explicable in terms of CHAPMAN'S classical theory of ionized layer formation. It has been found that, although this theory certainly gives a broad general explanation of E-layer phenomena, a number of anomalies have been identified which indicate the operation of certain perturbing factors not envisaged in its formulation. Some of these anomalies have already been reported (APPLETON, 1955). The present note deals with the results of a further examination of one of them, namely the effect of electron transport phenomena on diurnal variations of maximum electron density.

We may first note that the theory of CHAPMAN is based on the simple continuity equation

$$\frac{\partial}{\partial t} N(h) = q(h, \chi) - \alpha[N(h)]^2 \tag{1}$$

where $N(h)$ is the electron density at height h and time t; $q(h, \chi)$ expresses the rate of electron production per unit volume, and so is a function of both height and solar zenith distance χ, while α is the recombination coefficient, assumed independent of both h and t. In a recent paper (APPLETON and LYON, 1955), two of us have shown how, under quasi-stationary conditions such as are believed to obtain in the E-layer, the layer maximum electron density N_m itself obeys an equation, similar to (1), as follows:—

$$\frac{dN_m}{dt} = q_0 \cos \chi(1 - a^2) - \alpha N_m{}^2 \tag{2}$$

where a^2 is a quantity which can be estimated, but which is usually negligible compared with unity except when χ is large.

An approximate solution of (2), with "a" neglected, suffices to express the theoretical variation of N_m over the central part of the day. This solution shows, as we should expect, that N_m should generally be greater in the afternoon than in the forenoon. Also, the maximum of N_m should occur shortly after noon. Methods of determining α' the *effective* value of the recombination coefficient, can be based on each of these two effects (APPLETON, 1937 and 1953). In the first method we compare the values of N_m for similar values of χ, before and after noon. The value of α' is then given by

$$\alpha' = \frac{\left|\frac{dN_m}{dt}\right|}{\Delta N_m \cdot N_m} \tag{3}$$

where $\left|\frac{dN_m}{dt}\right|$ and N_m are the numerical averages of the a.m. and p.m. values, and ΔN_m is the amount by which the p.m. value of N_m exceeds the corresponding a.m. value. In the second method we use the theoretical relation expressing the delay (Δt) of the N_m maximum relative to noon. This is

$$\Delta t = \frac{1}{2\alpha' N_m} \tag{4}$$

We may note, then, that, if simple Chapman theory were applicable to the E-layer, we should find α', as determined by way of (3) or (4), to be constant under all conditions.

As a result of the examination of a great wealth of diurnal data in relation to equations (3) and (4), we can report the most flagrant disagreements between experiment and theory, if α' is assumed constant. Indeed, estimates of the values of α', using (3) and (4), reveal relatively enormous variations of that quantity. Since it seems clear (APPLETON, 1955) that the value of the natural recombination coefficient α could not possibly vary by the amounts exhibited by α', we believe that the non-constancy of the latter is an expression of the existence of some form of perturbation. An obvious type of perturbation to be considered is the effect of electron transport.

Let us suppose, for example, that electron transport of some kind is present. Our basic continuity equation must then be written

$$\frac{dN}{dt} = q(t) - \alpha N^2 - \text{div}\,(Nv) \tag{5}$$

where v is the velocity of electron transport. If electron transport occurs in the form of vertical drift, as suggested by MARTYN (1947, 1948), it is possible to calculate theoretically (APPLETON and LYON, 1955) how the height of maximum ionization, and the value of maximum ionization itself, are altered under E-layer conditions. It is also possible to estimate the effect of vertical drift on the value of the effective recombination coefficient, α', as determined by either (3) or (4).

For example, when α' is obtained from the noon delay Δt via the simple relation (4), it may be shown that

$$\alpha' = \alpha + \frac{1}{2N}\frac{\partial(v'N)}{\partial N} \tag{6}$$

where v' is $\dfrac{\partial v}{\partial h}$, and the second term is to be evaluated at noon. If, then, the vertical drift gradient is appreciable we should expect the value of α' (and the corresponding value of Δt) to vary with latitude and any other conditions which may affect electron transport.

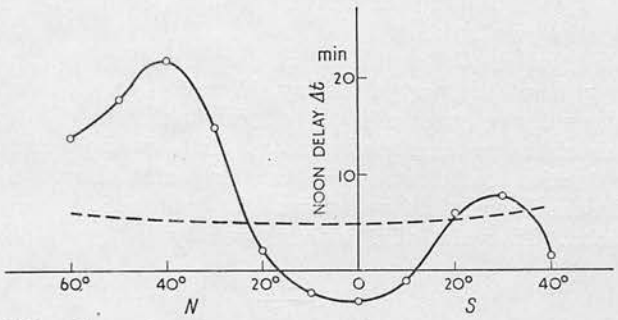


Fig. 1. The full line shows the relation between Δt (noon delay of maximum N_m) and geographical latitude for June (mean of values for the years 1953 and 1949). The broken line shows the theoretical variation for $\alpha = 10^{-8}$ cm^3/sec.

We now turn to a sample of our experimental results. In Fig. 1 the full line shows the variation of Δt (noon delay of N_m) over a wide range of latitudes for the same period of observation, and the broken line indicates the theoretical values of Δt for $\alpha' = 10^{-8}$ cm^3/sec. It will be seen that although Δt does not vary in the manner to be expected if α' were constant, it nevertheless has a definite pattern. We suggest that the pattern is related to

the S_q overhead current system which, by way of the earth's magnetic field, would be expected to cause a vertical drift pattern having a configuration related to it.

Thus MARTYN (1953) has shown that "in a given region of the ionosphere the vertical drift of ionization depends entirely on the *east-west current in that region*." Now, near the S_q current focus the east-west component of current is always small, but at higher and lower latitudes it becomes large near noon, and the directions of flow are opposite in the two cases. Hence "it seems certain that, at the ionospheric levels where flow the currents mainly responsible for the solar and lunar magnetic variations, the daily oscillations of these regions at low latitudes must be *in phase opposition* to those at high latitudes" (MARTYN, *ibid*).

Now, if we examine the perturbation of Δt shown in Fig. 1 (relative to its theoretical value for a simple Chapman layer) we see at once that it shows precisely such a phase opposition between low and high latitudes. Increasing or decreasing α will, of course, lower or raise the broken curve in Fig. 1, but this statement will remain true for all values of α greater than about $0{\cdot}5 \times 10^{-8}$ cm^3/sec. It might be objected that the transition latitudes indicated by Fig. 1 do not agree very precisely with those obtained for the S_q current system from geomagnetic data. Thus CHAPMAN and BARTELS (1940, Fig. 16, p. 229) show the main current focus in summer at 30° N, whereas the cross-over point in Fig. 1 lies between 20° N and 25° N (depending on the value of α chosen). There are, however, several reasons why a very close agreement in this respect is not to be expected.

First, as noted by CHAPMAN and BARTELS (*ibid.*), there is very considerable variability (of the order of 10° to 15°) in the latitude of the current focus as shown by geomagnetic data, and the average appropriate to our data may differ appreciably from the value quoted.

Secondly, the geomagnetic data refer to the integrated value of the currents throughout the whole ionosphere, whereas our results refer to the maximum of the E-layer only.

Thirdly, the perturbation we have measured will depend not so much on v as on $\partial v/\partial h$, and it is not to be expected that these will necessarily be zero at the same latitude.

Fourthly, the angle between the plane of the earth's field and that of the geographical meridian is not in general negligible and may well affect the latitude of zero drift effect.

So far as the southern (winter) hemisphere is concerned, it is clear that there is agreement at any rate to the extent that both the currents flowing and the perturbations observed are much smaller. On the other hand, since the focus in this hemisphere is near 50° S, it is doubtful if the apparent change of phase at 20° S is significant. (As there are fewer stations in operation in the southern hemisphere, our values of Δt are less reliable in this region.)

Further confirmation of the above interpretation has been obtained, using some equinox results. As has already been reported, Slough equinox diurnal measurements of fE indicate that α', as determined by (3), increases rapidly as χ increases. Now, Slough is situated north of the S_q northern-current focus, and so the ionosphere there is subject to an east-west overhead current. We have, however, found that, near the latitude of the current focus itself (say between latitudes 25° and 40°), at which location the S_q currents do not run athwart the horizontal component of the earth's magnetic field in the daytime period, the application of (3) leads to sensibly constant values of α' over a considerable range of χ. In other words, the diurnal variation of E-layer ionization appears to be most nearly Chapman-like near the S_q current focus, just where we should expect vertical drift to be at a minimum. On the other hand, corresponding results of the application of (3), at both higher and lower latitudes, are subject to marked perturbations, but of opposite sense in the two cases, expressed in the inconstancy of α'. The average value of α', which

we have obtained, as indicated above, from stations near the current focus, is $1{\cdot}0 \times 10^{-8}$ cm^3/sec; but it should be mentioned that a relatively small residual perturbation in this region (from whatever cause) might well render this value considerably in error.

The above remarks refer to the *latitude* of the S_q current focus. We have, however, been able, by further studies of the diurnal asymmetry in E-layer ionization, to identify the *longitude* of the current focus, (15° to 30° W), relative to that of the noon meridian.

In conclusion, however we may choose to interpret points of detail, it seems clear that the S_q current system exerts a small but readily detectable influence on the diurnal variation of E-layer ionization, and most probably does so by way of some form of vertical drift mechanism.

The above results relate, as stated, to the E-layer. We find that the F_1-layer exhibits somewhat similar phenomena.

Acknowledgement—This work has been supported by way of a grant from the Department of Scientific and Industrial Research and has been carried out as part of the programme of the Radio Research Board of that Department.

E. V. Appleton
A. J. Lyon
(Mrs.) A. G. Pritchard

The University,
Edinburgh

References

Appleton, E. V.	1955	Proceedings of 4th Meeting of the Mixed Commission on the Ionosphere, pp. 14 to 22, Brussels.
Appleton, E. V., and Lyon, A. J.	1955	The Physical Society, Report of 1954 Cambridge Conference on the Physics of the Ionosphere, pp. 20 to 39.
Appleton, E. V.	1937	*Proc. Roy. Soc.* **A162,** 451.
Appleton, E. V.	1953	*Journ. Atmos. Terr. Phys.* **3,** 282
Chapman, S., and Bartels, J.	1940	Geomagnetism (Oxford: at the Clarendon Press)
Martyn, D. F.	1947–8	*Proc. Roy. Soc.* **A189,** 241, and **190,** 273; and **194,** 429 and 445
	1953	*Phil. Trans.* **A246,** 281

THE PHYSICAL SOCIETY
REPRINTED FROM
THE PHYSICS OF THE IONOSPHERE—REPORT OF 1954 CAMBRIDGE CONFERENCE, p. 20

Ionospheric Layer Formation under Quasi-Stationary Conditions

By Sir EDWARD APPLETON and A. J. LYON
University of Edinburgh

Abstract. A modified theory of ionospheric layer formation is described in which the value of the spatial maximum of electron density N_m and its atmospheric height $z(N_m)$ are given explicitly. The work owes much to Chapman's classical treatment of the theory of ionospheric layer formation, but yields results which are more directly applicable to the type of problems encountered in practice. It may be used, for example, in cases where the atmospheric properties, such as temperature and recombination coefficient, vary with height, or where the layer is distorted by atmospheric tides.

It has long been recognized that the theory of ionospheric layer formation is chiefly complicated by the fact that, as the sun's zenith distance χ alters in the course of the day, the level of maximum ionization N_m lags behind the level of maximum electron production q_m. It is shown that the distance a between these two levels is equal to $(\tan\chi\, d\chi/dt)/2\alpha N_m$, where α is the recombination coefficient and t is the time.

An approximate solution of the continuity equation is used to investigate detailed features of the N, or N_m, curves near noon, and also to derive second-order approximations for a and for N_m. With the aid of these results, theoretical curves of the diurnal variations of N_m and the height at which it occurs may readily be computed for any latitude and time of year. The formulae are valid for the greater part of the solar day and include 'Chapman function' corrections where necessary.

A number of examples illustrating the significance of the 'time of relaxation' of an ionized layer $1/2\alpha N_m$ are quoted. The distorting influence of vertical drift on both N_m and $z(N_m)$ for an ionized stratum of E layer type is shown, for example, to depend on two factors only, namely the vertical drift characteristics and the time of relaxation of the ionized medium.

§ 1. Introduction

The theoretical problems examined in this paper arose in the course of an attempt to discover how far the actual behaviour of the E layer of the ionosphere could be interpreted in terms of the existing theory of ionized layer formation, particularly as classically expounded by Chapman (1931 a, b). For this purpose we used routine experimental determinations of N_m, the spatial maximum of electron density, obtained by radio sounding, to compare the actual variations of N_m in the E layer with those predicted by theory. It was soon found however that considerable discrepancies existed, suggesting that, possibly, some of the simplifying assumptions of Chapman's theory are not in fact justified at E layer heights.

These assumptions are: first, that the ionizing radiation is effectively monochromatic, being absorbed according to a mass-absorption law, by one atmospheric

constituent only; second, that the temperature and composition of the atmosphere in the region of the layer are constant both in space and time; third, that electron disappearance obeys a recombination law with a recombination coefficient α independent of atmospheric pressure; and, fourth, that the effects of electron drift whether due to atmospheric tides, diffusion or meteorological causes, are negligible. Clearly the non-fulfilment of some of these conditions may well account for the observed discrepancies.

A major difficulty impeding our investigation was, however, the prohibitive labour involved in computing accurate theoretical diurnal curves of N_m by the methods used by Chapman (and subsequently by Wilkes (1939)). Without such curves the precise assessment of the discrepancies in question is impossible. The difficulties are even greater when we desire to know exactly what perturbations will result if one or other of the above assumptions breaks down—if, for example, variability of temperature or recombination coefficient or the effects of atmospheric tides have to be allowed for.

Our primary aim is therefore to derive simple approximate formulae for N_m, and the corresponding height $z(N_m)$, valid at least for the greater part of the day, sufficiently accurate for comparison with current experimental data, reasonably convenient for computation, and yet flexible enough to permit modifications for the various perturbing factors (such as tides, temperature variations, dependence of α on pressure, etc.) which we anticipate will require investigation.

§2. The Essential Features of Chapman's Theory

As stated above, Chapman (1931 a, b) has considered the effect of ionizing solar photons, impinging on a rotating atmosphere, in which the electron disappearance rate is proportional to the square of the electron population density. The basic relation of the theory, for any point in the atmosphere, is therefore the continuity equation

$$dN/dt = q - \alpha N^2, \qquad \ldots\ldots(1)$$

where N is the electron density, q is the rate of production of electrons per unit volume per second and t the time.

For a plane-stratified, isothermal atmosphere (in which the density of the ionizable constituent falls off exponentially with height) Chapman (1931 a) has shown that the relation between rate of electron production q, height h and solar zenith distance χ is given by

$$q = q_0 \exp(1 - z - e^{-z} \sec \chi), \qquad \ldots\ldots(2)$$

where q_0 is the maximum rate of electron production, occurring at height h_0, say, when the sun's rays are vertical ($\chi = 0$), and where

$$z = (h - h_0)/H, \qquad \ldots\ldots(3)$$

H being the scale height of the atmospheric component ionized.

The spatial maximum of q at any time, and the height at which it occurs, are given by

$$q_m = q_0 \cos \chi \qquad \ldots\ldots(4)$$

and

$$z(q_m) = \ln(\sec \chi), \qquad \ldots\ldots(5)$$

whilst the particle density of the atmospheric component ionized, at the same height, is

$$n_m = \frac{\cos\chi}{AH}, \qquad \ldots\ldots(6)$$

where A is the absorption cross section of the component in question.

These equations are generally sufficient for low and medium values of χ but, for the higher range of values (e.g. exceeding 70°), in order to take into account the sphericity of the earth and its atmosphere, the term sec χ in (2) must be replaced by the Chapman function, $\mathrm{Ch}(R+z, \chi)$, where R is defined by

$$R = (a+h_0)/H, \qquad \ldots\ldots(7)$$

a being the radius of the earth and H, as before, the scale height. Putting $R+z=x$, Chapman (1931 b) has proved that

$$\mathrm{Ch}(x, \chi) = x\sin\chi\int_0^{\chi} \exp\{x(1-\sin\chi/\sin\lambda)\}\,\mathrm{cosec}^2\lambda\, d\lambda. \qquad \ldots\ldots(8)$$

Comprehensive tables of this function have recently been given by Wilkes (1954). However when x is large—as it usually is in practical cases—two useful approximations are available:

(a) for $\chi = 90°$, $\qquad \mathrm{Ch}(x, \tfrac{1}{2}\pi) \simeq (\tfrac{1}{2}\pi x)^{1/2} \qquad \ldots\ldots(9)$

(b) for χ not too large ($\tan^2\chi \ll x$)

$$\mathrm{Ch}(x, \chi) \simeq \sec\chi\left(1-\frac{\tan^2\chi}{x}\right). \qquad \ldots\ldots(10)$$

Comparison with the accurate tables computed by Wilkes (1954) shows that (9) is correct to within 1% for $x>50$, while (10) is within 1% (a) for $x=600$ (a representative value for the E layer) if $\chi \gtrsim 80°$ and (b) for $x=150$ (a possible value for the F layer) if $\chi \gtrsim 70°$.

In practical cases, $R \gg z$, so that R and x are very nearly equal. We may therefore write equations (2), (4), (5) and (6) as

$$q = q_0 \exp\{1-z-\mathrm{e}^{-z}\mathrm{Ch}(R, \chi)\}, \qquad \ldots\ldots(11)$$

$$q_m = q_0/\mathrm{Ch}(R, \chi) \qquad \ldots\ldots(12)$$

$$\simeq q_0\cos\chi\left(1+\frac{\tan^2\chi}{R}\right), \qquad \ldots\ldots(12\,a)$$

$$z(q_m) = \ln \mathrm{Ch}(R, \chi) \qquad \ldots\ldots(13)$$

$$\simeq \ln\sec\chi - \frac{\tan^2\chi}{R}, \qquad \ldots\ldots(13\,a)$$

and

$$n_m = 1/AH\,\mathrm{Ch}(R, \chi). \qquad \ldots\ldots(14)$$

The approximations (12 a) and (13 a) are valid under the conditions mentioned above.

Substituting for q from (2) or (11) in the continuity equation (1), Chapman showed that this equation can be solved numerically for particular values of z, giving a family of (N, t) curves at various heights. From these a second family of (N, z) or 'layer profile' curves at various times may then be derived. Finally, from these curves the values of N_m and $z(N_m)$ at any time of the day can be read off. Evidently however this is a lengthy and tedious process, and attempts have been made by Millington (1932, 1935) to simplify the necessary procedure for practical calculations of N_m. In his first paper Millington, noting that the (N_m, t) curve

is in fact the envelope of all the (N, t) curves, pointed out that, approximately, this "envelope is given by the curve corresponding to the height at which the rate of ionization is a maximum at noon". In other words, if χ_0 is the value of χ at noon, the envelope is considered to be identical with the (N, t) curve for the height z_0 where

$$z_0 = \ln \mathrm{Ch}(R, \chi_0) \qquad \ldots\ldots(15)$$

so that

$$q = q_0 \exp\{1 - z_0 - \exp(-z_0)\,\mathrm{Ch}(R, \chi)\}$$
$$= \frac{q_0}{\mathrm{Ch}(R, \chi)} \exp\left\{1 - \frac{\mathrm{Ch}(R, \chi)}{\mathrm{Ch}(R, \chi_0)}\right\}. \qquad \ldots\ldots(16)$$

Only one numerical solution of (1) is now required, using the values of q given by (16).

In his second paper, Millington (1935) used an even simpler expression in place of (16), namely

$$q = q_0/\mathrm{Ch}(R, \chi). \qquad \ldots\ldots(17)$$

Comparison with (12) shows that this in fact involves the assumption that the maximum of electron density N_m occurs at the level of the maximum electron production rate q_m. Now this, as may be seen from Chapman's curves, cannot strictly be the case. In figure 3(*b*) below, for example, the crosses show values of $z(N_m)$, the height of maximum N, at various times throughout the day, derived from a set of layer profiles given by Chapman (1931 a), whilst the dotted line shows the corresponding curve of $z(q_m)$, the height of maximum q. It is clear that, except near noon, the N_m level lags appreciably behind the q_m level, being higher in the forenoon and lower in the afternoon. One of the chief problems examined in this paper is that of determining the magnitude of this height-lag, and of improving Millington's approximation (17) in such a way as to take such height-lag into account.

§ 3. The Diurnal Variation of Ionization in a Layer at Constant Height

Before proceeding to our main task of evaluating N_m, we have found it instructive to find first an approximate solution of (1) which will be valid provided only that q and N do not vary too rapidly. Recognizing that q and N are functions of both z and t, we must now write (1) as

$$\frac{\partial N(z, t)}{\partial t} = q(z, t) - \alpha\{N(z, t)\}^2. \qquad \ldots\ldots(18)$$

We can, of course, reduce this to an equation in one independent variable t by restricting our attention, as Chapman did, to a particular height so that (18) becomes

$$dN/dt = q(t) - \alpha N^2, \qquad \ldots\ldots(19)$$

which has now the well-known form of a Riccati equation.

It is possible to achieve the same result by making z a suitable function of t. Thus, in general,

$$\frac{dN}{dt} = \frac{\partial N}{\partial t} + \frac{\partial N}{\partial z}\frac{dz}{dt},$$

and if we put $z = z(N_m)$, that is make z follow the N_m level where $\partial N/\partial z = 0$ we find that

$$\frac{dN_m}{dt} = \left(\frac{\partial N}{\partial t}\right)_{z = z(N_m)}$$
$$= q\{z(N_m), t\} - \alpha N_m^2. \qquad \ldots\ldots(20)$$

Since $z(N_m)$ is a function of t this equation is identical in form with (19). In other words, N_m, the spatial maximum of N, obeys the continuity equation, provided only that q is evaluated at the level of N_m. Hence all the results of this section, while applying primarily to N at a fixed height, will also be valid for N_m.

We are now in a position to obtain a solution of (19), or of its equivalent form (20), by the method of successive approximations. As was pointed out by Appleton (1937), it is true, at any rate in the case of the E layer, that the numerical magnitude of dN/dt in (19) is small compared with the numerical magnitudes of the other two terms; and, of course, near noon this is notably the case. We may therefore take as our zero-order approximate solution of (19)

$$N = \left(\frac{q(t)}{\alpha}\right)^{1/2}. \qquad \ldots\ldots (21)$$

Differentiating this equation with respect to t and inserting the resulting expression for dN/dt in (19) yields the first-order approximation

$$N = \left(\frac{q(t)}{\alpha}\right)^{1/2} - \frac{1}{4\alpha}\frac{q'(t)}{q(t)} \qquad \ldots\ldots (22)$$

and, similarly, we can calculate the second-order approximation†

$$N = \left(\frac{q(t)}{\alpha}\right)^{1/2} - \frac{1}{4\alpha}\frac{q'(t)}{q(t)} + \frac{1}{8\alpha^{3/2}}\frac{q''(t)}{\{q(t)\}^{3/2}} - \frac{5}{32\alpha^{3/2}}\frac{\{q'(t)\}^2}{\{q(t)\}^{5/2}}. \qquad \ldots\ldots (23)$$

This equation enables us to demonstrate, *inter alia*, certain interesting and important features of the variation of N (or N_m) near noon.

3.1. *Characteristics of the Diurnal Curves of N, or N_m, near Noon*

At noon $q'(t)$ is zero, and (23) becomes

$$N_0 = \left(\frac{q_0(t)}{\alpha}\right)^{1/2} + \frac{q_0''(t)}{8\alpha^{3/2}\{q_0(t)\}^{3/2}} \qquad \ldots\ldots (24)$$

where the subscript indicates noon conditions. Since, in practice, $q_0''(t)$ is negative we note that N_0 is less than the quasi-stationary value relating to noon, namely $[\{q_0(t)\}/\alpha]^{1/2}$. Let us now suppose that the value of the N maximum occurs τ seconds after noon, the time of q maximum. At that time the value of $q(t)$ is given by

$$q(t) = q_0(t) + q_0'(t)\tau + \tfrac{1}{2}q_0''(t)\tau^2. \qquad \ldots\ldots (25)$$

However we know that $q_0'(t)$ is zero, so that

$$q(t) = q_0(t) + \tfrac{1}{2}q_0''(t)\tau^2. \qquad \ldots\ldots (26)$$

Now, since at N maximum, dN/dt is zero, we can substitute from (26) for $q(t)$ in (19) to find

$$N_{max} = \left(\frac{q_0(t)}{\alpha}\right)^{1/2}\left[1 + \frac{1}{2}\frac{q_0''(t)\tau^2}{q_0(t)}\right]^{1/2} \qquad \ldots\ldots (27)$$

or, approximately,

$$N_{max} = \left(\frac{q_0(t)}{\alpha}\right)^{1/2} + \frac{q_0''(t)\tau^2}{4\alpha^{1/2}\{q_0(t)\}^{1/2}}. \qquad \ldots\ldots (28)$$

Now it has already been shown by Appleton (1953) that τ^2 is equal to $1/4\alpha q_0(t)$ so that, substituting in (28), we have

$$N_{max} = \left(\frac{q_0(t)}{\alpha}\right)^{1/2} + \frac{q_0''(t)}{16\alpha^{3/2}\{q_0(t)\}^{3/2}}. \qquad \ldots\ldots (29)$$

† The validity of this process and the magnitude of the residual errors are discussed in the Appendix. The same kind of result can be obtained, as Rydbeck and Wilhelmsson (1954) have shown, using the Jeffreys or W.K.B. solution of the wave equation, into which equation (19) can be transformed.

We therefore have, in (24) and (29), expressions giving us, respectively, the values of N at noon and at the time of the subsequent N maximum. We note that they differ from the noon quasi-stationary value $[\{q_0(t)\}/\alpha]^{1/2}$ by simply related quantities.

These results are illustrated graphically in figure 1 where the continuous line indicates the variation of N, or N_m, near noon. The dotted line indicates the time variation of the quasi-stationary quantity $[\{q(t)\}/\alpha]^{1/2}$ which reaches its maximum value $[\{q_0(t)\}/\alpha]^{1/2}$ at noon. This corresponds to point C.

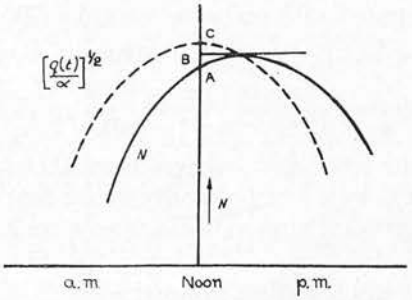


Figure 1. Illustrating variation of N with time near noon. The continuous curve indicates the trend of N which is a maximum slightly after noon. The broken curve indicates the variation of the quasi-stationary value $[\{q(t)\}/\alpha]^{1/2}$. The curves do not relate to actual values, the noon delay of N being exaggerated to illustrate the physical points discussed in the text.

The actual noon value of N is shown by A, while its subsequent maximum value is shown at an intermediate level indicated by B. Now from (24) and (29) we can see at once that

$$\mathrm{AC} = \frac{q_0''(t)}{8\alpha^{3/2}\{q_0(t)\}^{3/2}} \qquad \ldots\ldots(30)$$

and

$$\mathrm{AB} = \frac{q_0''(t)}{16\alpha^{3/2}\{q_0(t)\}^{3/2}} \qquad \ldots\ldots(31)$$

so that

$$\mathrm{AB} = \tfrac{1}{2}(\mathrm{AC}) = \mathrm{BC}.$$

As already explained this result will be valid either for N at any constant height or for the spatial maximum N_m. If, therefore, we know the values of N_m at noon and at the time of maximum N_m we can calculate the quasi-stationary value at noon and from this the value of the fundamental physical quantity $\{q_0(t)\}/\alpha$.

§4. The Height and Magnitude of Maximum Ionization Density N_m

The condition for a spatial maximum of N at any particular time t (which we denote by N_m) is, of course,

$$\frac{\partial N(z,t)}{\partial z} = 0, \qquad \ldots\ldots(32)$$

and if the form of $N(z,t)$ were known this would determine the height of N_m which we denote by $z(N_m)$. Now differentiating the continuity equation (19) with respect to z we have

$$\frac{\partial^2 N}{\partial z \partial t} = \frac{\partial q}{\partial z} - 2\alpha N \frac{\partial N}{\partial z}, \qquad \ldots\ldots(33)$$

so that, for maximum N conditions

$$\frac{\partial q}{\partial z} = \frac{\partial^2 N}{\partial z \partial t}. \qquad \ldots\ldots(34)$$

If circumstances are such that we can neglect the second differential on the right-hand side of (34) we have $\partial q/\partial z$ equal to zero, so that, in such a case, the spatial maximum of q occurs at the same level as that of N: that is, $z(q_m)$ and $z(N_m)$ are identical.

Let us suppose, however, that the N maximum does lag appreciably behind the q maximum, in which case, assuming that q is given by (2) and so $z(q_m)$ by (5), we may write

$$z(N_m) = \ln \sec \chi - a. \qquad \ldots\ldots(35)$$

Then, if a is small compared with unity we have, by (2),

$$\begin{aligned} q\{z(N_m), t\} &= q_0 \exp\{1 - \ln \sec \chi + a - e^a\} \\ &= q_0 \cos \chi \exp(-\tfrac{1}{2}a^2 - \ldots) \\ &\simeq q_0 \cos \chi (1 - \tfrac{1}{2}a^2). \qquad \ldots\ldots(36) \end{aligned}$$

In other words if the level $z(N_m)$ differs from the quasi-stationary level (as we may term it), ln sec χ, by a small quantity of the first order, the value of q at that level differs from its maximum value, $q_0 \cos \chi$, by a quantity of the second order only.

It follows from (20) and (36) that, approximately,

$$\frac{dN_m}{dt} = q_0 \cos \chi (1 - \tfrac{1}{2}a^2) - \alpha N_m^2. \qquad \ldots\ldots(37)$$

We can now state with more precision the condition under which the continuity equation reduces to the convenient approximate form used by Millington (1935) namely

$$dN_m/dt = q_0 \cos \chi - \alpha N_m^2 \qquad \ldots\ldots(38)$$

(cf. (17)—we here ignore the refinement of using the Chapman function). The required condition, by (36), is simply that

$$a^2/2 \ll 1. \qquad \ldots\ldots(39)$$

To make this condition still more precise and to enable us to determine the necessary correction when it is not fulfilled, we must calculate the magnitude of the height-lag a. We shall see, in §4.2, that this can be done, analytically, to any required degree of accuracy, by the use of equation (34) and the approximations of §3. In the following section, however, we shall derive the first approximation for a by a more intuitive method which, we believe, brings out very clearly the physical significance of the height-lag phenomenon.

4.1. *A First Estimate of the Magnitude of the Height-Lag a*

Suppose first that at some arbitrary time $t = 0$ the electron production rate q is a maximum (i.e. $\partial q/\partial z = 0$) at a height $z = z_0$ say, and suppose further that after Δt seconds the electron density N is a maximum (i.e. $\partial N/\partial z = 0$) at the same height. The situation is illustrated, diagrammatically, in figure (2) drawn for a case when both the N_m and q_m levels are descending, i.e. for some time before noon.

If Δt is a small interval then at $z = z_0$ and $t = \Delta t$, we have

$$\frac{\partial N}{\partial z} = 0 = \left(\frac{\partial N}{\partial z}\right)_0 + \left(\frac{\partial^2 N}{\partial z \partial t}\right)_0 \Delta t, \qquad \ldots\ldots(40)$$

where the subscript denotes $z = z_0$ and $t = 0$. Now by (1),

$$\left(\frac{\partial^2 N}{\partial z \partial t}\right)_0 = -2\alpha N \left(\frac{\partial N}{\partial z}\right)_0 \qquad \ldots\ldots(41)$$

and hence, by (40),

$$\Delta t = 1/2\alpha N. \qquad \ldots\ldots(42)$$

Now (42), giving the delay between the occurrence of the *spatial* maxima of q and N at a fixed height but at any time of day, is a similar expression to that found by Appleton (1953) for the delay between the *temporal* maxima of the same quantities near noon. Evidently both phenomena, and the height-lag, are manifestations of the same physical characteristic, called by Appleton (1953) the 'sluggishness' of the ionosphere, and they all result from the finite magnitude of the recombination coefficient α. If α were infinite Δt would be zero and there would be no height-lag—the N_m and q_m levels would be identical.

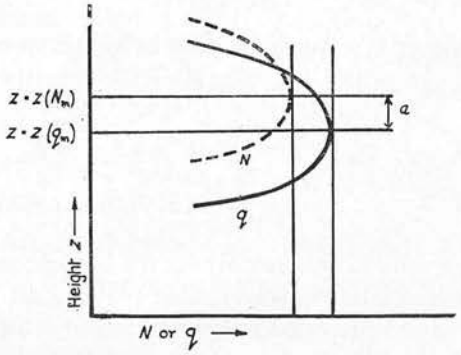


Figure 2. Illustrating the phenomenon of height lag for forenoon conditions. The height of maximum production rate q_m is at $z=z(q_m)$, while the height of maximum electron density N_m is at $z=z(N_m)$. After $1/2\alpha N_m$ seconds the N maximum has dropped through the vertical distance $a=z(N_m)-z(q_m)$. The curves are diagrammatic only.

Equation (42) enables us to determine a, for the height-lag is simply, to a first approximation, the delay-time Δt multiplied by the speed of descent of the q_m level, that is

$$a \simeq \frac{d\{z(q_m)\}}{dt}\Delta t \qquad \ldots\ldots(43)$$

$$\simeq \frac{d}{dt}(\ln \sec \chi)\frac{1}{2\alpha N_m},$$

by (42), or

$$a \simeq \frac{\tan \chi d\chi/dt}{2\alpha N_m}. \qquad \ldots\ldots(44)$$

Equation (44) shows that a is negative in the morning and positive in the afternoon. (It is convenient to treat χ as negative and $d\chi/dt$ positive before noon.) This means, by (35), that $z(N_m)$ is, as we should expect, above the q_m level before noon and below it after. It is also evident from (44) that as $\chi \to 90°$ our expression for a tends to infinity—and, of course, the conditions for its validity will no longer be satisfied.

The first approximation for N_m, derived from (37), is

$$N_m \simeq N_0(\cos \chi)^{1/2} \qquad \ldots\ldots(45)$$

where

$$N_0 = (q_0/\alpha)^{1/2}, \qquad \ldots\ldots(46)$$

and hence

$$\frac{dN_m}{dt} = -\frac{1}{2}N_m \tan \chi \frac{d\chi}{dt}, \qquad \ldots\ldots(47)$$

so that (44) may also be written

$$a \simeq -\frac{dN_m/dt}{\alpha N_m^2} \simeq -\frac{dN_m/dt}{q}. \qquad \ldots\ldots(48)$$

If further we restrict ourselves to the equator, at an equinox, so that

$$\chi = \omega t, \qquad \ldots\ldots(49)$$

where $\omega = 15°/\mathrm{h} = 7{\cdot}27 \times 10^{-5}$ rad/sec,
then, using Chapman's notation

$$\sigma_0 = \omega/\alpha N_0 \qquad \ldots\ldots(50)$$

(44) becomes

$$a \simeq \frac{\sigma_0 \tan\chi}{2(\cos\chi)^{1/2}}. \qquad \ldots\ldots(51)$$

The condition for Millington's approximation to hold is now, from (39) and (48),

$$\frac{1}{2}\left(\frac{dN_\mathrm{m}/dt}{\alpha N_\mathrm{m}}\right)^2 \ll 1, \qquad \ldots\ldots(52)$$

or, using (51),

$$\frac{\sigma_0^2 \tan^2\chi}{8\cos\chi} \ll 1. \qquad \ldots\ldots(53)$$

All these results are valid of course only for the isothermal conditions assumed in (2). It will be proved later, however, that (44) is still true when there is a linear temperature gradient provided a is measured on a suitable scale.

4.2. *The Second-order Approximation for the Height-Lag a*

To obtain a closer approximation for the quantity a we begin with the fundamental condition (34) which defines the N_m level in relation to the q_m level, that is

$$\frac{\partial q}{\partial z} = \frac{\partial^2 N}{\partial z \partial t}.$$

To evaluate the right-hand side of this equation with the required degree of accuracy it is sufficient to use (22), that is

$$N = \left(\frac{q}{\alpha}\right)^{1/2} - \frac{1}{4\alpha}\frac{q'}{q}.$$

For the present we confine our considerations to an isothermal plane-stratified atmosphere so that, as before,

$$q = q_0 \exp\{1 - z - \mathrm{e}^{-z}\sec\chi\},$$

and to the equator-equinox case defined by $\chi = \omega t$. Hence substituting in (22) for q and q' we have

$$N = \left(\frac{q}{\alpha}\right)^{1\,2} + \frac{\omega}{4\alpha}\mathrm{e}^{-z}\sec\chi\tan\chi$$

so that

$$\frac{\partial N}{\partial z} = \frac{1}{2}\left(\frac{q}{\alpha}\right)^{1/2}(-1 + \mathrm{e}^{-z}\sec\chi) - \frac{\omega}{4\alpha}\mathrm{e}^{-z}\sec\chi\tan\chi$$

and

$$\begin{aligned}\frac{\partial^2 N}{\partial z \partial t} = {} & \frac{\omega}{2}\left(\frac{q}{\alpha}\right)^{1/2}\mathrm{e}^{-z}\sec\chi\tan\chi \\ & - \frac{\omega}{4}\left(\frac{q}{\alpha}\right)^{1/2}\mathrm{e}^{-z}\sec\chi\tan\chi(-1 + \mathrm{e}^{-z}\sec\chi) \\ & - \frac{\omega^2}{4\alpha}\mathrm{e}^{-z}\sec\chi(2\tan^2\chi + 1). \qquad \ldots\ldots(54)\end{aligned}$$

We now define the quantity ϵ as

$$\epsilon = \frac{\tan\chi}{2(q\alpha)^{1/2}} \frac{d\chi}{dt} \qquad \ldots\ldots(55)$$

and, when $\chi = \omega t$, this reduces to ϵ_0 say, where

$$\epsilon_0 = \frac{\sigma_0 \tan\chi}{2(\cos\chi)^{1/2}} \qquad \ldots\ldots(56)$$

by (36), (46) and (50).

Also, by (35) $\qquad e^{-z}\sec\chi = e^{a} \simeq 1 + a + \frac{1}{2}a^2, \qquad \ldots\ldots(57)$

so that (54) now becomes

$$\frac{\partial^2 N}{\partial z \partial t} = q\epsilon_0(1+a) - \tfrac{1}{2}q\epsilon_0(1+a)a - 2q\epsilon_0^2(1+a) - \frac{q\sigma_0^2(1+a)}{4\cos\chi}, \qquad \ldots\ldots(58)$$

while

$$\frac{\partial q}{\partial z} = q(-1 + e^{-z}\sec\chi) = q(a + \tfrac{1}{2}a^2 + \ldots). \qquad \ldots\ldots(59)$$

Thus our fundamental equation (34) becomes

$$a + \tfrac{1}{2}a^2 + \ldots = \epsilon_0(1+a) - \tfrac{1}{2}\epsilon_0 a(1+a) - 2\epsilon_0^2(1+a) - \sigma_0^2(1+a)/4\cos\chi.$$

If we neglect terms of the second order in a and ϵ_0 we obtain the first approximation

$$a \simeq \epsilon_0 \qquad \ldots\ldots(60)$$

in agreement with equation (51) of the previous section. Or if we put $a = \epsilon_0$ in second-order terms and neglect third-order terms in a and ϵ_0, we have, as our second approximation,

$$a = \epsilon_0 - 2\epsilon_0^2 - \sigma_0^2/4\cos\chi. \qquad \ldots\ldots(61)$$

4.3. *A Second-order Approximate Equation for the Maximum Electron Density* N_m

By a similar process we may determine the second-order approximation to N_m using the approximate solution (23) of the basic continuity equation (1), given in § 3 above, which we may rewrite in the form

$$N = \left(\frac{q}{\alpha}\right)^{1/2}\left\{1 - \frac{1}{4(q\alpha)^{1/2}}\frac{q'}{q} + \frac{1}{32q\alpha}\frac{4qq'' - 5q'^2}{q^2}\right\}.$$

We can evaluate q' and q'' from (2), or we can proceed directly from (36), and we get, after some rearrangement

$$N_m = N_0(\cos\chi)^{1/2}\{1 + \tfrac{1}{2}\epsilon_0 - \tfrac{7}{8}\epsilon_0^2 - \sigma_0^2/8\cos\chi\}, \qquad \ldots\ldots(62)$$

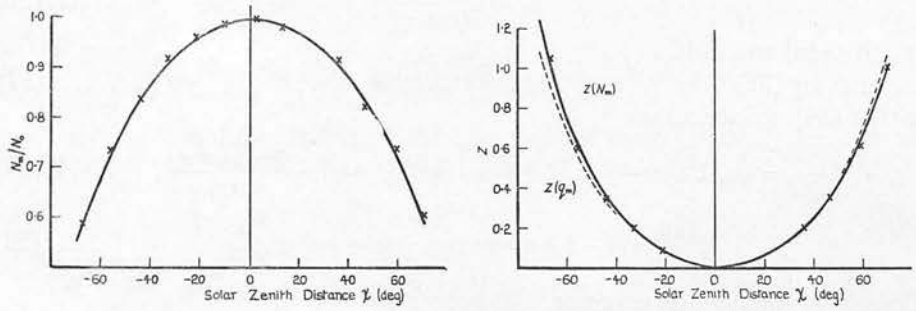
where, as before, ϵ_0 and σ_0 are given by (56) and (50) respectively. Also from (5), (35) and (61)

$$z(N_m) = \ln\sec\chi - \epsilon_0 + 2\epsilon_0^2 + \sigma_0^2/4\cos\chi. \qquad \ldots\ldots(63)$$

These two equations, (62) and (63), provide us with the formulae we set out to find, at any rate for the simplest case, that is they enable us to compute, very easily, diurnal curves of N_m and $z(N_m)$ over a substantial part of the solar day. The actual range of validity depends on the value of σ_0, and this matter is discussed in some detail in the Appendix. The equations will perhaps be most useful when the ϵ^2 terms are negligible, say for $|\epsilon| < \frac{1}{10}$. For the E layer if $\alpha \simeq 10^{-8}$, $\sigma_0 \simeq 0{\cdot}04$ and $|\epsilon_0| < \frac{1}{10}$ when $\chi < 70°$. A higher level of accuracy is possible with a smaller range of χ.

To illustrate the use of the equations (62) and (63) we have used them to calculate N_m and $z(N_m)$ with the above value, $\sigma_0 = 0{\cdot}04$. This value is one also used by Chapman in his classical investigations and enables us to compare his

results with ours. This is done in figure 3 where the smooth curves have been plotted from our computed values and the crosses indicate the values we have derived from Chapman's curves in the manner already described in § 2. It will be seen that the agreement is excellent: any small discrepancies can be attributed to the difficulty of using printed curves.



(*a*) (*b*)

Figure 3. (*a*) Day-time trend of N_m as calculated from the approximate formula (62). (*b*) Day-time trend of $z(N_m)$ as calculated from the approximate formula (63). In each case the crosses indicate values of the same quantity derived from Chapman's curves. To illustrate the phenomenon of height-lag, the levels of maximum electron production $z(q_m)$ are indicated by the broken line in (*b*).

Equation (62) also enables us to calculate the delay of maximum N_m after noon. Differentiating (62) yields the result

$$\Delta t = 1/2\alpha N_0, \qquad \ldots\ldots(64)$$

and so $$(N_m)_{max} = N_0(1-\sigma_0^2/16),$$

whilst the value of N_m at noon (i.e. $\chi = 0$) is

$$(N_m)_0 = N_0(1-\sigma_0^2/8).$$

It follows that with the notation of figure 1

$$AB = BC = \sigma_0^2/16 \qquad \ldots\ldots(65)$$

in agreement with the more general results of § 3.1.

4.4. *Modifications for any Season or Latitude*

To modify the above analysis for the general case, we must now write

$$\cos\chi = \sin\phi\sin\delta + \cos\phi\cos\delta\cos\omega t$$

whence

$$\tan\chi\frac{d\chi}{dt} = \omega\cos\phi\cos\delta\frac{\sin\omega t}{\cos\chi}. \qquad \ldots\ldots(66)$$

If we now repeat the analysis of §§ 4.2 and 4.3 using (66) we readily find that

$$N_m = N_0(\cos\chi)^{1/2}\{1+\tfrac{1}{2}\epsilon-\tfrac{7}{8}\epsilon^2-\tfrac{1}{2}A\} \qquad \ldots\ldots(67)$$

and $$z(N_m) = \ln\sec\chi - \epsilon + 2\epsilon^2 + A, \qquad \ldots\ldots(68)$$

where $$\epsilon = \tfrac{1}{2}\sigma_0\cos\phi\cos\delta\sin\omega t/\cos^{3/2}\chi \qquad \ldots\ldots(69)$$

and $$A = \tfrac{1}{4}\sigma_0^2\cos\phi\cos\delta\cos\omega t/\cos^{3/2}\chi. \qquad \ldots\ldots(70)$$

If tables of $\cos\chi$ as a function of solar hour angle are available, or once these have been computed, these equations enable us to calculate N_m and $z(N_m)$ for any latitude at any time of the year with no more labour than is required for the simple case $\chi = \omega t$.

4.5. *Modifications for Sphericity of Layer*

As we have seen, it is necessary, for a spherically stratified atmosphere, to replace sec χ in the basic equation (2) by the appropriate Chapman function $\mathrm{Ch}(R, \chi)$. For the normal range of validity of our results the approximate formula (10) for $\mathrm{Ch}(R, \chi)$ will usually be adequate so that

$$\frac{\partial}{\partial\chi}\{\mathrm{Ch}(R,\chi)\}=\sec\chi\tan\chi\left\{1-\frac{3\tan^2\chi+2}{R}\right\}. \qquad \ldots\ldots(71)$$

Equations (67) and (68) now become

$$N_{\mathrm{m}}=N_0(\cos\chi)^{1/2}\left\{1+\frac{\tan^2\chi}{2R}+\frac{1}{2}\epsilon\left[1-\frac{3\tan^2\chi+2}{R}\right]-\frac{7}{8}\epsilon^2-\frac{1}{2}A\right\}, \qquad \ldots\ldots(67\,a)$$

$$z(N)_{\mathrm{m}}=\ln\sec\chi-\frac{\tan^2\chi}{R}-\epsilon\left[1-\frac{3\tan^2\chi+2}{R}\right]+2\epsilon^2+A. \qquad \ldots\ldots(68\,a)$$

Alternatively, the appropriate correction factors can be derived from accurate tables of the Chapman function such as have been given by Wilkes (1954).

§ 5. Some Applications of the Preceding Analysis

The formulae derived in the foregoing sections solve the first part of our problem, that is, they make it possible to compute directly diurnal curves of N_{m} and $z(N_{\mathrm{m}})$ according to Chapman's classical theory of ionized layer formation. Accurate comparison of the theoretical and observed variations of these quantities becomes therefore a relatively simple matter. The remaining part of our problem is to find out how the theoretical curves of N_{m} and $z(N_{\mathrm{m}})$ would be altered if we now disallow one or other of the underlying assumptions of the theory, listed in § 1. We shall consider three such cases: (*a*) temperature varying with height, (*b*) recombination coefficient α varying with height and (*c*) the presence of vertical electronic drift due to tidal motions of the atmosphere.

5.1. *The Effects of Variation of Temperature with Height*

Nicolet (1951), and others, have examined the modifications of Chapman's theory when the scale height (or temperature) varies linearly with height. It is useful however to examine the matter somewhat more generally. Let us suppose that the scale height H of the absorbing gas, given in the usual notation by

$$H=\boldsymbol{k}T/\boldsymbol{mg},$$

varies quite arbitrarily with the height h. Since H is variable we must now replace z by the analogous height parameter ζ defined by

$$d\zeta=dh/H. \qquad \ldots\ldots(72)$$

It readily follows from this that

$$nH=n_1H_1\mathrm{e}^{-\zeta} \qquad \ldots\ldots(73)$$

where n is the particle density of the ionizable constituent and the subscript refers to an arbitrary reference level. Chapman's analysis now yields the results

$$q=nAS=[n_1AH_1S_\infty\mathrm{e}^{-\zeta}/H]\exp\{-n_1AH_1\sec\chi\,\mathrm{e}^{-\zeta}\}, \qquad \ldots\ldots(74)$$

$$\zeta(q_{\mathrm{m}})=\ln\frac{n_1AH_1}{1+\lambda}+\ln\sec\chi \qquad \ldots\ldots(75)$$

and

$$q_{\mathrm{m}}=\frac{S_\infty(1+\lambda)\cos\chi}{H\exp(1+\lambda)} \qquad \ldots\ldots(76)$$

where S is the photon flux and λ the scale height gradient, i.e.

$$\lambda = \frac{dH}{dh} = \frac{1}{H}\frac{dH}{d\zeta}. \qquad \ldots\ldots(77)$$

In the above results λ is not assumed constant so that both H and λ are effectively functions of χ. We shall now suppose however that for small enough values of ζ, λ may be regarded as constant and equal to λ_1, say. Then, by (77),

$$H = H_1 \exp(\lambda_1 \zeta). \qquad \ldots\ldots(78)$$

Let us also suppose that $\zeta(q_m)$ is zero for $\chi = \chi_1$, and equations (74), (75) and (76) now take the simple forms

$$q = q_1 \exp\{(1+\lambda_1)(1-\zeta-e^{-\zeta}\sec\chi/\sec\chi_1)\}, \qquad \ldots\ldots(79)$$

$$\zeta(q_m) = \ln\sec\chi - \ln\sec\chi_1 \qquad \ldots\ldots(80)$$

and

$$q_m = q_1(\cos\chi/\cos\chi_1)^{1+\lambda_1} \qquad \ldots\ldots(81)$$

where

$$q_1 = S_\infty(1+\lambda_1)\cos\chi_1/H_1\exp(1+\lambda_1). \qquad \ldots\ldots(82)$$

Provided that any variations of λ are relatively slow, we can cover the whole range of χ by taking a series of reference levels (defined by corresponding values of χ_1) near which λ is sensibly constant. If λ is constant throughout the region, then we can take χ_1 equal to zero and equations (78) to (82) become identical with the results obtained by Nicolet. In that case we evidently have, from (78),

$$\lambda\zeta = \ln(1+\lambda z), \qquad \ldots\ldots(83)$$

so that ζ measures height in units of a constant scale height, H_0 say, as z does, but on a logarithmic scale defined by (83).

From equations (43) and (75) we can obtain a generalized expression for the height-lag in units of ζ, namely,

$$a(\zeta) = \frac{\tan\chi\, d\chi/dt}{2\alpha N_m\{1+(1+\lambda)^{-1}d\lambda/d\zeta\}}. \qquad \ldots\ldots(84)$$

If λ is constant this becomes identical in form with (44). In that case, using (83), the magnitude of the height-lag may be expressed in terms of the linear height parameter z by the relation

$$a(z) = (\sec\chi)^{\lambda} a(\zeta). \qquad \ldots\ldots(85)$$

Thus the presence of a temperature gradient may appreciably increase the magnitude of the height-lag for high values of χ.

The second-order approximations for N_m and $z(N_m)$ near the arbitrary reference level associated with χ_1 are now

$$N_m = N_0(\cos\chi/\cos\chi_1)^{(1+\lambda_1)/2}\{1+(1+\lambda_1)[\tfrac{1}{2}\epsilon - \tfrac{7}{8}\epsilon^2 - \tfrac{1}{2}A]\}, \qquad \ldots\ldots(67\,b)$$

$$\zeta(N_m) = \ln\sec\chi - \ln\sec\chi_1 - \epsilon + (2+\tfrac{1}{2}\lambda_1)\epsilon^2 + A. \qquad \ldots\ldots(68\,b)$$

Thus while the expression for $\zeta(N_m)$ is virtually unaffected by the temperature gradient, both the exponent of $\cos\chi$ and the correction terms in the formula for N_m are increased by the factor $(1+\lambda_1)$.

Corrections for a varying λ, say for

$$\lambda = \lambda_1 + \lambda'\zeta$$

can be calculated from the general equations (74) to (77) if required.

5.2. *The Effects of Variation of Recombination Coefficient with Height*

If α is a function of height throughout the region of ion production the approximate solutions of the continuity equation at a constant height, equations (21),

(22) and (23), will still be valid for an appropriate part of the solar day, but the variation of α will cause changes in the height-distribution of N which may completely dwarf the distortion due to height-lag. In such a case the difference in height between the levels of N_m and q_m can no longer be treated as a small quantity even around noon. The magnitude of the distortion can be estimated from equation (21) which we write now as

$$N=\left[\frac{q_0 \exp(1-z-e^{-z}\sec\chi)}{\alpha(z)}\right]^{1/2}. \qquad \ldots\ldots(86)$$

For example, if α were directly proportional to atmospheric pressure, say

$$\alpha=\alpha_0 e^{-z} \qquad \ldots\ldots(87)$$

then

$$N=(q_0/\alpha_0)^{1/2}\exp\tfrac{1}{2}(1-e^{-z}\sec\chi). \qquad \ldots\ldots(88)$$

In this case, for a given χ, N increases monotonically with z and has no true spatial maximum. At a given height N has a temporal maximum at noon, but at the q_m level given by

$$z(q_m)=\ln\sec\chi$$

we have $N=(q_0/\alpha_0)^{1/2}$, independent of χ.

It is of some interest to consider whether if α is a function of height the method given by Appleton (1937) for the experimental determination of α remains valid or not. According to this method we can find both α and q_0 by comparing conditions for the same χ in the forenoon and afternoon, provided we assume that q is symmetrical about noon. If we take the height-lag into account, clearly the heights of the N_m level in the two cases will now differ by $2a$. Naturally if α changes appreciably within this distance, the above method is no longer applicable.

If however α is substantially constant over the height range $2a$, equation (36) shows that (neglecting the cube and higher powers of a) q remains symmetrical in χ despite the asymmetry in height, and the continuity equation (37) again yields Appleton's formula for α, namely

$$\alpha=\frac{(dN_m/dt)_1-(dN_m/dt)_2}{N_2^2-N_1^2}. \qquad \ldots\ldots(89)$$

5.3. *The Effect of Vertical Drift on the Height and Density of the Layer Maximum*

Martyn (1947 a, b, 1948) has suggested that ionospheric layers may be subject to distortion due to the influence of what he terms vertical drift. The general effect of this is to distort the layer in vertical section if the drift velocity varies with height. In such a case our continuity equation (1) becomes

$$\frac{\partial N}{\partial t}=q-\alpha N^2-\frac{\partial(Nv)}{\partial z}, \qquad \ldots\ldots(90)$$

where v is the vertical drift velocity at height z. Expanding the drift velocity term we have

$$\frac{\partial N}{\partial t}=q-\alpha N^2-N\frac{\partial v}{\partial z}-v\frac{\partial N}{\partial z}$$

and thus

$$\frac{\partial^2 N}{\partial z\partial t}=\frac{\partial q}{\partial z}-2\alpha N\frac{\partial N}{\partial z}-\frac{\partial N}{\partial z}\frac{\partial v}{\partial z}-N\frac{\partial^2 v}{\partial z^2}-\frac{\partial v}{\partial z}\frac{\partial N}{\partial z}-v\frac{\partial^2 N}{\partial z^2}.$$

At the spatial maximum of N, we have $\partial N/\partial z$ equal to zero, and thus

$$\frac{\partial^2 N}{\partial z \partial t} = \frac{\partial q}{\partial z} - N\frac{\partial^2 v}{\partial z^2} - v\frac{\partial^2 N}{\partial z^2} \qquad \ldots\ldots(91)$$

which, of course, reduces to (34) if there is no vertical drift.

How we proceed now depends on the form of v assumed. For example, in an attempt to explain the actual departure of E layer behaviour from the static layer theory predictions of Chapman, Kirkpatrick (1948) has assumed

$$v = -v_0 \sin(2\omega t + \beta z) \qquad \ldots\ldots(92)$$

where, numerically, $v_0 = 6{\cdot}25 \times 10^{-4}$, $\beta = 0{\cdot}07$, $\alpha N_0 = 1{\cdot}5 \times 10^{-3}$ and

$$\sigma_0 = \omega/\alpha N_0 = 0{\cdot}0485.$$

In such a case it can readily be shown that the second term on the right-hand side of (91) may be neglected relative to the others and we therefore have, as our condition for $z(N_m)$

$$\frac{\partial q}{\partial z} = \frac{\partial^2 N}{\partial z \partial t} + v\frac{\partial^2 N}{\partial z^2}. \qquad \ldots\ldots(93)$$

We require therefore to determine the magnitude of the new term $v\partial^2 N/\partial z^2$. We can find $\partial^2 N/\partial z^2$ with sufficient accuracy from the quasi-stationary value

$$N \simeq (q_0/\alpha)^{1/2} \exp \tfrac{1}{2}(1 - z - e^{-z}\sec\chi),$$

from which we find that

$$\frac{\partial^2 N}{\partial z^2} = -\frac{1}{2}\left(\frac{q}{\alpha}\right)^{1/2} e^{-z}\sec\chi = -\frac{1}{2}\left(\frac{q}{\alpha}\right)^{1/2}(1+a). \qquad \ldots\ldots(94)$$

Also, by (58) and (59)

$$\frac{1}{q}\frac{\partial^2 N}{\partial z \partial t} \simeq \epsilon_0 = \frac{1}{2}\sigma_0 \frac{\tan\chi}{(\cos\chi)^{1/2}}$$

while

$$\frac{\partial q}{\partial z} = q(-1 + e^{-z}\sec\chi) \simeq qa.$$

Equation (93) now becomes

$$qa = q\epsilon - \tfrac{1}{2}v(1+a)(q/\alpha)^{1/2}$$

or

$$a = \epsilon - \eta \qquad \ldots\ldots(95)$$

where

$$\eta = \tfrac{1}{2}v(1+a)/(q\alpha)^{1/2} \simeq \tfrac{1}{2}v/(q\alpha)^{1/2} \simeq \frac{v}{2\alpha N_m}. \qquad \ldots\ldots(96)$$

This result could clearly have been obtained directly by an argument similar to that used in § 4.1.

Equations (95) and (96) apply to any drift function provided that, as in the case considered by Kirkpatrick, $[\partial^2 v/\partial z^2]/\alpha N$ is negligible compared with unity. Now Kirkpatrick, by a series of somewhat lengthy calculations, found the alteration in height of the N_m level, $\delta z(N_m)$, due to vertical drift, and also the corresponding change δN_m in the spatial maximum of electron density.

We see at once that our quantity η gives us $\delta z(N_m)$. Taking Kirkpatrick's assumed expression for vertical drift this becomes

$$\delta z(N_m) \simeq \eta = \frac{1}{2}\frac{v_0}{\alpha N_0}\frac{\sin 2\chi}{(\cos\chi)^{1/2}} = -0{\cdot}21\frac{\sin 2\chi}{(\cos\chi)^{1/2}} \qquad \ldots\ldots(97)$$

This quantity is plotted over a solar day in figure 4 (*a*) the trend of η being shown by the continuous line. The crosses show points resulting from Kirkpatrick's computations.

There remains, however, the estimation of the change of maximum electron density δN_m due to the effect of vertical drift. Now our approximate value of N_m must be obtained from the equation

$$\frac{dN_m}{dt} = q - \alpha N_m^2 - N_m \frac{\partial v}{\partial z}. \qquad \ldots\ldots(98)$$

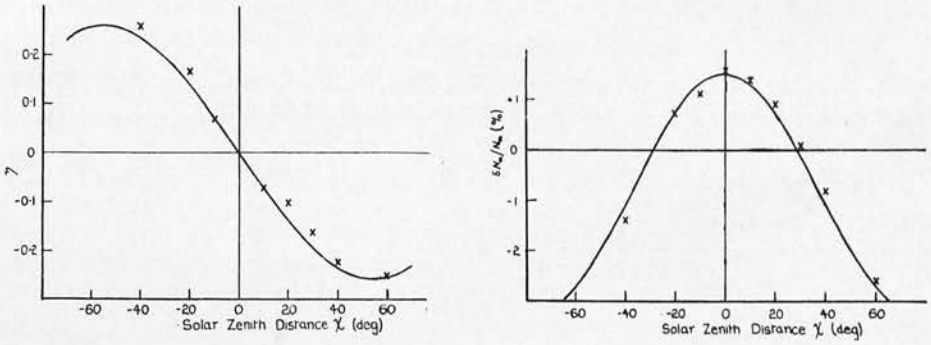
If we assume the drift term and the dN/dt term to be small relative to the other two we have

$$N_m \simeq \left(\frac{q}{\alpha}\right)^{1/2} \left\{1 - \frac{1}{2}\frac{dN/dt}{q} - \frac{\partial v/\partial z}{2(q\alpha)^{1/2}}\right\}. \qquad \ldots\ldots(99)$$

Now
$$q = q_0 \cos\chi(1 - \tfrac{1}{2}a^2) \simeq q_0 \cos\chi(1 - \tfrac{1}{2}\eta^2) \qquad \ldots\ldots(100)$$

if, as is true in Kirkpatrick's case, we may neglect ϵ^2 and $\epsilon\eta$ relative to η^2; while, as shown previously,

$$\frac{1}{q}\frac{dN}{dt} \simeq \epsilon.$$



(*a*) (*b*)

Figure 4. (*a*) Day-time trend of the change η in the height of N_m, calculated from (97), for the particular case of vertical drift considered by Kirkpatrick. (*b*) Day-time trend of the corresponding fractional changes in N_m (i.e. $\delta N_m/N_m$) calculated from (101). In each case the crosses indicate points resulting from Kirkpatrick's computations.

We therefore have from (99) and (100),

$$N_m \simeq \left(\frac{q}{\alpha}\right)^{1/2} \{1 + \tfrac{1}{2}\epsilon - \tfrac{1}{4}\eta^2 - \xi\} \qquad \ldots\ldots(101)$$

where
$$\xi = \frac{\partial v/\partial z}{2(q\alpha)^{1/2}} \simeq \frac{\partial v/\partial z}{2\alpha N_m}. \qquad \ldots\ldots(102)$$

The fractional correction $\delta N_m/N_m$ is therefore given by

$$\delta N_m/N_m \simeq -\xi - \tfrac{1}{4}\eta^2 \qquad \ldots\ldots(103)$$

where η and ξ are given by (96) and (102). It will be seen, from (96), (97), (102) and (103), that vertical drift influences on both N_m and $z(N_m)$ are determined by two factors only, (*a*) the drift characteristics and (*b*) the quantity $1/2\alpha N_m$ which has previously been identified (Appleton 1953) as the 'time of relaxation' of the layer.

Now, in actual magnitude, the appropriate values of η and ξ for the case of vertical drift considered by Kirkpatrick are given by (97) and

$$\xi \simeq -\frac{\beta v_0}{2\alpha N_0}\frac{\cos 2\chi}{(\cos\chi)^{1/2}} = -0{\cdot}015\frac{\cos 2\chi}{(\cos\chi)^{1/2}}. \qquad \ldots\ldots(102a)$$

In Figure 4 (*b*) the continuous line shows the trend of $(-\xi-\frac{1}{4}\eta^2)$ throughout the day, using these numerical values. Individual points for $\delta N_m/N_m$ resulting from Kirkpatrick's computations are again shown by crosses in the same diagram.

It will be seen that equations (97) and (101) enable us to estimate rapidly the effect of any vertical drift on the height and density of the layer maximum to a degree of accuracy which will usually be sufficient for practical purposes. There is no inherent difficulty in calculating the second-order approximation when required, but this is best done for a particular range of numerical values since the latter affect the relative magnitudes of the various terms.

ACKNOWLEDGMENT

This work was carried out as part of the programme of the Radio Research Board of the Department of Scientific and Industrial Research and was supported by way of a grant from that Department.

REFERENCES

APPLETON, E. V., 1937, *Proc. Roy. Soc.* A, **162,** 451; 1953, *J. Atmos. Terr. Phys.*, **3,** 282.
CHAPMAN, S., 1931 a, *Proc. Phys. Soc.*, **43,** 26; 1931 b, *Ibid.*, **43,** 483.
KIRKPATRICK, C. B., 1948, *Aust. J. Sci. Res.*, Series A, **1,** 423.
MARTYN, D. F., 1947 a, *Proc. Roy. Soc.* A, **189,** 241; 1947 b, *Ibid.*, **190,** 273; 1948, *Ibid.*, **194,** 429, 445.
MILLINGTON, G., 1932, *Proc. Phys. Soc.*, **44,** 580; 1935, *Ibid.*, **47,** 263.
NICOLET, M., 1951, *J. Atmos. Terr. Phys.*, **1,** 141.
RYDBECK, O. E. H., and WILHELMSSON, H., 1954, *Chalmers tek. Högsk. Handl.*, No. 149.
WILKES, M. V., 1939, *Proc. Phys. Soc.*, **51,** 138; 1954, *Proc. Phys. Soc.* B, **67,** 304.

APPENDIX

The Range of Validity of the Method of Successive Approximations

If the process described in § 3 is continued a step further we obtain the third-order approximation

$$N_{(3)} = \left(\frac{q}{\alpha}\right)^{1/2}\left\{1-\frac{1}{4(q\alpha)^{1/2}}\frac{q'}{q}+\frac{1}{32q\alpha}\frac{4qq''-5q'^2}{q^2} - \frac{1}{128(q\alpha)^{3/2}}\frac{8q^2q'''-36qq'q''+30q'^3}{q^3}\right\}. \qquad \ldots\ldots(104)$$

When the terms containing q'' and higher derivatives are negligible (cf. equation (114)) this simplifies and continuing the process yields

$$N_{(5)} = \left(\frac{q}{\alpha}\right)^{1/2}\left\{1+\frac{1}{2}\epsilon-\frac{5}{8}\epsilon^2+\frac{15}{8}\epsilon^3-\frac{69}{8}\epsilon^4+53\epsilon^5\right\}$$
$$\simeq \left(\frac{q}{\alpha}\right)^{1/2}\left\{1+\frac{1}{2}\epsilon-\frac{5}{8}\epsilon^2+2\epsilon^3-9\epsilon^4+50\epsilon^5\right\} \qquad \ldots\ldots(105)$$

where, as before,

$$\epsilon = -\frac{1}{2(q\alpha)^{1/2}}\frac{q'}{q}.$$

The additional terms introduced by continuing the process still further increase by successive factors which are greater than $14\epsilon/2$, $17\epsilon/2$, $20\epsilon/2$ etc. and hence the series will ultimately diverge no matter how small ϵ is.

Our problem is to find limits for the error at any given stage of the process or, more precisely, given a maximum error δ, to determine for what range of ϵ (and hence of χ, or t) the error in some specified approximation $N_{(r)}$ will remain less than δ. We do not attempt here a rigorous solution to this problem, but we can obtain a semi-empirical rule, sufficient for most practical purposes, in the following way.

The method of successive approximations begins by taking

$$\alpha N_{(0)}^2 = q. \qquad \ldots\ldots(106)$$

At this stage the error in the differential equation is

$$\alpha N_{(0)}^2 + \frac{dN_{(0)}}{dt} - q = \frac{dN_{(0)}}{dt} = E_0, \text{ say.} \qquad \ldots\ldots(107)$$

In the next stage we make the reasonable assumption that this error E_0 is mainly an error in $\alpha N_{(0)}^2$ and that the error in $dN_{(0)}/dt$ is much smaller. If this is true E_0 gives the order of magnitude of the error in the zero-order approximation (106). We now test our assumption by writing

$$\alpha N_{(1)}^2 = q - E_0 \qquad \ldots\ldots(108)$$

i.e. by placing the whole of the error on αN^2, and then calculating the new error

$$\alpha N_{(1)}^2 + \frac{dN_{(1)}}{dt} - q = E_1, \text{ say.} \qquad \ldots\ldots(109)$$

If our original assumption was correct then $N_{(1)}$ will be a much better approximation than $N_{(0)}$, i.e. E_1 will be much less than E_0, and conversely.

In general we can test whether the error in $\alpha N_{(r)}^2$ is of order E_r by finding whether or not $E_{r+1} \ll E_r$. This rule may be stated in an alternative form by writing

$$N_{(r)} = N_{(0)} + C_1 + C_2 + C_3 + \ldots + C_r \qquad \ldots\ldots(110)$$

where C_1, C_2 ... are the successive corrections of which the first three are given explicitly in equation (104). Each correction is related to the preceding error by the approximate equation

$$C_r = \frac{E_{r-1}}{2\alpha N_{(r-1)}}. \qquad \ldots\ldots(111)$$

Our rule now states that the error, e_r say, in $N_{(r)}$ is of order C_{r+1} provided that

$$C_{r+2} \ll C_{r+1}. \qquad \ldots\ldots(112)$$

At this point we must distinguish between forenoon when q' is positive and afternoon when q' is negative. In the former case the successive corrections are all negative whereas in the latter case they alternate in sign. This means that in the afternoon our rule implies that $e_r < C_{r+1}$ but in the forenoon it tells us only that $e_r < 2C_{r+1}$.

Numerical tests indicate that condition (112) is sufficiently satisfied if

$$C_{r+2} < \tfrac{1}{3} C_{r+1}. \qquad \ldots\ldots(112\,a)$$

We can now use (105) to state the semi-empirical rules that, in the afternoon

the error in $N_{(1)}$ is less than $\frac{5}{8}\epsilon^2$ provided $\epsilon < \frac{1}{9}$;

$N_{(2)}$ is less than $2\epsilon^3$ provided $\epsilon < \frac{1}{15}$;

$N_{(3)}$ is less than $10\epsilon^4$ provided $\epsilon < \frac{1}{20}$;

whilst in the forenoon the limits for the errors are double these amounts.

To test these rules numerically we choose $q = q_0 \cos \omega t$ and put $\nu = N/N_0$ so that, following Chapman, the differential equation becomes

$$\sigma_0 d\nu/d\chi = \cos\chi - \nu^2 \qquad \ldots\ldots(113)$$

and we have solved this equation numerically for $\sigma_0 = 0{\cdot}04$ (a suitable value for the E layer if α is of the order 10^{-8}). The table compares the results of this computation with those obtained from equation (104), which now takes the form

$$\nu_{(3)} = (\cos\chi)^{1/2} + \tfrac{1}{4}\sigma_0 \tan\chi \left(1 - \frac{5}{4}\epsilon + \frac{15}{4}\epsilon^2\right) - \frac{\sigma_0^2}{8\cos\chi}(1 - \frac{7}{2}\epsilon) \qquad \ldots\ldots(114)$$

with $\epsilon = \frac{1}{2}\sigma_0 \tan\chi/(\cos\chi)^{1/2}$. (The last term in (114) contains the terms neglected in (105).)

We can sum up the results of this test by saying that, if $|\epsilon| < \frac{1}{10}$, the method of successive approximations gives an accuracy to the order $\pm 1\%$; but if we require a much higher accuracy, say to $\pm \frac{1}{10}\%$, then we must ensure that $|\epsilon| < \frac{1}{15}$. To give the method a severer test we have made a similar computation with $\sigma_0 = 0{\cdot}1$ and working to five decimal places. The results confirm the above conclusions.

Forenoon.	(1) $\chi = -70°$	$\epsilon = 0{\cdot}093$	$\nu_c = 0{\cdot}5517$	
	$\nu_0 = 0{\cdot}5856$	$e_0 = 0{\cdot}0339 = 6\%$	$C_1 = -0{\cdot}0274$	
	$\nu_1 = 0{\cdot}5582$	$e_1 = 0{\cdot}0065 = 1{\cdot}2\%$	$C_2 = -0{\cdot}0035$	
	$\nu_2 = 0{\cdot}5547$	$e_2 = 0{\cdot}0030 = 0{\cdot}5\%$	$C_3 = -0{\cdot}0011$	
	(2) $\chi = -67\frac{1}{2}°$	$\epsilon = 0{\cdot}076$	$\nu_c = 0{\cdot}5965$	
	$\nu_0 = 0{\cdot}6240$	$e_0 = 0{\cdot}0275 = 4{\cdot}6\%$	$C_1 = -0{\cdot}0237$	
	$\nu_1 = 0{\cdot}6003$	$e_1 = 0{\cdot}0038 = 0{\cdot}6\%$	$C_2 = -0{\cdot}0026$	
	$\nu_2 = 0{\cdot}5977$	$e_2 = 0{\cdot}0012 = 0{\cdot}2\%$	$C_3 = -0{\cdot}0007$	
	$\nu_3 = 0{\cdot}5970$	$e_3 = 0{\cdot}0005 = 0{\cdot}1\%$		
	(3) $\chi = -60°$	$\epsilon = 0{\cdot}053$	$\nu_c = 0{\cdot}6722$	
	$\nu_0 = 0{\cdot}6924$	$e_0 = 0{\cdot}0202 = 3\%$	$C_1 = -0{\cdot}0183$	
	$\nu_1 = 0{\cdot}6741$	$e_1 = 0{\cdot}0019 = 0{\cdot}3\%$	$C_2 = -0{\cdot}0015$	
	$\nu_2 = 0{\cdot}6726$	$e_2 = 0{\cdot}0004 = 0{\cdot}06\%$	$C_3 = -0{\cdot}0003$	
	$\nu_3 = 0{\cdot}6721$	$e_3 = 0{\cdot}0001 = 0{\cdot}01\%$		
Afternoon.	(4) $\chi = +75°$	$\epsilon = 0{\cdot}147$	$\nu_c = 0{\cdot}5400$	
	$\nu_0 = 0{\cdot}5078$	$e_0 = -0{\cdot}0322 = 6\%$	$C_1 = +0{\cdot}0375$	
	$\nu_1 = 0{\cdot}5453$	$e_1 = +0{\cdot}0053 = 1\%$	$C_2 = -0{\cdot}0073$	
	$\nu_2 = 0{\cdot}5380$	$e_2 = -0{\cdot}0020 = 0{\cdot}4\%$	$C_3 = +0{\cdot}0034$	
	$\nu_3 = 0{\cdot}5414$	$e_3 = +0{\cdot}0014 = 0{\cdot}3\%$		
	(5) $\chi = +70°$	$\epsilon = 0{\cdot}093$	$\nu_c = 0{\cdot}6108$	
	$\nu_0 = 0{\cdot}5862$	$e_0 = -0{\cdot}0246 = 4\%$	$C_1 = +0{\cdot}0273$	
	$\nu_1 = 0{\cdot}6135$	$e_1 = +0{\cdot}0027 = 0{\cdot}4\%$	$C_2 = -0{\cdot}0035$	
	$\nu_2 = 0{\cdot}6100$	$e_2 = -0{\cdot}0008 = 0{\cdot}1\%$	$C_3 = +0{\cdot}0011$	
	$\nu_3 = 0{\cdot}6111$	$e_3 = +0{\cdot}0003 = 0{\cdot}05\%$		

In this table ν_0, ν_1, ν_2 . . . denote successive approximations, e_0, e_1 . . . are the errors compared with the computed value ν_c from the numerical solution and C_1, C_2 . . . are the successive corrections derived from equation (104).

One further difficulty remains to be considered. So far we have made no mention of the initial conditions and these will certainly affect the situation, at any rate in the early forenoon. Now if we have a solution of (113), valid up to some known value of ϵ, we could certainly extrapolate it backwards to sunrise

by a suitable numerical method. Let us then postulate a 'quasi-stationary solution' $\nu = \nu_q$ valid for the whole day and to which $\nu_{(0)}$ $\nu_{(1)}$ $\nu_{(2)}$ are successive approximations within the limits stated above. If ν differs from ν_q due to the initial conditions or to some transient effect let us write

$$\nu = \nu_q + \mu.$$

Then

$$\sigma_0 \frac{d\mu}{d\chi} = -2\mu\nu_q - \mu^2 \qquad \ldots\ldots(115)$$

so that μ necessarily tends to zero as χ increases. If μ^2 is negligible there is a quasi-exponential decline with a variable time-constant $\sigma_0/2\nu_q$ which is simply the familiar $1/2\alpha N$ once again.

Numerical tests indicate that by the time the quasi-stationary approximations are valid any remaining effects of the initial conditions will usually be negligible.