INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION

INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
34100 TRIESTE (ITALY) - P.O.B. 586 - MIRAMARE - STRADA COSTIERA 11 - TELEPHONES: 224281/2/3/4/5/6
CABLE: CENTRATOM - TELEX 460392-I

SMR/92 - 31

AUTUMN COURSE ON GEOMAGNETISM, THE IONOSPHERE
AND MAGNETOSPHERE

(21 September - 12 November 1982)

IONOSPHERIC PREDICTION

K. Davies

Space Environment Laboratory
National Oceanic and Atmospheric Administration
325 Broadway
Boulder, Colorado 80303
U.S.A.

International Centre for Theoretical Physics

Autumn Course on Geomagnetism, the Ionosphere and Magnetosphere

Workshop on Radio Propagation in the Tropics

Lecture III

Ionospheric Prediction

by

Kenneth Davies

3.1 History

The early history of exploration of the ionosphere was governed by the need to understand the long distance propagation of radio waves. While the existence of electric currents in the upper atmosphere had been postulated in the nineteenth century it was the success of Marconi in 1901, in establishing radio connection across the Atlantic, that led to a sustained interest in the ionosphere.

I should like to point out that many discoveries about the ionosphere have been made in the face of erroneous predictions. Thus, prior to 1901, the theorists had predicted that radio waves could not cross the Atlantic. During the early years of radio it was discovered that, for long-distance propagation, long waves gave higher signal strengths than short waves so that the longer wavelengths were coveted by government and commercial stations. Governments felt free, therefore, to leave the short waves to amateur radio operators for short-range communications. In the early 1920's the amateurs found that they could communicate across the North Atlantic on wave-lengths around 100 meters and discovered that the higher the frequency the stronger the signal. In the 1950s it was incorrectly predicted that high frequency communications would shortly be superceded by other, more reliable, frequency systems that would be free of disturbance: such as VHF forward scatter from the D region, reflection of VHF from meteor trails, satellite systems on VHF. The feeling always has been that we know enough about the ionosphere to adequately predict, or extrapolate, into another part of the radio spectrum only to find that we were wrong.

High-frequency radio communications were found to be sensitive to time of day, season, to the 11 year sunspot cycle, see Figure 3.1, and to the occurrence of solar flares (the Mogel-Dellinger effect) and geomagnetic storms. Thus, a need arose to predict optimum radio frequencies that could be used to communicate on HF. For a given circuit such optimum frequencies varied with time and a set of frequencies were needed for adequate coverage. The prediction of these frequencies involved two distinct procedures: (1) a knowledge of the condition of the ionosphere and (2) a method of converting vertical soundings to oblique parameters. Information on the ionosphere was available from vertical soundings at a small number of stations and many assumptions were necessary for global interpolation. A basic prediction was that the maximum or critical frequencies depended on the solar zenith angle. This prediction turned out to be reasonably good for the E and F1 layers but is generally unreliable for the F2 layer which is by far the most important layer from the viewpoint of radio transmission. Turning to the second step, there are two methods of converting from vertical to oblique propagation. The first is to assume a parabolic layer and calculate the height of maximum density and the half-thickness of the parabola from vertical soundings and then calculate the maximum frequency for appropriate ranges (e.g. 3000 km). The second method is the use of transmission curves that are applied directly to ionograms to give factors by which the critical frequency must be multiplied to give the maximum frequency.

World War II provided a great impetus to develop ionospheric predictions both for planning and operational purposes. A much more extensive global network was established and this revealed the important geomagnetic control of the F2 layer dominated by the equatorial anomaly in maximum reflected frequencies -- Figure 3.2.

3.2 Needs for Ionospheric Predictions

Society, in general, is concerned with the ionosphere because of its role in radio communications, navigation (e.g. OMEGA), surveillance, early warning (OTH) and others listed in Figure 3.3. A major societal concern is the effective management of and efficient utilization of the limited radio frequency spectrum (JATC, 1964). The high-frequency band is highly congested because its use is stimulated by the relatively low cost of terminal equipment which is a major consideration for new developing countries. This band is also valuable for mobile communications in sparsely populated areas such as in the Arctic and in Antarctica. Basically, while newer technologies have enabled more intensive use of the spectrum this has been more than offset by the increased demand. Society is particularly concerned with the following three aspects of communications: (1) safety (2) commercial and (3) personal. The value of ionospheric predictions in saving lives is inestimable, the commercial value is a trade-off between costs and benefits whereas the personal value is more a matter of convenience.

In the utilization of the frequency spectrum the following considerations are important:

(1) This resource is used, not consumed, and is wasted when not used.

(2) It has the dimensions of space, time and frequency that can be shared.

(3) It is an international resource.

(4) The spectrum is wasted when its characteristics are not utilized properly.

(5) It is subject to pollution and needs policing.

Probably the greatest social pressure is that resulting from the increasing mobility of the world's population. There are more people on the move, at higher speeds and radio provides the only efficient means of communication. Communications insure more efficient business, public and private safety, and national security. Thus any medium that affects radio propagation is of vital concern to society.

The main categories of users of ionospheric predictions are: (1) planners and designers of radio circuits who require long term predictions for the determination of system coverage and frequency allocation for communications, broadcasting, early warning and surveillance systems, (2) users of navigation and timing systems who depend on the stability of the reflecting layer for reliable time and position determination, (3) researchers who study the ionosphere e. g. with rockets and satellites and (4) those whose activities are adversely affected by the ionosphere such as radio astronomers, satellite geodesists, etc.

3.3 Morphological Predictions

The primary need for long-term (months to years) predictions is in the planning of radio circuits: frequency allocations, antenna design, power requirements etc. For this purpose the primary data concern the critical frequencies of the E, F1 and F2 layers, the layer heights and the absorption of radio waves in the D region as functions of time and global positions. The parameters of the E and F1 layers are, relatively, well behaved and can be represented by simple formulas involving the solar zenith angle, the 12-month running average sunspot number and the layer critical frequencies for an overhead sun and zero sunspot number. Measurements of D region absorption are few so that, although the layer varies considerably with time and location, the absorption is usually determined from oversimplified formulas (Davies, 1965).

From the point of view of radio communications the F2 layer is by far the more important of the ionospheric layers. It is also very variable and the parameters cannot, in general, be described by simple analytical formulas. From critical frequencies measured at about 150 ionospheric stations the world-wide and temporal variations of foF2 have been expressed in terms of mathematical functions (Jones and Gallet, 1962) and stored on computers (Haydon

et. al 1976, CCIR 1978). Such a representation of the F2 critical frequency is often called a numerical map and a sample graphical output is shown in Figure 3.2 which is dominated by the equatorial anomaly. Such maps are available for sunspot numbers zero and 100 and the critical frequencies for other sunspot numbers are found by linear interpolation and/or extrapolation. The diurnal variations are represented by maps for different universal times.

The variation of the critical frequencies of the E and F2 layers over Washington, D. C. is shown in Figure 3.4. To a first approximation the critical frequencies of the F2 layer can be represented by a formula of the form

$$foF2 = a\,(1 + bR) \qquad 3.1$$

Close inspection of Figure 3.4 shows that the gradient 'b' varies from one cycle to another. Both 'a' and 'b' vary with location. Near the geomagnetic equator the gradient 'b' is small.

3.4 Oblique Propagation

Up to now we have discussed vertical propagation but in practical communications it is necessary to determine the relationship between vertical propagation and oblique propagation. We have seen already that ionospheric absorption decreases with increasing frequency so that communications tends to be better on higher frequencies. Hence, a communicator wants to use the highest frequency possible. In this he is restrained by two facts: (a) the frequency allocation (2) the maximum frequency reflected from the ionosphere. The former is partly an administrative matter the second a technical matter.

While the conversion from vertical to oblique propagation is straightforward for a plane ionosphere, with no magnetic field, no electron collisions etc, such is not the case when the magnetic field, collisions and horizontal variations especially are involved. Fortunately, because higher frequencies can be used with oblique propagation than with vertical propagation the effects of the

magnetic field and of collisions are either negligible or may be treated as perturbations. As mentioned before there are two methods of converting from vertical to oblique propagation (A) models (B) transmission curves.

A. Model

Here we shall consider only the parabolic layer which is the most common model. Its use is based on the fact that near the maximum of a Chapman layer the distribution is parabolic viz

$$f_N^2 = f_c^2 \left\{ \frac{2(h-h_o)}{y_m} - \frac{(h-h_o)^2}{y_m^2} \right\} \qquad 3.2$$

where f_N is the plasma frequency (see equation 1.13) at a height h in a parabolic layer with base height h_o and semi-thickness y_m -- see Figure 3.5. The virtual height of reflection of a radio wave of frequency f in the absence of a magnetic field and of collisions is given by

$$h' = h_o + \frac{1}{2} y_m \frac{f}{f_c} \ln \frac{f_c + f}{f_c - f} \qquad 3.3$$

The parameters h_o and y_m can be determined from ionogram data and by plotting h' versus

$$\frac{1}{2} \frac{f}{f_c} \ln \frac{f_c + f}{f_c - f} \quad .$$

From Figure 3.5 we have that the ground distance

$$D = 2 h_o \tan \phi_o + 2 \sin \phi_o \int_{h_o}^{h_r} \frac{dh}{\mu^2 - \sin^2 \phi_o} \qquad 3.4$$

where h_r is the height of reflection and the refractive index, μ, is given by

$$\mu^2 = 1 - a(h-h_o) + b\,(h-h_o)^2 \qquad 3.5a$$

where

$$a = 2 f_c^2 / y_m f^2 \qquad 3.5b$$

and $$b = f_c^2 / y_m^2 f^2 \qquad 3.5c$$

Substitution and integration gives

$$D = 2 h_o \tan \phi_o + Q y_m \qquad 3.6a$$

where $$Q = \sin \phi_o \frac{f}{f_c} \ln \frac{f_c + f \cos \phi_o}{f_c - f \cos \phi_o} \qquad 3.6b$$

Before proceeding further with the parabolic layer theory let us consider what happens to the range D as we increase the elevation angle (decrease ϕ_o) from a small value (Figure 3.6). The range first decreases, reaches a minimum and then increases very sharply. This is shown in the D versus ϕ plot in Figure 3.7 in which we note that, for a given D, there are two rays with different ϕ's and there is a minimum distance, the skip distance, within which normal ray transmission is not possible. In another way, for a given distance, there is a maximum usable frequency or MUF. The two rays, for a given D, are called the low-angle ray and the high-angle or Pedersen ray. From equation 3.6 we see that there is a lower limit ϕ_c of ϕ_o given by

$$\cos \phi_c = f_c/f \qquad 3.7$$

below which penetration occurs. Note that ϕ_c does not correspond to the skip ray. The condition for the skip is $d D/d\phi_o = 0$ which gives

$$\frac{y_m}{h_o} u \ln \frac{1+u}{1-u} + 2 = 2 \tan^2 \phi_s \left(\frac{u^2}{1-u^2} \frac{y_m}{h_o} - 1\right) \qquad 3.8a$$

where $$u = f_m \cos \phi_s / f_c \qquad 3.8b$$

where f_m is the MUF and ϕ_s is the angle of incidence corresponding to the skip.

Equation 3.8 can be solved for ϕ_s for given u, from ϕ_s we can get f_m. The ratio $f_m/f_c = M(D)$ is the MUF factor for a distance D. For given values of y_m/h_o and f_c the M factor can be plotted against D and, hence, against ϕ_o.

This theory has been extended to include the curvatures of the earth and ionosphere, e. g. see Appleton and Beynon (1940, 1947).

B. Transmission Curves

An alternative to the parabolic method is the use of transmission curves developed in the USA. These are based on the following theorems:

The secant law which relates the frequencies f_v and f_o reflected at the same height with vertical propagation and at an incidence ϕ_o viz

$$f_o = f_v \sec \phi_o \qquad 3.9$$

Breit and Tuve's theorem which states that the time t to traverse an actual curved path is the same as that of the equivalent triangle TAR in Figure 3.8

i.e. $$c t = TA + AR \qquad 3.10$$

Martyn's theorem states that the virtual heights of reflection of a vertically reflected and an obliquely reflected wave are equal provided the frequencies are related by equation 3.9

$$h'_{ob} = \frac{1}{2} \cos \phi_o P_{ob}' = (TP) \cos \phi_o = h_v' \qquad 3.11$$

Here P_{ob}' is the group or virtual path.

Now $$\sec \phi_o = \sqrt{(D/2)^2 + (h'_v)^2} / h'_v \qquad 3.12$$

To find the oblique frequency corresponding to a given f_v we can solve equations 3.11 and 3.12 simultaneously by a graphical technique called transmission curves. One type of transmission curve plot is shown in Figure 3.9. Another type is illustrated in Figure 3.10 in which sec ϕ_o has been corrected for the curvatures of the earth and the ionosphere. The transmission curves (Davies, 1965) are plotted on the same virtual height scale as a vertical ionogram. At certain distances the transmission curve will intersect the ionogram trace, for a given layer, at two heights corresponding to the low angle and high angle rays.

The curve corresponding to the skip distance will be tangent to the ionogram. By recording the ionogram and the transmission curve on the same logarithmic frequency scale one can be slid over the other until trangency occurs for a given distance, Figure 3.10. The maximum usable frequency is read off the frequency scale at the 1.0 reading on the $\sec\phi_o$ scale and the M(D) factor found from the $\sec\phi_o$ scale at f_c.

3.5 Vertical to Oblique Conversion

From the theorems of Section 3.3 it is possible to sketch the appearance of an oblique ionogram where the transmitter and receiver are separated. For a given f_{ob} there are two values of h' and, hence, two values of P' the group path. As the frequency is increased the two values of P' approach and meet at the MUF or junction frequency, JF, as shown in Figure 3.11. Experimental oblique ionograms are shown in Figures 3.12 and 3.13. Figure 3.13 shows the extension of the trace to very high frequencies by scatter propagation.

3.6 Circuit Predictions

3.6.1 MUF predictions

From the numerical maps of MUF(zero)F2 and MUF(4000)F2 it is possible to interpolate numerically to find MUF(D)F2. In earlier days this was done manually, using graphical overlays and nomograms, for each hour of the day, for each month of the year and for a whole range of sunspot numbers. In recent years the whole procedure is done by computer (see Haydon et. al. 1976). The basic data required, such as: geographical and time variations of the ionosphere, radio noise, the relationships between the ionospheric characteristics and the path geometry, signal attenuation, and the theoretical performance of common antenna systems, are stored as part of the general data base.

For simple predictions, such as the highest useful frequencies, only the geographical locations of the circuit terminals and the times of interest are required. For more complex predictions, e.g. reliability, additional information is needed such as: the frequency of operation, the type of antenna, the transmitter power, the type of service required, and man-made noise environment at the receiver etc.

Because the MUF varies from hour to hour on a given day, from day to day at a given hour, with season and sunspot number and because the D region absorption also varies it is necessary to have several (at least two) frequencies for operation of a given circuit.

For prediction purposes monthly median MUFs are obtained but this implies that on half the days the MUF is below this median so that communications on the MUF is possible 50% of the days. To enable communication on 90% of the days a rule of thumb is to use the optimum working frequency FOT which is

$$FOT = 0.85 \ MUF \qquad 3.13$$

A sample FOT plot is shown in Figure 3.14 which indicates that frequencies near 6 MHz and near 15 MHz are required for 24 hour operation of this circuit.

3.6.2 Transmitter power

Having selected a set of frequencies the design engineer must next calculate the transmitter power, P, required for his particular type of service. The distance attenuation is calculated on the basis of $P_r \propto d^2$. The ionospheric absorption can be obtained from the computer program which is based on a formula of this type :

$$L = 615.5 \ n \sec\phi \ (1 + 0.0037R)(\cos 0.881\chi)^{1.3}/(f+f_H)^{1.98} \ dB \qquad 3.14$$

The 0.881χ is used to allow for the fact that the absorption does not go to zero when the zenith angle $\chi = 90°$ -- see Davies (1965), chapter 9. In equation 3.14 ϕ is the angle of incidence on the D region and n is the number of hops. The antenna gain, relative to an isotropic radiator, must be obtained from the

known antenna configurations or the reference antennas used in the programs. The gyrofrequency f_H is obtained from a numerical model of the geomagnetic field. The ambient radio noise is obtained from worldwide numerical maps of radio noise on 1 MHz such as that of Figure 3.15. From the frequency dependence of the noise, such as is shown in Figure 3.16, the noise on a frequency f is determined. To take into account the uncertainty in the noise level due to the spread of the data a correction σ is included. In a computer program these curves are expressed as mathematical functions.

Calculation procedures can be divided into two distance regimes: (a) less than 4000 km (b) greater than 4000 km. The 4000 km distance is about the maximum one-hop range for the F2 layer at 320 km altitude.

(a) Distances less than 4000 km

From the geographical coordinates of the terminals the circuit length and midpoint coordinates are calculated. The appropriate UT at the midpoint is determined. The critical frequencies, the M factor, the solar angle, layer heights and the gyrofrequency are determined from the numerical models stored in the computer. Hence we determine the MUF and, hence, the FOT.

Next we determine the effects of the E layer of which there are three possibilities:

(1) The normal E layer determines the MUF

(2) The sporadic E layer determines the MUF

(3) E layer cut-off is present

The procedure is: determine the radiation angle for one-hop E propagation for the appropriate distance, if > 0 proceed to calculate the E layer MUF for various hops to see if MUF (E) > < MUF (F2). Do the same for E_s. To determine the E, or E_s, cut-off find the one-hop E distance d_E corresponding to the F2 radiation, or elevation, angle and the coordinates of the ray at the E layer.

If MUF(F2) < MUF(E) for the distance d_E, cut off will occur and E-layer propagation dominates, otherwise F2-layer propagation only need be considered. Note that E cut-off essentially determines the low frequency limit because if the propagation involves several E-hops the signal suffers heavy D region absorption (see Davies 1965, Chapter 7).

The power loss is taken as the sum of the losses due to: the ionosphere, distance, ground reflection (for multiple hops) antenna gain or loss, all of which are available in the program or must be supplied (e.g. if a non-standard antenna is used). The required carrier/noise ratio must be specified and the noise power calculated for the appropriate bandwidth (Davies 1965, Chapter 7) in order to determine the required power at the receiver and, hence, of the transmitter. For more details the interested reader should consult the reports by Barghausen et al (1969) and Haydon et al (1976).

(b) Distances greater than 4000 km

As the distance is increased beyond the one-hop F2 limit ($\tilde{\sim}$ 4000 km) it has been found empirically that the propagation does not change discontinuously from one hop to two hops. Some reasons for this include: ionospheric scatter, ground scatter, high-angle ray and ionospheric tilts. Experience has shown that propagation fails only when the ionosphere cannot support the propagation at two "control points" on the great circle path at 2000 km from each end. In the practical determination of the F2 MUF the control points 2000 km from each end are found. The MUF for the path is the lower of the two MUFs, for a distance of 4000 km, with each control point as the center of a 4000 km path. When it is desired to apply the procedure to the E, E_s, and F1 layers control points 1000 km each end may be used (see Davies 1965, Chapter 7).

3.7 Choice of Operating Frequency

While the FOT, selected by calculations such as those described, may be optimum from the point of maximum signal strength and number of days/month reception, it does not necessarily yield the best communications e.g. minimum errors. A major reason for this is the congested nature of the spectrum. Because the world's population and wealth is concentrated in a few regions (e.g. Western Europe, Eastern North America, India, East Asia) the FOT for most users tend to be about the same value. Hence, interference from other users in these regions is intense. In fact a frequency well below the FOT may be best for an individual circuit because, although the signal is weaker, the man-made interference is much less.

References

Appleton, E. V. and W. J. G. Beynon, 1940, The application of ionospheric data to radio communications problems, Pt. I, Proc. Phys. Soc. 52, 518; 1947, Pt. II, Proc. Phys. Soc., 59, 58.

Barghausen, A. F., J. W. Finney, L. Lee Proctor and L. D. Schultz, 1969, Predicting long term operational parameters of high-frequency sky-wave telecommunications systems, ESSA Tech Rep. ERL 110-ITS-78, U. S. Department of Commerce.

CCIR, 1978, CCIR atlas of ionospheric characteristics, Report 340-3, Int. Telecommunications Union, Geneva.

Davies, K., 1965, Ionospheric Radio Propagation, NBS Monograph 80, U. S. Gov't Printing Office, Washington, D. C.

Haydon, G. W., M. Leftin and R. Rosich, 1976, Predicting the performance of high frequency sky-wave telecommunications systems (the use of the HF MUFES 4 program), OT Report 76-102, Office of Telecommunications, U. S. Department of Commerce.

JATC, 1964, Radio Spectrum Utilization, Joint Technical Advisory Committee, Inst. of Electr. and Electron Eng., New York.

Jones, W. B. and R. M. Gallet, 1962, Representation of diurnal and geophysics variations of ionospheric data by numerical methods, Rad. Sci. 660 (4), 419-438.

Figure Captions

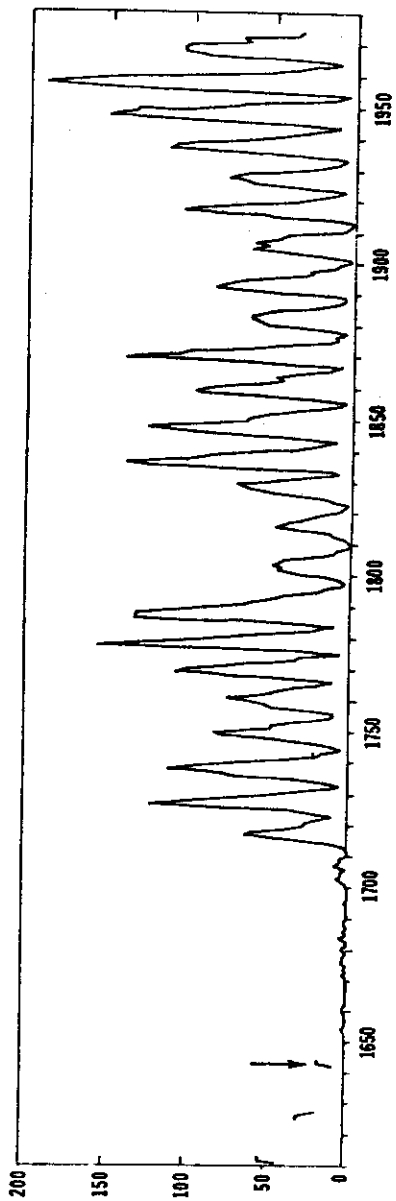

Fig. 3.1 Sunspot numbers 1610-1970

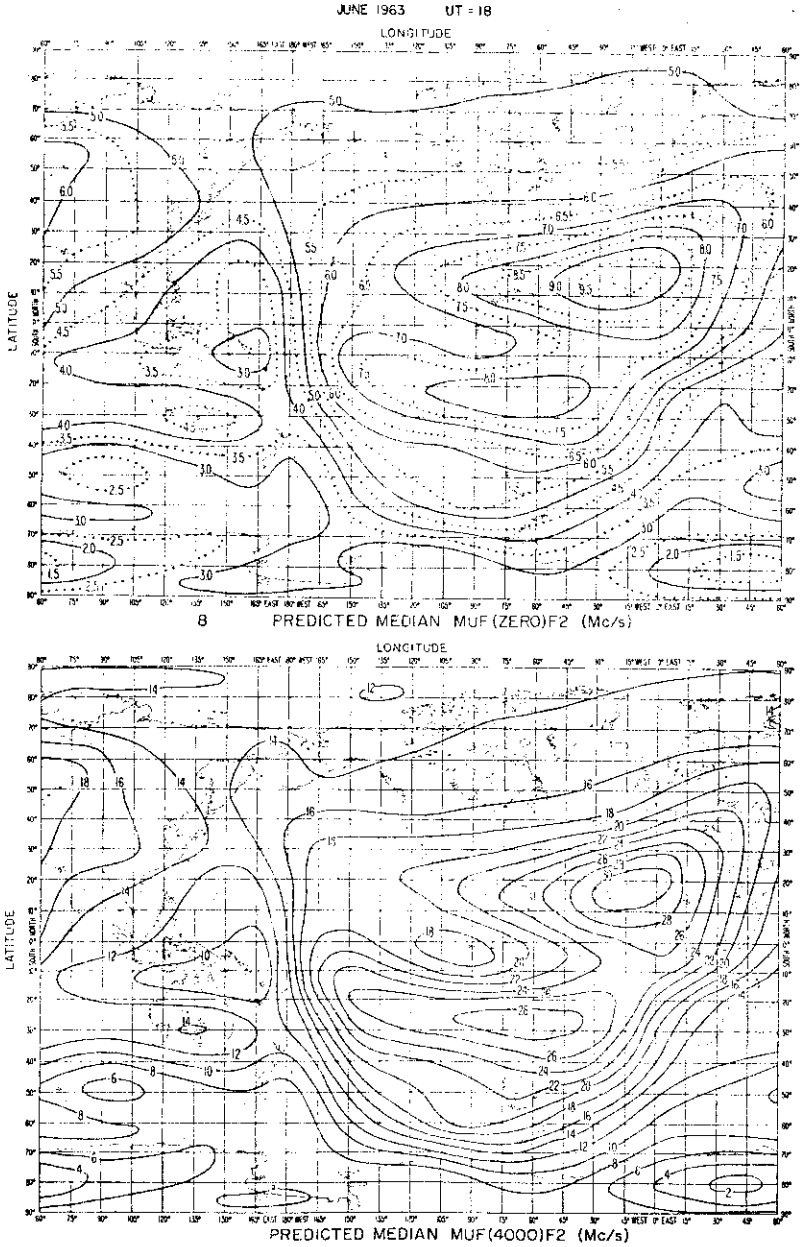

Fig. 3.2 Contour maps of MUF(zero)F2, fxF2, and MUF(4000)F2

Typical Users of the Ionosphere

Primary customer application areas and times when their systems are most affected by environmental anomalies, primarily local.

customer	day or night effects*	type of activity producing effect
Civilian satellite communication	night	Magnetic storms
Commercial aviation—mid-latitude communication (VHF)	day	Solar radio emissions
Commercial aviation—polar cap communication (HF)	day & night	PCA, magnetic storms
Commercial aviation navigation (VLF)	day & night	PCA, magnetic storms
Electric power companies	day & night	Magnetic storms
Long line telephone communication	day & night	Magnetic storms
High altitude polar flights, radiation hazards	day & night	Solar proton events
Civilian HF communication Coast Guard, GSA, commerical companies, VOA	day & night	X-ray emission, U.V. emission, magnetic storms
Geophysical exploration	day	Magnetic storms
Satellite orbital variation military and civilian	day & night	U.V. emission, magnetic storms
DOD SATCOM communication	night	Magnetic storms
DOD HF communication	day & night	X-ray emission, U.V. emission, PCA, magnetic storms
DOD reconnaissance	day & night	PCA, magnetic storms
DOD navigation	day & night	X-ray emission, U.V. emission
ERDA communication prospective customers	day & night	X-ray emission, U.V. emission, magnetic storms
International community	day & night	All
Scientific satellite studies IMS, Solar Maximum mission, Shuttle, solar constant measurements, stratospheric ozone variation, interplanetary missions	day & night	Optical solar flares, magnetic storms, X-ray emission, U.V. emission, solar proton events, solar features
Scientific rocket studies IMS, magnetosphere, ionosphere, upper atmosphere, sun	day & night	Optical solar flares, solar features, magnetic storms, solar proton emission, X-ray emission
Scientific ground studies IMS, sun, interplanetary, magnetosphere, ionosphere, upper atmosphere, stratosphere, troposphere, seismological/geomagnetic	day & night	Optical solar flare, magnetic storms, solar proton emission, X-ray emission, U.V. emission, solar features

*at or near solar maximum.

Fig. 3.3 Users of ionospheric forecasts

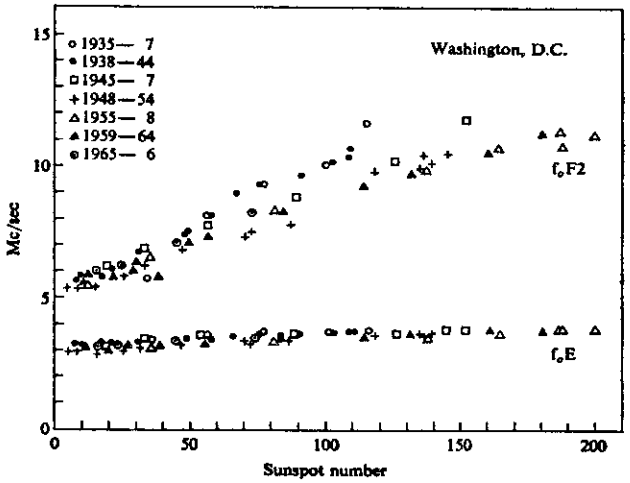

Fig. 3.4 Variation of E and F2 critical frequencies at Washington, D. C. over 3 sunspot cycles

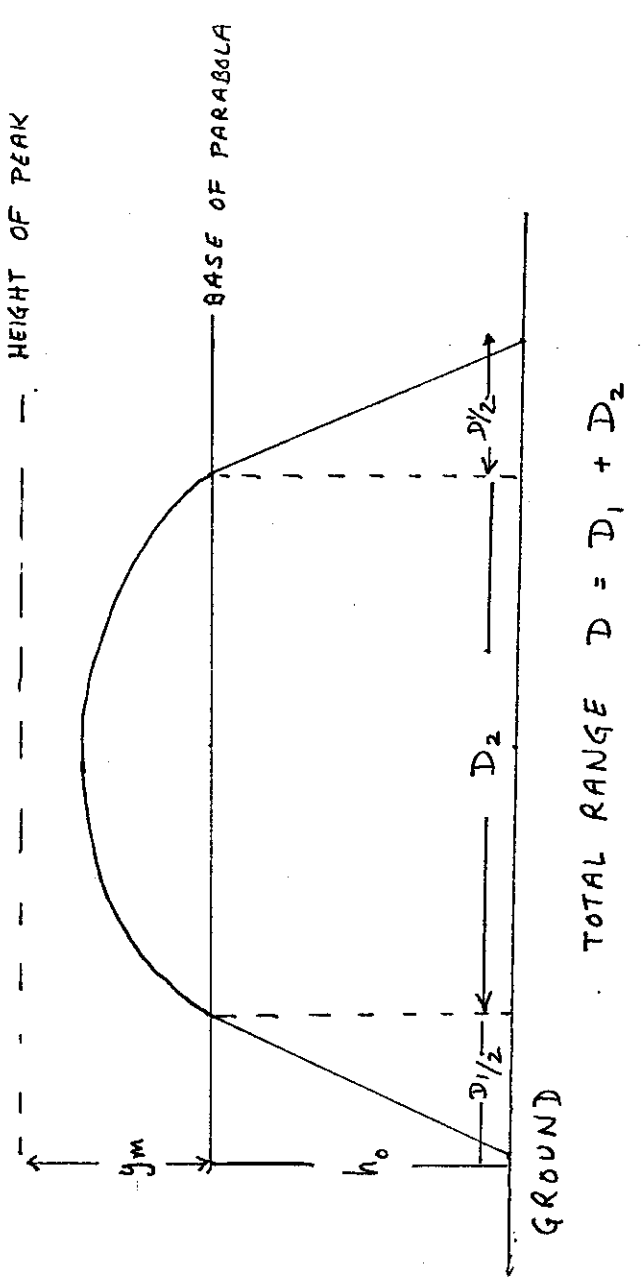

Fig. 3.5 Ray path in a plane parabolic layer

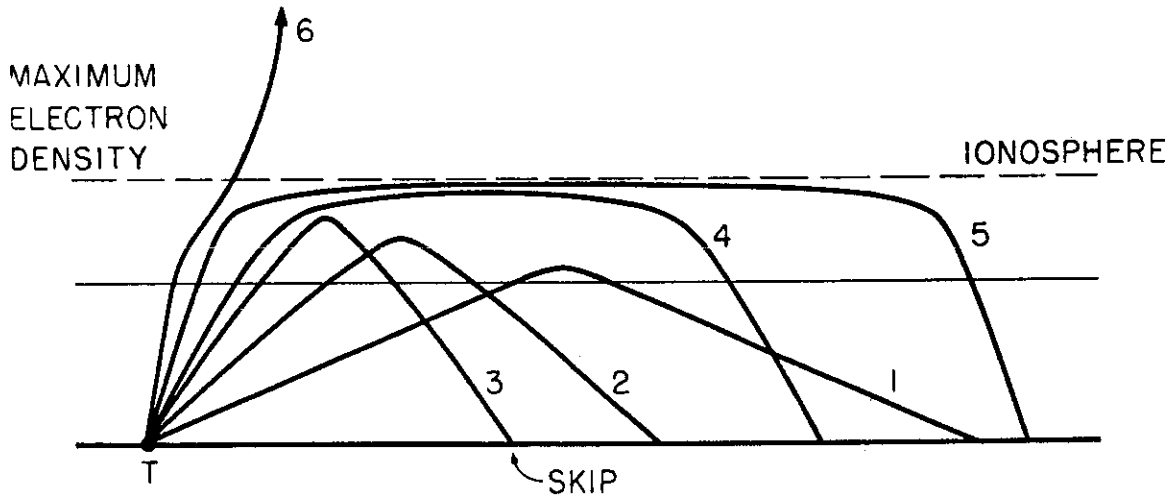

Fig. 3.6 Ray paths on fixed freqeuncy, with varying elevation

VARIATION OF RANGE (D) WITH ANGLE OF INCIDENCE (ϕ_o) FOR A PARABOLIC LAYER (PLANE EARTH AND IONOSPHERE)

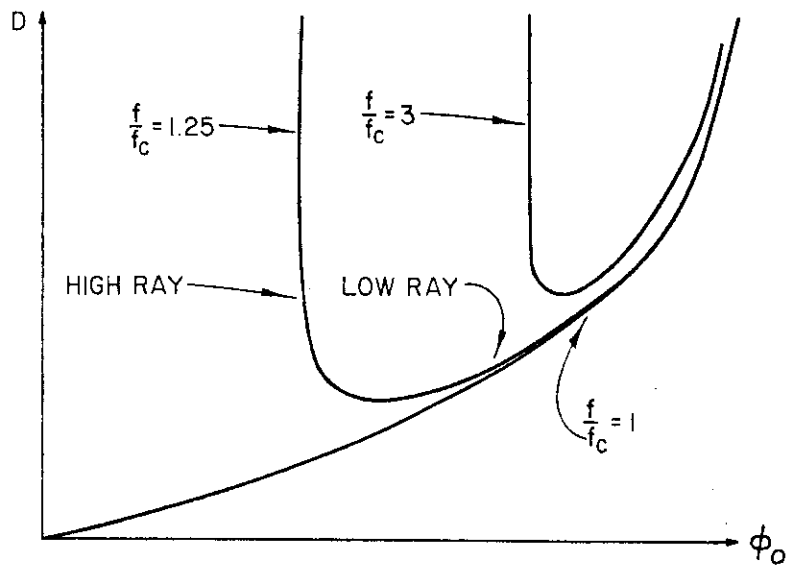

Fig. 3.7 Variation of range D with angle of incidence ϕ_o for a plane parabolic layer

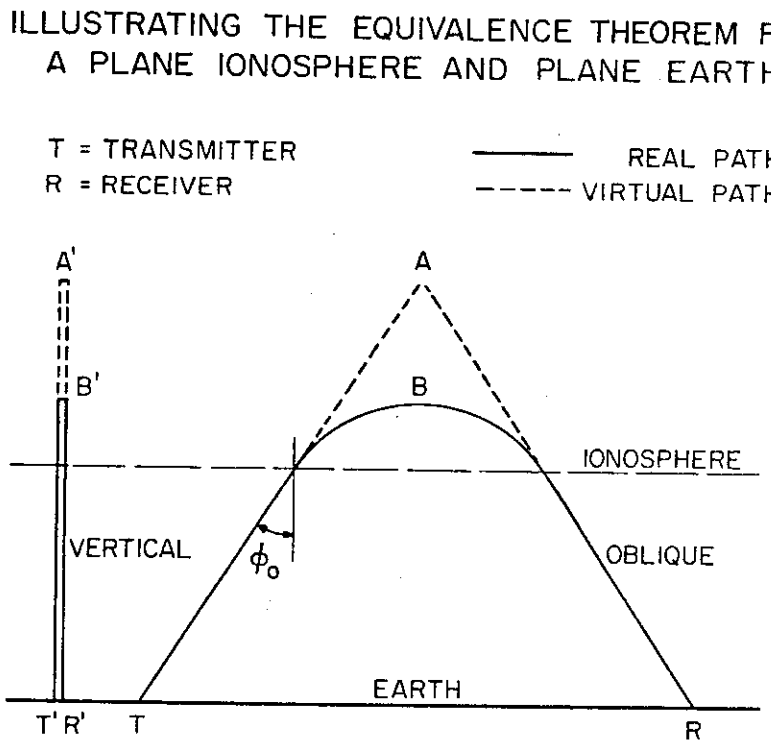

Fig. 3.8 Illustrating the equivalence theorem for a plane ionosphere and plane earth

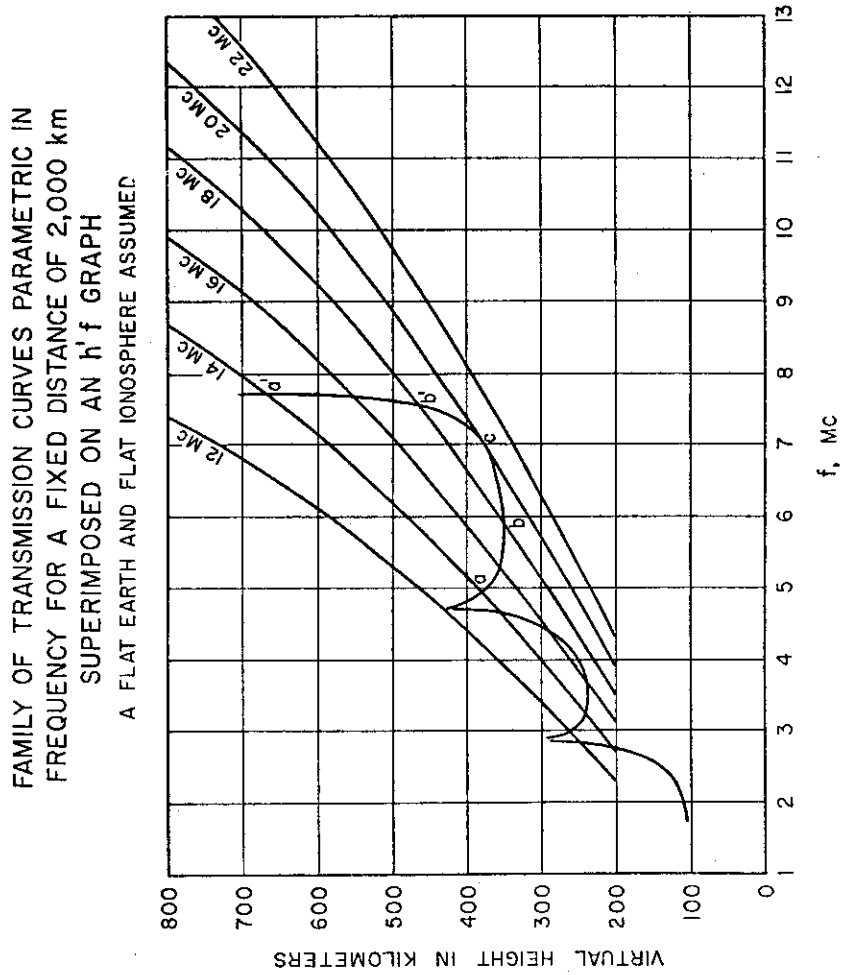

Fig. 3.9 Transmission curves superimposed on ionogram

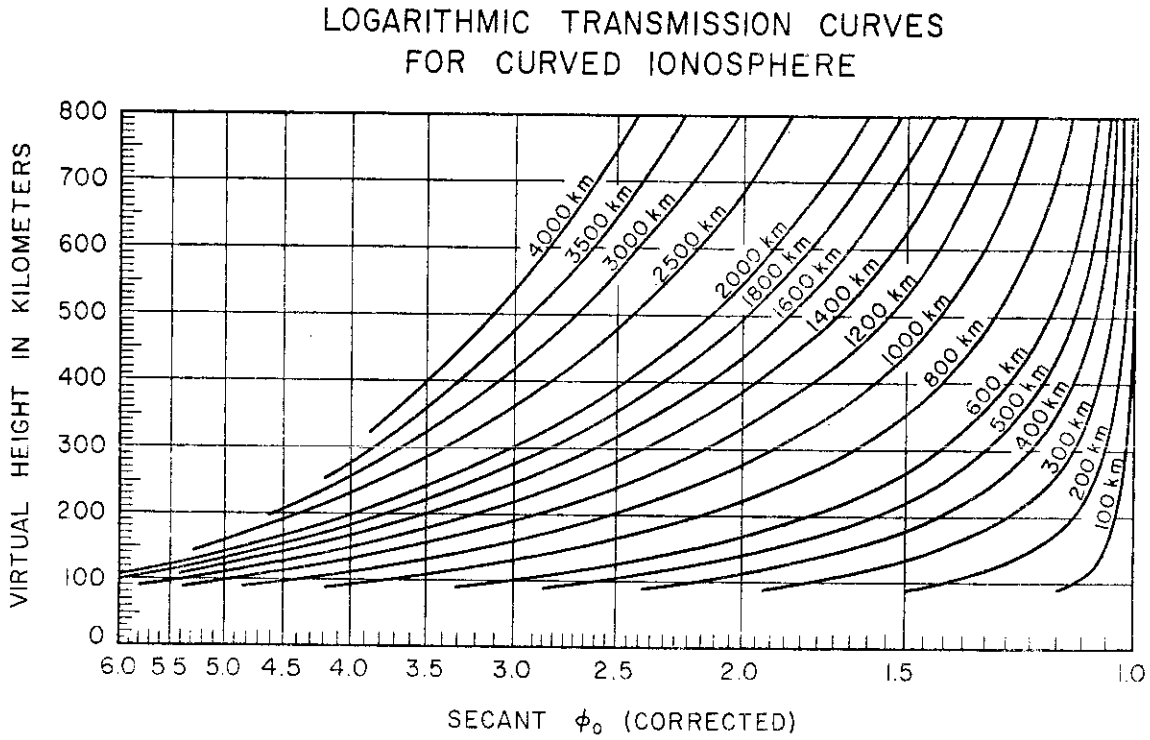

Fig. 3.10 Logarithmic transmission curves for curved ionosphere

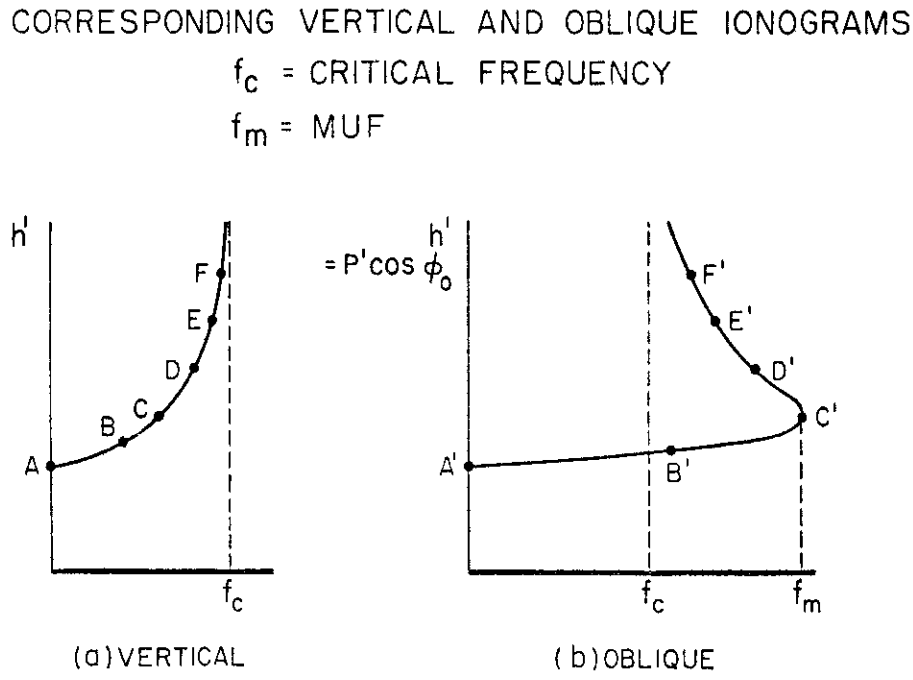

Fig. 3.11 Corresponding vertical and oblique ionograms

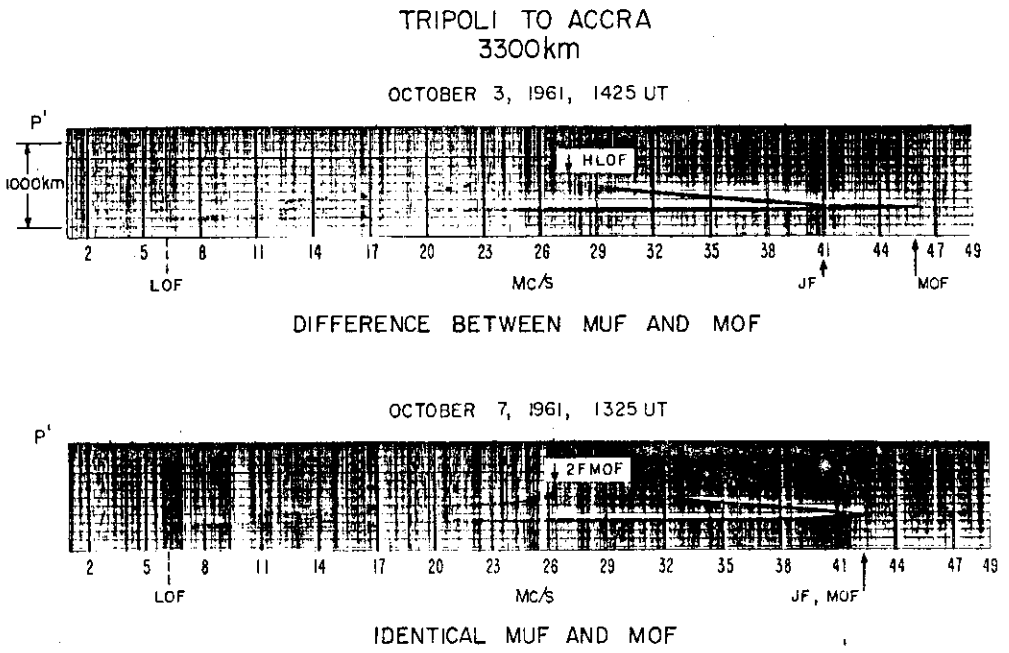

Fig. 3.12 Low-latitude oblique ionograms over the 3300 KM path Tripoli-Accra

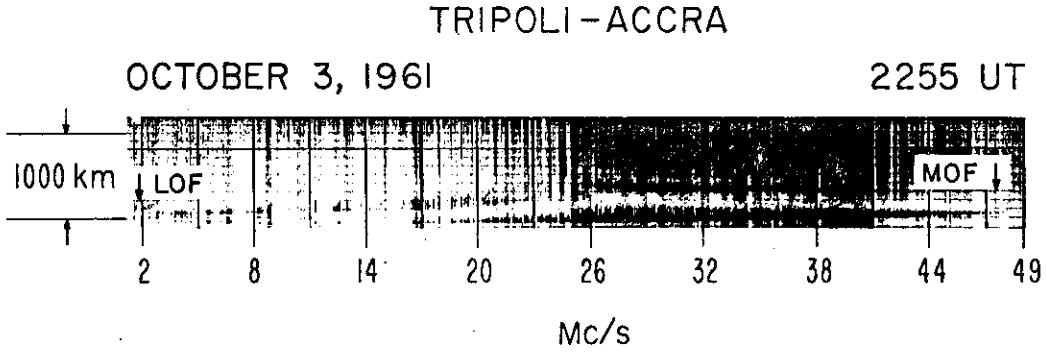

Fig. 3.13 Oblique ionogram for the 330 km path, Tripoli to Accra, cshowing effect of scatter

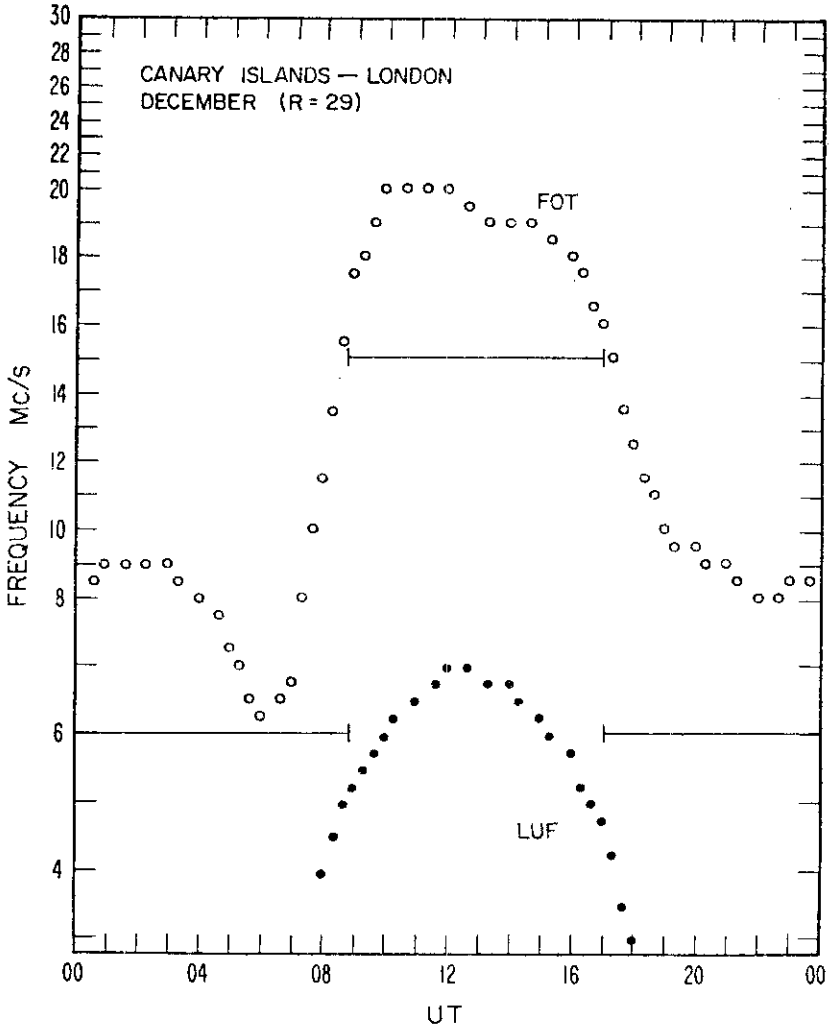

Fig. 3.14 Dirunal variation of optimum working frequency, the lowest useful frequency and a two frequency assignment

Fig. 3.15 Expected values of atmospheric radio noise, F_{am} in db above $kT_o b$ at 1 Mc/s for winter season, 0400-0800 hrs.

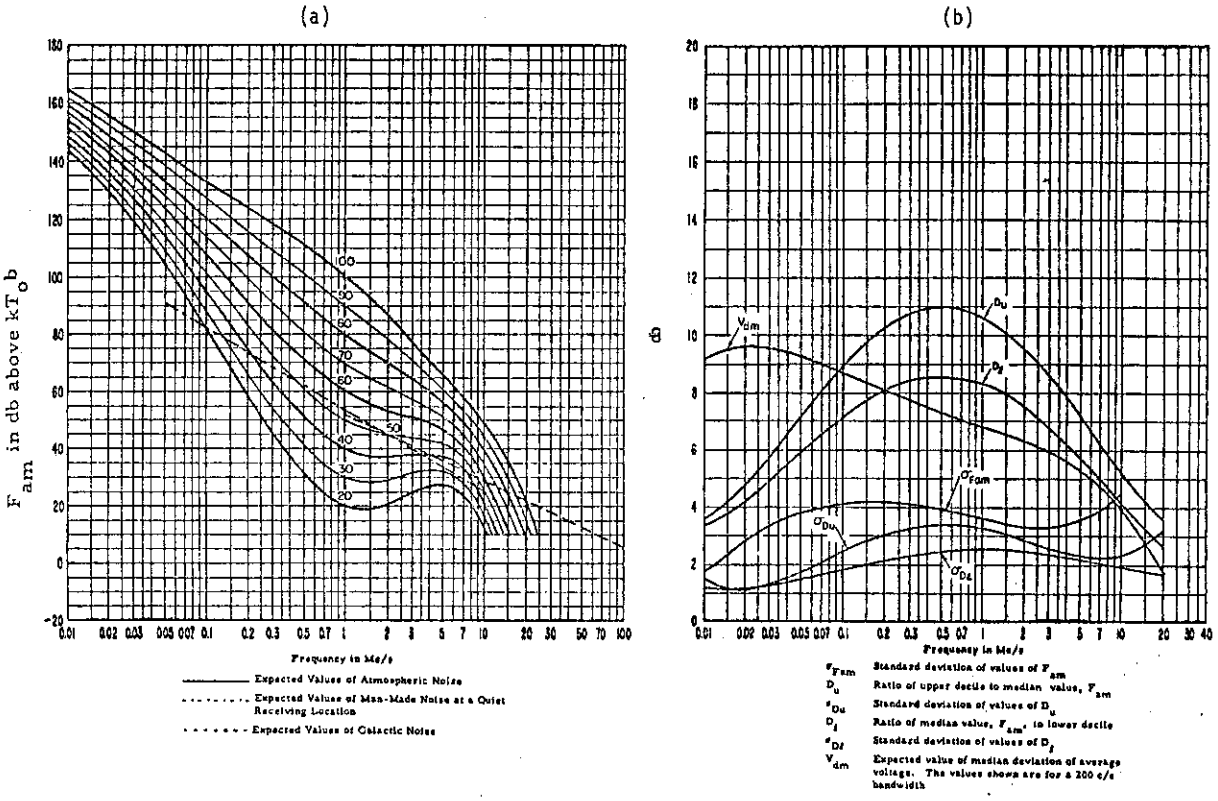

Fig. 3.16 (a) Variation of radio noise with frequency winter season 00-04 hrs
(b) Data noise variability and character winter season 00-04 hrs