6. STARTING PROCEDURES FOR THE ORDINARY RAY

6.1 Outline

Ionograms do not provide measurements at frequencies below some minimum value f_1. Thus there is an "unobserved" section of the electron density profile, for plasma frequencies FN from 0 to f_1. When extraordinary-ray measurements are available at low frequencies, the amount of ionisation in the unobserved region can be estimated and the remainder of the profile correctly calculated. This is described in Section 8. When only ordinary-ray data are available there is an infinite number of profiles, with different variations in the unobserved region, which can exactly reproduce the observed virtual heights (e.g. curves d and e in Fig. 5, Section 7). Physical arguments and previous experience must therefore be used to obtain a reasonable starting assumption. Different possible methods have been reviewed previously (Titheridge, 1975a). When suitable extraordinary ray measurements are not available, the best procedure for obtaining a consistent series of results seems to be the use of a starting point (h_s,f_s) based on a synoptic model of the underlying region.

Four distinct starting procedures are available in POLAN for use in the absence of extraordinary-ray data. These are fully defined by the single input parameter START, as described in Section 6.2. The methods are:

- (a) The normal default, using a start based on extrapolation of the observed virtual heights.
- (b) The preferred method, using previous experience to define an appropriate mean starting height at a fixed plasma frequency of 0.5 MHz.
- (c) An alternative model start, in which the plasma frequency is chosen at a fixed height.
- (d) A direct start, primarily for test purposes; this assumes that there is no ionisation with FN less than the lowest scaled frequency f_1.

When f_1 is large the first polynomial calculated in starts (a) to (c) covers a large frequency range in which there are no observations. Some unphysical variations could then be obtained. To ensure a smooth, monotonic start giving approximately the same results with all modes of analysis and with changes in f_1, two constraints are added to define the profile shape in the unobserved region. These constraints are:

(i) An additional virtual height point is included at the frequency $f_0 = (f_s + f_1)/2$. The corresponding virtual height h'_0 is obtained by extrapolating the initial virtual heights back to this frequency, as shown in Fig. 2.

(ii) The initial gradient dh/dFN, at the point (f_s,h_s), is given a value that would produce a virtual height of about h'_0 at frequencies just above h_s. This is obtained from the relation (Titheridge, 1985b).

$$dh/df_N = (1 + 1.8/f_1).(h'_0 - h_s).$$

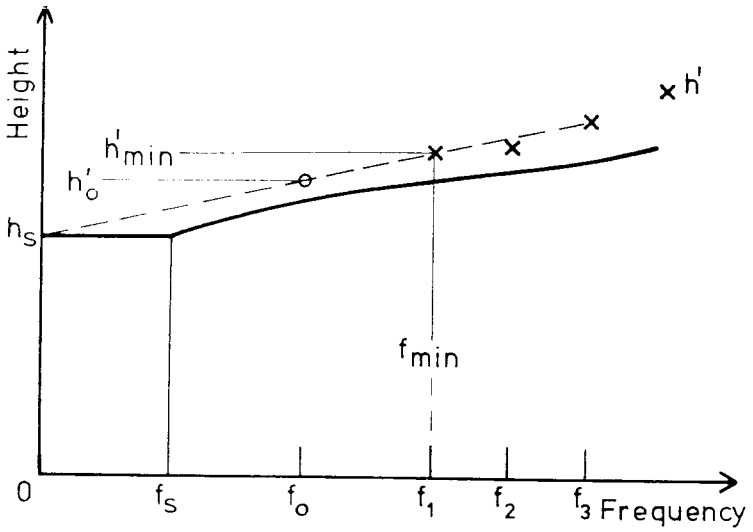

Figure 2. The start procedure using ordinary-ray data only. h_s is obtained by linear extrapolation of the (maximum) slope indicated by the first three virtual heights. The circled virtual-height point at f_0 is added to ensure a smooth, physically acceptable profile in the "unseen" region from f_s to f_1.

(a) START = 0.0

If START is zero the analysis begins from an initial real-height point (f_s,h_s), where h_s is determined by extrapolating the initial section of the ordinary-ray trace linearly down to zero frequency. This is the default procedure, illustrated in Fig. 2. The starting frequency f_s is normally 0.5 MHz, but is kept less than $0.6f_1$ to ensure a sufficient range for the starting correction. Extrapolation is based on the first three ordinary-ray points, and the absolute value of $(h'_3 - h'_1)/(f_3 - f_1)$ is used so that curling up of the trace at low frequencies (which is produced by an underlying layer with FN close to f_1) also gives a downwards correction to the starting point. If h'min is the lowest of the first 3 virtual heights, the starting height is effectively

$$h_s = h'min - fmin|dh'/df|.$$

This gives a low value of h_s, showing the presence of appreciable underlying ionisation, if the O-ray trace is turning downwards (due to reducing real heights) or upwards (due to increasing group retardation) as we approach fmin. To prevent unreasonable extrapolations, upper and lower limits are placed on the value of h_s using the expression

$$h'min/4 + 55 < h_s < h'min/2 + 60 \quad km$$

where h'min is the lowest observed virtual height. These limits agree roughly with the night-time models of Section 6.4. Careful use of a limited extrapolation of this type generally gives a worthwhile improvement, although results are less consistent than with the model procedure (b).

(b) START > 44.0

When START is greater than 44 it gives a model starting height to be used. Thus the analysis begins from a real-height point (f_s,h_s) where h_s = START. f_s is nominally 0.5 MHz, but is kept below $0.6f_1$, as in (a). The value of START will depend on some mean diurnal model, and varies in general from about 90 km near noon to 150 km after midnight. The selection of a suitable value of START, for different times, seasons and latitudes, is discussed in Section 6.3(b). The given value of START is modified within POLAN, if required, to be compatible with the observed virtual heights. This is done by making h_s less than $0.6h'min + 0.4h'_s$ where h'min is the minimum observed virtual height and h'_s is the extrapolated value obtained in (a) above.

(c) 0.0 < START < 44.0

Positive values of START less than 44.0 are used to define an initial plasma frequency f_s from which to start the analysis, at a fixed height h_s which is usually 90 km. This allows use of the (fixed-height, variable-FN) synoptic starting model of McNamara (1979). In POLAN the fixed height can also be set at 110, 130, 150 or 170 km, to allow greater accuracy with night ionograms where a starting height of 90 km (with a very low density) is too far below the base of the observed F layer to be useful or reliable. The different heights and densities are specified by the parameter START as follows:

For:		Use:		
0 < START < 10.,		f_s = START MHz,	h_s = 90 km.
10 < START < 20.,		f_s = START - 10 MHz,	h_s = 110 km.
20 < START < 30.,		f_s = START - 20 MHz,	h_s = 130 km.
30 < START < 40.,		f_s = START - 30 MHz,	h_s = 150 km.
40 < START < 44.,		f_s = START - 40 MHz,	h_s = 170 km.

(d) START < 0.

If START = -1.0, underlying ionisation is ignored and calculations begin directly from the first scaled point (f_1,h_1). This gives the ionosphere a flat base at the height h_1 = h'min, and variations due to underlying ionisation are excluded from the analysis. START = -1.0 can also be used to obtain any starting point not covered by options (a) to (c) above, by entering the desired starting point as the first data point (f_1,h_1). Calculations differ from use of the standard options in that the additional point (i) in Section 6.1 is not added, and the constraint (ii) is not applied to the initial gradient.

Values of START less than -1.0 are used to specify a polynomial start for the analysis of combined ordinary and extraordinary ray data, as described in Section 8.6.1.

6.3 Day-time Starting Models

(a) The data base

Ionograms normally contain no reflections from heights below about 100 km, and seldom show extraordinary-ray traces for the lower E region. Corrections for group retardation occurring in the daytime D region must therefore be based on other information. 700 experimental electron-density profiles of the D and lower E regions have been compiled into a computer accessible library by McNamara (1978a). Profiles corresponding to undisturbed conditions were used in a multiple regression analysis to obtain expressions giving the mean variations (at a fixed height or density) with solar zenith angle, season, latitude and solar cycle (McNamara, 1979). From these expressions mean results have been derived for use in the model starts (b) and (c) of Section 6.2, as described below. Note that the data base is heavily weighted to the North American region, so the models will be less representative elsewhere.

(b) The Variable Starting-Height Model

The recommended procedure for the incorporation of model starting data is use of a variable starting height h_s at a fixed starting frequency f_s. From the D-region data base McNamara (1979) obtained the height at a fixed plasma frequency of 0.5 MHz as

$$h_{0.5} = 91.5 - (5.6 \cos\chi \pm 1.4 \cos S + 0.027R) \qquad (10)$$

with a residual standard deviation of 3.8 km. The + and - signs apply to the Northern and Southern Hemispheres respectively. χ is the solar zenith angle, and R is the 12-month smoothed sunspot number. Seasonal changes depend on the parameter $S = 30M - 15$ degrees, where M is the month. For a station at latitude ϕ, and a local time of T hours from noon, the zenith angle is given by

$$\cos\chi = \sin D.\sin\phi + \cos D.\cos\phi.\cos A$$

where the hour angle $A = 15.T$ degrees and the solar declination $D = 23.44 \cos S$ degrees.

At $R = 50$ this expression gives the heights shown in Table 4. At each latitude the change from summer to equinox to winter has been linearised so that it may be expressed in the form $h_s = h_{eq} +/-\Delta h$, where $h_{eq} + \Delta h$, h_{eq} and $h_{eq} - \Delta h$ apply to midsummer, equinox and midwinter respectively. This makes interpolation more convenient, and introduces an additional error of not more than 0.3 km at any point. Note that the sign of the seasonal variation changes at high latitudes near noon.

Table 4. Starting heights h_s for use at a fixed plasma frequency f_s of 0.5 MHz, in the analysis of daytime ionograms. Main tabulated values apply to the equinoxes. These are followed by the corrections required at midsummer and at midwinter, using the upper and lower signs respectively.

Latitude	(a) Hours from Local Noon					
	0	2	3	4	5	hr
±10°	84.8±1.0	85.6±1.0	86.4±1.0	87.5±1.0	88.8±1.0	km
±30°	85.6 ±.2	86.2 ±.2	86.9 ±.2	87.8 ±.3	89.0 ±.3	
±50°	86.7 ∓.3	87.2 ∓.3	87.8 ∓.3	88.5 ∓.3	89.3 ∓.2	
±70°	88.3 ∓.7	88.6 ∓.7	88.9 ∓.7	89.3 ∓.7	89.7 ∓.6	

Latitude	(b) Hours after Sunrise or before Sunset					
	2.0	1.5	1.0	0.5	0.0	hr
±10°	87.5±1.3	88.1±1.4	88.8±1.4	89.5±1.4	90.2±1.4	km
±30°	87.8±1.2	88.4±1.3	89.0±1.4	89.6±1.4	90.2±1.4	
±50°	88.5±1.1	89.0±1.2	89.4±1.3	89.8±1.4	90.2±1.4	
±70°	89.4±0.9	89.6±1.0	89.8±1.2	90.0±1.3	90.2±1.4	

Table 4 is used only when a normal E layer trace is visible. At local times more than two hours from sunrise or sunset, h_S is obtained from the upper section (a) of the table. Within two hours of sunrise or sunset the lower section (b) should be used. The two sections agree closely near the equinoxes but the seasonal variation increases near sunrise/sunset, becoming +/-1.4 km (at all latitudes) at $\chi = 90°$. The final solar-cycle term of equation (10) is not included in Table 4. It adds a further change of +1.4 km at solar minimum, and -1.4 km at solar maximum (R = 100).

Use of Table 4 to estimate the required starting height h_S, to the nearest km, is straightforward. The rise in h_S towards sunrise or sunset follows the natural movement of the E layer. Thus the height range over which the correction is applied is rather more constant than for procedure (c). The frequency range is also more constant, extending from 0.5 MHz to the value of fmin on the ionogram; this is typically about 1.5 MHz (the top of the broadcast band) throughout the day. For most work it is therefore recommended that the parameter START in POLAN be set equal to a height obtained from Table 4, for the analysis of daytime ionograms when the E layer trace is visible. This automatically gives the fixed f_S (of 0.5 MHz), variable h_S starting procedure of Section 6.1(b).

(c) The Variable Starting-Frequency Model

From the same D-region data base McNamara (1979) used undisturbed densities at a height of 90 km to derive an expression giving the mean variation of log(electron-density) with solar zenith angle and with season. There were no significant overall changes with latitude or with solar cycle. Expressing these results in terms of the peak plasma frequency FN, and making some simplifications (which introduce errors of less than 5%) gives the result

$$(FN)_{90km} = 0.43\ (1 + 2.5\ \cos^2\chi)\ (1 + 0.2\cos S) \qquad (11)$$

where $S = 30M - 15$ degrees and M is the month. For the Southern Hemisphere the last factor becomes $(1 - 0.2\ \cos S)$. Results from (11) for summer, equinox and winter conditions are plotted in Fig. 3.

The height of 90km is always slightly below the traces obtained from a normal E layer. This makes it a good overall choice for a daytime starting height (e.g. Wright et al., 1975). For N(h) calculations during the day, when a normal E layer is observed, the parameter START in POLAN may be set equal to the estimated value of FN at a height of 90 km. This automatically gives the fixed h_S (90 km), variable FN starting procedure of Section 6.2(c). In the absence of other information, FN is obtained from (11) or from the curves in Fig. 3. The latter are calculated from (11) and represent the current best estimate of the mean value of $(FN)_{90}$ under quiet conditions. An indication of the changes which occur in various types of disturbance is given by McNamara (1978c). At night, the value of 90 km is much too low for reliable calculations with a fixed-height starting model. The starting height can therefore be altered to 110, 130, 150 or 170 km if desired, by appropriate choice of the parameter START, as described in Section 6.2(c).

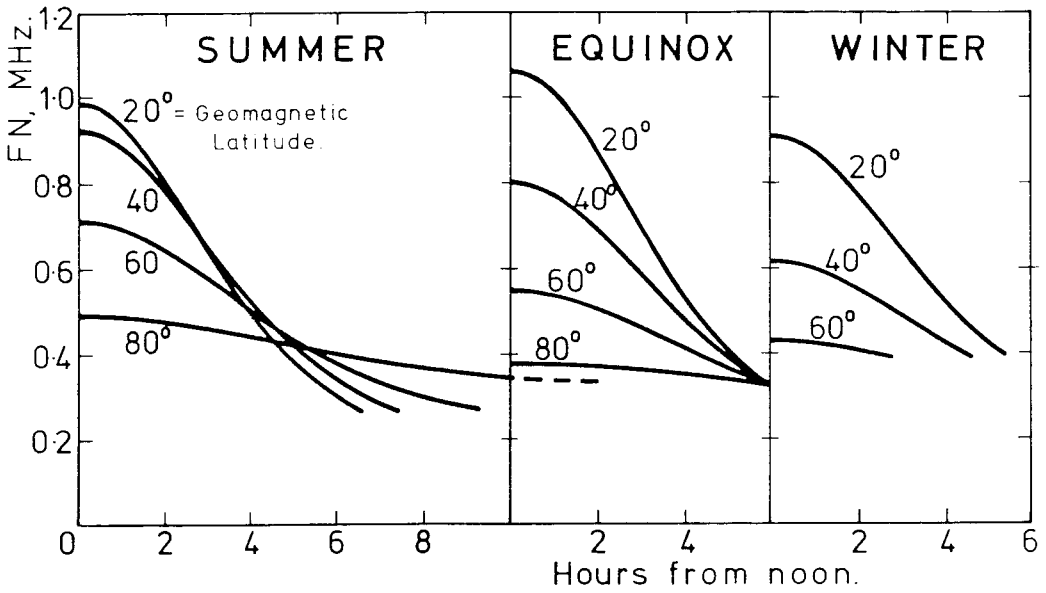

Figure 3. Mean starting frequencies (FN) for the analysis of daytime ionograms, using a fixed starting height of 90 km.

6.4 A Night-time Starting Model

During the night the base of the F layer is generally at a height near 200 km. A starting point near 90 km then leaves a large region with no observations. Calculated profiles will be ill-defined over this region, giving variable and unreliable results. Observations of the night E layer provide no help, since these are still at a comparatively low height with a wide valley between the night E and the low F layers.

In the absence of extraordinary-ray data, some mean model should therefore be assumed to provide a starting point in the height range 130 to 180 km. A procedure described previously (Titheridge, 1967b, 1975a) used a constant starting frequency f_s of 0.5 MHz. The corresponding height h_s was taken as 90 km during the day (agreeing with the mean daytime results of Section 6.2(b)). h_s becomes about 120 km after sunset, when the E layer trace is not visible. It increases to about 150 km near midnight, and remains at this value until shortly before sunrise. Results give a reasonably continuous and consistent sequence of profiles, with an approximately correct allowance for mean changes in the night E region. The recommended values of h_s were, however, based on comparisons with a fairly small number of profiles obtained using good ordinary and extraordinary-ray traces; primarily the mean night-time sequence for Slough described in Titheridge (1959c).

McNamara (1979) has considered this problem using the collection of about 100 experimental night-time profiles, for heights from about 80 to 280 km, given by Knight (1972). He gives starting points (h_s, f_s) which lie on the mean electron density profiles, with h_s equal to or greater than 150 km. Thus they make no allowance for ionisation below h_s. Correcting for this effect gives the circles plotted in Fig. 4. The zenith angle scale, at the top of the figure, corresponds exactly to the lower hourly scale at a latitude of 48° (at the equinoxes), and the scales are approximately correct for all low and middle latitudes. The drop of 60 km in h_s between $\chi = 100°$ and $\chi = 90°$ is similar to the drop of 50 km in the real height of the F region profile at FN = 1.5 MHz. The rise of 150 km in the value of h_s in the two hours after sunset also roughly parallels the rise of 30 km in the real height at 1.5 MHz. With a fixed starting frequency f_s of 0.5 MHz the extent of the underlying "unseen" section of the calculated profiles is therefore roughly constant, in both height and frequency range, throughout the night. Figure 4 is also in reasonable agreement with the previously recommended starting points at 0.5 MHz, of 90 km during the day (when the E layer is visible), 120 km after sunset or before sunrise, and 150 km near midnight.

It is therefore recommended that starting heights from Fig. 4 be used for the analysis of night-time ionograms in the absence of extraordinary-ray data. In the program POLAN this is achieved by setting the parameter START equal to the height read from Fig. 4, at the appropriate hour T. When the day-time E layer trace can be scaled, START is equal to about 90 km as shown by the straight lines for summer and winter in Fig. 4. These lines correspond to the heights of Table 4(b). The abrupt decrease in h_s when the E layer trace disappears at sunset counteracts the effect of the sudden increase in the lowest measured virtual height, and so helps to maintain reasonable continuity in the calculated heights of the F layer.

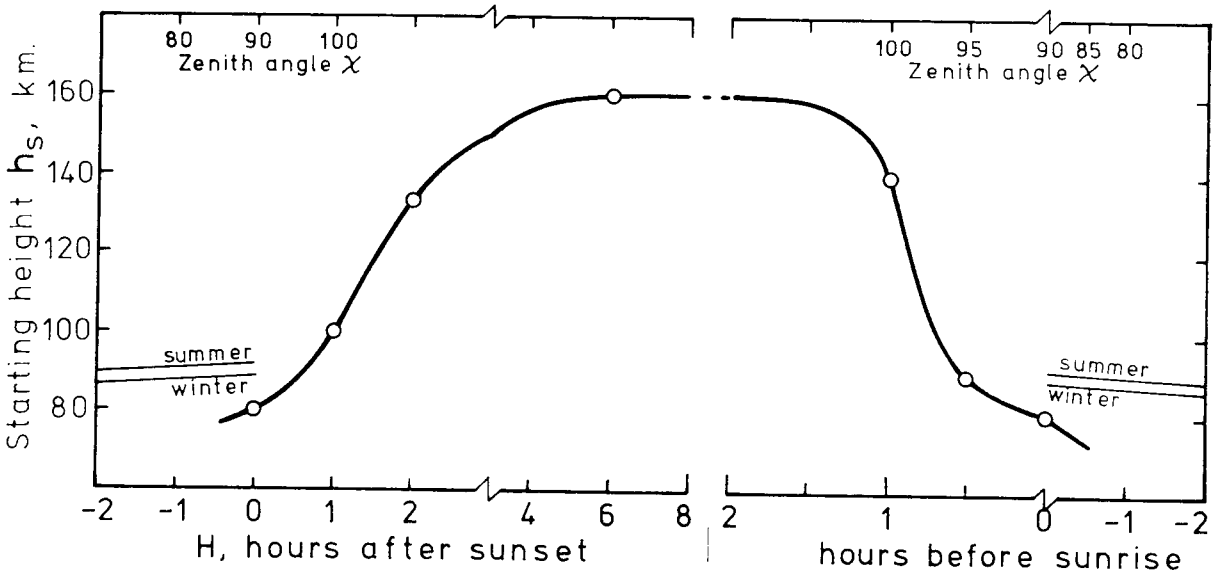

Figure 4. The starting height h_s for use at a fixed frequency f_s of 0.5 MHz, at night (heavy line) and when the normal E layer is visible near sunrise and sunset (thin lines labelled for summer and winter).

7. VALLEY PROCEDURES FOR THE ORDINARY RAY

7.1 Outline

Where the virtual-height trace on an ionogram is discontinuous we have the possibility of a non-monotonic N(h) profile. This problem arises primarily in the analysis of daytime ionograms, when a valley or decrease in electron density commonly occurs between the E and F layers. Uncertainty as to how this "unseen" region should best be treated, in routine ionogram analysis, has caused most workers to ignore the problem. Thus most of the profiles which have been calculated to date give the monotonic real height which satisfies the ordinary-ray virtual-height data. This procedure has the merit of being straightforward and well-defined. It has the fault of producing an answer which is an extreme case - the lower limit to the range of possible profiles - and so is always too low. Using POLAN, the monotonic profile can be obtained by setting the parameter VALLEY equal to 10.0 (as in Section 7.4).

There is not and cannot be any purely mathematical procedure for determining the "correct" profile from the range of possible solutions, using the ordinary ray only. Thus curves a, b, and c in Fig. 5 give real heights for the lower F region which vary over a range of 100 km. All three curves agree exactly with the given virtual-height curve and so are equally valid mathematical solutions. The full range of possible real-height curves extends from the monotonic (no valley) solution to the largest valley which is consistent with a positive value of dN/dh at the base of the upper layer. Selection between different solutions must be based purely on physical grounds, having regard to the reasonableness or physical likeliness of different types of profile. To allow meaningful comparison of results, the selection should follow some consistent procedure.

When suitable extraordinary-ray data are available, some information can be obtained about the valley region. At best two parameters can be estimated, relating basically to the overall width of the valley and to the relative amount of high density ionisation (Section 9.1 and Appendix B.2). In most cases only one parameter can be reliably determined. We then use some model for the shape of the valley, to give suitable proportions of high and low density ionisation, and determine the single valley-width parameter VWIDTH.

Ordinary-ray calculations use a similar approach. A standard valley shape is defined as in Section 7.2. This has physically reasonable values for the decrease of FN above the peak of the lower layer, for the mean depth of the valley, and for the gradient dh/dFN at the top of the valley region. The valley problem then reduces to the choice of a suitable value for VWIDTH. This can be specified directly, for an individual profile or for a whole run, by appropriate choice of the parameter VALLEY in POLAN. Upper and lower limit profiles can also be obtained as described in Section 7.4. For most work, however, it seems preferable to obtain VWIDTH by including physically-desirable conditions as part of the least-squares analysis. This seems the best general approach, and perhaps the only feasible or useful one, for selecting between the wide range of possible valley profiles in the absence of extraordinary-ray data.

The construction of POLAN makes direct incorporation of constraints on the real-height profile no more difficult than the use of standard virtual-height data. Any number of constraints can be used, as necessary conditions or as desirable attributes of the profile. Necessary conditions specify limits on the allowable range for one or more of the real-height parameters q; these conditions are described in Section 7.3.3 and are imposed after the normal solution of the virtual-height equations. Desirable attributes are included directly, with some appropriate weight, in the set of equations for the least-squares solution.

Section 7.3.2 describes the physical constraints used in POLAN for valley calculations. The first constraint requires the valley width to be about twice the neutral scale height; this condition is given a weight such that it strongly influences the calculation of VWIDTH, but does not predetermine it. Further equations, with smaller weights, require a smooth transition from the valley region to the following profile. All equations are combined in the normal (default) valley procedure to determine VWIDTH. Different valleys, and upper or lower limit profiles, can also be obtained using a non-zero value of the input parameter VALLEY as described in Section 7.4.

7.2 The Standard Valley

The vertical scale of an electron density profile is governed primarily by the local scale height SHA of the neutral atmosphere. Thus for both E and F layers the peak thickness is proportional to the scale height of neutral gases which govern the height variations of production, loss and diffusion (e.g. Titheridge, 1973). The vertical separation of the E and F layers, and the width of the valley between them, will also (in the idealised case) be proportional to the mean scale height of the neutral gas.

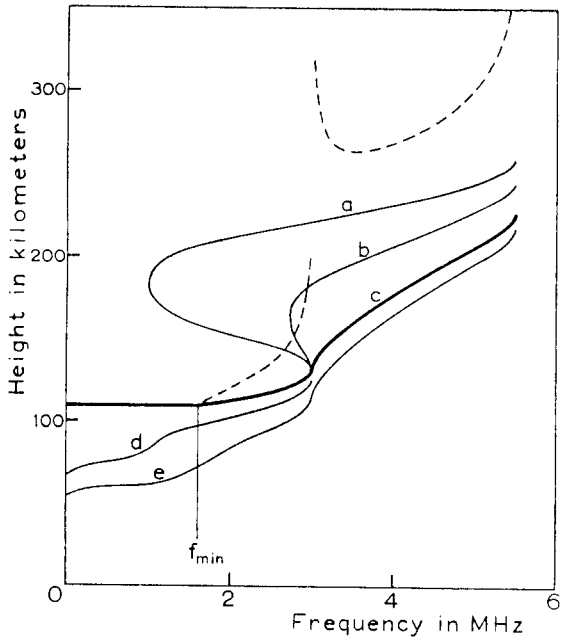

Figure 5. Some of the infinite number of possible real-height profiles (solid lines) corresponding to the same virtual-height curve (broken lines).

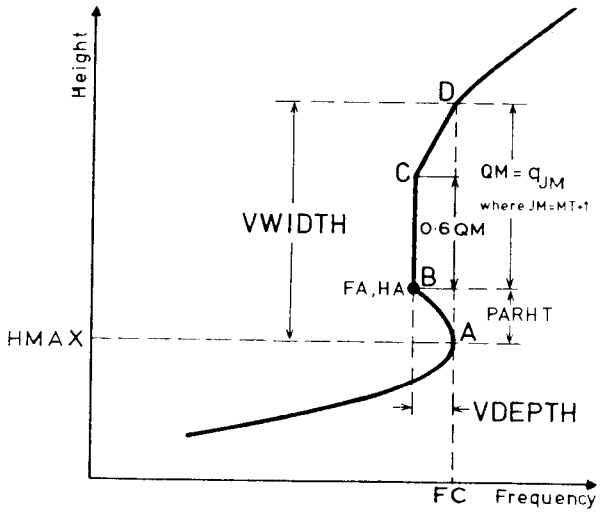

Figure 6. The form and notation of the standard valley.

POLAN uses a model atmospheric scale height SHA related to true height h by

$$SHA = h/4 - 20 \text{ km.} \tag{12}$$

The valley width is taken as

$$VWIDTH = 2SHA = HMAX/2 - 40 \text{ km} \tag{13}$$

where HMAX is the height of the underlying peak. These relations give reasonable values for the scale height and for the valley width, in both the E and F1 regions, as shown by Table 5.

Table 5. Parameters for the standard valley defined by equations (12) to (14).

Height of underlying peak:	HMAX	=	100	110	120	160	180	200 km.
Model scale height:	SHA	=	5	7.5	10	20	25	30 km.
Standard Valley:	VWIDTH	=	10	15	20	40	50	60 km.
	VDEPTH	=	0.03	0.05	0.08	0.21	0.29	0.36 MHz.

The shape of the assumed valley is shown in Fig. 6. The peak section calculated for the lower layer is extrapolated upwards to the point where the plasma frequency drops to FC - VDEPTH. The scale height SH used in this extrapolation is 40% greater than the value calculated for the lower part of the peak. This makes allowance for the normal increase in SH with height, and gives a shape which closely matches the peaks obtained with overlapping Chapman layers. From the top of the parabolic section the electron density profile consists of two linear segments reaching a plasma frequency FC at a height HMAX + VWIDTH. Occasionally insertion of this model valley will result in negative values of the gradient dh/DFN just above the valley; the valley width is then automatically reduced, as described in Section 7.3.3.

Rocket and backscatter profiles indicate nearly full daytime valleys under most conditions. The default depth for the standard E/F region valley is therefore taken as VDEPTH = 0.05 MHz, at VWIDTH = 20 km. The mean depth will vary with the valley width, as in Fig. 7. When VWIDTH becomes small, the approximately parabolic peak of the underlying layer forces VDEPTH to decrease rapidly, approximately as $VWIDTH^2$. For larger valleys VDEPTH will change more slowly with the overall valley width, to agree with rocket and backscatter results (summarised in Lobb and Titheridge, 1977a) which indicate a maximum depth of about 0.2foE. Widest valleys should still be deepest on the average, however, or the irregular variations shown by most direct measurements in the valley region would lead to the common appearance of intermediate traces on the ionograms. Valley depth is very much less important than valley width in obtaining correct heights for the F region (Section 9), so the model needs only approximate values for VDEPTH.

At small values of valley width, VDEPTH is made proportional to $VWIDTH^2$. This keeps an approximately constant shape for the valley, with the initial parabolic section (a range PARHT in Fig. 6) a constant fraction of the total width. For larger valleys the depth is made proportional to width. This is accomplished using

$$VDEPTH = 0.008\ VWIDTH^2/(20 + VWIDTH) \qquad (14)$$

where VWIDTH is obtained from (13). Valley constants calculated from (13) and (14) are shown in Table 5. The constants in (14) can be changed if required, as described in Section 9.4. To ensure that valley depths never become impossibly large, the actual depth used in the calculations is VDEPTH.FC/(VDEPTH + FC) where FC is the critical frequency of the underlying peak. For normal valleys with VDEPTH much less than FC this has no significant effect. It can be important, however, when attempting to calculate the valley depth using extraordinary-ray data (Section 9.3). The constants in (14) can be changed if required, to use different valley shapes, as described in Section 7.4.

7.3 Addition of Physical Constraints

7.3.1 The real-height equation

Throughout most of a calculated profile, each real-height section is defined by the relation

$$H - HA = \sum_{j=1}^{NT} q_j\ (FN - FA)^j$$

This gives a polynomial passing through a known real-height point (HA,FA). The coefficients q_j are obtained by solving a number of virtual and real-height equations as described in Section 5.1. For general start or valley calculations we do not have a known starting height for the polynomial segment, so an additional (constant) term is included in the real-height expression giving

$$H - HA = \sum_{j=1}^{NT} q_j\ (FN - FA)^j + q(JM) \qquad (15)$$

where JM = NT + 1.

The term q(JM) gives a constant offset to the calculated real heights. This is negative when used to allow for a starting correction (Section 8), and positive to allow for a valley above the peak of a layer. The constant term is added at the end of the real-height expression so that

(a) the form of the first NT terms is unchanged;

(b) the value of q(JM) can be adjusted readily after the initial solution, without affecting any other constraints which may have been added; and

(c) other constraints which may be applied after the initial solution (Section 7.3.3) will over-ride a previous definition of q(JM).

The values of q_j (j = 1 to JM) are obtained by solving the set of simultaneous equations stored in the matrix B of Section 4.2. The equations are specified by coefficients B(i,j) where i = 1 to M and j = 1 to JM+1; the terms B(i,JM+1) contain the right-hand sides of the equations (7), corresponding to the left side of equation (15). The M rows of coefficients are calculated by a subroutine COEFIC (Appendix D.1) which also sets following rows of B to zero. Any required constraints can be specified in these following rows, and the value of M incremented by the number of added constraints. The subroutine SOLVE then calculates the coefficients q_j in (12) as a least-squares solution of the M simultaneous equations, including the added constraints.

7.3.2 Added valley equations

Physically desirable features of the valley region, and of its relation to other parts of the profile, are expressed analytically and included with the virtual-height equations in the least-squares solution. These added equations relate the real-height coefficients q(j) to the standard valley parameters VWIDTH and VDEPTH. POLAN currently adds 4 equations of this sort for ordinary-ray valley calculations, representing the following three conditions:

(a) The valley width.

The overall width of the standard valley in Fig. 6 is VWIDTH = q(JM) + PARHT. The parabolic distance PARHT is calculated from the value of VDEPTH and the scale height of the underlying peak, as in Section 7.2. The valley-width equation added to the set of simultaneous linear equations is then:

$$q(JM) = VWIDTH - PARHT \qquad (16)$$

This equation has a weight 1.0, as for the virtual-height data points, so the equality is indicated but not enforced in the least-squares solution. The weight is increased to 100 when the parameter VALLEY (or HVAL, Section 7.4) is less than -2.0, indicating that a specific valley width is required. When extraordinary-ray data are used to determine the valley width, as in Section 9, the weight given to the condition (16) is reduced to 0.04.

(b) Gradient continuity

With a reasonably realistic model for the variation of electron density with height in the valley region, we expect a reasonable match between the gradient at the top of the valley and the initial real-height slope q(1), at the point D in Fig. 6. An exact match is obtained if q(1) = 0.4q(JM)/VDEPTH, since the section CD covers a range of 0.4q(JM) in height and VDEPTH in frequency. There is normally some negative curvature of the real-height profile in this region, so the relation used in POLAN is q(1) = 0.25q(JM)/VDEPTH. Using a weight of 0.4^2 for the least-squares solution, the added equation becomes:

$$0.4q(1) - 0.1q(JM)/VDEPTH = 0. \qquad (17)$$

(c) Profile smoothness

Implausible E/F valleys can require rapid changes in dh/dFN, at the bottom of the F layer, to give agreement with observed virtual-height curves. The true profile is unlikely to have large fluctuating gradients in this region, where the scale height is reasonably large and the electron density is increasing rapidly. Thus we would like the initial real-height section above a valley to be represented by the lowest order of polynomial which gives reasonable agreement with observed virtual heights. Linear and parabolic sections are not sufficiently flexible in general for reliable results. When four or more terms are used in the real-height polynomial, the preference for a low-order representation in (15) is expressed by the relation

$$0.5q(NT) = 0. \qquad (18a)$$

With higher order methods, using NT > 4, POLAN adds in addition the relation

$$0.15q(NT-1) = 0. \qquad (18b)$$

The effective weights of 0.5^2 and 0.15^2 given to these relations express a desirable tendency rather than a requirement.

7.3.3 Limiting constraints

After the valley calculation, results are checked to ensure that profile parameters are within physically reasonable limits. Unacceptable parameters are fixed at the limiting value, and the other parameters adjusted to obtain a new least-squares solution (using the subroutines ADJUST and SOLVE). Limits are imposed in the following order:

(a) F-region gradient

The gradient dh/dFN of the real-height profile is large at the top of a valley (the point T in Fig. 7). This is partly a consequence of fairly shallow valleys. Examination of theoretical and experimental profiles suggests dh/dFN > 10 km/MHz at all times, and dh/dFN > 20 km/MHz nearly always.

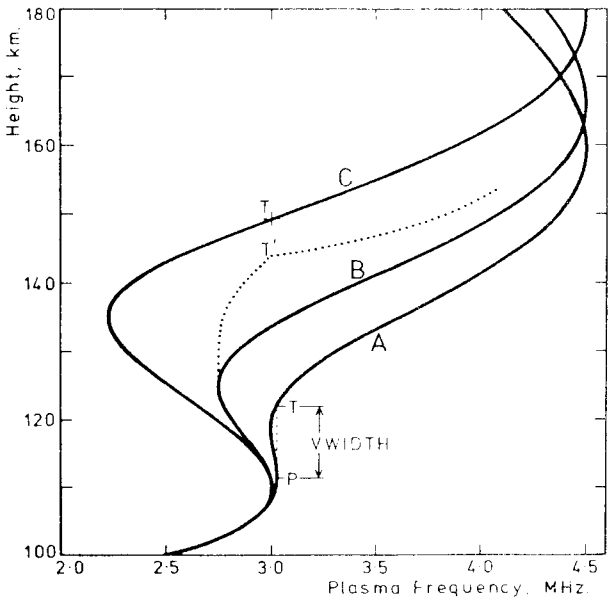

Figure 7. Profiles formed by the superposition of a Chapman E layer, with HMAX = 110 km, SH = 10 km and FC = 3.0 MHz, and a Chapman F1 layer with SH = 20 km.

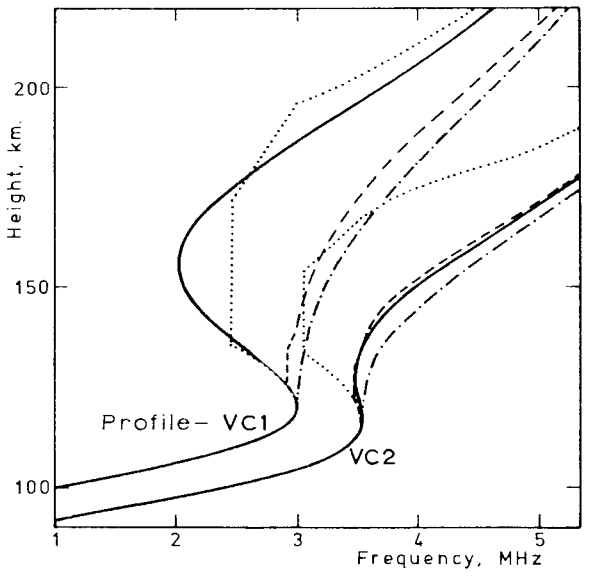

Figure 8. Dual-Chapman profiles for conditions near sunrise or sunset (VC1), and near noon (VC2). Broken lines give the profile calculated by POLAN at a dip angle of 50°, using the standard valley (HVAL=0 or 1). Upper and lower limit calculations (obtained with HVAL= 5 or 10 respectively) are shown by dotted and chain lines.

An absolute minimum slope can be calculated directly. For a given valley size the gradient is least when the ratio of scale heights for the upper and lower layers is least. For the F1 and E layers, consisting of O^+ and O_2^+ ions respectively, the scale heights have a ratio of 2:1 for isothermal conditions; in fact the temperature increases with height so the scale height ratio will always exceed 2:1. Using SH = 10 km for the upper E region, the smallest scale height for the F1 layer is 20 km. With this value, varying the height of the F region peak gives the family of valleys shown in Fig. 7. These are calculated for foE/foF1 = 0.66 since minimum gradients are obtained near this (minimum) condition. Curves A, B and C in Fig. 7 have gradients dh/dFN of 56, 17 and 11.5 km/MHz respectively, at the points T. The last figure is for a valley width of 39 km, and an unreasonably large depth of 0.77 (due to the neglect of other ionising processes in the valley region).

The six model test profiles used by URSI W.G. G/6/2 (McNamara and Titheridge, 1977) have gradients at the top of the valley equal to 100, 90, 70, 62, 41 and 19 km/MHz. The last value corresponds to an overlapping-Chapman model with a large valley depth of 0.88 MHz. The condition dh/dFN > 15 km/MHz therefore seems a reasonable requirement at the top of the E/F1 valley. For an F1/F2 valley the slope will be considerably greater. For valleys within the E region the slopes may be less. The local scale height SHA (Section 7.2) is therefore taken as the lower limit for the value of dh/dFN above a valley. At heights of 120, 140, 160 and 180 km this gives minimum gradients of 10, 15, 20 and 25 km/MHz respectively.

Values of dh/dFN become too small when the calculated valley is too wide. After each valley calculation in POLAN, results are checked to see if the gradient term q_1 is greater than SHA. If not, the equation $10q_1 = 10SHA$ is added. The new least-squares solution then has a gradient $q_1 \simeq SHA$; a decreased curvature term q_2; and a decreased valley width parameter q(JM). q_2 and q(JM) are further checked as in (b) and (c) below. The altered valley width is then used to calculate a new value for the valley depth, and the complete calculation is repeated once (as in Section 7.3.4).

(b) F-region curvature.

The curvature of the real-height profile at the top of the valley region is negative for physically reasonable profiles. For the profiles A, B and C of Fig. 7 the values of d^2h/dFN^2 at the point T are -640, -2.5 and -0.1 km/MHz^2 respectively. The deep V-shaped valley of C seems quite untypical of experimental profiles. Shallower valleys with the same width will have negative curvatures, similar to profile B, above the valley. The constraint currently adopted in POLAN is $d^2h/dFN^2 < -3$ km/MHz^2, or $q_2 < -1.5$. The purpose of this constraint is to rule out profiles of the type shown by the dotted line in Fig. 7. This is an alternative solution of the virtual-height trace corresponding to profile B; the too-small value of dh/dFN above the valley, the positive curvature, and the gradient discontinuity at T' all suggest that this solution is physically implausible. In practice the limit on q_2 is seldom invoked by POLAN, since condition (a) above normally ensures compliance with (b).

(c) Valley Width.

Adjustments (a) and (b) above reduce the valley width to an acceptable value, if it was initially too large. Occasionally the data lead to a too-small valley width, giving q(JM) negative so that VWIDTH < PARHT. The equation $10q(JM) = 1.0$ is then added to the least-squares solution, forcing $q(JM) = 0.1$ and VWIDTH - PARHT = 0.1 km. The decreased value of VWIDTH gives a smaller valley depth, and hence a smaller VWIDTH, in the next iteration.

7.3.4 Valley depth iteration

The first valley calculation uses a depth VDEPTH obtained from (14), assuming the standard value of VWIDTH. Solution of the augmented set of equations gives the profile parameters q(1) to q(JM), and a new value for VWIDTH. To maintain approximately the standard valley shape, the valley depth must now be altered to correspond to the calculated width. This is done by inserting the new VWIDTH in (14) to give a new value for VDEPTH. The full set of simultaneous equations is then recalculated and solved to obtain final values for the profile parameters. Further iterations are not necessary since valley depth is less important than width in defining the heights of the upper layers (as discussed in Section 9).

7.4 Valley Options in POLAN

Section 7.3 describes the normal valley calculation used in POLAN when extraordinary-ray data are not provided. The result is obtained by a least-squares solution of the virtual-height equations along with several additional equations which bias the results towards the standard valley model. This is the default procedure, obtained when:

(i) the final parameter VALLEY in the subroutine call to POLAN is zero, and

(ii) the virtual height h'(FC) corresponding to the critical frequency FC of the lower layer is zero. This data point is used to mark the transition between layers; the value of FC may be scaled or zero, as in Section 10.2.

Other options are obtained by specifying a non-zero value for VALLEY or for h'(FC). The type of analysis is controlled by a parameter HVAL within POLAN. This is normally set by the parameter VALLEY in the call to POLAN. If in any data set the height h'(FC) is non-zero, however, HVAL is set equal to h'(FC) for that analysis. Thus to use a particular valley analysis throughout a run, the main program sets the required constant in VALLEY. For individual profiles the type of analysis can be changed by giving a non-zero value for h'(FC).

The different types of analysis are defined below. In all cases positive values of VALLEY (or of h'(FC)) apply scaling factors to the standard valley, to increase or decrease the width and depth. Values of 5.0 or 10.0 give the extreme upper and lower limit profiles. Negative values of VALLEY specify absolute values of valley width and/or depth to be used in the analysis. The value of $|h'(FC))|$ must always be less than 30, to show that it is not a measured virtual height.

(a) VALLEY = 0.0 or 1.0 gives the normal analysis, with additional equations included in the least-squares solution (Section 7.3.2) so that the result will tend towards the standard valley model. (The added physical constraints can be omitted by using a negative value for AMODE, as described in Section 10.2.2.)

(b) VALLEY = 0.1 to 5.0 uses the normal valley analysis, but with the standard valley width multiplied by the factor VALLEY. If the resulting width proves too large for compatibility with the virtual-height data, the initial gradient (q_1) at the start of the following profile section is too small; this will be corrected automatically by setting q_1 equal to the minimum allowed value (Section 7.3.3), and adjusting the other real-height parameters to obtain a new least-squares solution. Thus the width is reduced as required until an acceptable gradient is obtained above the valley. A new (smaller) value of VDEPTH is then calculated, and the analysis is iterated once.

(c) VALLEY = 5.0 is used to give a maximum-valley result, and an upper-limit profile. This is the extreme case of (b). A valley width of 5 times the standard value is almost always impossibly large, and so will be reduced to obtain reasonable positive values of dh/dFN above the valley. Thus the final result has the greatest width which will give a physically acceptable profile for the upper layer. Typical results are shown by dotted lines in Fig. 8.

(d) VALLEY = 10.0 is used to specify a monotonic result, with no valley. This gives the lower limit to the range of possible real-height profiles (the chain lines in Fig. 8).

(e) VALLEY = -0.01 to -0.99 specifies that the calculations are to use a valley depth VDEPTH = |VALLEY| MHz, instead of the value obtained from equation (14). The valley width remains standard.

(f) VALLEY = -1.0 requests a two-parameter calculation of valley depth and width, when extraordinary ray data are available, as described in Section 9.3. Using VALLEY = -1.X forces the depth iteration to begin from VDEPTH = 0.X MHz, and is used primarily for tests of the type described in Section 9.4.

(g) VALLEY = -N, where N is an integer in the range 2 to 30, specifies a valley width of 5N km. This value is enforced by using a weight of 10 with the corresponding equation (16) in the least-squares solution.

(h) VALLEY = -W.D, where W and D are non-zero, is a combination of (e) and (f). The calculation uses a valley width of 5W km and a depth of 0.D MHz. Thus VALLEY = -6.75 gives a profile with VWIDTH = 30 km and VDEPTH = 0.75 MHz.

When the value of VALLEY (or of h'(FC)) is less than -2.0, the weight given to the valley-width equation in the least-squares solution is increased from WVAL = 1.0 to WVAL = 10.0. Results will then conform closely to the specified values of VWIDTH and VDEPTH.

The effects of changes in valley width on the calculated F-region real heights are shown in Fig. 9. Since the equations defining the real-height profile are linear (Section 3), the changes in real height are proportional to the changes in valley width. This is strictly true only for a fixed value of valley depth. The effect of depth variations is small, however, so the curves of Fig. 9 apply with reasonable accuracy for all but very deep valleys. The curves shown are calculated for typical E-layer critical frequencies of 3.1 MHz at 30° dip and 1.4 MHz at 70° dip. Ordinary ray calculations are not very sensitive to changes in magnetic field and, when variations in FB are ignored, depend only on the ratio f/FC. Fig. 9 may therefore be used to estimate the effect of changes in the assumed valley width, on the calculated real heights of the upper layer, under most conditions. Using the broken line in Fig. 9 gives the following approximate result: the change in calculated real height $\triangle h$ above a valley, due to a change DWIDTH in the assumed valley width, is

$$\triangle h = 0.8\ \mathrm{DWIDTH}.(\mathrm{FC}/f)^2 \qquad (19)$$

where f is the plasma frequency at which the real height is calculated and FC is the critical frequency of the underlying peak. For plasma frequencies greater than 2FC the real-height error decreases approximately as 1/f, as shown by the dotted line in Fig. 9, giving

$$\triangle h = 0.4\ \mathrm{DWIDTH}.\mathrm{FC}/f \qquad (19b)$$

Changes can be made in the shape of the valley used by POLAN, by modifying the valley constants which are stored as DATA statements in the program STAVAL (Appendix F.2). The valley has a parabolic section covering a height range PARHT in Fig. 6. This distance is calculated using a scale height which is greater than that of the sub-peak profile by a factor VPEAK. Above this is a flat valley base extending over a distance (1-VBASE).q(JM). The constants normally used are VPEAK = 1.4 and VBASE = 0.6. By altering these values a wide range of different shapes can be obtained. Thus a simple triangular valley is obtained if VPEAK = VBASE = 0.0, and a rectangular valley if VPEAK = 0 and VBASE = 1.0. The relation between the depth and width of the valley is set by equation 14 which uses the constants VDEEP = 0.008 and VCONST = 20. Variation of these parameters (by changing the DATA statement in STAVAL) will change the ratio of depth to width for all valleys. For example, setting VDEEP = 0.016 will double the depths of all the valleys shown in Fig. 17 (section 9.2). Setting VCONST = 0.0 will make the valley depth directly proportional to the calculated width.

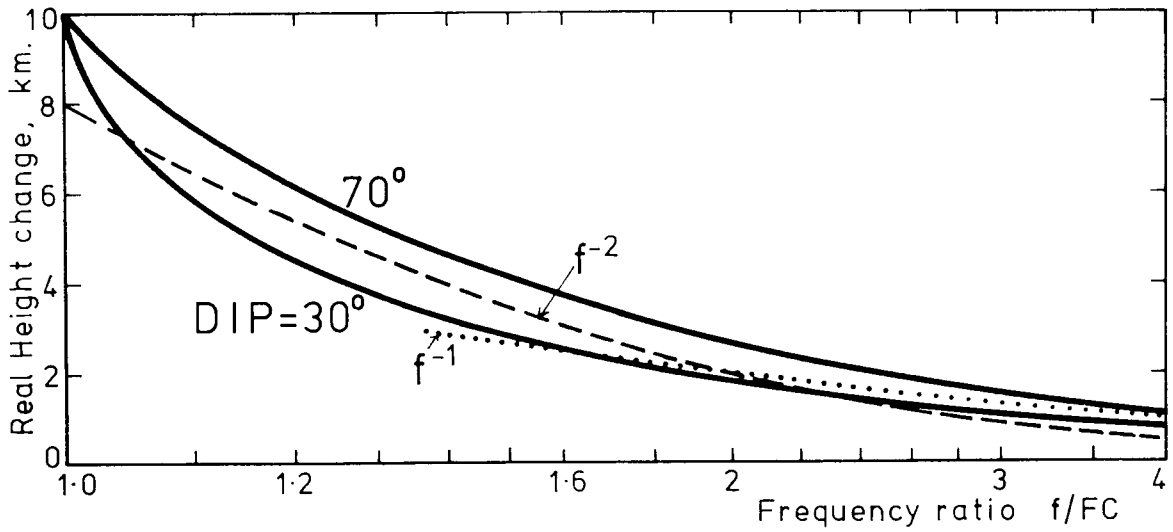

Figure 9. Changes in the calculated real heights of the F layer, due to an increase of 10 km in the assumed width of the E/F valley, at magnetic dip angles of 30° and 70°. FC is the E layer critical frequency. Broken and dotted lines show, for comparison, real height changes proportional to $1/f^2$ and $1/f$ respectively.

8. START CALCULATIONS USING THE EXTRAORDINARY RAY

8.1 Practically Obtainable Information

Ionograms give the virtual heights of the ordinary (O) and extraordinary (X) rays only down to some limiting frequency fmin. For the X ray we define fmin as the plasma frequency FN at which the lowest observed frequency is reflected, and will assume in the meantime that this is similar to the frequency of the lowest O ray. Reflections are not obtained from the ionised region with FN < fmin. Information about this underlying region can however be obtained from the difference in the virtual heights of observed O and X rays. For this purpose we consider "corresponding" O and X frequencies, which are reflected at the same real height and have the same value of plasma frequency at reflection FR. The wave frequencies f_O and f_X of the corresponding rays are related by

$$FR = f_O = (f_X(f_X - FH))^{0.5} \tag{20}$$

where FH is the gyrofrequency. The group retardations for corresponding O and X rays are then

$$h'_O(FR) - HR = \int (\mu'_O(FR,FN) - 1)\,(dh/dFN).dFN \tag{21a}$$

$$h'_X(FR) - HR = \int (\mu'_X(FR,FN) - 1)\,(dh/dFN).dFN \tag{21b}$$

The group refractive index μ' is a function of the wave frequency f (or equivalently the plasma frequency FR at reflection), the plasma frequency FN, and the magnetic field constants gyrofrequency (FH) and dip angle (I). For vertical propagation at a given site the dip angle is constant. The height variations of FH are considered in Appendix C, and will be ignored here. μ' then depends only on FR and FN. The form of this dependence is quite different for the O and X rays. The ratio

$$R_{X,O} = (\mu'_X - 1)/(\mu'_O - 1) \tag{22}$$

also varies considerably with FR, at fixed FN (Appendix B.1). Thus the effective weights $(\mu'-1)$ in (21a) and (21b) are different functions of FN, and the form of these functions changes as FR varies.

Ionisation in the unobserved region at FN < fmin is represented by an equivalent monotonic profile, which has only one height at each value of FN. It can then be shown that an exact knowledge of the functions $h'_X(FR)$ and $h'_O(FR)$ over any finite interval FR_1 to FR_2 is sufficient (in theory) to define the function dh/dFN for all FR < FR_2. In practice we do not have an exact mathematical function for the virtual heights, but a small number of data points with finite errors. The extent of the information which can be obtained about the underlying region then depends on the size of the difference $\mu'_O - \mu'_X$, and the amount by which it varies with FN and with FR. Observed values of h'_X and h'_O give no information about the distribution of ionisation in any region where $R_{X,O}$ (equation 22) does not change significantly with FN.

Group index calculations (Appendix B.1) show that $R_{X,O}$ is approximately independent of FN, at a given FR, when FN < 0.7 FR. This is true for all frequencies FR at dip angles less than 30° or greater than 60°. At intermediate dip angles it is true for FR > 2FH. Under these conditions, and taking FR as the minimum observed frequency fmin, no information about the distribution of ionisation at FN < 0.7fmin can be obtained from the observed O and X-ray virtual heights. This is true for any method of analysis. Thus a basic requirement for a stable analysis procedure, producing consistent results, is that it should not attempt to determine the distribution of ionisation in the region where FN < 0.7fmin.

For dip angles near 67°, the variations of μ'_O and μ'_X with FN can be closely approximated by expressions of the form $\mu' = 1 + a.FN^2 + b.FN^8$, for FN < 0.9FR (Appendix B.2). a and b are independent of FN, at a given frequency FR. Group delays due to this region are therefore defined (to within about 1%) by the integrals of FN^2 and FN^8. So for the region with FN < 0.9fmin only these two parameters can be determined.

At dip angles near 80° the ratio $R_{X,O}$ in (22) is approximately equal to 8.5FH/FR, for 0 < FN < 0.98FR (Appendix B.1). Observed delays then give only one piece of information about the ionisation with FN < 0.98fmin. At dip angles near 30° the ratio $R_{X,O}$ is approximately constant, at a given FR, for all plasma frequencies FN less than FR (Titheridge, 1974b). Only one parameter can then be obtained relating to the unseen ionisation. This single parameter is, however, sufficient to correct the real heights at FN > fmin for the effect of the underlying region.

We conclude that, under most conditions, not more than two independent parameters can be obtained for the underlying region at FN < 0.9fmin. One of these parameters normally defines the total electron content of this region, while the second gives some distribution function. Analysis of practical ionograms (Titheridge, 1959b,c) shows that the total content is generally well defined, while the second parameter is obtainable with only marginal accuracy.

For the region with 0.9fmin < FN < fmin a single parameter will normally suffice, to define the mean gradient just below fmin. Thus a total of two or three parameters is fully sufficient to fit all available information about the underlying region. The only exception to this (apart possibly from calculations at dip angles near 50°) occurs when the group retardations become very large near fmin, showing the presence of an underlying peak -- as discussed in Section 8.5.2. Paul and Smith (1968) also conclude that generally only 1 or 2 parameters can be determined for 'unseen' regions. Tests carried out by Gulyaeva (1973) showed clearly that the use of many parameters for the unobserved regions led to unreliable or ambiguous results in many cases; restriction to a one-parameter model avoided most of these problems. Note that a special problem arises at dip angles near 35°, such that start and valley calculations of reasonable accuracy become almost impossible for ionograms taken at dip angles between 33° and 38° (Appendix B.2).

8.2 Inclusion of Extraordinary Ray Data

8.2.1 The general approach

Real-height calculations based primarily on O-ray data can be adjusted iteratively to optimise agreement with a number of X-ray measurements. The process consists essentially of making successive modifications to the profile shape at FN < fmin, to obtain best agreement between observed and calculated X-ray virtual heights. The final result is a profile which exactly fits the O-ray data, and gives some best overall fit to the X-ray data. Iterative adjustment of the profile in this way is described by Lockwood (1969, for the topside case when X data are analysed and iteratively adjusted to fit O-ray observations), by Gulyaeva (1972, 1973), by Becker (1978) and by Paul (in Gulyaeva et al, 1978). Howe and McKinnis (1967) use a least-squares fit to O and X data, repeating this with different frequency intervals for the underlying region until results appear satisfactory.

Iteration can be avoided if the model for the underlying ionisation does not include any variable frequencies. A single least-squares solution can then be obtained incorporating both O and X data. This is the method used in POLAN. Neither O nor X rays are fitted exactly, but the overall RMS error will be appreciably less than if the fitting errors were confined to only one component. The two components may be given different weights, by adjustment of the constant WVIRTX in POLAN; the relative weight given to X-ray measurements varies as $(WVIRTX)^2$. In a combined O and X ray fit, the total number of real-height terms being calculated is normally greater than the number of O rays used. A small value for WVIRTX will then give a result in which the O-ray data are fitted (almost) exactly, and the RMS deviation is minimised for the X-ray data. Thus we can obtain, in one step, the types of solution provided by most of the iterative procedures mentioned above.

In general, when highly retarded X traces are avoided, it can be shown that use of similar weights for both the O and X data gives maximum stability and insensitivity to measurement errors. POLAN therefore begins an analysis with WVIRTX = 1.0. If there is an apparent conflict between O and X data, WVIRTX is reduced to 0.5 so that O-ray virtual heights are fitted 4 times more accurately than the X data. This gives results similar to those from the iterative approach, with a decreased sensitivity to (for instance) errors in FH or the effects of magneto-ionic path splitting. Thus the analysis is automatically adjusted to reflect the relative degree of confidence placed in the X-ray data.

8.2.2 Implementation in POLAN

The virtual-height equations in the set (7), used to define a given polynomial, can include both ordinary (O) and extraordinary (X) rays. X-ray data are identified in the calculations by using negative frequencies; this automatically gives the right sign at all points in the group index subroutine (Appendix D.3). Calculation of the coefficients B(i,j) then requires no special considerations other than using the correct upper limit of integration in (4) - this must be the plasma frequency at reflection FR, equal to $(f_i(f_i + FH))^{0.5}$ when the wave frequency f_i is negative. The optimum value of FH for use in start calculations is discussed in Appendix C.

With O-ray data only, the frequency up to which the starting polynomial extends is determined by the number of data points (NV) used by the specified mode of analysis (Section 5.2). If X-ray data are also available the range of the polynomial generally is not altered; those X-ray points which fall within this range (or within 0.1 MHz of the top frequency) are included, and any outside this

range are ignored. This is in agreement with the general philosophy of building choices into the program wherever possible, rather than depending on the judgement of an operator.

If the number of accepted X-ray points is NX, an additional NX virtual height equations appear at the beginning of the set (7). Thus the total number of equations to be solved becomes NV = NF + NX where NF is the number of O-ray data points. The value of NT (the number of terms in the fitted polynomial) is increased by (NX + 1)/2 over the value originally specified. One further term is added corresponding to the constant in the polynomial expansion, to allow adjustment of the starting height (Section 8.3.1). Thus for the standard analysis (MODE = 5) a start with O-ray data only uses a 4-term polynomial fitted to the first 5 virtual heights. If one or two X-ray points are added, this becomes 6 terms fitting 6 or 7 heights. With equally-spaced O and X data we get NX = 5, and POLAN fits 8 real-height terms to 10 virtual heights (5 O-ray plus 5 X-ray). Equations which bias an ill-defined result towards an acceptable model starting height are also included in the least-squares analysis, as described in Section 8.4.

The number of real heights determined from the first polynomial is always less than the number of O-ray data points used to calculate the polynomial coefficients. Thus for a Mode 5 analysis, the start calculation is used to determine only the real heights h_1 to h_3 at the O ray frequencies f_1 to f_3. These heights depend on both O and X-ray data. The next step in the analysis calculates a polynomial passing through h_1, h_2 and h_3 and giving a least-squares fit to the O-ray virtual heights h'_3, h'_4, h'_5 and h'_6. This polynomial defines the next real height h_4 at the frequency f_4. The accurate fit to h_1, h_2 and h_3 ensures a smooth continuation, so that h_4 still has a considerable dependence on the X-ray data used to determine the first three heights. O-ray virtual heights h'_4 and h'_5 are used in both the first step of the analysis (in conjunction with X-ray data) and in the second step (without X-ray data). This gives a smooth transition from initial heights, which depend on both O and X data, to a profile variation (at high frequencies) governed by the O-ray only. The net result is similar to the procedure used by Wright (1967), who instructs operators to scale additional O-ray data at the top of an O/X fitting region.

8.2.3 Data selection within POLAN

With data scaled at a normal frequency interval Δf of about 0.1 MHz (Section 8.5), the points used in a starting calculation cover a frequency range of about 0.4 MHz. When virtual heights are changing only slowly above fmin the value of Δf may be increased, giving a corresponding increase in the fitted range. When virtual heights increase rapidly near fmin it is not correct to use smaller values of Δf, in an attempt to obtain greater accuracy by defining the changes in h' more closely. Such changes are due to the proximity of an underlying peak, and this section of the trace is best omitted (Section 8.5.2). Values of Δf appreciably less than 0.1 MHz may, however, be desirable for two reasons:

(a) Experimental errors in the measurement of h' and of f can be reduced by scaling additional points to be smoothed in the least-squares fitting process.

(b) Virtual heights may begin to increase rapidly with increasing frequency, within 0.4 MHz of fmin, due to the presence of a cusp or peak on the ionogram. Such data should not be used in the start calculation, since changes in h' will depend more on the changing gradient near reflection than on the density of underlying ionisation. Thus at least 5 O-ray points should always be scaled before appreciable peak retardation begins. (An example of the large errors produced by including cusp data which does not exactly fit the assumed value of FH is given in Section 8.4.2(b).)

Starting data may be provided at frequency intervals appreciably less than 0.1 MHz for either of the above two reasons, or through an operator mistakenly trying to define a rapid increase in h' near fmin. These cases are treated appropriately in POLAN by the following steps.

(i) The specified number of O-ray frequencies (NF, equal to 5 for the normal analysis) is selected. Each point is checked, in the normal way, to ensure a monotonic frequency variation and an absence of 'singular' points corresponding to a cusp, restart or peak. If [h'(NF) - h'(NF-1)] > 200[f(NF) - f(NF-1)] the last point is deleted, in accordance with (b) above.

(ii) If the number of initial X rays (NX) is not zero, and the NF selected O-ray points cover a frequency range of less than 0.4 MHz, the next O-ray point f(NF+1) is checked. If satisfactory, and if the value of [h'(NF+1) - h'(NF)]/[f(NF+1) - f(NF)] is less than 30 km/MHz, NF is incremented by one to include the additional point.

(iii) Step (ii) is repeated until either the range of frequencies exceeds 0.4 MHz, or the gradient of the virtual height curve exceeds 30 km/MHz. Any X-ray points which are reflected at plasma frequencies more than 0.05 MHz greater than the last included O ray are then deleted.

Two constants are used in this selection process. These are the minimum desirable frequency range for the start calculation (FFIT, normally 0.4 MHz), and the maximum positive virtual-height gradient to be allowed at the top of the range (GFIT, normally 30 km/MHz). The 'normal' values are set by a DATA statement at the beginning of POLAN, and may be changed as required for particular studies.

8.3 The Slab Start in POLAN

8.3.1 The start model

The main start model used in POLAN for X-ray calculations is shown in Fig. 10. F1 is the lowest plasma frequency from which reflections were obtained. The starting model consists of a polynomial beginning at FA = 0.6F1, and an underlying slab with FN increasing linearly from 0.3F1 to 0.6F1. This gives two basic parameters for the underlying ionisation, of the general type described in Section 8.1. The linear slab section serves to produce the correct total amount of low-density ionisation. The initial section of the polynomial from 0.6F1 to F1 defines the mean gradient dh/dFN near 0.8F1, and gives a continuous variation of all derivatives through the lowest observed frequency F1. After a slab start real heights are calculated at FN = 0.3F1, 0.6F1 and 0.8F1.

With O-ray data only, calculations start from a height obtained by linear extrapolation of the first three O-ray measurements (Section 6.2). With X-ray data the same linear extrapolation is used, but it is continued only down to the frequency 0.8F1. The height HA obtained in this way is taken as the origin for the starting calculation, at the frequency FA = 0.6F1. This limited extrapolation gives a minimum amount of correction, so that HA gives a reasonable upper limit for the true height at 0.6F1. Physically reasonable results can then be assured by requiring that the final starting height be less than HA, as described in Section 8.4.

The real-height section above the starting frequency FA is represented by

$$h - HA = \sum_{j=1}^{MT-1} q(j).(f-FA)^j + q(MT+1) \qquad (23)$$

The term q(MT) corresponds to the thickness of the underlying slab, from FN = 0.3F1 to 0.6F1. This term does not appear in (23) since the corresponding coefficients B(i,MT) are zero in the real height arrays (equations 5 to 8). For the virtual-height arrays the coefficients B(i,MT) give the mean value of $(\mu' - 1)$ in the linear slab, at the frequency F_i. The calculated value of q(MT) then gives the thickness of the slab.

The final term in (23), at j = MT+1, provides the constant term required to allow a shift in the starting height of the polynomial. Thus the calculated value of q(MT+1) gives the real-height correction at the frequency FA (Fig. 10). For this term both the real and virtual height coefficients B(i,MT+1), in equations 5 to 8, are equal to unity. The constant term is placed at the end of the expansion so that:

(a) the initial terms are the same for both O and X-ray calculations, and

(b) the coefficient q(MT+1) is readily modified or removed from the least-squares solution, if the calculated value is physically unreasonable.

8.3.2 Test examples

Two test models were developed for evaluating start procedures. These are based on Arecibo backscatter data (McNamara and Titheridge, 1977) and are shown by the solid lines in Fig. 11. Corresponding virtual heights h'_O and h'_X, for a dip angle of 20°, are given in Fig. 15 (Section 8.6.1). The first model incorporates a slowly-changing gradient above 0.5 MHz, and much of the group retardation (h' - h) is due to the fairly large gradient dh/dFN in the reflecting region. The second model (Fig. 11b) has a smaller and more constant gradient in the reflecting region, with a rapid increase in the gradient (and in the resulting group retardation) at frequencies below 1.0 MHz.

The dotted section in Fig. 11(a) is the equivalent monotonic representation for the underlying ionisation. This gives exactly the same virtual heights for both O and X rays (at frequencies above 0.6 MHz) as the true profile. Thus it is an exactly equivalent solution of the given virtual height data, and is the result towards which an ideal analysis should tend. The slab start calculation with fmin = 1.0 MHz, giving the circled real heights at 0.3, 0.6 and 0.8 MHz, does represent this equivalent profile quite accurately.

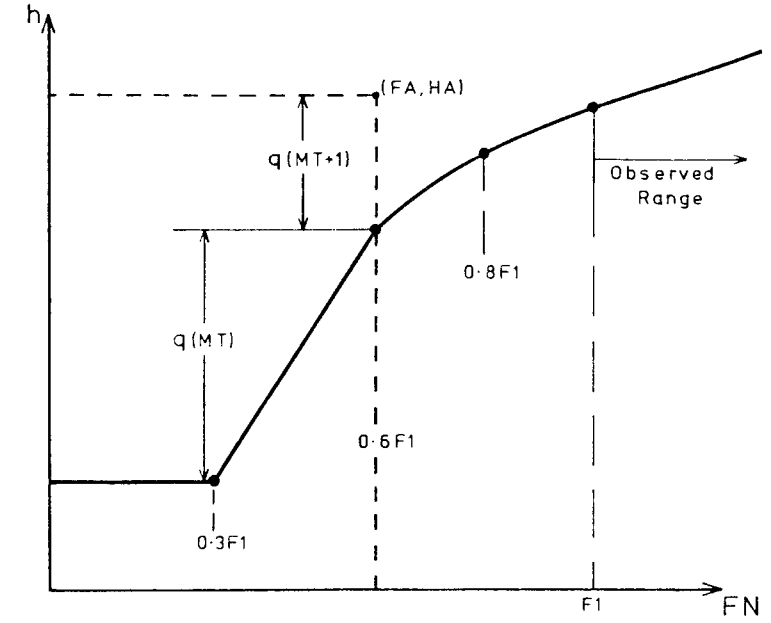

Figure 10. The general form of the Slab Start profile, using virtual-height data down to a minimum (plasma) frequency F1.

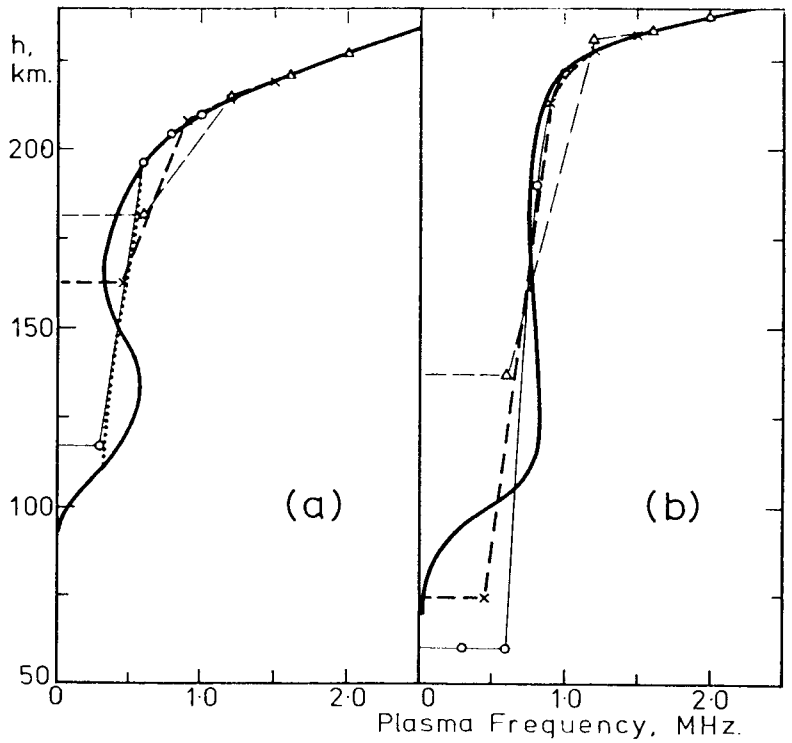

Figure 11. Solid lines show the model real-height profiles 3A and 3B. Circles (joined by continuous lines), crosses (with heavy broken lines) and triangles (with thin broken lines) give the calculated real-height profiles at an angle of 20° using data down to fmin = 1.0, 1.5 and 2.0 MHz respectively. Results at 70° dip are shown by solid symbols wherever these can be distinguished from the 20° results.

Calculations using fmin = 1.5 or 2.0 give a simplified representation of the underlying ionisation, shown by the broken lines in Fig. 11(a). These approximate profiles still give quite closely the correct value for the integral of FN^2dh. They fit the observed virtual heights, at all frequencies, to well within experimental error. The data set for profile 3B at Dip 70° would not be analysed from 1.0 MHz, as discussed in Section 8.5.2. The maximum R.M.S. fitting error for the other 11 calculations in Fig. 11 is DEVN = 0.01 km. The mean value is DEVN = 0.003 km. So no further information is obtainable about the unobserved region, at FN < fmin, even if all virtual heights are known to an accuracy of about 3 metres (i.e. to 1 part in 10^5).

Using O and X data at the normal frequency spacing of about 0.1 MHz, the default mode 5 in POLAN selects 5 O-ray points and 5 X-ray points for the slab start calculation (Section 8.2.3). This gives an 8-term polynomial. Calculations were also repeated using 7 terms. For each model and value of fmin (apart from the case discussed in Section 8.4.2(b)) the four independently calculated results for the thickness of the low-density slab agreed to about ±10%. Values for the electron content at FN < 0.8fmin agreed to within ±1.1%. This confirms that only the total amount of ionisation, and not the distribution, is important at FN < 0.8fmin. Calculated heights at 0.6fmin had a mean spread of ±1.8 km within each group. The solution is therefore stable and well-defined in the unobserved region. In the observed region all calculated heights were accurate to within about 0.1 km at a dip angle of 20°, and 0.5 km at Dip 70°, for all values of fmin.

Values of h'_X differ appreciably at dip angles of 20° and 70° -- particularly for model 3B. The change of dip angle gives a large change in the relative importance of different densities of underlying ionisation, as discussed in Appendix B.1. The good consistency of the calculated model for the unseen region, with changes in dip angle and in the degree of the fitted polynomial, shows that we have a reliable, well-defined result. Calculations under a wide range of conditions show that the real-height errors vary smoothly with dip angle (apart from an increase near dip 35°, as in Appendix B.2). For dip angles near 29° good accuracy is still obtained in the observed region although the distribution at FN < fmin can be quite unreliable, as discussed in Section 8.1. The mean accuracy of the calculated heights in the observed region is slightly better than for the polynomial start of Section 8.6. Variations in the unseen region are, however, only about half as large. This increased consistency is the main reason for preferring the slab start in POLAN.

8.4 Physical Constraints and Height Iteration

8.4.1 Constraints included in the least-squares solution

Start calculations begin from a height HA at a starting frequency FA (Fig. 10). With O-ray data only, HA is set equal to an estimated starting height HS. At START = 0., HS is calculated from a limited extrapolation of the initial virtual-height trace. For more reliable and consistent results, START is used to enter a suitable model starting height (obtained from the results of Sections 6.3 and 6.4).

When X-ray data are used in a start calculation, HA is set equal to a reasonable upper limit for the real-height at the frequency FA (as described in Section 8.3.1). The combined O/X analysis then calculates the real-height correction q(MT+1) at this frequency, and the thickness q(MT) of a linear slab representing the total amount of underlying ionisation (as in Fig. 10). This calculation makes no direct use of the model starting height HS. With good data the least-squares solution of the virtual-height equations gives well-defined results for the parameters q_j, corresponding to a sharp minimum in the RMS virtual-height fitting error. Data are normally less than ideal, however. The absence of X-ray data at low frequencies, off-vertical propagation of some rays, and other errors produce increased uncertainty in the least-squares solution. Results still give the values of q for which the real-height profile best fits the virtual-height data, but this fit may be nearly as good (or as bad) for a considerable range of values.

With poor data we would like results to be biased towards the O-ray model, by an amount which increases as the broadness of the minimum in the least-squares solution increases. This is achieved by adding further equations to the least-squares solution, setting q(MT) and q(MT+1) to values which agree with the model O-ray starting height HS. These equations are given some low weight, so that they have little effect on a well-defined X-ray calculation. An ill-defined solution will, however, be shifted towards the model values of q(MT) and q(MT+1).

To include physical conditions in a combined O and X-ray starting calculation, the following equations are added to the set of virtual-height equations used in the least-squares solution.

$$q(MT+1) = 1.6\,(HS - HA) \qquad (24)$$

$$0.3q(MT) = 0.1HS - 6 \qquad (25)$$

$$0.3q(MT-1) = 0. \qquad (26)$$

The first two conditions define a slab start profile (Fig. 10) which corresponds approximately to the O-ray profile with a starting height HS (at the starting frequency f_s, normally 0.5 MHz). The third condition indicates that we would prefer the last term in the polynomial real-height expression to be small (giving a lower-order solution, if the data do not require otherwise). The factor 0.3 in (26) gives this relation a small effective weight in the analysis, only one tenth of that for equations (24) and (25).

Both sides of equations (24) to (26) are multiplied by a weighting factor WS when they are added to the set of simultaneous equations in the matrix B (Section 7.3). When HS is estimated by extrapolation of the initial O-ray trace, we use WS = 0.1. The added conditions then have a weight of 0.01, so that they normally have little effect on the solution. When HS is defined by an entered model value of START, the weight WS is increased to 0.5. The calculated profile will then be biased considerably towards the starting model, whenever the data by themselves would give an ill-defined or ambiguous result.

Note that the above constraints have NOT been included in any of the examples discussed within Section 8 of this report. Thus Figs. 11 to 16 show results which depend only on the virtual-height data. Incorporation of equations 24 to 26 in the least-squares solution, using any reasonable value for HS, considerably reduces the variations obtained with poor data.

8.4.2 Limits placed on the calculated profile

With a simple two-parameter representation of the underlying ionisation, as in Fig. 10, physically reasonable virtual-height data always produce physically reasonable results. We can therefore require that basic physical criteria are satisfied in the unobserved region. Occasionally, with bad input data, a physically acceptable shape cannot be obtained without incurring a large increase in the virtual-height fitting error. In such cases some unphysical variations are allowed in the unobserved region, to give a better-defined profile at f > fmin. Adjustment of the solution to impose some physical restriction is done accurately and rapidly by adding a further equation into the previous least-squares solution, using the subroutine SOLVE (Appendix F.4). This gives new values for all real-height parameters q(j), and for the virtual-height fitting error DEVN. Several modifications of a calculated profile can therefore be tried, adding or adjusting constraints until the solution is physically acceptable, with no significant increase in computation time.

Application of physical limits gives an automatic reduction in the effects of bad data in almost all cases. Conditions (i) to (iii) below are applied by the subroutine ADJUST immediately after the first real-height solution. (iv) and (v) are checked in section CX2 of the subroutine STAVAL. Conditions are applied in the order shown, so that (iii) is not checked until a solution has been obtained which satisfies (i) and (ii). In most cases curing one bad condition (e.g. a negative initial gradient) will also bring about conformity with the following conditions. The real-height parameters directly involved in checks (i), (ii) and (iii-iv) are q(1), q(MT) and q(MT+1) respectively. When applied in this order the addition of one constraint (by the subroutine SOLVE) retains all previous constraints. Each constraint is added to the set of simultaneous equations with a weight of 20 so that the new condition is closely met by the new least-squares solution. The large weight ensures that addition of subsequent constraints does not significantly alter the previously-constrained parameters. The constraints applied are:

(i) The initial gradient q(1) = dh/dFN must be greater than 2.0 km/MHz. Thus if initially q(1) < 2.0, the condition q(1) = 2.0 is imposed on the solution.

(ii) The thickness of the slab of underlying ionisation, q(MT), must be positive. If initial calculations give a negative result the condition q(MT) = 0.1 is imposed.

(iii) The calculated height at the frequency FA must be less than HA. This requires that the last parameter q(MT+1) in the real-height expansion must be negative.

(iv) Within the observed range, successive calculated real heights must give a mean gradient $\Delta h/\Delta f$ greater than 2.0 km/MHz.

(v) Finally, if a height-dependent gyrofrequency is being used, the entire starting calculation is repeated using the previously-calculated real heights to obtain a more accurate value of FH for each ray. Iteration continues until the changes in real height are less than 2 km (giving an accuracy of better than 0.1% in the value of FH at reflection). Optimum values of FH for use in the unseen region are described in Appendix C.5.

With the overlapping-polynomial approach, the highest frequencies employed in each step of the calculation serve only to ensure the correct overall trend for the upper part of the real-height segment. Thus the number NH of final real heights determined in each step is normally less than the

number of O-ray data points. Check (iv) is applied only to the first NH calculated heights, to ensure an acceptable profile. If the check fails, the start offset q(MT+1) is reduced by an amount equal to the greater of (10 km, 0.5|q(MT+1)|). Programming is arranged so that application of this new constraint automatically removes any previous constraints on q(MT+1). Real heights are then recalculated. Check (iv) is iterated (a maximum of 10 times) lowering the starting height until each final real-height step corresponds to a mean gradient of greater than 2.0 km/MHz.

The above checks take place automatically, adjusting the start solution to obtain a result giving the best (least-squares) match with the given data, subject to the requirement of a physically reasonable profile in both the observed and unobserved regions. With good data correctly scaled, the first direct solution of the virtual-height equations seldom requires any adjustment. With poor data the calculated profile is not well-defined; this is indicated by a large value for the RMS virtual-height fitting error DEVN. The adjustments necessary to obtain a physically meaningful profile then cause little increase in DEVN, so that the final result represents an almost equally good mathematical solution of the given data.

A physically acceptable profile is always demanded in the observed region at FN > fmin. On rare occasions the initial least-squares solution is physically impossible in the unobserved region, but is well-defined in the sense that any changes cause a large increase in the fit parameter DEVN. This implies that the true distribution of underlying ionisation involves large gradient changes just below fmin. These cannot be adequately modelled by a monotonic two-parameter distribution (Section 8.5.2). This case is allowed for by examining the increase in DEVN caused by an application of constraint (ii). If DEVN increases by more than 25% when the slab thickness is set to zero, the constraint is removed to allow a negative value of q(MT). Real heights in the unseen region then decrease from 0.3fmin to 0.6fmin, and increase rapidly at FN > 0.6fmin. Real heights in the observed range (FN > fmin) are calculated using this unphysical variation of unseen ionisation. The true calculated start parameters, with a negative value for SLAB, are listed by POLAN. To avoid upsetting operators or plotting routines, however, the final real-height and frequency arrays returned by POLAN are modified to a monotonic variation with the same total electron content. This entails replacing the initial points by a flat base, extending up to a plasma frequency slightly greater than the original starting frequency FA (as in the example of Fig. 12).

POLAN normally lists only the final start, valley and peak parameters. When the input parameter LIST is non-zero, additional information is printed as detailed in Section 10.3.2. This includes the initial parameters q, any adjustments made (in checks (i) to (v) above), the reason for the adjustment and the resulting increase in DEVN. Thus with poor data, for which the best mathematical solution is physically undesirable, we get an overall best-guess solution and a listing of the adjustments which have been made. If these adjustments are large, the possibility of scaling errors should be examined.

8.4.3 Examples

Note that all the calculations described within Section 8 are made without including the physical constraints of Section 8.4.1 in the least-squares solution; thus the variations shown in Figs. 12 and 13 have not been restricted by any assumptions about the expected size of the starting correction.

(a) Using accurate data.

Of the calculations described in Section 8.2.2, only analysis of the difficult model 3B with fmin < 1.1 MHz gives unphysical start parameters. The extreme case occurs at dip 70° and is shown in Fig. 12. Direct application of the slab start calculation gives the variation shown by the continuous line. This attempts to simulate peak retardation by giving the maximum amount of high density ionisation (compatible with a smooth polynomial variation down to 0.6 MHz), plus a large negative amount of ionisation at FN < 0.6 MHz. The result fits O and X virtual-height data to well within experimental error, and gives good accuracy in the observed region, as shown by the "direct" results in Table 6.

TABLE 6. Results from the slab start calculation of Figure 12, where "Direct" results are plotted as a continuous line and "Adjusted" results as a broken line.

Analysis	q(MT)	Virtual-height error DEVN	Real-height error at f = 1.0	1.2	1.5	2.0	MHz
Direct	-310	0.07	-0.42	-0.32	-0.23	-0.13	km
Adjusted	0	0.34	+3.07	+0.12	-0.43	-0.65	km

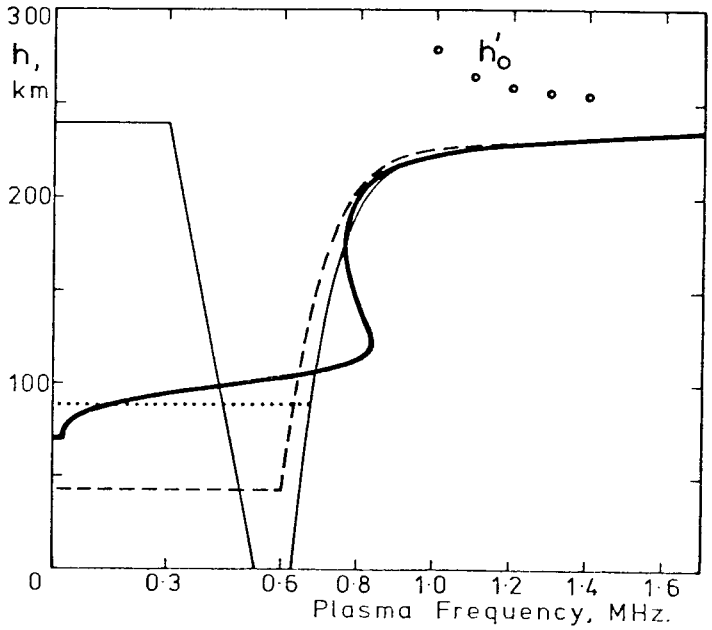

Figure 12. The test profile 3B (heavy line) analysed at dip 70°, using O and X-ray data at intervals of 0.1 MHz in FN from fmin = 1.0 MHz. The fine line is the direct result of a 'slab start' calculation. Imposing physical constraints on this result gives the broken-line variation.

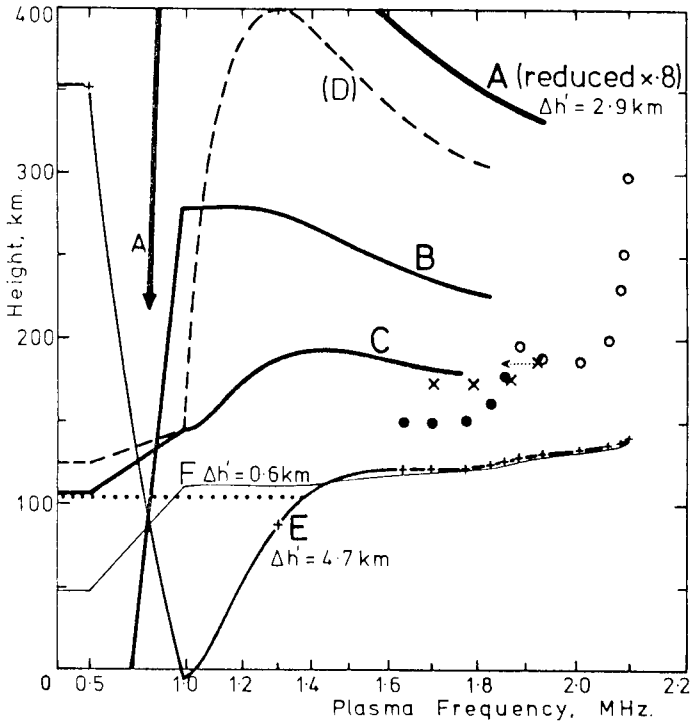

Figure 13. Analysis of ionogram data at a dip angle of 57°, using a too-small value of the gyro-frequency FH. The horizontal scale is proportional to FN^2. Profile A is the first result of the start calculation. Automatic corrections for initial gradient and start offset produce B and C. Elimination of negative gradients then gives E, where the calculated real heights are marked +. Analysis using the correct value of FH, with (automatic) rejection of the fifth O-ray point from the start calculation, gives the profile F which fits all virtual heights to within 1 km.

The calculated parameter q(MT), giving the thickness of the low-density slab, has a physically impossible negative value in the first calculation. POLAN then automatically adjusts the solution so that q(MT) is zero, giving the broken-line profile in Fig. 12. As shown by the bottom line of Table 6, this appreciably increases the real-height errors and considerably increases the virtual-height fitting error DEVN. The latter effect shows that the original unphysical profile was not an ill-defined solution, but a genuine (and worthwhile) attempt to cope with a difficult distribution of underlying ionisation. The constraint q(MT) = 0 is therefore removed, and the original solution accepted. The start constants are listed as:-

"START CORRECTION = -289 km, SLAB = -310 km".

Data in the real-height array are, however, modified as shown by the dotted line in Fig. 12. This gives a flat base with a height and plasma frequency (0.67 MHz) calculated to give the same total electron content as the original (continuous line) calculation.

(b) With bad data.

Common data errors, and the resulting treatment, are illustrated in Fig. 13. Circles and crosses show O and X ray data, scaled from a Canberra ionogram, plotted against the plasma frequency at reflection. This was done using the given ground value of FH (1.52 MHz) extrapolated to 140 km. The main data error is immediately apparent; while the O ray shows considerable cusp retardation at 3.5 MHz, the corresponding X-ray cusp occurs at a considerably higher plasma frequency. Thus the value of FH is incorrect. The "ground" value of 1.52 is in fact the E-region value. Allowance for this error moves the X-ray points down (in plasma frequency) as shown by the dotted line, so that they become compatible with the O data; direct analysis then gives the profile F shown by the fine line.

A second error in the data of Fig. 13 is the scaling of only three O-ray points before cusp retardation becomes important. This violates condition (iv) of 8.5.3.

A normal start calculation (with POLAN Mode 5) uses 5 values of h'_0, and the conditions of Section 8.2.3 include all 4 values of h'_X in Fig. 13. Direct formulation and solution of the virtual-height equations gives the real-height curve A (plotted on a reduced scale). Application of check (i) above sets q(1) = 2.0 km/MHz giving curve B. Check (iii) reduces the starting height of the polynomial to give curve C. Application of these two checks in the reverse order would give D; a worse result than C since imposition of check (iii) removes the initial negative gradient, and q(1) is not set equal to 2.0. In the considerable change from A to C or D the virtual-height fitting error DEVN increases by only 20% (from 2.9 to 3.5 km) showing that the original highly distorted profile was not well-defined.

Curve C is further adjusted downwards by condition (iv) above to give the result E. This increases the virtual-height fitting error by a further 30%, and gives (with no operator intervention) a reasonable result from very bad data. The final downwards adjustment of the polynomial starting point (HA,FA) has produced a physically impossible height increase at lower frequencies. This is well below the observed part of the profile, and has been required to cope with the (physically impossible) data. It is therefore not deleted from the mathematical solution, which is used to calculate following heights. The final output arrays are, however, adjusted as shown by the dotted line to give a more usable variation at low frequencies (as in 8.4.1).

The current version of POLAN checks the mean gradient dh'/dFN corresponding to the highest two frequencies used in a start calculation. If this gradient exceeds 200 km/MHz, the last point is rejected (Section 8.2.3). With the data of Fig. 13 this automatic check limits the start calculation to four O-rays and three X-rays. Analysis with the incorrect value of FH then gives a result which is appreciably more acceptable in the unseen region. In addition, inclusion of the physical constraints described in Section 8.4.1 greatly reduces unphysical variations of the type shown in Fig. 13.

8.5 The Choice of Scaling Frequencies

Corrections for underlying ionisation are based on the difference $h'_X - h'_O$ for O and X rays reflected at approximately the same plasma frequency FR. A first order correction for delay in the underlying region is often all that can be reliably obtained. In this case only a single value of h'_X is necessary. Provided the profile is not changing rapidly near fmin, and there are no underlying peaks with plasma densities greater than about 0.7fmin, this single measurement is sufficient to determine the overall effect of the low-density ionisation. Scaling of additional points then contributes little to the theoretical accuracy, but will help to minimise the effects of experimental errors.

With difficult ionograms (as in 8.5.2 below) accurate results must use two parameters for the underlying ionisation. This requires a minimum of two scaled values of h'_X, which define the rate at which $h'_X - h'_O$ changes with frequency. To determine this change accurately, measurements must cover a minimum frequency range of about 0.2 Mhz. The difference $h'_X - h'_O$ due to underlying ionisation decreases rapidly with increasing frequency, so that use of a frequency range greater than 0.7 Mhz adds little information about the unseen region, and has the disadvantage of including greater variations in the gradient at reflection.

The test models 3A and 3B used previously are derived from real data and represent the two basic types of underlying profile. They are therefore appropriate for investigating the effect of different scaling rules. Model 3A gives slowly-varying virtual heights near fmin, as considered in 8.5.1. For model 3B the virtual heights increase rapidly at low frequencies, and scaling rules for this case are discussed in Section 8.5.2. Recommended scaling rules derived from these and other tests, and from the studies of Appendix B, are summarised in Section 10.4.

8.5.1 Ionograms with h' varying slowly near fmin

Using either the polynomial start (Section 8.6.1) or the slab start in POLAN, calculated real heights for model 3A are accurate to within about 0.1 km for all dip angles and for values of fmin from about 0.8 to 2.5 MHz. This is shown, for representative low and high dip angles, in Fig. 11. The virtual-height data are fitted with an RMS error of less than 5 metres over the above range of fmin and for most reasonable scaling intervals. Results are then limited by experimental accuracy, and scaling requirements are: (i) use data at the lowest available frequencies, where $h'_X - h'_O$ is greatest, and (ii) use a frequency interval $\triangle f$ sufficiently large that successive scalings represent essentially independent measurements.

O-ray data are normally scaled at a frequency interval $\triangle f$ of about 0.1 MHz. When the trace is irregular or ill-defined, additional points should be scaled in the range from fmin to fmin + 0.4 MHz. Such points are automatically included in a start calculation by POLAN, without any increase in the degree of the fitted polynomial, to give increased smoothing of random errors (Section 8.2.3). Scaling intervals less than 0.1 MHz are also used (throughout the ionogram) whenever maximum accuracy is required regardless of scaling and computer time. Such data would usually be analysed using MODE > 5 in POLAN to give additional point-to-point smoothing of the data. On some occasions small initial values of $\triangle f$ are required to obtain at least 5 virtual heights before the retardation due to an observed cusp or peak becomes appreciable; this is discussed in Section 8.2.3.

With slowly-varying virtual heights, scaling 3 to 5 X-ray points is adequate. The plasma frequencies at reflection (FR_X) for these points should conform to one or both of the relations

$$FR_X < \text{fmin} + 0.45 \quad \text{MHz} \tag{27a}$$

or

$$FR_X < \text{FM} + 0.05 \quad \text{MHz} \tag{27b}$$

where fmin and FM are the frequencies of the first and fifth scaled O rays. Thus X-ray data should normally begin at a wave frequency f_X within 1 MHz of fmin. As with the O ray, only points well below observed cusps or peaks should be included.

Occasionally X-ray traces will be observed only at $FR_X >$ fmin + 0.45 MHz. About three points are then scaled, at intervals of 0.05 to 0.1 MHz from the lowest usable frequency. A careful operator will also increase the frequency interval for the scaled O-ray points, so that FM satisfies (27b). If this is not done, however, POLAN will automatically include additional O-ray points in the start calculation until the highest O-ray frequency (FM) is greater than the first value of FR_X. All X rays with $FR_X >$ FM + 0.05 are then accepted. This automatic adjustment brings at least one, and generally 2 or 3, X-ray points into the range of the start calculation.

8.5.2. Ionograms with h' increasing near fmin

The occurrence of a large, rapid increase in group retardation for both the O and X rays near fmin indicates the presence of an underlying peak, with a critical frequency FC_S exceeding about 0.8fmin. In the equivalent monotonic profile this corresponds to an abrupt change from a small to an infinite value of dh/dFN. The assumption of a smoothly-changing gradient below fmin is therefore violated, and additional parameters must be introduced if this part of the trace is to be analysed.

If the underlying critical frequency FC_S is greater than about 0.9fmin, observed O and X traces near fmin are highly retarded. Analysis of these sections could yield information on the density and thickness of the peak. However the situation is essentially the same as that described by Lobb and Titheridge (1977a) in connection with "restart" calculations above an E/F layer valley. The use of data at frequencies within 0.1 MHz of foE gives increased errors in the calculated F-layer

heights, unless the value of foE is known to within 1%. This precision is barely obtainable when E-layer virtual heights are observed; it is quite unobtainable when foE < fmin. Traces which are highly retarded by the presence of an underlying peak should therefore not be included; whatever the method of analysis, uncertainties in the critical frequency of this unseen peak will decrease the overall reliability. X-ray data at low frequencies can also be unreliable because of errors in the assumed value of the gyrofrequency.

A large amount of ionisation with FN just less than fmin causes h'_X, in particular, to become very large at the lowest observed frequencies. Several different approaches could be adopted to the scaling of such records.

(a) Use normal procedures, scaling at a frequency interval of about 0.1 MHz from fmin.

(b) Include additional values of h'_X to define the variation near fmin more accurately.

(c) Use smaller frequency intervals Δf for both O and X rays, to define the initial retarded traces accurately.

(d) Use increased values of Δf so that results depend more on the overall shape of the virtual height curves, and less on the detailed variation of retardation at low frequencies.

(e) Omit the initial, highly-retarded values of h'_X, which depend primarily on the parameters of the underlying peak. Retain all O-ray data, since this is less affected by underlying retardation and relates more directly to the real-height gradients near reflection.

(f) Retain low frequency X-ray data, which is more directly related to the unseen ionisation, but omit the O-ray data which contains a mixture of underlying and reflection-gradient effects.

(g) Omit initial sections of both O and X traces.

The test model 3B of Fig. 11, analysed from fmin = 1.0, provides a realistic example of the most difficult type of profile. The relative contributions to h'_O and h'_X of different parts of an underlying profile change appreciably from low to high dip angles (Appendix B.1). Calculations using model 3B at dip angles of 20° and 70° therefore provide a reasonable basis for comparing the overall merits of different scaling procedures.

Figure 14 shows the virtual-height curves for model 3B, at the two dip angles, on a logarithmic frequency scale from 0.9 to 2.7 MHz. Points corresponding to virtual-height gradients of 100, 200 and 500 km/MHz are labelled. Broken lines show the values of h'_X plotted against the plasma frequency at reflection (FR). 42 different sets of accurate virtual-height data were calculated from model 3B, corresponding to different scaling rules. Each set was analysed by the normal Mode 5 slab start in POLAN, using five O-ray points. Thus the check in POLAN which includes additional points if the fitted range is less than 0.4 MHz (Section 8.2.2) was disabled to show directly the effects of changing f. The physical conditions of Section 8.4.1 were also omitted. Conclusions from different groupings of the calculations are summarised in (i) to (iv) below.

(i) For the first group of results FRmin was constant at 1.0 MHz, and a different (fixed) value of Δf was used for each calculation. The real-height errors Dh vary smoothly above 1.0 MHz. Results are summarised in Table 7 in the form Dh_2 (DEVN), where Dh_2 is the real-height error at 2.0 MHz and DEVN is the RMS error with which the virtual-height data is fitted.

The real-height errors increase considerably at small and large values of Δf, and are a minimum near Δf = 0.1 MHz. The sign of the errors changes from 20° to 70° because of the dip angle variations discussed in Appendix B.1. DEVN also has a minimum, giving the most stable and accurately-defined solution, for intervals near 0.1 MHz. At the smallest frequency intervals DEVN tends to zero since the length of the fitted profile segment tends to zero, and accuracy degenerates to that of a one-point fit; thus Dh increases, even although DEVN is small, when Δf gets too small. A frequency interval Δf of about 0.1 MHz therefore appears optimum, when a fixed number of points (5 for each ray) is scaled from the lowest observed frequencies. Even with precise data, errors increase if we use small scaling intervals to define accurately a small section of the highly retarded trace near fmin. With practical data the errors will increase even more. Use of a large fitting range gives too large a mixture of underlying retardation and reflection gradient variations, and accuracy is again reduced. These considerations rule out approaches (c) and (d) above.

TABLE 7. Real-height errors Dh_2, and the RMS virtual-height fitting errors (DEVN) using a fixed minimum plasma frequency of 1.0 MHz for both O and X rays.

Scaling Interval Δf =	.05	.07	.10	.14	.20	MHz
DIP 20: Dh_2 (DEVN) =	+.90 (.03)	-.29 (.04)	+.23 (.01)	-.29 (.02)	-.46 (.06)	km
DIP 70: Dh_2 (DEVN) =	-4.4 (.04)	1.23 (.23)	-.13 (.07)	2.54 (.22)	5.49 (.42)	km

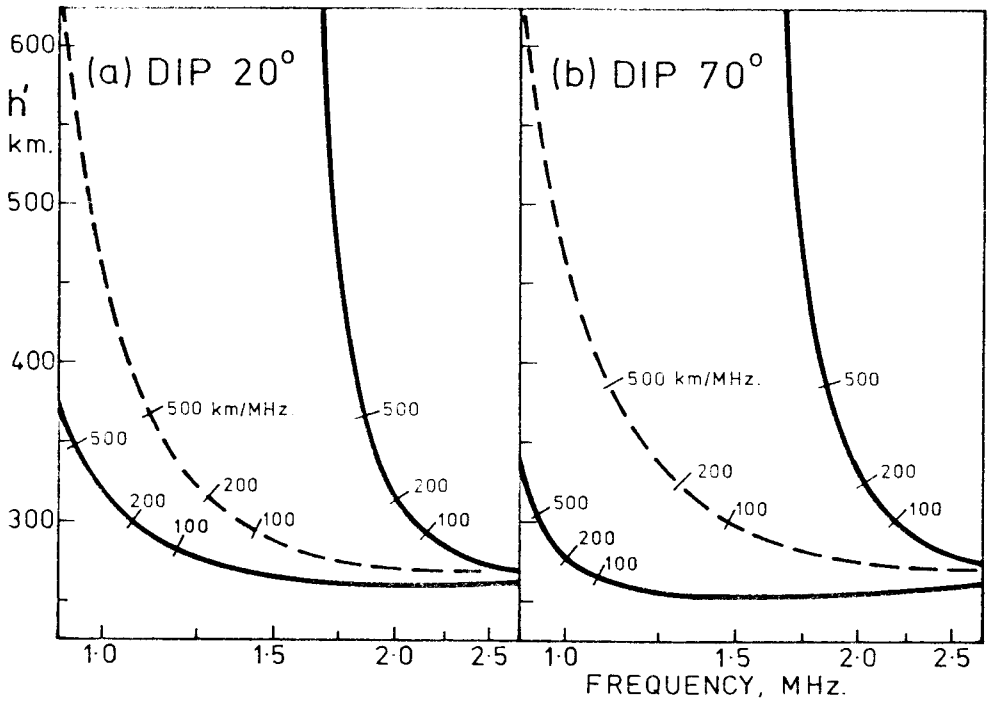

Figure 14. Virtual heights for the ordinary and extraordinary rays, calculated using the model profile 3B, at dip angles of 20° and 70°.

(ii) The second group of calculations used a few closely spaced points near FRmin, with Δf = 0.1 MHz at higher frequencies. This scaling attempts to maintain approximately constant virtual-height changes between data points, giving both adequate detail in the retarded sections and a reasonable overall frequency range. Results show an overall increase in both DEVN and Dh as additional points are added, at FRmin = 1.0 and at FRmin = 1.2. Thus approach (b) above does not help.

(iii) Omitting the largest virtual height (the X-ray point at FR = 1 MHz) gives good results at dip 20°, and bad at dip 70°. Increasing FRmin to 1.1 MHz for both O and X rays gives results which are less dependent on Δf. Omission of the next X-ray point, at FR = 1.1 MHz, makes things worse again. Thus approach (e) above, where only the largest (X-ray) virtual heights are omitted, is incorrect.

(iv) Increasing the minimum scaled frequency for both O and X rays, and retaining a constant frequency interval Δf of 0.1 MHz, gives the results summarised in Table 8. With FRmin = 0.9 we see clearly the bad effect of including a greatly retarded trace (which would not normally be visible). Increasing FRmin above 1.0 MHz gives a general increase in accuracy at dip 20°, and little overall change at 70°. The virtual-height fitting error DEVN does, however, decrease considerably at both dip angles. This indicates a well-defined, stable solution which is not sensitive to changes in Δf.

It should be noted that, with accurate data and an accurate analysis procedure, there is NO loss of real-height accuracy (in the observed region) for values of FRmin up to 1.5 MHz. At higher frequencies, however, the difference between the group retardations of the O and X components decreases and results are limited by experimental errors in $h'_X - h'_O$.

TABLE 8. Real-height errors Dh at 2.0 MHz, and the virtual-height fitting errors (DEVN, in brackets) for calculations using different minimum plasma frequencies FRmin, with Δf = 0.1 MHz.

FRmin =	0.9	1.0	1.1	1.2	1.5	2.0 MHz
DIP 20:	3.11 (20)	.23 (.01)	-.08 (.01)	-.22 (.00)	.00 (.00)	.07 (.00) km
DIP 70:	-3.49 (.59)	-.13 (.07)	1.18 (.02)	.81 (.02)	.54 (.01)	-.80 (.01) km

We conclude that low-frequency points should be omitted when (and only when) they show retardation effects increasing rapidly towards low frequencies. Scaling rules are of importance primarily with difficult data - of which the model 3B analysed above is a good example. For this model, virtual-height data are best omitted at frequencies below about 1.3 MHz. The curves of Fig. 14 show a gradient dh'/dFN of -200 km/MHz for the X ray at this frequency, at both dip angles. This gradient is therefore adopted as defining the lowest frequency at which X-ray data should be scaled. Gradients are generally less for the O ray. O-ray data should, nevertheless, not be scaled to appreciably lower plasma frequencies than the X-ray data. This has been shown in the discussion under (iii) above, and is also seen in other results. Using FRmin = 1.2 MHz for the X ray, addition of an O-ray data point at 1.1 MHz increases the errors by a factor of 1.7 at dip 20° and a factor of 6.9 at dip 70°.

8.6 Other Start Procedures

8.6.1 The polynomial start in POLAN

A simple procedure for including ionisation with FN < fmin is to start the first real-height polynomial from some frequency f_S (typically about 0.5 MHz) which is well below fmin. No other changes in the normal analysis of Section 5 are required, apart from the addition of a constant term to equation (9) to allow calculation of the starting height h_S at the frequency f_S. The result is a single analytic expression describing the ionisation from FN = f_S up to the highest frequency FM used in the start calculation. This procedure is requested by using a negative value for the parameter START in POLAN. The absolute value of START specifies the mean field height, as described in Appendix C.5.

The broken lines in Fig. 15 show the real heights obtained by the polynomial start procedure with f_S = 0.5 MHz and fmin equal to the common value of 1.5 MHz. In the observed region (FN > fmin) the accuracy is good. In the unobserved range the calculated polynomial cannot be expected to follow the true profile closely. It does, however, give approximately correct values for the total amount of underlying ionisation and for the gradient at fmin. As has been shown (Section 8.1) these two quantities contain all that can normally be determined about the underlying ionisation.

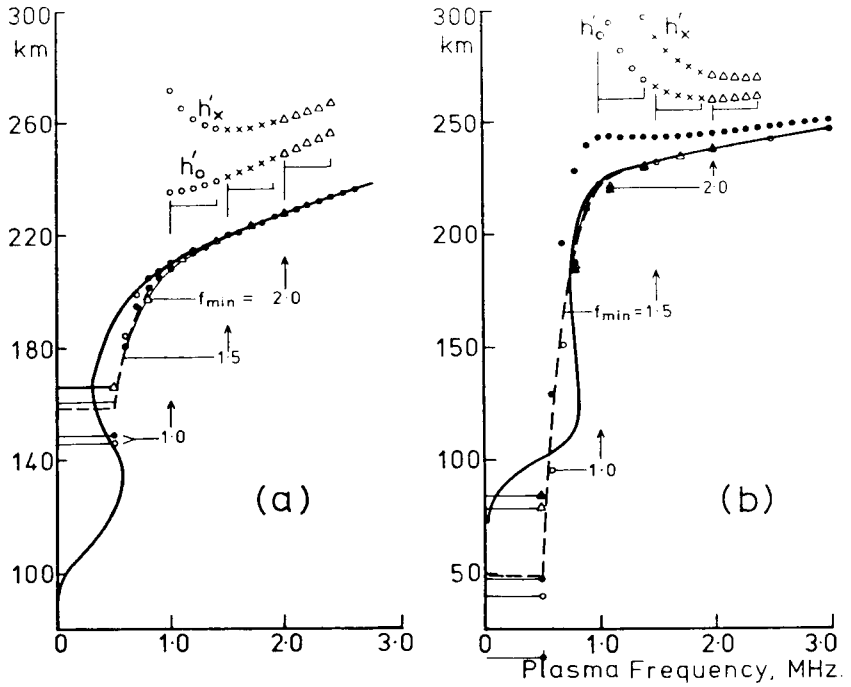

Figure 15. Polynomial start results for two model profiles. The virtual heights h'_O and h'_X are plotted as a function of the plasma frequency at reflection, for a dip angle I of 20°. Real heights calculated from fmin = 1.0, 1.5 and 2.0 MHz are shown as circles, broken lines and triangles respectively. Corresponding results at I = 70° are shown by solid dots, continuous lines and solid triangles.

Circles and triangles are used in Fig. 15 to show the virtual-height data used, and the calculated real heights, for values of fmin equal to 1.0 and 2.0 MHz respectively. Data points are used at a fixed interval of 0.1 MHz in FN. The RMS errors with which the calculated start polynomials fit the virtual-height data, and the real-height errors at 3 fixed frequencies, are given in Table 9 using three values of fmin for each model. At a dip angle of 20°, 5 of the 6 calculated profiles are accurate to within 0.1 km at frequencies above 1.5 MHz, and match the given virtual height data to within a few metres. At 70° dip angle the real-height errors increase to about 0.5 km - again with the exception of model 3B analysed from 1.0 MHz.

Table 9 shows that in all cases the accuracy at a given frequency increases as fmin is increased, from 1.0 to 1.5 to 2.0 MHz. This occurs because the assumption of a simple distribution for the underlying ionisation, with no large fluctuations in gradient at $0.7\text{fmin} < \text{FN} < \text{fmin}$, becomes progressively more accurate. The one unacceptable result is the high dip angle analysis of model 3B from fmin = 1.0 MHz. In this case the underlying ionisation has a broad peak with FN exceeding 0.8fmin. This violates the basic assumption that the underlying ionisation can be represented by a smooth variation of dh/dFN at $\text{FN} > 0.7\text{fmin}$. The single-polynomial representation is unable to produce the sharp corner required, just below fmin, to match the test profile. The occurrence of this condition is clearly signaled by the rapid increase of the X-ray virtual height near fmin, showing that a higher value of fmin should be used (Section 8.5.2).

The polynomial start provides a single analytic expression extending from the beginning of the real-height profile, at (h_s, f_s), to above fmin. It gives reliable results with good accuracy under most conditions, and reasonable results under all conditions (when physical constraints are applied, as in Section 8.4, to avoid implausible results due to bad data). However, as shown in Section 8.1, not more than two (or occasionally three) independent parameters relating to the underlying ionisation can normally be found. So for reliable and consistent results it is preferable to use only two or three variables below fmin. A polynomial start used with the normal mode 5 analysis in POLAN fits an 8-term polynomial to 5 O-ray and 5 X-ray data points. The effective number of free parameters below fmin is difficult to define, and does not appear constant. Calculated profiles suggest up to 4 effective variables in this region on occasion. The slab start model is therefore the preferred procedure in POLAN. However, the continuously varying gradient of the polynomial start is desirable for some work. In this case the parameter START in POLAN is made negative (less than -1.0). For $1.0 < (\text{-START}) \leq 3.0$, the calculated start profile is of the form illustrated in Fig. 15 with $f_s = (\text{-START})\text{-1}$ MHz and an additional real height calculated at a frequency of $0.5(f_s + \text{fmin})$. If the value of START is less than -3.0, POLAN uses a polynomial start from the frequency f_s used for the normal O-ray start.

TABLE 9. Polynomial start results. The RMS deviation in the virtual-height fitting error (DEVN), and the calculated real-height errors at 1.5, 2.0 and 2.5 MHz, resulting from analysis of the two test profiles using $f_s = 0.5$ MHz.

MODEL	fmin	Errors at 20° dip angle				Errors at 70° dip angle				
		DEVN	1.5	2.0	2.5	DEVN	1.5	2.0	2.5	MHz
3A	1.0 MHz	0.01	0.12	0.10	0.08	0.03	-1.01	-0.68	-0.51	km
	1.5	0.00	0.05	0.06	0.03	0.00	-0.46	-0.30	-0.23	
	2.0	0.00	0.01	0.03	0.04	0.00	-0.21	-0.18	-0.13	
3B	1.0 MHz	0.04	-0.75	-0.77	-0.69	0.40	9.58	6.34	4.64	km
	1.5	0.00	-0.02	-0.05	-0.04	0.00	0.79	0.52	0.32	
	2.0	0.00	-0.15	0.03	0.04	0.00	-0.38	-0.16	-0.11	

8.6.2 The single-point starting correction in SPOLAN

The ordinary ray analysis (Section 6) uses an artificial starting point (h_s, f_s), at a frequency f_s which is nominally 0.5 MHz but constrained to lie between 0.35 and 0.6 times fmin. The calculated profile then includes enough of the "low-density" region at $\text{FN} < 0.7\text{fmin}$ to represent adequately any reasonable amount of low-density ionisation. When only O rays are available, the starting height h_s is obtained from synoptic models. A single X-ray measurement may however be used to determine h_s such that an approximately correct allowance is made for both the total amount of underlying ionisation and the ionisation gradient near fmin (Titheridge, 1975b).

The single-point correction is invoked in SPOLAN by providing $(h'_X,-f_X)$ as the first point in the input data. The negative frequency is used to denote an X-ray measurement. SPOLAN then calculates the starting height h_S and sets (h_S,f_S) as the first data point. This is followed by an additional point (h'_0,f_0), where $f_0 = 1/2(f_S + \text{fmin})$ and $h'_0 = h'\text{min}$. The correct value of h_S is obtained from the relation

$$h_S = h'\text{min} - C.(FH/f_S)(h'_X - h'\text{min}) \tag{28}$$

where

$$\begin{aligned} C &= A + B \cos(0.041I - 0.25) \\ A &= -0.265 + 0.19(\text{fmin}/FH) + 1.08(\text{fmin}/FH)^2 \\ B &= 1.40 - 2.80(\text{fmin}/FH) + 1.47(\text{fmin}/FH)^2 \end{aligned} \tag{29}$$

and the dip angle I is in degrees.

When group retardation near fmin is caused primarily by low-density ionisation, h'_X does not vary rapidly with frequency and the precise value of f_X is not important. To allow in addition for gradients near fmin, the value of f_X must correspond to a plasma frequency $FN_X = \text{fmin} + DF$ where the value of DF is obtained from Table 10. For given values of FH and I, this Table is used to determine DF as a function of fmin. The frequencies f_X at which h'_X is scaled are then determined from the normal relation

$$f_X = (FN^2 + FH^2/4)^{0.5} + FH/2 \tag{30a}$$

where

$$FN = \text{fmin} + DF. \tag{30b}$$

For best accuracy the above correspondence between f_X and fmin should be maintained. This is generally straightforward since night-time X-ray traces commonly extend to a lower plasma frequency (although a higher wave frequency) than for the O-ray. When an X trace does not continue down to the desired scaling frequency f_X, we may (i) extrapolate the observed trace down to f_X, or (ii) scale the lowest observed frequency, as f_X, and commence O-ray scalings from the corresponding value of fmin.

Procedure (i) is appropriate when the X trace varies smoothly and the extrapolation required is not unreasonable. When h'_X is beginning to increase rapidly at the lowest frequencies, approach (ii) should be used. In many cases a compromise is adopted: h'_X is extrapolated as far as seems safe, and O rays are scaled from the corresponding value of fmin given (in accordance with (30)) by

$$\text{fmin} = (f_X (f_X - FH))^2 - DF. \tag{31}$$

When h'_X begins to increase rapidly at low frequencies, and disappears (through absorption) before reaching the value of f_X required by (30), some available O-ray data are ignored. However, it is shown in Section 8.5.2 that scaling of virtual heights which are strongly affected by the presence of an underlying peak does not increase the accuracy of calculated real heights in the observed range.

Calculated real heights with and without the single-point correction are shown by the lower and upper dotted lines in Fig. 16. In the unobserved region at $FN <$ fmin the corrected profile may bear little resemblance to the true variation. The total electron content of this region is reproduced reasonably well, however, so that the F layer heights are obtained with good accuracy. Note that while the upper curves in Fig. 16 show the complete virtual-height traces for the X ray (plotted as a function of the plasma frequency at reflection), only the single value marked by a cross is used in the analysis. The calculated profile continues smoothly below fmin, with a monotonically increasing gradient, down to the starting frequency f_S. The result has approximately the correct values for the total electron content and for the gradient near fmin, and agrees exactly with all the given O-ray data and with the single X-ray measurement.

TABLE 10. The frequency shift DF/FH required to determine the extraordinary ray frequency f_X in equation (30).

Dip angle I =	5	20	30	40	50	60	70	80°
fmin = FH	-0.032	-0.026	-0.013	0.020	0.077	0.162	0.256	0.355
fmin = (7/6)FH	-0.078	-0.054	-0.016	0.048	0.134	0.233	0.329	0.401
fmin = (8/6)FH	-0.115	-0.075	-0.016	0.071	0.174	0.281	0.379	0.444
fmin = (9/6)FH	-0.145	-0.088	-0.011	0.092	0.205	0.317	0.413	0.480

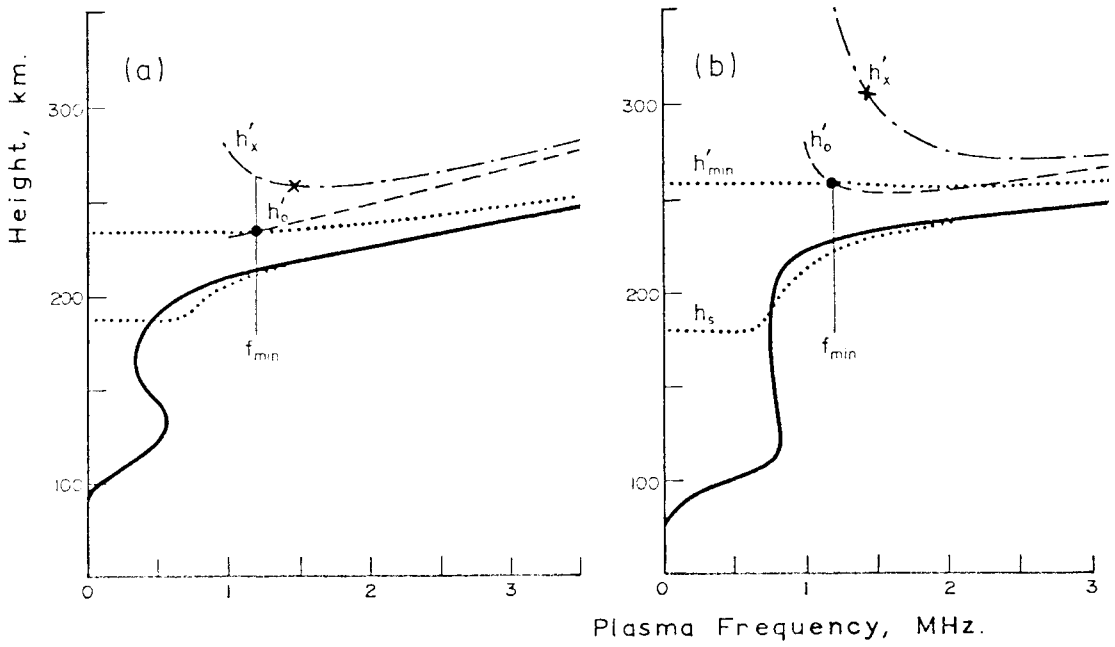

Figure 16. Analysis of the two basic types of start profile, at a dip angle of 70°. Upper dotted lines show the result of a direct O-ray analysis from fmin = FH = 1.2 MHz. The lower dotted lines are obtained using the single-point procedure in SPOLAN, with a starting frequency f_S of 0.6 MHz.