Chapter 17

Skywave Communications

www.skywave-radio.org

17 Skywave Radio Communications

An HF radio signal transmitted at an oblique elevation angle will be bent back to Earth by the refractive properties of the ionosphere, returning to Earth some distance from the transmitting station (Figure 1). HF signals are generally refracted in the F region of the ionosphere. However, under certain conditions they can also be refracted back to Earth from the E layer as shown in Figure 2.

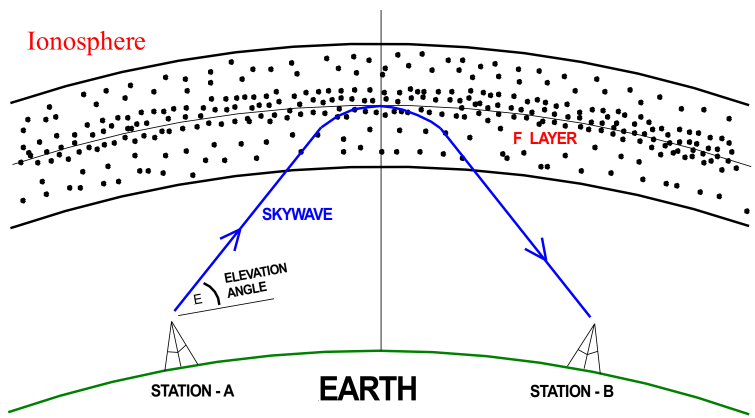

Figure 1 Radio waves passing through the ionosphere refract back to Earth (source: author)

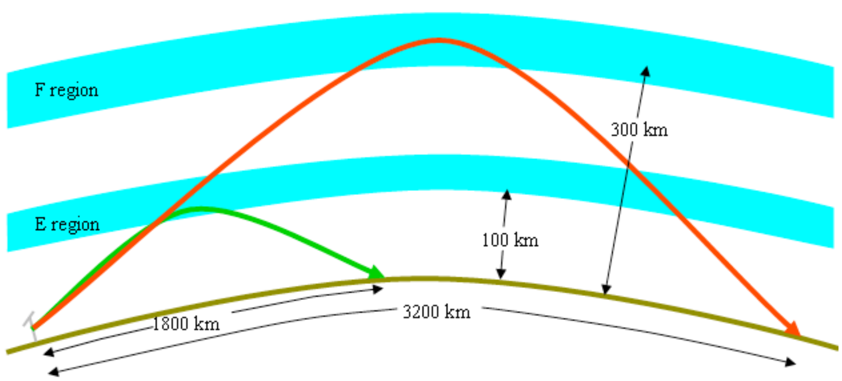

Figure 2 Radio waves refract from both the E and F regions (source: Space Weather Service)

In general, several hops through the ionosphere (Figure 3) are required to reach the destination radio station. The length of each hop depends on the altitude at which the signal is refracted. Signals refracting in the F layer travel further than signals refracted in the E region (Figure 2).

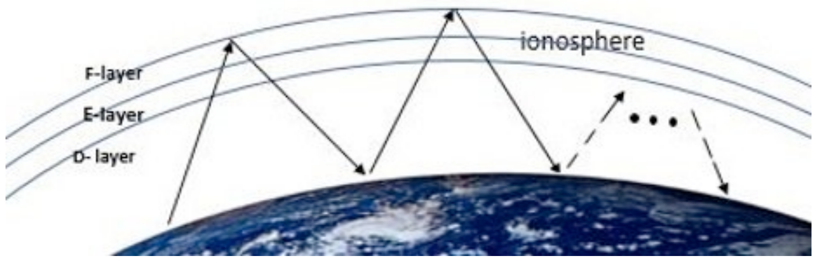

Figure 3 Multiple hops required to reach distant stations (source: ResearchGate)

Frequency dispersion in the ionosphere's index of refraction causes short wavelength signals to be refracted higher in the ionosphere than long wavelength signals. For example, the 20 meter signal in Figure 4 refracts higher in the ionosphere than an 80 meter signal. Consequently, the hop distance of short wavelength signals is greater than long wavelength signals. For this reason, the short wavelength 20 through 10 meter frequency bands are excellent long distance (DX) bands while the longer wavelength 80 and 40 meter bands are not. To achieve the best DX, we want to operate at the highest possible frequency. However, if a signal's wavelength is too short for current conditions, it will penetrate the ionosphere and be lost to outer space. The 15 meter signal in Figure 4 experiences this problem.

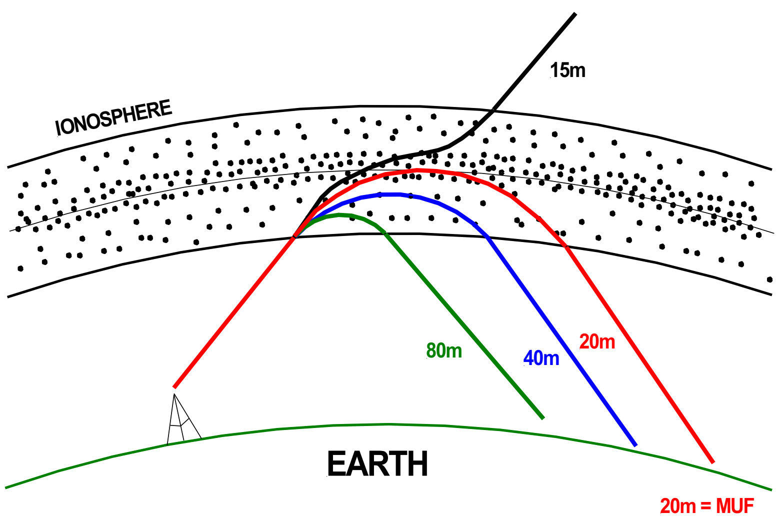

Figure 4 Frequency dispersion in the ionosphere (source: author)

During the day radio waves are partially absorbed as they pass through the ionosphere's D layer in route to and from the E and F regions (Figure 3). Long wavelength 80 through 40 meter signals are more heavily absorbed than shorter wavelength 20 through 10 meter signals. Typically, signals in the 80 meter band are completely absorbed during the day preventing the band from being used for skywave communications. The band can, however, be used for local communications by utilizing line of sight and ground wave propagation. During the day 40 meter signals are at least partially absorbed in the D layer. 20 through 10 meter signals experience little if any D layer absorption and so are excellent day time bands. At night the D region disappears so absorption is not a problem on 80 and 40 meters. Throughout the night 80 and 40 meter signals continue to refract back to Earth, but do so from higher in the ionosphere resulting in longer hop distances. At night 20 through 10 meter signals are lost to outer space due to the ionosphere's declining levels of ionization. Consequently, 80 and 40 meters are excellent night time bands.

17.1 Signal Absorption

As we know, Extreme Ultra Violate (EUV) energy from the Sun ionizes some of the atoms and molecules in Earth's upper atmosphere. The ionization process creates free electrons (electrons that have broken away from their parent atoms) and ions (atoms and molecules that have lost one or more electrons). Ions are massive compared to electrons. In fact, ions are 20,000 times more massive than electrons. Ions are too massive to be affected by a feeble radio wave that may pass by. However, low mass free electrons readily interact with passing radio waves. As they do so they absorb energy from the radio waves. The absorbed energy is not lost. Instead, it causes the free electrons to vibrate. Vibrating electrons produce electromagnetic waves which in this case are at the same frequency as the passing radio wave. The energy absorbed by the electrons is reradiated in random directions altering the direction of the radio wave, that is, causing the radio wave to bend. Because the absorbed energy is reradiated, little radio wave energy is lost as it travels through the E and F regions of the ionosphere.

The D region, illustrated in Figure 5, is different. The concentration of neutral atoms and molecules in the D region is very high compared to the E and F regions. Because of the high atmospheric density, free electrons recombine with ions so fast they do not have time to reradiate energy that they have absorbed from a radio wave. Instead, the absorbed energy is passed on to the massive ions in the form of heat.

D Layer absorption is inversely proportional to frequency squared. That is

$$Absorption \propto \frac{1}{f^2}$$

D layer absorption is greatest in the 160 and 80 meter frequency bands. The absorption on 40 meters is 1/4 that on 80, while absorption on 20 meters is only 1/16 that occurring on 80 meters. D layer absorption on 15 meters is insignificant. To avoid absorption, we want to operate at the highest frequency possible.

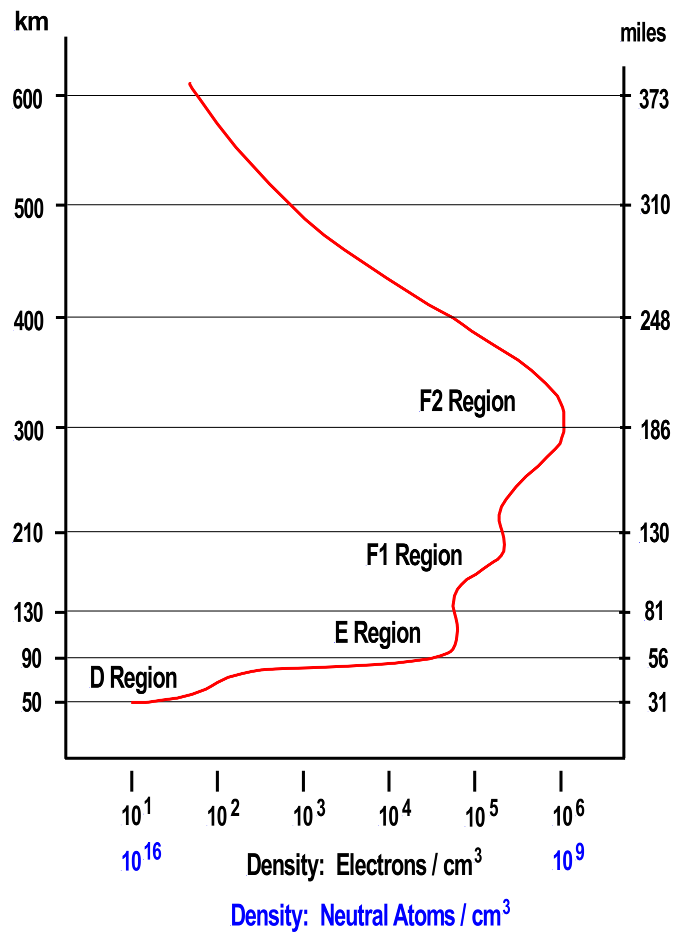

Figure 5 The ionosphere's ionization profile, electron density vs altitude (source: author)

17.2 Maximum Usable Frequency

Maximum Usable Frequency (MUF) is the highest frequency radio signal that can be transmitted through the ionosphere from one **specific** radio station to another (Figure 6). Transmitting at a frequency higher than the MUF will cause the radio signal to penetrate through the ionosphere and be lost to outer space.

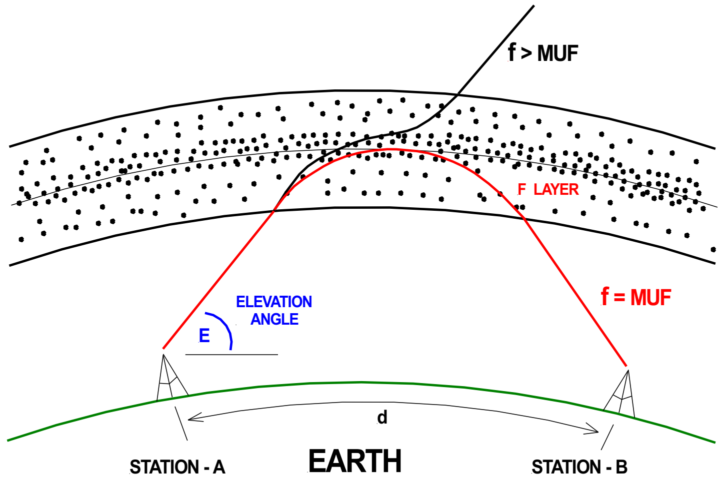

Figure 6 Illustration of Maximum Usable Frequency (source: author)

MUF depends on where the communicating stations are located relative to each other. For example, at any given time, the maximum usable frequency for communicating between Los Angeles California and Denver Colorado is different than the MUF for communicating between Los Angeles and San Diego California.

17.3 Graphically Determining MUF

The red trace in Figure 7 illustration the propagation path of a signal transmitted at an oblique angle by a radio station at site T. At an altitude of h km, refraction in Earth's ionosphere bends the signal back toward Earth where it is received by the radio station at site R. While the red trace accurately portrays the path of the radio signal through the ionosphere, its curved shaped can complicate propagation analysis without adding any additional useful information. In these situations it is easier to assume that radio signals are reflected back to Earth by the ionosphere. For example, in Figure 3 the multiple hop propagation path typical of most radio transmissions is easier to illustrate assuming that radio signals are reflected by the ionosphere.

The blue trace in Figure 7 shows the path that the signal would have to follow from radio station T to the receiving station at R if the signal were reflected in the ionosphere. Notice that the reflection point A at an altitude h'_o is greater than the signal's actual refraction altitude h.

The magenta trace on the right side of Figure 7 illustrates a signal transmitted vertically straight up into the ionosphere. Such a signal is typically transmitted by an ionosonde for the purpose of generating an ionogram (the red trace in Figure 8) depicting current ionospheric conditions. The altitude at which the vertical signal is reflected back to Earth is designated as h'_v. For example, in Figure 8 the refection altitude h'_v for a 6 MHz signal transmitted by an ionosonde is around 320 km.

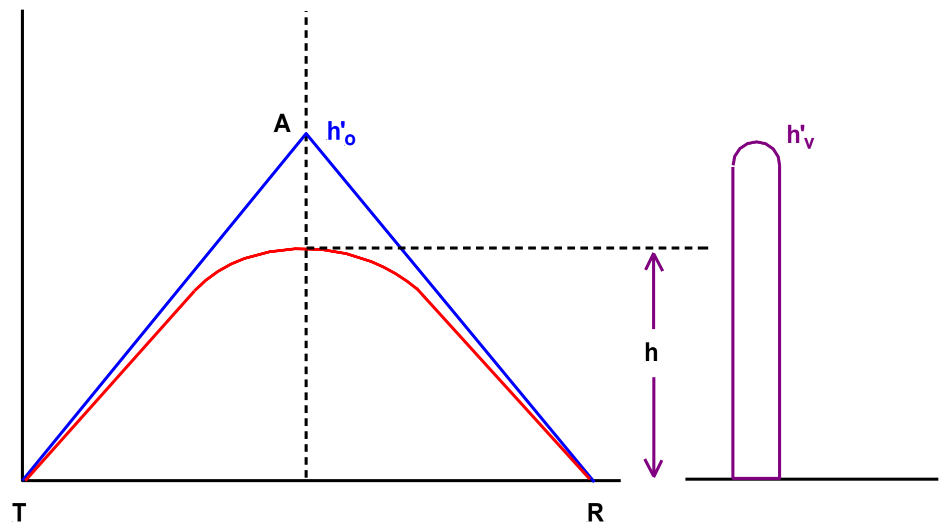

Figure 7 Ionospheric Propagation Virtual Height (source: author)

In 1935 D. F. Martyn showed that the virtual reflection heigh h'_o of an oblique wave is the same as the reflection height h'_v of a wave transmitted vertically. Martyn's theorem permits an ionogram (red trace in Figure 8) to be superimposed on transmission curves (blue traces in Figure 8) to graphically determine the MUF between two stations D km apart. The blue transmission curves in Figure 8 are obtained from the equation

$$h' = \frac{D}{2\sqrt{(f_o/f_v)^2 - 1}}$$

by holding the distance D between two stations constants, selecting a value for the oblique transmission frequency f_o, and varying the vertical incident frequency f_v over a range of values to produce a single transmission curve. The value of f_o is incremented and the process repeated forming the next transmission curve. The procedure is continued over the range of f_o values of interest forming a family of transmission curves.

The shape of the red ionogram trace in Figure 8 indicates that the F2, F1, and E regions of the ionosphere are present. The intersection of a blue transmission curve having a frequency f_o with the red ionogram trace gives the virtual height of reflection for that transmitted signal. For example, a signal transmitted at a frequency of 14 MHz intersects the ionogram trace at 5 places, points A and A' in the F2 region, d and d' in the F1 region, and point e in the E region. These 5 points are the only points in the ionosphere at which the 14 MHz signal in this example can reflect and reach the receiving site R located a distance of D km from the transmitting station T. Each of these reflection points identifies a particular propagation path from T to R. Thus 5 different propagation paths are possible. A different set of transmission curves, reflection points, and propagation paths are required to reach some other receiving station R' located a distance E km from the transmitter.

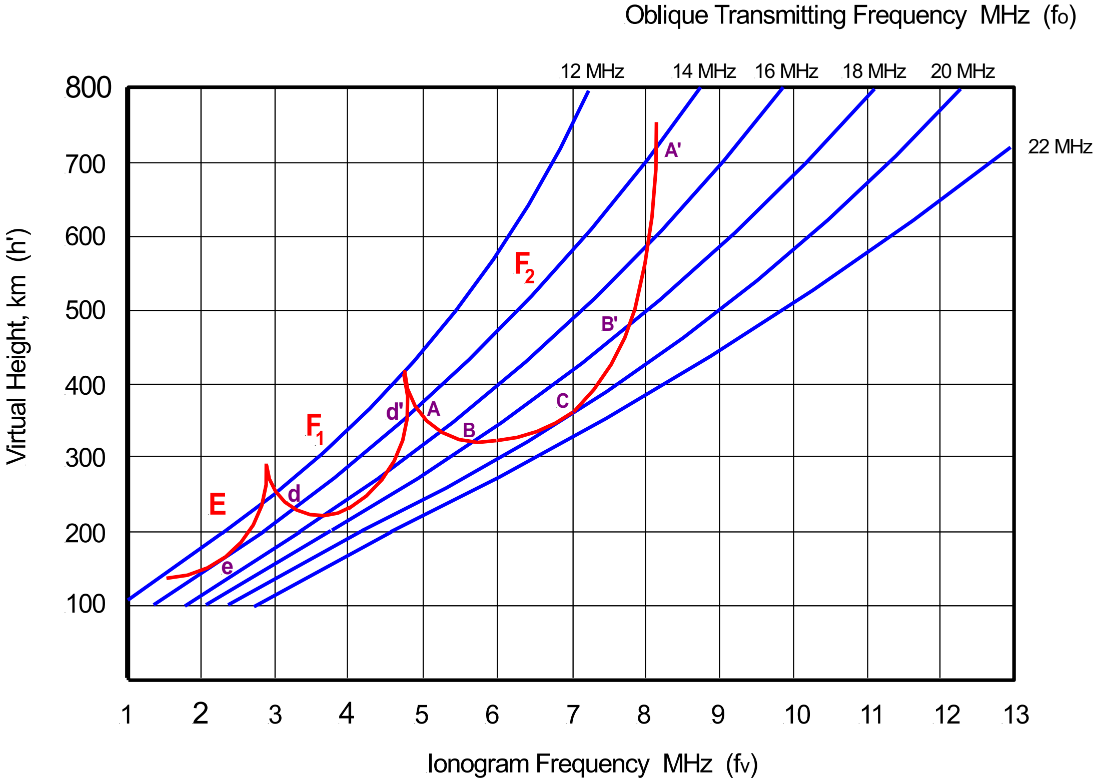

Figure 8 Ionogram superimposed on transmission curves (source: derived from Davies)

There are two reflection points in the F2 region of the ionosphere (A and A'). The propagation path corresponding to the lower reflection point, point A, is called the low-angle path or ray while the signal reflecting at point A' is the high-angle propagation path. The high-angle path is frequently referred to as a Pedersen ray. There are also two reflection points in the F1 region (d and d') producing a low-angle propagation path and a high-angle path for that region as well. In this example, only a single reflection point occurs in the E region of the ionosphere.

A signal transmitted from the ground will be reflected by the first reflection point that it encounters. In Figure 8 the first 14 MHz reflection point encountered is in the E region of the ionosphere. While reflection is technically possible in both the F1 and F2 regions, no reflections occur at these higher altitudes since the signal is reflected by the E layer. The F1 and F2 regions are said to be screened by the E region since the E region prevents signals from reaching the higher F1 and F2 layers.

In this example, an 18 MHz signal intersects the ionogram curve only in the F2 region of the ionosphere. Consequently the 18 MHz signal can not reflect from either the E or F1 regions of the ionosphere. Moving to a transmitting frequency of 18 MHz or higher prevents the E and F1 regions from screening the F2 region.

Notice that the virtual heights for the two 18 MHz reflection points (B and B') are closer together than the 14 MHz reflection points A and A'. B and B' are separated by only 130 km in virtual height while A and A' are 310 km apart. Increasing the transmitting frequency further to 20 MHz causes the low-path and high-path reflection points to merge at a single point C. In fact, the 20 MHz transmission curve is tangent to the ionogram at point C. Increasing the transmitting frequency just a little more, say to 21 MHz, will cause the resulting transmission curve to miss the ionogram altogether. In this example, 20 MHz is the maximum usable frequency for signals transmitted a distance of D km from Station T to Station R. A signal with a frequency higher than 20 MHz will penetrate the ionosphere and be lost to outer space.

Any signal with a frequency lower than the F2 maximum usable frequency will intersect the F2 segment of the ionogram in two places, as illustrated in Figure 8. The same is true for the F1 and E maximum usable frequencies. In Figure 8, a 14 MHz signal intersects the ionogram trace at only one point, point e, in the E region of the ionosphere. Thus, in this example, 14 MHz is the MUF for signals reflecting in the E region.

The above observations are particularly important. All signals propagating through the ionosphere at a frequency below the MUF will follow two paths from the transmitter T to the receiving station R. The two paths are the high-angle path and the low-angle path. The two paths become closer together as the transmitting frequency is increased. The two paths merge, becoming a single path, at the signal's maximum usable frequency. Merging of the two paths enhances the strength of the received signal.

Both the low-angle and the high-angle signals arrive at the receiving station. Unfortunately, the path followed by the high-angle signal is longer than the low-angle path meaning that the two signals are out of phase when they arrive at the receiving station. The out of phase signals interfere with each other causing the resulting received signal to fade in and out.

The high-angle signal is more sensitive to small changes in ionospheric conditions than the low-angle signal. Consequently, the high-angle signal tends to become de-focused and weaker than the low-angle signal. The signal received at the receiving station is generally the low-angle signal with the contribution, and interference, from the high-angle signal usually being less important. However, in some cases, ionospheric conditions are such that it is the high-angle signal that is actually received.

As a final note before leaving this section. The ionogram shown in Figure 8 is for the ordinary or o-mode of propagation. Including the extra ordinary x-mode of propagation adds a second ionogram trace slightly displaced from the o-mode ionogram. In the equation

$$h' = \frac{D}{2\sqrt{(f_o/f_v)^2 - 1}}$$

there are now two ionosonde frequencies, f_{vo} for the vertical o-mode signal and f_{vx} for the vertical x-mode signal. This adds a second set of reflection points. Consequently, there are 4 propagation paths when transmitting at a frequency below the MUF. These paths are the low-angle and high-angle o-mode paths and the low-angle and high-angle x-mode paths. The o-mode and x-mode paths sometimes end up at different destinations.

17.4 Calculating MUF

Mathematically, maximum usable frequency is a function of the ionosphere's critical frequency at the center of the first hop and the transmitted signal's elevation angle. In equation form

$$MUF = \frac{f_c}{\sin E}$$

where

- MUF = Maximum Usable Frequency,
- f_c = The ionosphere's Critical Frequency, and
- E = Elevation angle of the signal radiating from the transmitting antenna.

17.4.1 Critical Frequency

As discussed previously, and illustrated in Figure 9, critical frequency $\mathbf{f_c}$ is the highest frequency signal that can be transmitted straight up and reflected back down to Earth.

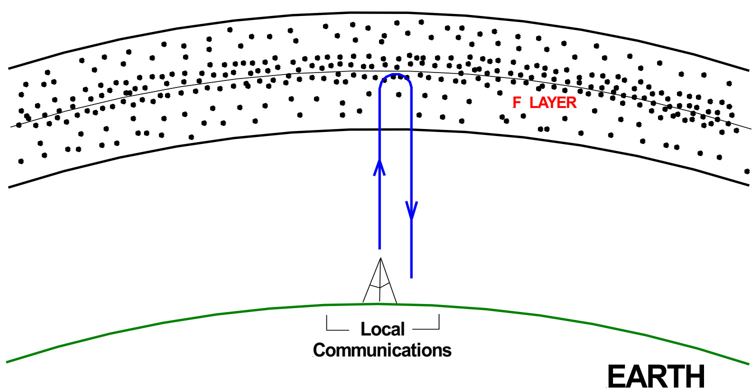

Figure 9 f_c highest frequency vertical signal reflected by ionosphere (source: author)

Each region of the ionosphere has its own critical frequency determined by the maximum electron density in that region. The critical frequencies for the various regions are designated as:

- F2 Region = f_cF2
- F1 Region = f_cF1, and
- E Region = f_cE

The critical frequencies appear as spikes in the ionogram shown in Figure 8. In this figure the critical frequencies are approximately f_cE = 2.9 MHz, f_cF1 = 4.8 MHz, and f_cF2 = 8.2 Mhz.

Maximum usable frequency is always equal to or greater than the critical frequency. For example, at an elevation angle of 90°, maximum usable frequency is equal to the critical frequency.

$$MUF = \frac{f_c}{\sin E} = \frac{f_c}{\sin 90°} = \frac{f_c}{1} = f_c$$

At an elevation angle of 45°, maximum usable frequency is greater than the critical frequency.

$$MUF = \frac{f_c}{\sin E} = \frac{f_c}{\sin 45°} = \frac{f_c}{0.707} = 1.414\, f_c$$

Since the ionosphere is formed by extreme ultra violate (EUV) radiation from the Sun, critical frequency and maximum usable frequency vary:

- In accordance with the 11 year solar cycle,
- Seasonally, and
- Throughout the day

17.4.1.1 Variations In Critical Frequency Due To The Solar Cycle

The E, F1, and F2 critical frequencies all vary with the solar cycle as shown in Figure 10. This is to be expected. An increase in solar activity produces an increase in EUV radiation responsible for ionizing Earth’s upper atmosphere.

The red trace in Figure 10 is the smoothed sunspot number (ssn) for the years 1988 through 2015. The vertical scale on the right is the smoothed sunspot number. The scale on the left is critical frequency in MHz. The graph shows three solar cycles. The solar cycles shown on the right and left are partial solar cycles while the complete Solar cycle 23 is shown in the middle. Solar cycle 23 began as a minimum in late 1996 with sunspot numbers around 10 or so. Sunspot maximum occurred in mid 2000 and again in 2002 before declining to the next solar minimum in 2009.

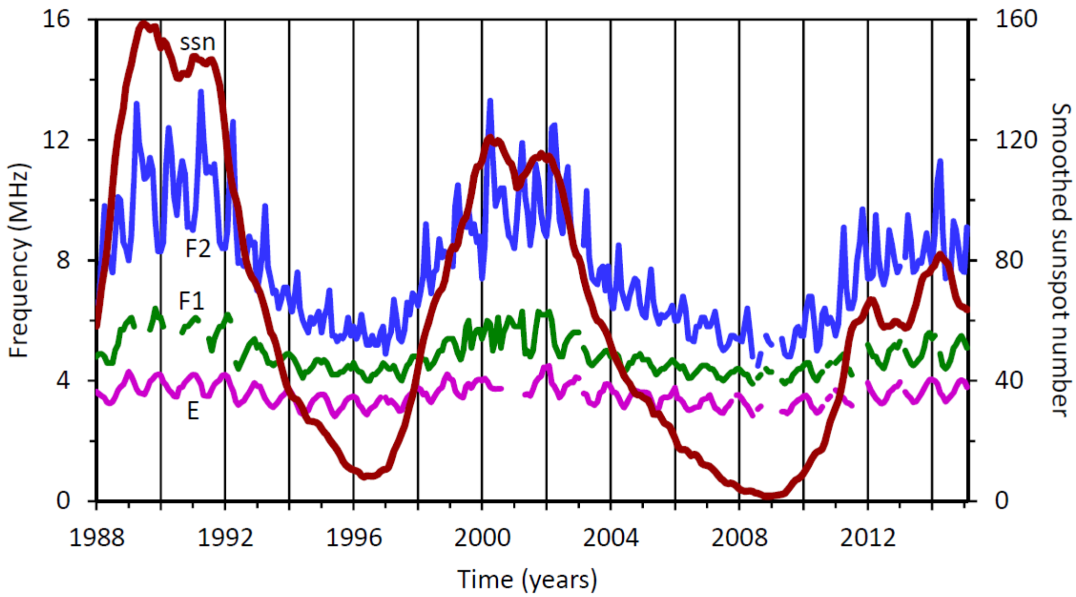

Figure 10 Variations of critical frequencies with the solar cycle (source: sws.bom.gov.au)

As can be seen in Figure 10, the solar cycle has the greatest effect on F2 critical frequencies. The highest daytime F2 critical frequencies occur during solar maximums with $8 < f_cF2 < 14$ MHz while at solar minimum the typical F2 critical frequency range is $5 < f_cF2 < 6$ MHz. The highest F1 and E critical frequencies also occur during solar maximum. However, the solar cycle has less of an effect on F1 and E critical frequencies with the E region critical frequency being affected the least.

17.4.1.2 Seasonal Variations In Critical Frequencies

Seasonal changes in critical frequencies are due primarily to:

- Seasonal changes in zenith angles, and
- Seasonal changes in the Earth's upper atmosphere.

Noon time zenith angles (Figure 11) are always less in summer when the Sun is more overhead. We would thus expect critical frequencies to be higher in summer than in the winter. And they are for the E and F1 zones, but not so for the F2 region. During solar maximum F2 critical frequencies are substantially higher in winter than in summer, as illustrated in Figure 12, despite the fact that in winter the Sun is low in the sky. This is known as the seasonal anomaly.

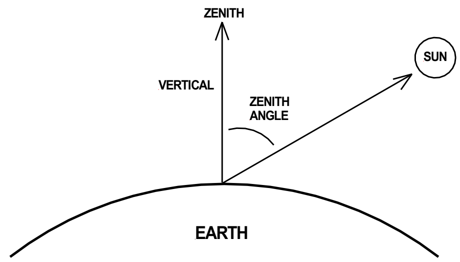

Figure 11 Zenith angle (source: author)

As expected, E region critical frequencies are slightly higher in summer during both solar maximum and minimum. The F1 critical frequency is also higher in the summer during solar minimum. However, the F1 region disappears in the winter during solar maximum.

All critical frequencies are lower during solar minimum as illustrated in Figure 13. Low activity on the Sun means lower levels of EUV radiation available to ionizes Earth's upper atmosphere. Consequently, critical frequencies are lower during solar minimum. In addition, F2 critical frequencies tends to peak in the late afternoon during solar minimum instead of around noon time (Figure 13). This is not what would be expected, but it is what happens. The F1 region can appear in both the winter and summer during solar minimum.

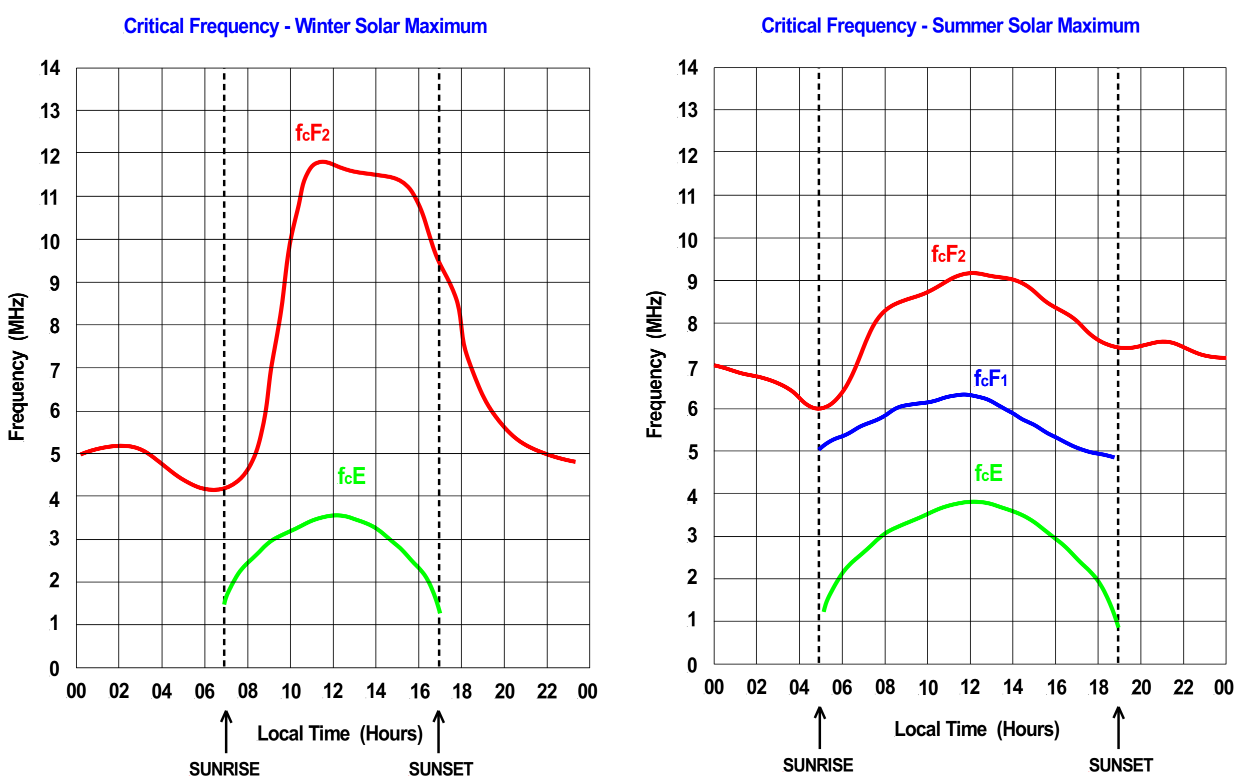

Figure 12 Winter and summer critical frequencies during solar maximum (source: author)

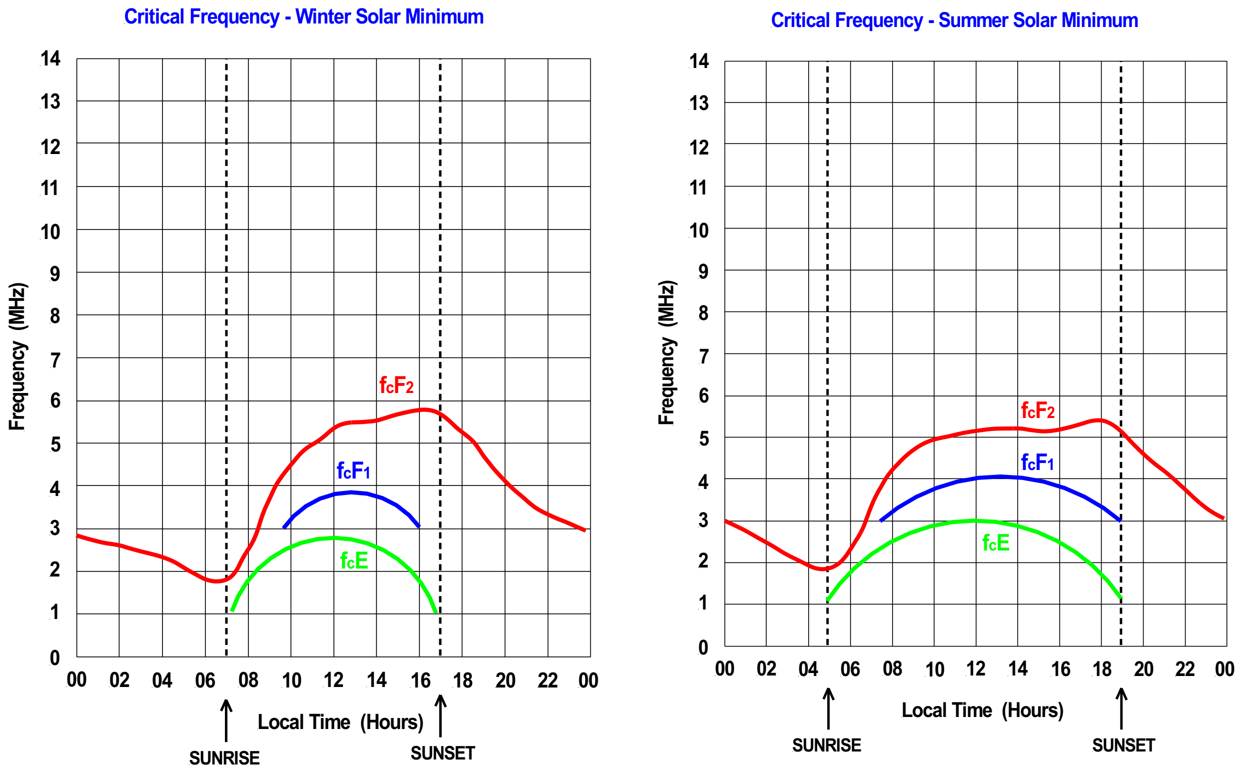

Figure 13 Winter and summer critical frequencies during solar minimum (source: author)

17.4.1.3 Diurnal Variations In Critical Frequencies

Critical frequencies vary throughout the day and night as shown in Figures 12 and 13. It is important to note that daily variation in F2 critical frequencies follow a general trend but do not exactly repeat from one day to the next, making prediction of F2 performance difficult.

At night the F1 and F2 regions combine forming a single F region. Ionization ceases at night, due to the lack of sunlight, while recombination continues. Consequently, electron densities and F region critical frequencies decline throughout the night. F region critical frequency drops to its lowest level just before sunrise with f_cF around 2 – 3 MHz during solar minimum and 4 – 6 MHz during solar maximum.

Ionization resumes at sunrise causing the F region to split into F1 and F2 layers, except in winter during solar maximum when the F1 layer does not form. F2 critical frequency quickly increases until around 10 AM to Noon when it plateaus as electron production and recombination reach equilibrium. During solar maximum the F2 region reaches its highest critical frequency around noon with critical frequencies from 9 – 12 MHz. It then declines in the late afternoon and throughout the evening as electron product decreases and finally stops while recombination continues. However, during solar minimum F2 critical frequency peaks around noon, levels off or slightly declines, and then peaks again at a higher frequency of around 5 – 6 MHz in the late afternoon.

During the day the F1 and E critical frequencies both peak at local noon. At night, the F1 region merges with the F2 region forming a single F layer. The E region does not completely disappear at night. At night $F_cE \sim 0.6$ MHz, a value which is too low to have any significant effect on HF communications. Thus, for practical purposes we assume that the E region also disappears at night.

17.4.1.4 F2 Critical Frequency Map

The global F2 critical frequency map shown in Figure 14 is provided by the Australian Government. It is an excellent tool for determining current F2 critical frequencies world-wide. The map is created automatically from reports received from ionosonde monitoring stations around the world and is updated every 15 minutes. The map is available by clicking on "Critical Frequency" under the "Current Conditions" tab on the www.skywave-radio.org website.

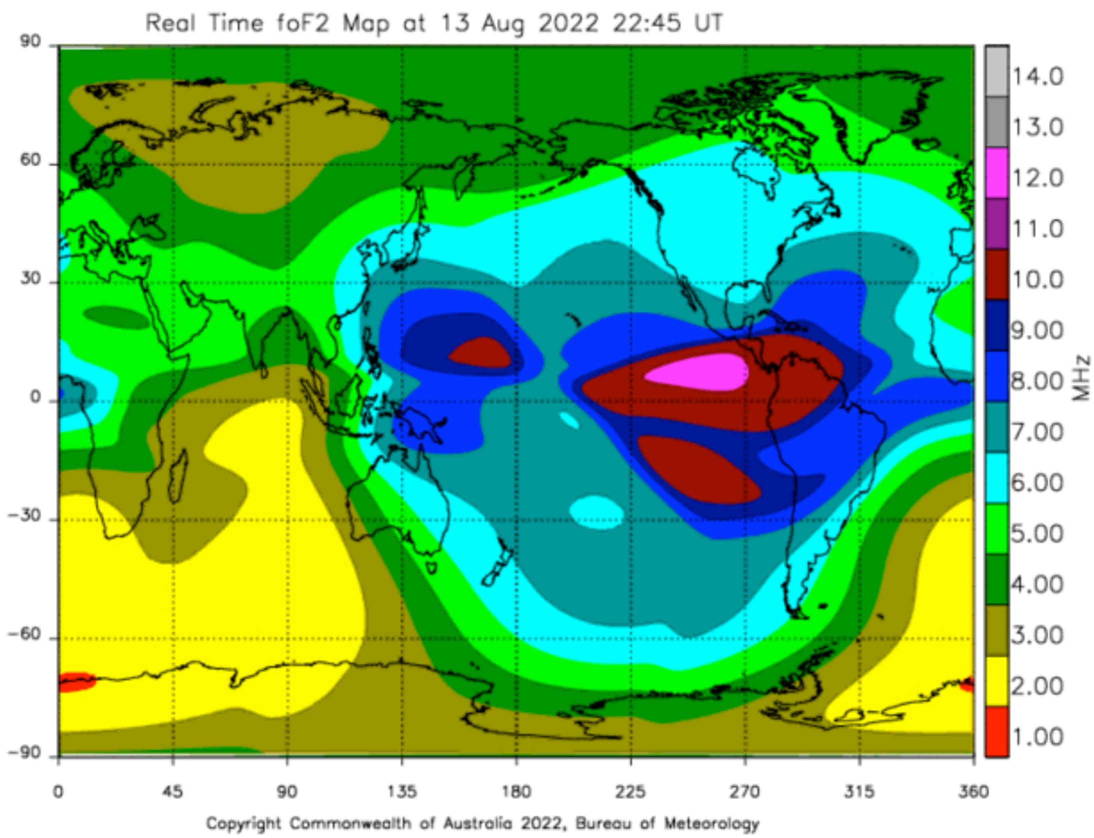

Figure 14 Critical Frequency Map (source: Australia Bureau of Meteorology)

The map shown in Figure 14 is for August 13, 2022 at 22:45 UT during the ascending phase of Solar Cycle 25. At this point in time the Sun was just emerging from solar minimum. Consequently, critical frequencies were close to solar minimum levels. The daytime F2 critical frequency for the United States was only 6 MHz. Consistent with solar minimum conditions, critical frequency peaked in the late afternoon and early evening. Off the east coast of the United States critical frequency reached 7 MHz at 7 PM local time while it had been around 6 MHz earlier in the afternoon.

Figure 14 clearly shows the winter anomaly. August is winter time in the southern hemisphere. Due to the winter anomaly, we would expect F2 critical frequencies to be higher in the southern hemisphere than in the northern hemisphere, and they are. In the southern hemisphere the 7 MHz critical frequency band stretches down almost to 60^o south latitude. However, in the northern hemisphere the 7 MHz critical frequency band extends only a little above 30^o north latitude. In addition, significant segments of the 5 and 6 MHz critical frequency bands extend beyond 60^o latitude in the southern hemisphere. Not so in the northern hemisphere. In the northern hemisphere only small segments of the 5 and 6 MHz critical frequency bands extend beyond 60^o latitude. Darkness in the southern hemisphere at the time this map was generated produced night time critical frequencies as low as 2 MHz while long hours of sunlight in the Artic region maintained critical frequencies at 4 MHz, just below daylight levels.

The highest critical frequencies generally occur in the tropics, as is the case in Figure 14. We would expect the highest critical frequencies to occur over the equator where radiation from the Sun is most intense. However, this is not the case. In Figure 14 patches of high critical frequency levels occurred on both sides of the equator. This is a typical example of the equatorial anomaly explained in some detail in Chapter 17 Equatorial Ionosphere.

17.4.1.5 Critical Frequency Summary

Critical frequency f_c in the maximum usable frequency equation

$$MUF = \frac{f_c}{\sin E}$$

varies over the course of the 11 year solar cycle, seasonally, and throughout the day. Consequently, the maximum usable frequency for communicating between two HF radio stations also varies over the solar cycle, from one season to another, and throughout the day.

17.4.2 Elevation Angle

Elevation angle E in the MUF equation is the angle at which a signal must be transmitted to reach a distant receiving station as illustrated in Figure 15. Elevation angle is measured relative to horizontal ground.

$$MUF = \frac{f_c}{\sin E}$$

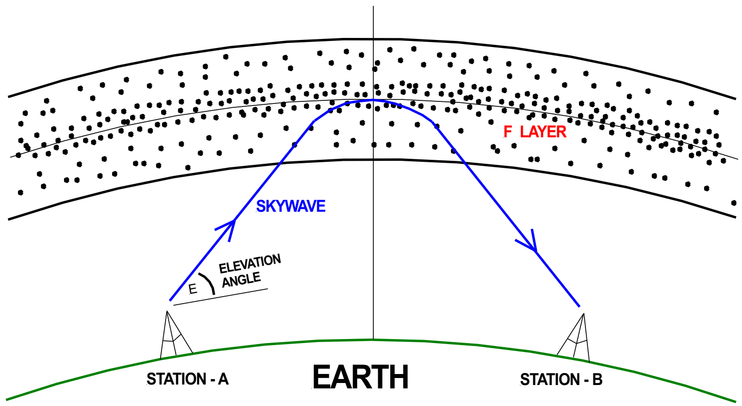

Figure 15 Elevation Angle

Figure 16 is a distance verses elevation chart for HF signals transmitted through the ionosphere. The x-axis is the distance in miles along Earth's surface from the transmitting to receiving station. The y-axis is the elevation angle at which the signal must be transmitted to reach the receiving station. In this figure, the Red trace is for a radio signal refracted in the ionosphere at an altitude of 125 miles above Earth's surface. The Blue trace is for a signal refracted at an altitude of 220 miles. As illustrated in Figure 17, an altitude of 125 miles is approximately equal to the lower edge of the F region while 220 miles is typically the F region upper boundary. Actual refraction of a radio signal occurs somewhere between these two limits.

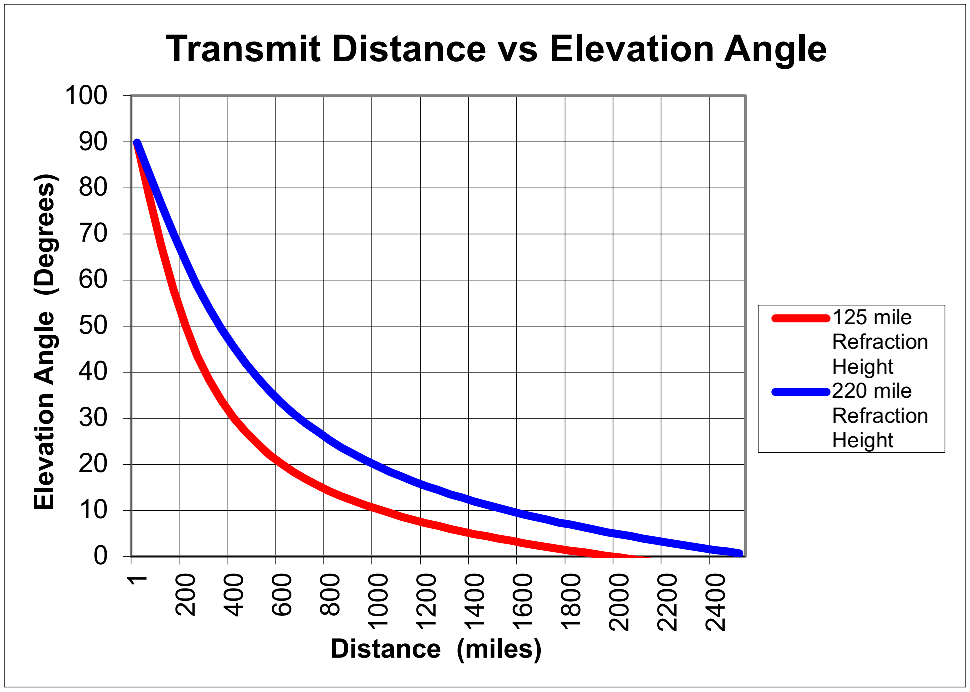

Figure 16 Transmit Distance vs Elevation Angle (source: author)

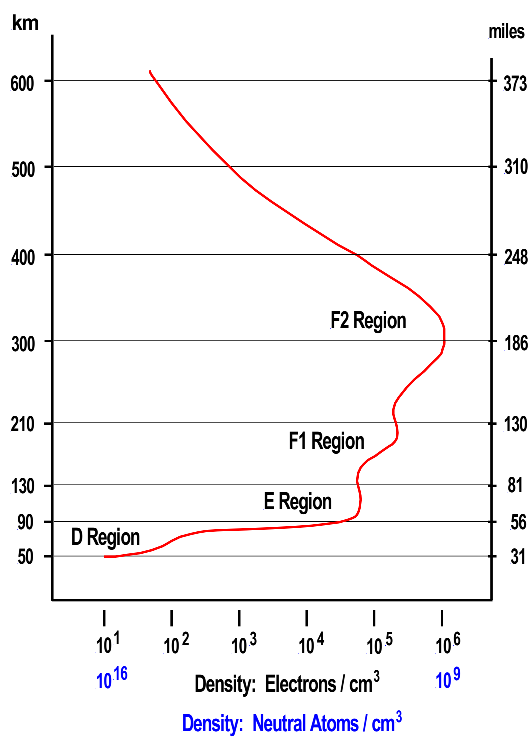

Figure 17 Regions of the ionosphere (source: author)

The equation required to produce the distance verses elevation angle chart is derived below using Figure 18.

In Figure 18, T is the location of the transmitting station and R the receiving station's location. The green arc labeled "D = Distance" is the ground distance between the transmitting and receiving sites.

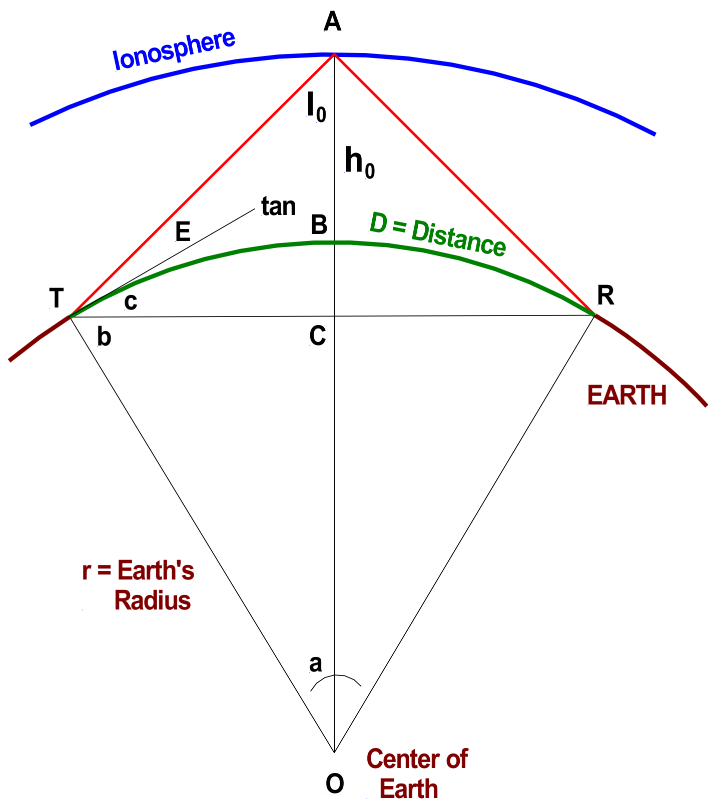

Figure 18 Deriving the distance verses elevation angle equation (source: derived from Davies)

The line segment BA is the virtual altitude h_0 of the ionospheric point at which the transmitted radio signal (red trace) is reflected back to Earth arriving at the receiving station R. The angle I_0 is the angle of incidence (with respect to vertical) of the radio signal with the ionosphere. The angle E is the elevation angle of the transmitted signal relative to ground at the transmitting site. Ground at the transmitting site is represented by the line tangent to Earth at the transmitting point T. In Figure 18 this line is labeled "tan".

The radius of the Earth is denoted by r (approximately 4,000 miles) and is measured from the center of the Earth (O) to the Earth's surface. The Distance D from the transmitting to the receiving site is the arc along the Earth surface from T to R. The length of this arc is

$$D = ra$$

were a is the angle in radians at the center of the Earth measured between the two radii OT and OR. Rearranging terms

$$a = \frac{D}{r}$$

The following trigonometric relations can be derived from the above figure

$$\sin\frac{a}{2} = \frac{TC}{r}$$

$$TC = r\sin\frac{a}{2}$$

$$\cos\frac{a}{2} = \frac{OC}{r}$$

$$OC = r\cos\frac{a}{2}$$

$$OB = r$$

$$BC = OB - OC = r - r\cos\frac{a}{2} = r\left[1 - \cos\frac{a}{2}\right]$$

$$h = h_0 + BC = h_0 + r\left[1 - \cos\frac{a}{2}\right]$$

$$\tan I_0 = \frac{TC}{h} = \frac{r\sin\frac{a}{2}}{h_0 + r\left[1 - \cos\frac{a}{2}\right]}$$

$$I_0 = \tan^{-1}\left[\frac{r\sin\frac{a}{2}}{h_0 + r\left[1 - \cos\frac{a}{2}\right]}\right]$$

$$E = 90° - I_0 - c$$

$b + c = 90° =$ the angle between the radius of a circle and the tangent to the circle (the Earth)

$$c = 90° - b$$

$$b = 90° - \frac{a}{2}$$

$$c = 90° - \left(90° - \frac{a}{2}\right)$$

$$E = (90° - I_0) - \left[90° - \left(90° - \frac{a}{2}\right)\right]$$

Since

$$a = \frac{D}{r}$$

$$E = (90° - I_0) - \left[90° - \left(90° - \frac{D}{2r}\right)\right]$$

$$I_0 = \tan^{-1}\left[\frac{r\sin\frac{a}{2}}{h_0 + r\left[1 - \cos\frac{a}{2}\right]}\right]$$

$$E = \left[90° - \tan^{-1}\left[\frac{r\sin\frac{D}{2r}}{h_0 + r\left[1 - \cos\frac{D}{2r}\right]}\right]\right] - \left[90° - \left(90° - \frac{D}{2r}\right)\right]$$

To evaluate this equation the 90º terms must be converted to radians. An angle $g°$ is converted to g radians by the following equation

$$g = \frac{\pi}{180} g° \quad and \quad g° = \frac{180}{\pi} g$$

If

$$g° = 90°$$

then

$$g = \frac{\pi}{180} g^\circ = \frac{\pi}{180} 90 = 1.57\, radians$$

Substituting 1.57 radians for 90°, the equation for elevation angle E becomes

$$E = \left[1.57 - \tan^{-1}\left[\frac{r \sin\frac{D}{2r}}{h_0 + r\left[1 - \cos\frac{D}{2r}\right]}\right]\right] - \left[1.57 - \left(1.57 - \frac{D}{2r}\right)\right] \; radians$$

Since we want elevation angle expressed in degrees, we must convert E radians to E°, so

$$E^\circ = \frac{180}{\pi} E = 57.296\, E$$

and

$$E^\circ = 57.296\, E = 57.296\left[1.57 - \tan^{-1}\left[\frac{r \sin\frac{D}{2r}}{h_0 + r\left[1 - \cos\frac{D}{2r}\right]}\right]\right] - \left[1.57 - \left(1.57 - \frac{D}{2r}\right)\right]$$

There are only three independent variables in this equation

r = the radius of the Earth,

h_0 = the altitude at which the radio wave is reflected from the ionosphere, and

D = the distance from the transmitting to the receiving station.

The graph for elevation angle $E°$ verses distance D is obtained by holding $r\ and\ h_0$ constant while varying D over a range from 1 to 2,500 miles. Doing this produces the graph shown in Figure 16.

17.4.3 Calculating MUF – Example

The following example explains how MUF is calculated and used in determining connectivity for a weekly 40 meter California Emergency Services Net (CESN). Figure 19 shows three of the stations participating in the net. Sacramento is CESN net control while San Bernardino is the backup net control. San Diego is one of the Emergency Operation Centers (EOCs) participating in the net.

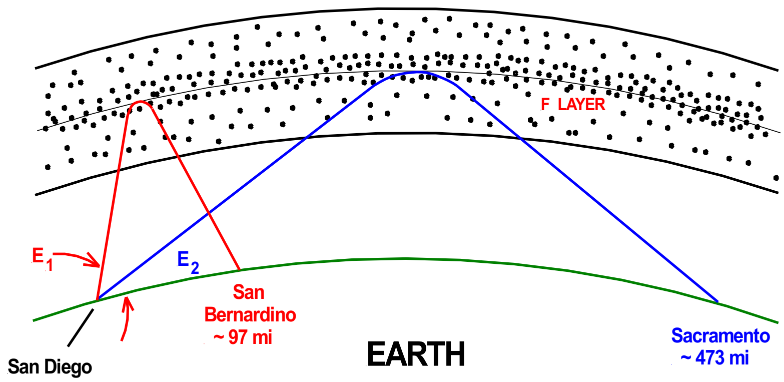

Figure 19 Portion of the California Emergency Services Net (source: author)

The question is this, can the San Diego EOC communicate with both the CESN net control station located in Sacramento and San Bernardino backup net control? To answer this question the maximum usable frequencies from San Diego to Sacramento and from San Diego to San Bernardino must be determined and compared to the nets operating frequency of 7.23 MHz, where MUF is

$$MUF = \frac{f_c}{\sin E}$$

The first step is to determine the elevation angle E_1 for communicating from San Diego to San Bernardino and E_2 for communications between San Diego and Sacramento. The distance from San Diego to San Bernardino is 97 miles while Sacramento is 473 miles from San Diego. The elevation angles are determined using the graph in Figure 20 and assuming that radio signals are refracting from the ionosphere at an altitude midway between the red and blue limits shown in the figure. At a distance of 97 miles an angle of about 75° is required to communicate between San Diego and San Bernardino. An angle of approximately 35° is required to communicate from San Diego to Sacramento, a distance of 473 miles.

The next step is to determine the ionosphere's critical frequency f_c above California at the time of the net. This is done by using the Critical Frequency Map in Figure 21. For this example, the net is assumed to have occurred May 28, 2020 at 17:15 UT. At that time the critical frequency across North America was 5 MHz.

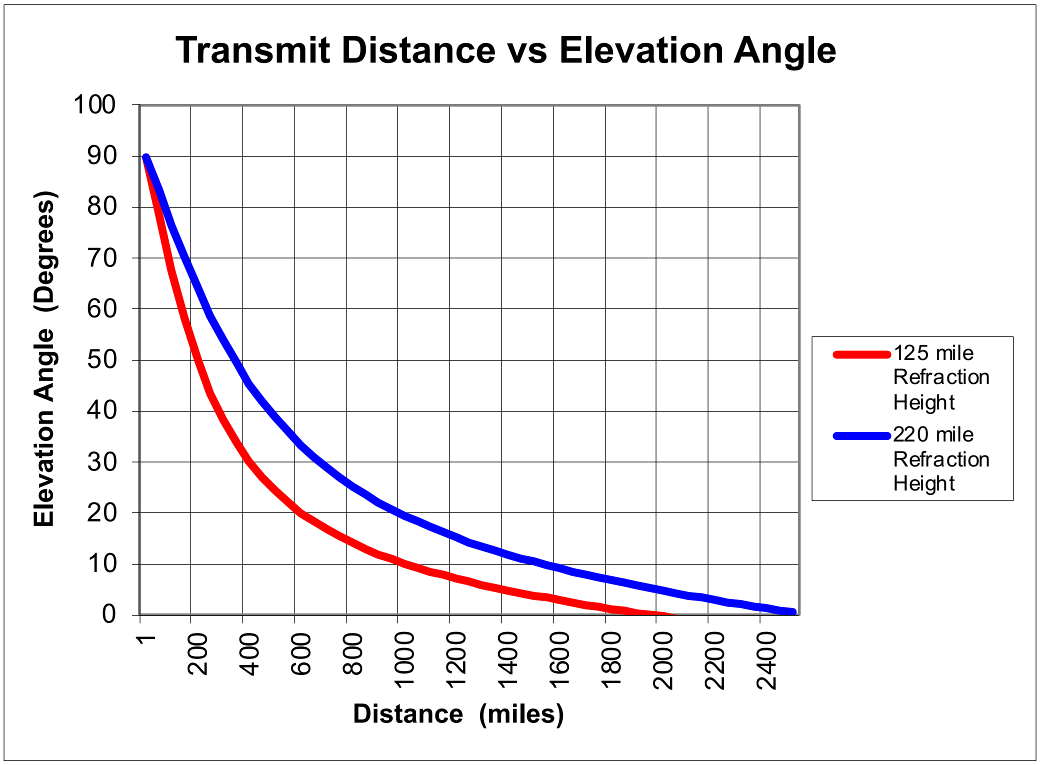

Figure 20 Distance verses Elevation Angle Graph (source: author)

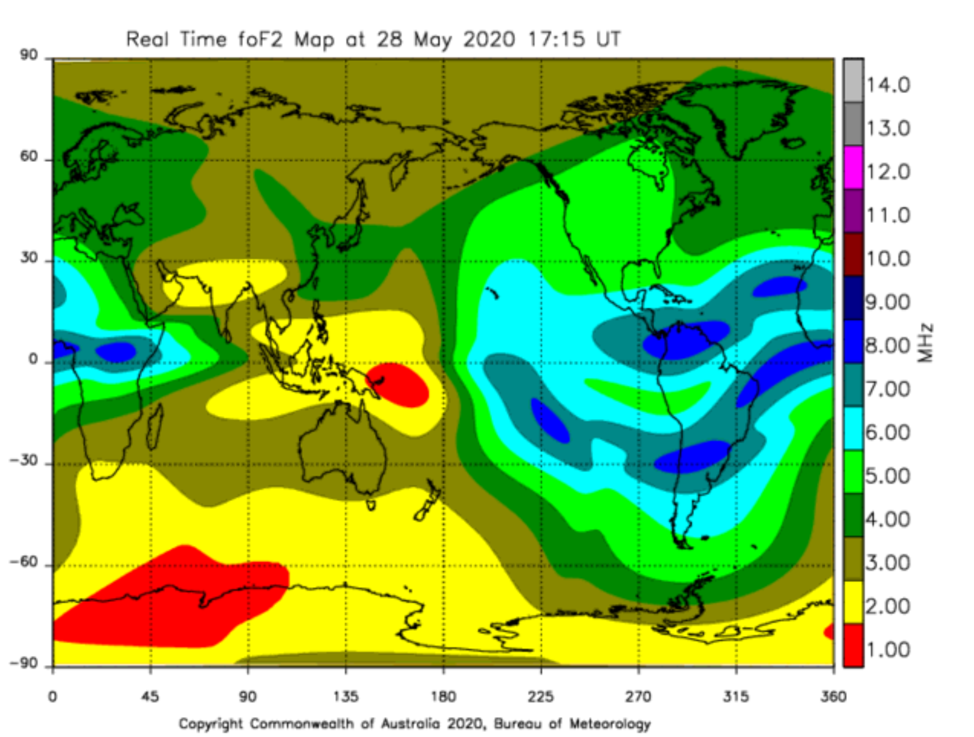

Figure 21 Critical Frequency Map for May 28, 2020 at 17:15 UT

With the critical frequency and the elevation angles E_1 and E_2 known, the maximum usable frequencies for San Diego to San Bernardino and San Diego to Sacramento can be calculated. Using the MUF equation, the results are

- MUF San Diego to San Bernardino = 5.18 MHz
- MUF San Diego and Sacramento = 8.72 MHz

In this example, San Diego can easily communicate with Sacramento since the 8.72 MHz MUF for this path is greater than the net's 7.23 MHz operating frequency. However, a problem exists with the San Diego to San Bernardino path. For this path the net's 7.23 MHz operating frequency is greater than the 5.18 MHz MUF. At an E_1 elevation angle of 75°, the signal from San Diego to San Bernardino penetrates the ionosphere and is lost to outer space as illustrated in Figure 22. Consequently, in this example, San Diego is only able to communicate with Sacramento.

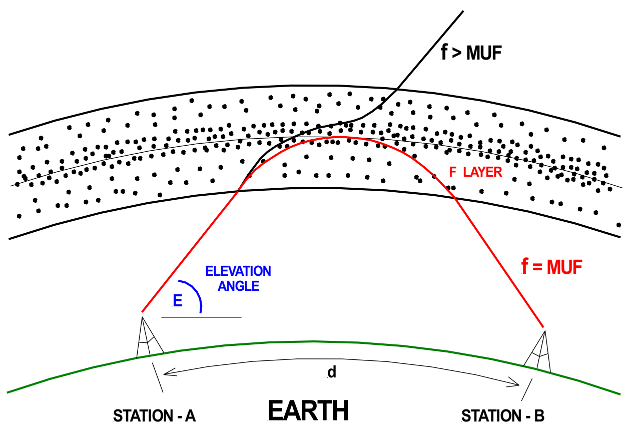

Figure 22 Signal with f > MUF is lost to outer space (source: author)

17.4.4 Frequency of Optimum Transmission (FOT)

Working at the Maximum Usable Frequency is literally "living on the edge". Small changes in critical frequency and other ionospheric parameters cause the MUF to be in a continuous state of change. At one moment a MUF signal refracts back to Earth like the red trace in Figure 22. At the next moment it is lost to outer space (the red-black trace). Consequently, signals transmitted at the MUF often fade in and out. Communications is generally more stable by transmitting at a slightly lower frequency than the MUF. It is generally accepted that the Frequency of Optimum Transmission (FOT) is 80 to 85% of the MUF.

In the example above, the MUF for the San Diego to Sacramento path is 8.72 MHz. The FOT for this path is approximately 7.3 MHz, explaining the excellent stable communications between San Diego and Sacramento predicted in the example.

FOT can be directly read from Hourly Area Prediction (HAP) charts provided by the Australian Bureau of Meteorology. HAP charts predict the optimum frequency for communicating between a specified city (the Base City) and a selected distant station. Los Angeles, California is the base city for the HAP chart shown in Figure 23.

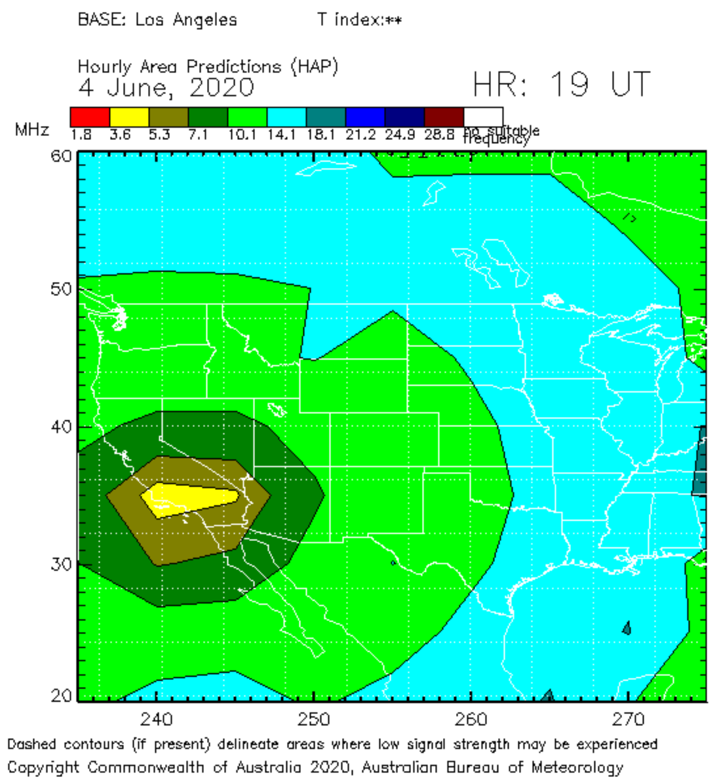

Figure 23 HAP Chart (source: Australian Bureau of Meteorology)

The vertical axis of the chart is degrees Latitude. The horizontal axis is degrees East Longitude (measured eastward around the Earth from the Prime Meridian). The color band at the location of the base city (yellow in this example) is by definition the critical frequency f_c at the time the chart was produced.

The HAP chart color bands represent the recommended HF frequency (FOT) for communicating between the base city and a selected distant location for a given date and hour. There are several different versions of the HAP chart available. The colored regions shown in the Figure 23 HAP chart represent different amateur radio frequency bands. For example, 30 meters (10.1 MHz) was the FOT for communicating between Los Angeles, California and Portland, Oregon at 1900 UT (noon local time) on June 4, 2020. The critical frequency at that time (during solar minimum) was around 4.5 MHz.

A HAP Chart is an estimate of the current FOT. Remember that FOT < MUF, specifically

$$FOT \equiv 0.825\, MUF$$

Consequently, for a given FOT shown in a HAP chart, the corresponding

$$MUF \approx 1.2\, FOT$$

In Figure 23 the MUF for Los Angeles to Portland was approximately

$$MUF \approx 1.2\, FOT = 1.2(10.1\, MHz) = 12.12\, MHz$$

Since FOT is simply an estimate, it is possible that communications between Los Angeles and Portland could have been conducted on 20 meters (14.2 MHz). However, it is unlikely that Portland could be reached on 17 meters (18.1 MHz).

A HAP Chart is a starting point in selecting a frequency band.

The FOT shown on the Figure 23 HAP chart for Los Angeles to Portland is 30 meters. Note that this is the highest frequency band for dependable communications from Los Angeles to Portland. Any frequency band lower than the FOT could also be used. For example, 40 meters could be used for communicating between Los Angeles and Portland. However, the 40 meter path could encounter multi-path interference and deep D Layer absorption not present on 30 meters. In general, the best communications between two stations is obtained using the highest possible frequency, i.e., the FOT frequency.

The HAP chart in Figure 24 could have been used to predict communication conditions between San Diego, Sacramento, and San Bernardino during the 6/3/2020 CESN. San Diego is not in the HAP chart data base so the map in Figure 24 is centered on the next closest large city (Los Angeles). The map must be visually shifted downward to be centered on San Diego. When this is done, the FOT to Sacramento is 40 meters (dark green region of the chart). Thus, San Diego should be able to hear Sacramento during the CESN net. The FOT to San Bernardino is 80 meters (the yellow region of the chart). Consequently, San Diego would not be able to hear San Bernardino on the 40 meter CESN net on this particular day. Instead, San Bernardino CESN traffic would have to be relayed to San Diego through Sacramento.

HAP charts, and instructions for using them, are found under the “Tools” tab of the www.skywave-radio.org website.

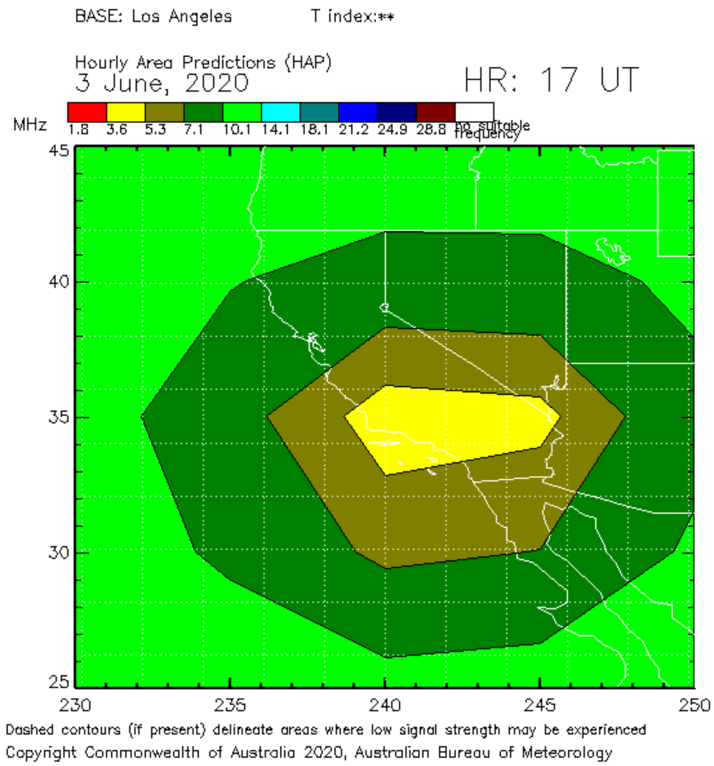

Figure 24 HAP chart for the CESN net on 6/3/2020

17.5 Maximum Usable Angle

Maximum Usable Angle (MUA) is the highest angle signal that can be transmitted at an operating frequency of f_o and still be refracted back to Earth when the critical frequency is f_c. Maximum usable angle E_M is illustrated by the red trace in Figure 25. Transmitting at an angle greater than MUA (the blue trace) will cause the signal to penetrate the ionosphere and be lost to outer space.

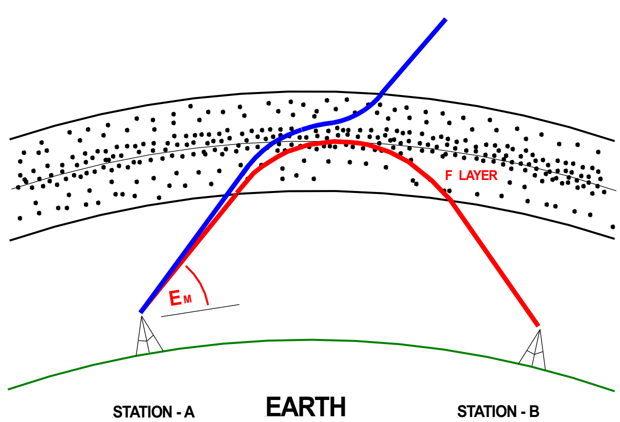

Figure 25 Maximum Usable Angle (source: author)

In equation form

$$MUA = E_M = \sin^{-1}\left(\frac{f_c}{f_0}\right)$$

For example, the MUA when operating late at night on 80 meters (f_0 = 3.8 MHz) with a critical frequency of f_c = 3.0 MHz is

$$MUA = E_M = \sin^{-1}\left(\frac{f_c}{f_0}\right) = \sin^{-1}\left(\frac{3.0\ MHz}{3.8\ MHz}\right) = 52°$$

Under these conditions, signals transmitted at an elevation greater than 52° will be lost to outer space. This could be a serious problem!

The radiation pattern for a typical 80 meter Inverted-V antenna $\frac{1}{8}\lambda$ above ground is shown in Figure 26.

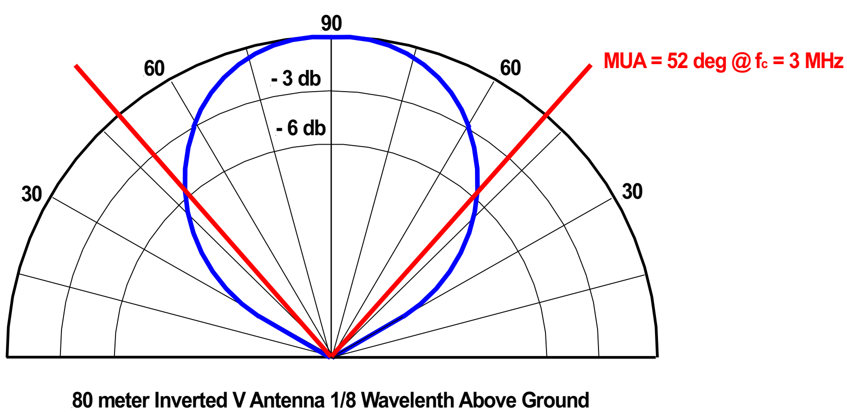

Figure 26 Radiation pattern for an 80 Inverted-V antenna (source: author)

Nearly all of the energy from the antenna is radiated at an angle greater than 60°. This generally is not a problem. The antenna is an excellent Near Vertical Incident Skywave (NVIS) antenna for local and regional communications, particularly for emergency communications work. However, on this particular night there is a problem. The MUA = 52°. Virtually all of the energy radiated by the antenna is lost to outer space. The antenna stops working late at night!

An 80 meter vertical antenna is required for operation throughout the night during years of solar minimum when the critical frequency is low. Figure 27 shows the radiation pattern for an 80 meter vertical antenna. This is an excellent night time 80 meter antenna. Nearly all of the antenna's radiation is at an angle below a MUA = 52°. In fact, the antenna works well all the way down to a critical frequency of 1 MHz (MUA = 15°). As illustrated in Figure 28, during solar minimum the critical frequency can easily get down to 2 MHz at night and at times down to even 1 MHz.

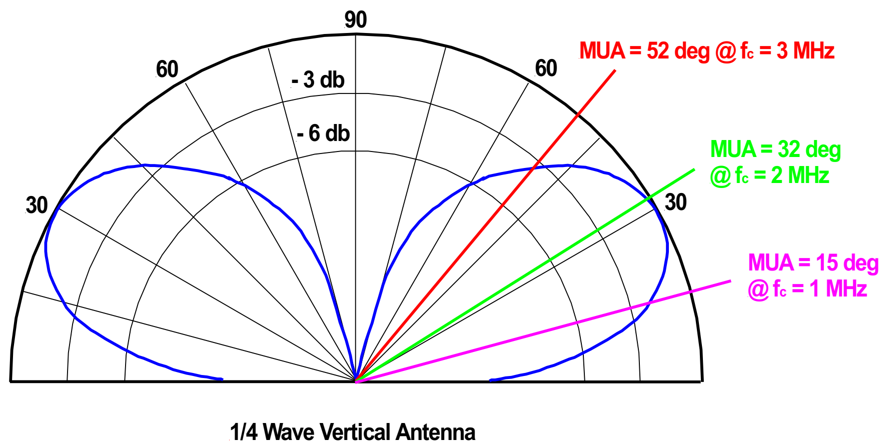

Figure 27 Radiation pattern for an 80 vertical antenna (source: author)

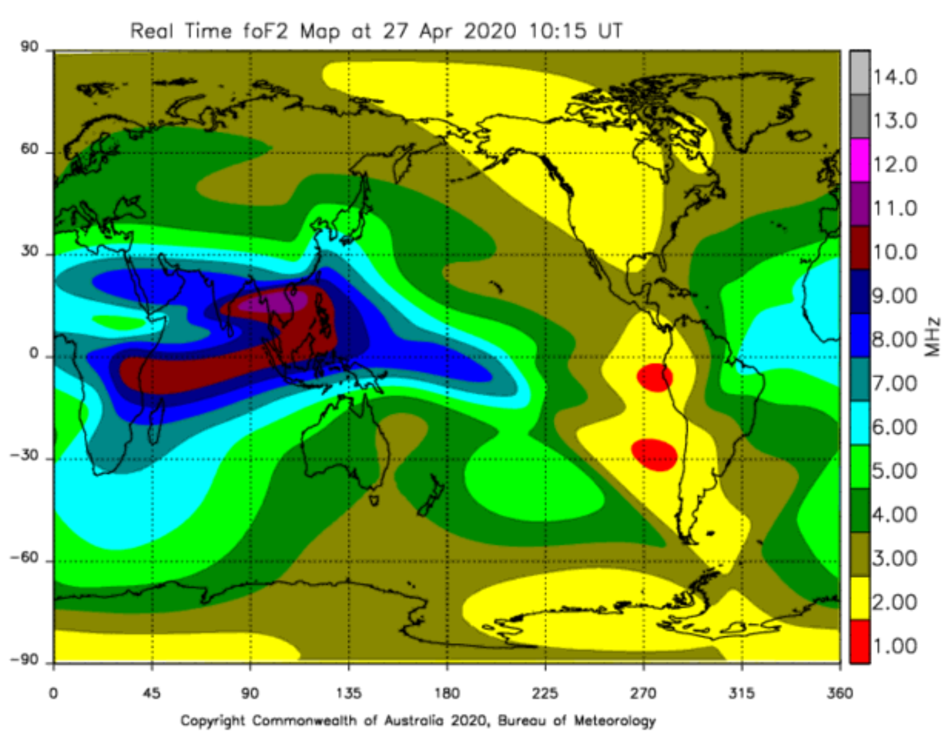

Figure 28 Typical critical frequencies during solar minimum

17.6 High And Low Propagation Paths

In Figure 29, increasing the elevation angle E shortens the distance transmitted in a single hop from point 1 to point 2 to point 3. Increasing E a little more causes a strange thing to happen. Instead of the distance becoming shorter, it becomes dramatically longer, reaching points 4 and 5 instead. Increasing E slightly more causes the signal to penetrate the densest part of the ionosphere and be lost to outer space as illustrated by Ray 6.

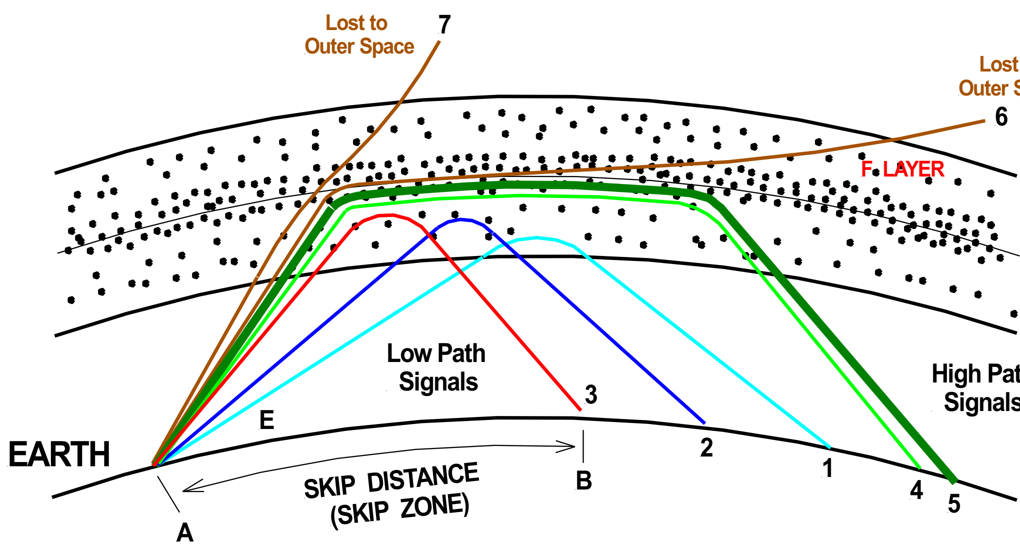

Figure 29 High and low propagation paths (source: author)

The elevation angle for Ray 5 (E_5) is thus the Maximum Usable Angle since any angle greater than E_5 will cause the transmitted signal to penetrate the ionosphere. The difference between E_5 and E_3 (the elevation angle for Ray 3) is very small. Ray 3 is particularly significant in that it is:

- The shortest possible ray,
- The ray at which the high and low paths coincide, and
- The ray having a relatively strong stable signal.

Consequently, E_3 is usually defined as the MUA.

Amateur radio antennas do not radiate narrow "laser beams" of energy. Instead, amateur radio antennas are relatively crude devices radiating energy over a wide range of elevation angles, as illustrated in Figure 30. Consequently, the various propagation paths 1 through 7 in Figure 29 all occur simultaneously resulting in a complex array of distant stations that can be reached.

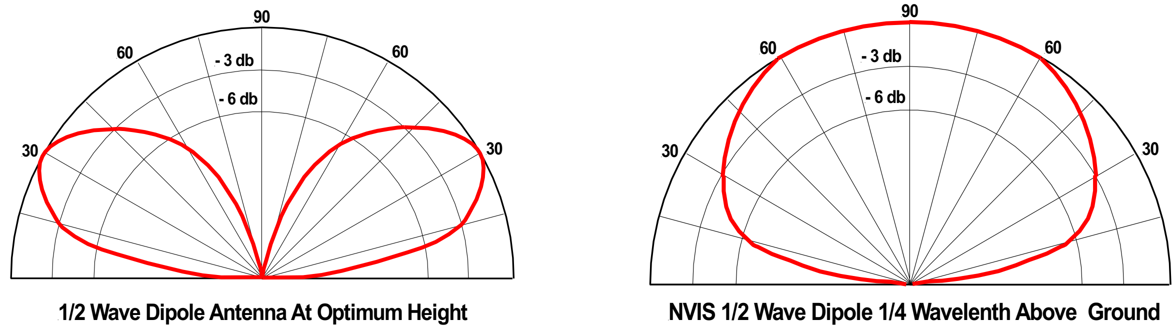

Figure 30 Amateur radio antennas illuminate large segments of the sky (source: author)

17.7 Skip Distance

In Figure 31 Ray 3 is the shortest possible path for a signal transmitted from Point A. Increasing E_3 slightly increases the hop distance to point 4. Decreasing E_3 slightly also increases the hop distance, this time to point 2. Because of this phenomena, Station A can not transmit a signal to any locations closer than Station B.

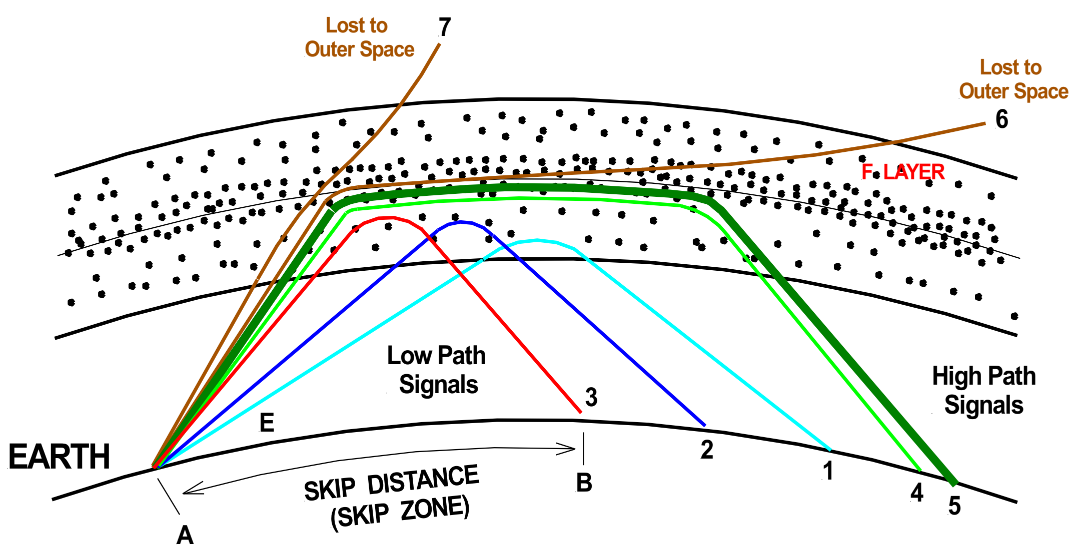

Figure 31 Skip distance (source: author)

The distance between Station A and Station B is defined as the skip distance. Stations closer to Station A (in the Skip Zone) can not be reached or heard, they are “skipped over”.

17.7.1 Skip Distance Charts

Skip distance charts, like the one shown in Figure 32, can be created for each frequency band. The chart provides an estimate of skip distance based on the current critical frequency F_c.

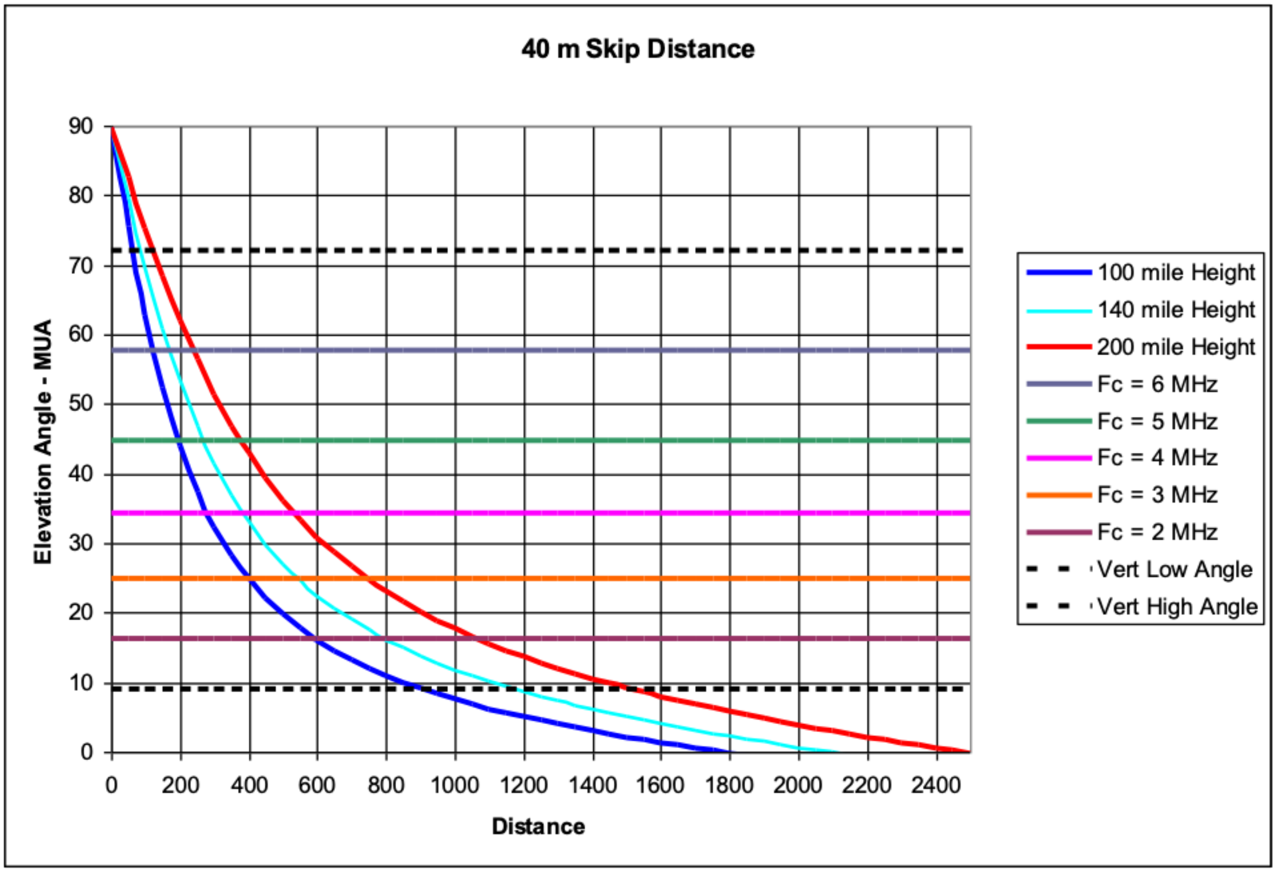

Figure 32 Skip Distance Chart (source: author)

The **Red** trace is the distance vs elevation angle for a hypothetical signal refracted back to Earth at an altitude of 200 miles, near the top of the F Layer. The **Blue** trace is the distance vs elevation angle for a signal refracted back at altitude of 100 miles, near the bottom of the F Layer. The horizontal lines show the Maximum Usable Angles for various critical frequencies F_c .

The shortest skip distance possible is the distance at which the lowest altitude refraction point in the ionosphere, the blue curve, intersects the current critical frequency F_c. This is the closest station that can be reached. On 40 meters at F_c = 3 MHz (the orange horizontal line) the shortest skip distance is about 400 miles. This is a rough estimate. However, under these circumstances (F_c = 3 MHz) one can be relatively sure that the actual skip distance is somewhere between 400 and 750 miles, represented by the intersection of the orange horizontal line with the blue and red curves. An estimate of around 600 miles would probably be fairly close. One can also be pretty sure that stations 300 miles away will be deep in the skip zone and can not be reached or heard.

As an example, we can assume that in mid afternoon the 40 meter critical frequency is around 5 MHz (dark green horizontal line) producing a MUA of about 45°. The resulting skip distance is in the neighborhood of 200 to 400 miles. In the evening, as the critical frequency F_c drops to around 3 MHz, the MUA decreases to about 25° producing a skip distance of roughly 600 miles. That is, "the skip goes long" as night approaches, a common phrased used on the amateur radio bands.

In the previous California Emergency Services Net (CESN) example, the critical frequency at net time was 5 MHz. Using the 40 meter Skip Distance chart in Figure 32, the skip distance at this particular critical frequency is 200 to 350 miles. San Diego, located 97 miles from San Bernardino, is in San Bernardino's skip zone. San Diego can not hear San Bernardino. However, San Diego is 473 miles from the primary net control station in Sacramento. San Diego is well outside Sacramento's skip zone and can hear Sacramento clearly.

In general, during the CESN net Sacramento transmits net traffic to Southern California, including to alternate net control in San Bernardino. San Bernardino then retransmits the net traffic to Northern California, to places like Redding that are in Sacramento's skip zone. Stations in Central California may be a problem. Some of these stations could be in the skip zones of both Sacramento and San Bernardino. In this situation San Bernardino would likely direct Redding to forward traffic to the Central California stations. The point of this example is that all of this planning can be done **prior** to the net by using the Figure 32 Skip Distance chart in conjunction with the current critical frequency, insuring a very successful net.

Skip Distance charts for 80, 40, 30, and 20 meters are available under the "Tools" section of the www.skywave-radio.org web site.

17.7.2 Reachable Stations

In Figure 33 the range of stations reachable in a single hop is bounded on top by the current critical frequency line, the dark blue curve to the left, the red curve to the right, and on the bottom by the lowest angle at which the transmitting antenna can radiate a signal. At $F_c = 3$ MHz, stations that can be reached lie in range from 400 to 1400 miles. At $F_c = 5$ MHz, stations that can be reached range from 200 to 1400 miles. Increasing the critical frequency shortens the skip zone but does not affect the furthest station that can be reached in a single hop. The furthest station that can be reached is limited by the transmitting antenna. The left antenna illustrated in Figure 30 is an excellent long distance DX antenna able to transmit signals down to an angle of 15°. In most cases the lowest angle at which an antenna can radiate energy is around 15 to 20°. In Figure 33 the longest hop possible for an antenna radiating at 10° is around 1,400 miles. Most antennas can not reach this distance in a single hop.

It is important to note that this discussion relates to stations that can be reached in a single hop. Multiple hops are used to reach stations further away.

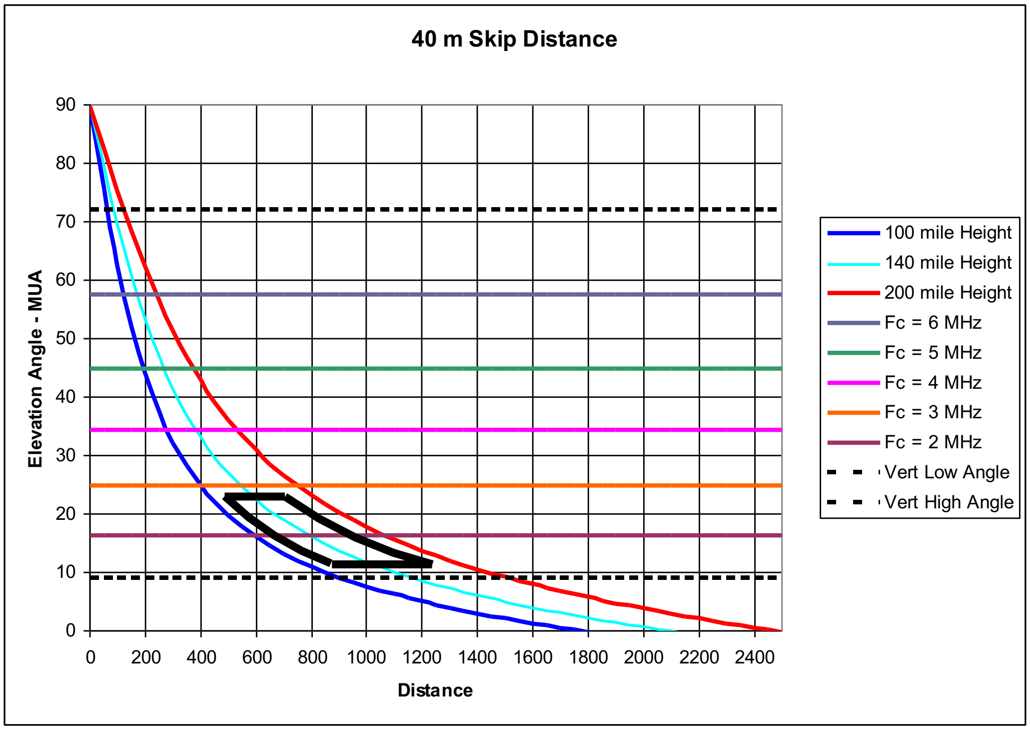

Figure 33 Reachable single hop stations (source: author)

Stations just beyond the skip zone are typically strong. At this distance the high and low path rays merge increasing signal strength. Also, the elevation angle at this distance is relatively high, nearly equal to the MUA, meaning that the signal passes through the D Layer quickly minimizing absorption. A low elevation angle long hop signal spends more time traversing the D Layer than a high angle short hop signal, as illustrated in Figure 34. Consequently, a low angle long hop signal is more likely to be absorbed by the D Layer than a high angle short hop signal.

This is important for emergency communications. Radio stations within a disaster area must communicate with sites outside the disaster region in order to coordinate disaster relief. The best stations to communicate with are those just beyond the skip zone since they will typically be the strongest and most stable stations to reach (assuming that the skip zone itself extends outside the disaster).

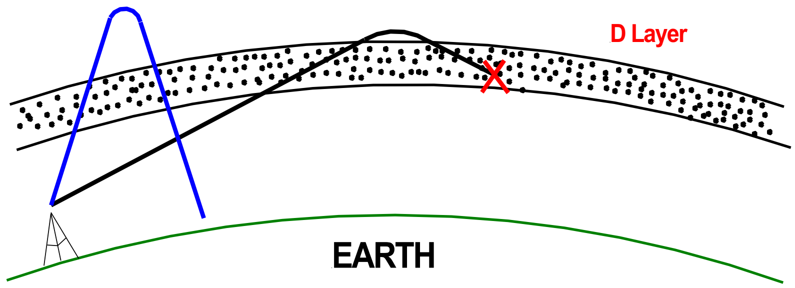

Figure 34 Long hop signals may be absorbed (source: author)

Two or more hops are required to reach many stations that we communicate with. For multi-hop communications, the first hop must occur within the reachable single hop range. In the example above with F_c = 3 MHz and the antenna radiating at a lower limit of 15°, the first hop must occur within a range 400 to 1100 miles from the transmitting station.

17.7.3 HAP Charts and Skip Distance

The HAP chart for 1700 UT June 4, 2020, centered on Los Angeles, California is shown in Figure 35. In this figure the yellow region at the center of the chart represents the critical frequency f_c at that time, approximately 4 MHz. The brown area is the maximum usable frequency (5.3 MHz) for communicating in southern California. Attempting to communicate with stations in the brown area on 40 meters will result in the 40 meter signals penetrating the ionosphere and being lost to outer space. The closest stations that can be contacted by Los Angeles on 40 meters lie along the boundary between the brown and dark green regions. This boundary is the 40 meter skip distance for Los Angles while the brown area is the 40 meter skip zone.

Knowing this, the HAP chart in Figure 35 can be used to calculate the 40 meter skip distance for Los Angeles at 1700 UT on June 4, 2020. Los Angeles is located at a Longitude of 241.73°. The boundary between the brown and dark green regions east of Los Angeles is at a Longitude of about 248°. The 40 meter skip distance for Los Angeles is thus

$$Skip\ Distance = 6.3° = 248° - 241.7°$$

Expressing skip distance in degrees Longitude is not very useful for us. We want the skip distance measured in miles. The procedure for converting from degrees Longitude into miles is as follows.

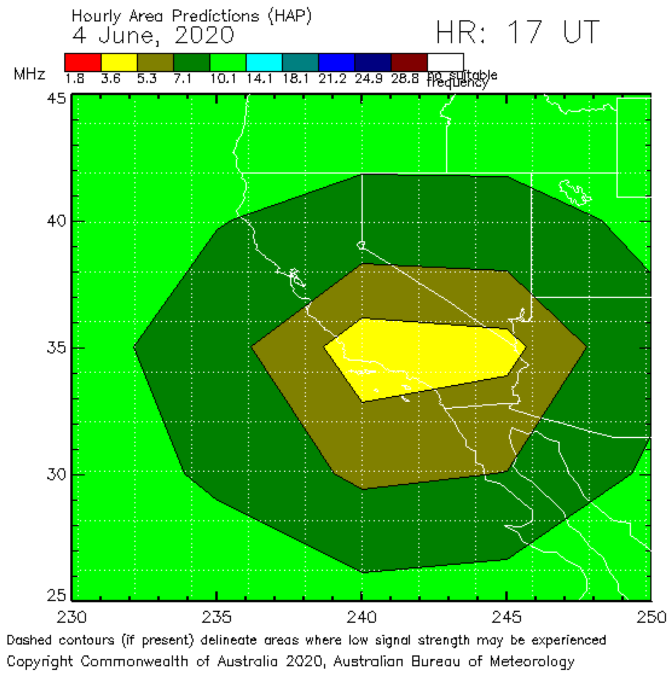

Figure 35 HAP Chart June 4, 2020

The distance in miles of one degree Longitude is

$$D_{1^\circ Long} = D_{E1^\circ Long}(\cos d_{Lat})$$

where

$D_{1^\circ Long}$ = Distance in miles of 1° Longitude at the transmitting site (Los Angeles)

$D_{E1^\circ Long}$ = Distance in miles of 1° Longitude at Earth's equator = 69.172 miles

d_{Lat} = Latitude in degrees of the transmitting site (Los Angeles = 34.05°)

Skip distance in miles is then equal to

$$S = [L_T - L_R]\, D_{E1^\circ Long}(\cos d_{Lat})$$

The 40 meter skip distance in Figure 35 is

$$S = [L_T - L_R]\, D_{E1^\circ Long}(\cos d_{Lat}) = [248^\circ - 241.7^\circ](69.172\,)\cos 34.05^\circ$$

$$S = 361\; miles$$

Note: to simplify the math this skip distance is calculated in the east-west direction. Since the skip zone is roughly circular, this skip distance is assumed to apply in the north-south direction as well.

The 40 meter skip distance shown in Figure 36 for a critical frequency f_c = 4 MHz is between 290 and 510 miles. The 361 mile skip distance calculated from the HAP chart is within this range.

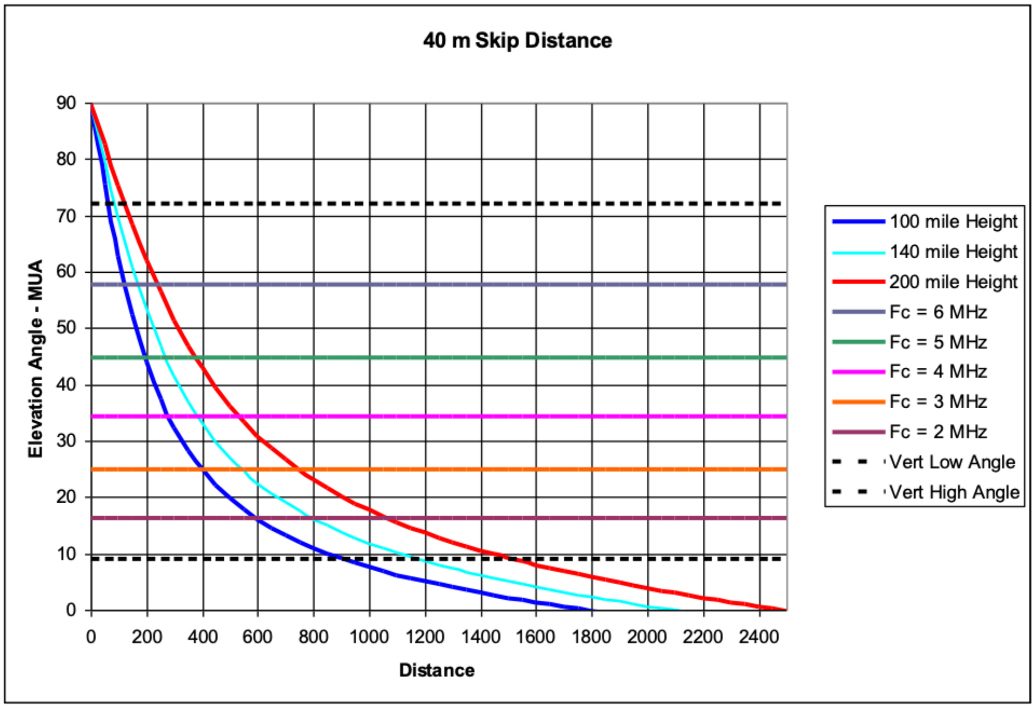

Figure 36 40 meter Skip Distance Chart (source: author)

17.7.4 Skip Distance Determined By Antenna

Skip distance will be determined by your antenna IF the maximum radiated angle of your antenna (MRA) is less than the MUA determined by the current critical frequency f_c. For example, at a critical frequency of f_c = 6 MHz the 40 meter MUA is around 60^o producing a skip distance of about 150 miles, as illustrated in Figure 36. If a vertical transmit antenna is being used, the antenna's maximum radiation angle MRA is probably about 45^o as shown in Figure 37. Referring back to Figure 36, the skip distance for an elevation angle of 45^o is at around 300 miles. Stations closer than 300 miles will be skipped over by the vertical antenna. Vertical antennas are long distance antennas not suited to close in Near Vertical Incident Skywave (NVIS) emergency communications.

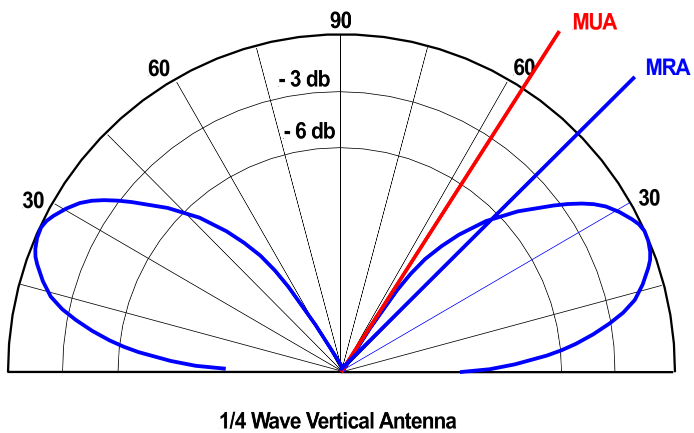

Figure 37 Vertical antenna elevation diagram (source: author)

17.8 Minimum Critical Frequency

Minimum Critical Frequency is useful in determining band openings. Minimum critical frequency $\mathbf{f_{cm}}$ is the lowest critical frequency capable of supporting communications from a particular type of antenna at a given operating frequency. In equation form minimum critical frequency is

$$f_{cm} = f_o \sin E_a$$

where

f_o = The operating frequency

E_a = The elevation angle of signals radiating from the transmit antenna.

Minimum critical frequency depends on the characteristics of the transmitting antenna, as illustrated in the following examples.

Figure 38 is the radiation pattern for a horizontal half wavelength dipole installed at its optimum height ($1/2\ \lambda$ above ground). At an operating frequency of f_o = 14.2 MHz the minimum critical frequency f_{cm} at various elevation angles E_a is

- **$\mathbf{f_o}$ = 14.2 MHz**
- f_{cm} = 7.1 MHz @ Ea = 30 deg
- f_{cm} = 4.9 MHz @ Ea = 20 deg
- f_{cm} = 3.7 MHz @ Ea = 15 deg (at the -3 db point on this antenna pattern)

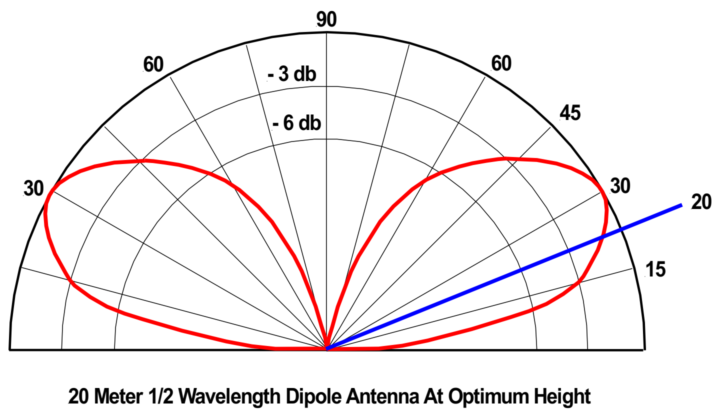

Figure 38 Minimum critical frequency 20 meter example (source: author)

For this particular type of antenna on 20 meters, f_{cm} is around one half of the operating frequency f_o at an angle of 30°. At an angle of 14.5°, f_{cm} is about one quarter of the operating frequency. On 10 meters (28 MHz) at an angle of 20° f_{cm} is 9.6 MHz. In very general terms, for this type of antenna, a band is probably open (usable) if f_{cm} is around 1/3 to 1/2 of the operating frequency. This assumes that others on the band are using similar antennas.

The minimum critical frequency for high elevation (NVIS) antennas is different. For the 80 meter antenna shown in Figure 39, the lowest practical elevation angle E_a is probably around 60°. At this angle, and an operating frequency f_o = 3.8 MHz, the minimum critical frequency f_{cm} = 3.3 MHz. Thus, the minimum critical frequency f_{cm} required to support a high elevation angle antenna is about 87% of the operating frequency f_o. This compares to around 33 to 50% of the operating frequency for antennas with lower angle radiation patterns.

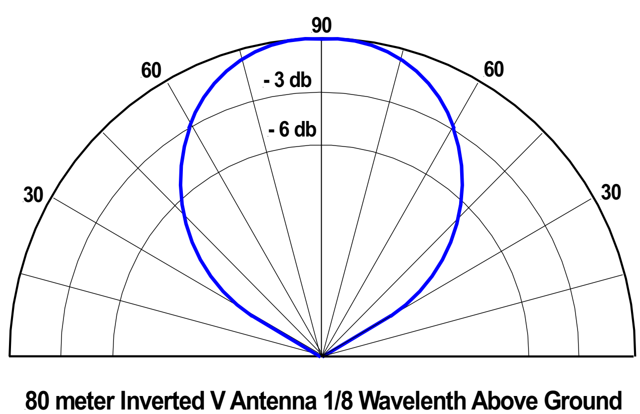

Figure 39 Minimum critical frequency 80 meter example (source: author)

On May 28, 2020 at 17:15 UT (during solar minimum) the critical frequency over the U.S. lower 48 states was 5.0 MHz. This is below the 20 meter 30º elevation angle minimum critical frequency of 7.1 MHz and equal to the minimum critical frequency for a 20º elevation angle. So, there were probably not many people on 20 meters on this date. If fact, the band may not have been open.

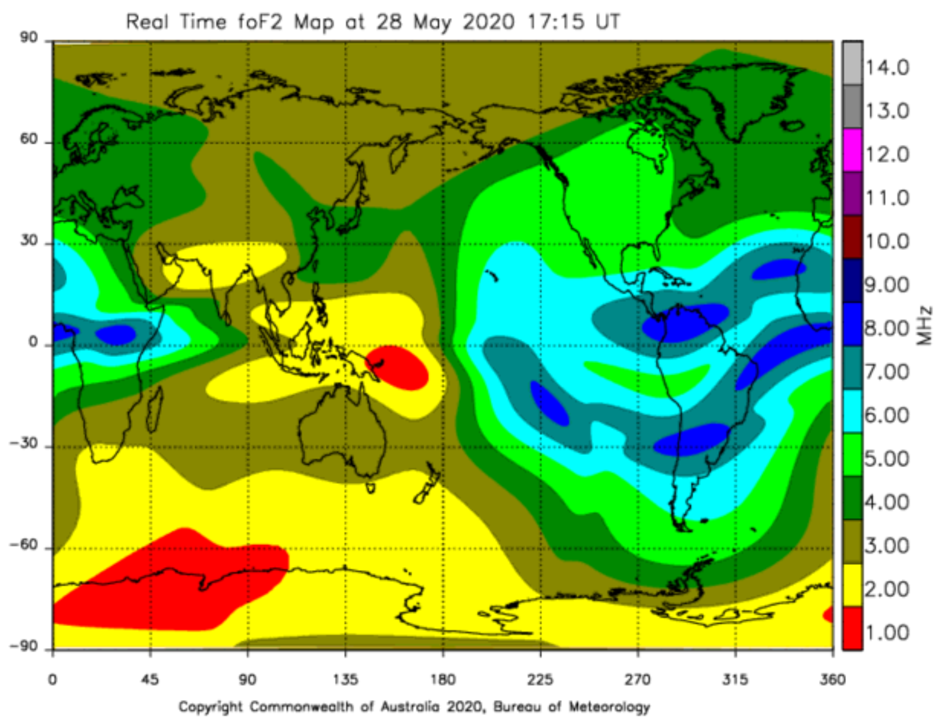

Figure 40 Critical Frequency Map for May 28, 2020 at 17:15 UT

Figure 41 shows the radiation pattern for a 40 meter NVIS antenna 1/4 wavelength above ground. For this antenna the lowest practical elevation angle E_a is probably around 30º.

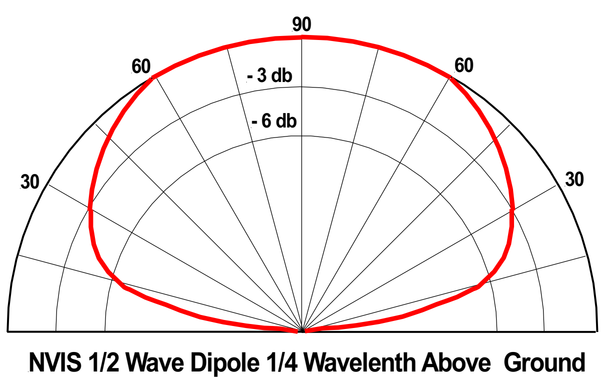

Figure 41 Antenna pattern for an NVIS 40 meter antenna (source: author)

The minimum critical frequency for this antenna operating at 7.2 MHz with an elevation angle E_a = 30 degrees is

$$f_{cm} = f_o \sin E_a = (7.2) \sin 30 = 3.6 \; MHz$$

Thus the 40 meter band should have been open on May 28, 2020 since the actual critical frequency for that date (f_c = 5.0 MHz) was greater than the minimum required frequency (f_{cm} = 3.6 MHz). As a general rule, a band will be open if

$$f_c \geq f_{cm} \quad for\ that\ frequency\ band$$

Just because a band appears to be open, that is $f_c \geq f_{cm}$, does not mean that anyone is operating on that band. For example, on one occasion the 15 meter band was completely dead. There was not a single station on 15 meters ……… except for one, W1AW coming into southern California loud and clear transmitting a morse code practice session.

17.9 Propagation Modes

Propagation modes describe the single and multi-hop paths through the ionosphere that radio signals traverse in route from the transmitting to the receiving radio station. The various propagation modes include:

- F and E propagation modes,
- Sporadic E propagation,
- Multi-path propagation,
- Backscatter,
- Great Circle propagation,
- Gray Line propagation,
- Equatorial Sporadic E propagation,
- Transequatorial Propagation (TEP), and
- Ionospheric Ducting

17.9.1 F and E propagation modes

We usually assume that long distance skywave communications consists of multiple hops through the F region of the ionosphere as illustrated in Figure 42. While this is the model that we assume, the actual propagation of a radio signal through the ionosphere can be more complex. Refractions occur in the E region of the ionosphere as well as in the F region. This leads to a variety of possible propagation modes.

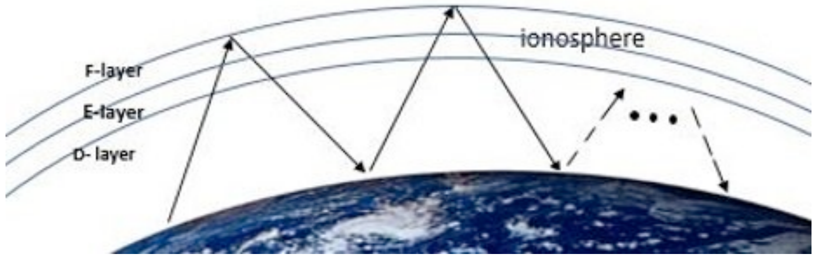

Figure 42 Multiple hop propagation through the ionosphere (source: ResearchGate)

In Figure 43 a signal reaches its intended destination in a single hop through the ionosphere's F region. This propagation path includes two passes through the D layer but no ground reflections. Each reflection from the Earth's surface attenuates the transmitted signal. Reflection from poorly conducting arid soil can result in a signal attenuation of 3 db or greater while a signal reflecting from the ocean's surface may only be attenuated 0.5 db.

In Figure 44 a signal requires 2 hops through the lower altitude E region to reach the same destination. This propagation path involving 4 passes through the D layer plus one ground reflection. Signal losses occur with each pass through the D region and reflection from Earth's surface. Consequently, propagation through the E region results in higher losses and a weaker received signal than a single hop through the F region.

As described above, ground reflection losses depend on soil conditions at the ground reflection point. Significant signal loss will occur if ground conditions are poor. The situation is even worse in the western United States. Not only are soil conditions poor (semi- arid), reflection in a mountainous area can result in the signal being scattered to unplanned destinations.

Clearly, we want to minimize the number of hops required to reach a receiving station by transmitting as high as possible into the F region of the ionosphere.

We do have some control over F verses E mode propagation by selecting the proper operating frequency. The maximum usable frequency for the E mode is lower than for the F mode of propagation. In Figure 45 the peak E region critical frequency during the summer is $f_cE = 3$ MHz while the F_2 region critical frequency f_cF_2 is a little over 5 MHz. Calculating the MUFs for both E and F_2 mode propagation at a transmit elevation angle $E = 30^o$ we get

$$[MUF]_E = \frac{f_{cE}}{\sin E} = \frac{3\ MHz}{\sin 30} = 6\ MHz$$

$$[MUF]_{F2} = \frac{f_{cF2}}{\sin E} = \frac{5\ MHz}{\sin 30} = 10\ MHz$$

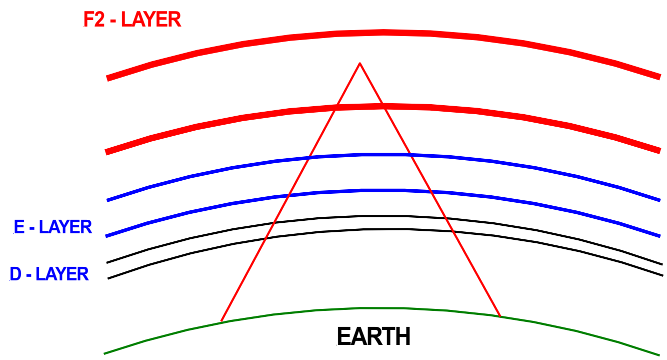

Figure 43 F mode propagation (source: author)

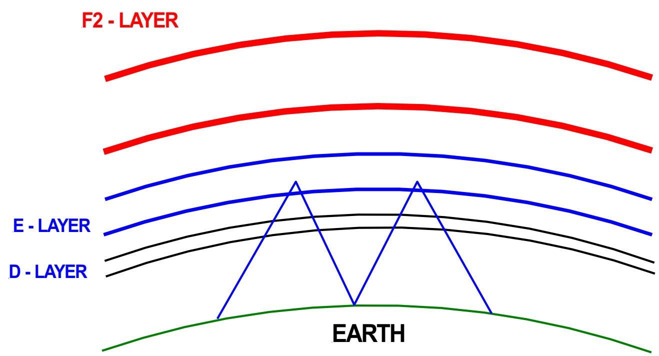

Figure 44 E mode propagation (source: author)

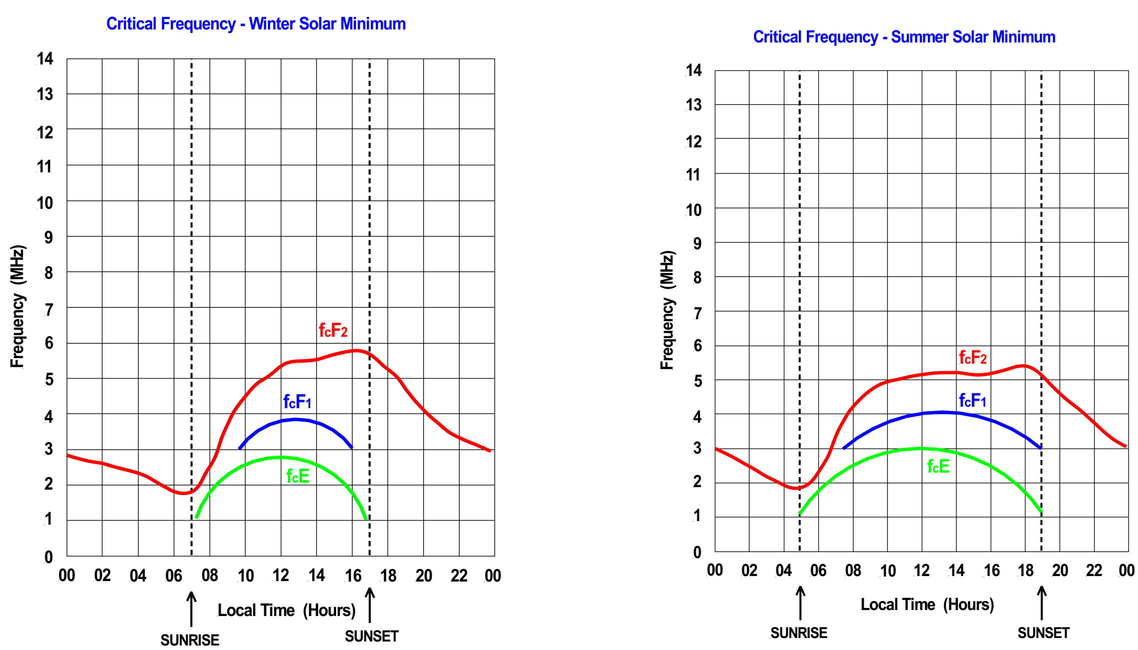

Figure 45 Winter and summer critical frequencies during solar minimum (source: author)

Signals transmitted at a frequency greater than the E region MUF of 6 MHz will pass through the E region without being refracted back to Earth. In this example, transmitting at a 40 meter frequency of 7.2 MHz will result in the signal penetrating the E region and following a multi-hop F mode propagation path. An 80 meter 3.8 MHz signal is below the 6 MHz E region MUF. Consequently, an 80 meter signal will be refracted back to Earth in the E region incurring more hops, more passes through the absorbing D layer, and more ground reflections than a 40 meter signal in route to the same destination. In addition, absorption in the D layer is inversely proportional to frequency squared according to

$$Absorption \propto \frac{1}{f^2}$$

An 80 meter signal will incur 4 times more absorption with each pass through the D layer than a 40 meter signal. A signal requiring two hops through the E region to reach the intended destination will incur one ground reflection loss plus 8 times more D layer absorption than a one hop 40 meter signal (4 times more absorption on each pass plus twice as many trips through the D layer). In this example, we want to operate on 40 meters to deliver the best signal to the receiving site. This illustrates what is becoming a recurring theme. Optimum communications through the ionosphere is achieved by operating at the highest possible frequency.

17.9.2 Sporadic E Propagation

Sporadic E (E_s) formations of abnormally high ionization within the E region (Figure 46) are important because they can reflect HF radio signals at frequencies up to about 100 MHz. They are called sporadic E because they randomly appear in various sizes and shapes, persist for minutes to hours, and occur from one day to the next with little predictability. Sporadic E zones are relatively large structures about 2 kilometers thick with horizontal dimensions stretching hundreds of kilometers. In general, sporadic E appearances seem to have little direct relationship to the ionization processes responsible for the E region itself.

Sporadic E zones often have electron densities far greater than normal E region levels and at times even greater than in the F region. Sporadic E patches can appear opaque to radio waves, reflecting waves that normally would have been refracted high in the F2 layer. This can seriously impact HF radio circuits. Instead of a single hop through the ionosphere multiple hops, with more ground reflections and more passes through the attenuating D region, may be required to reach a destination as illustrated on the left in Figure 47. This can seriously degrade signal levels at the receiving site. Worse yet, the intended receiving location could be missed altogether as occurs on the right in Figure 47. At times sporadic E patches are partially transparent or patchy permitting radio waves to penetrate through the gaps. However, a partially transparent sporadic E patch often leads to weak or fading signals as the sporadic E zone evolves.

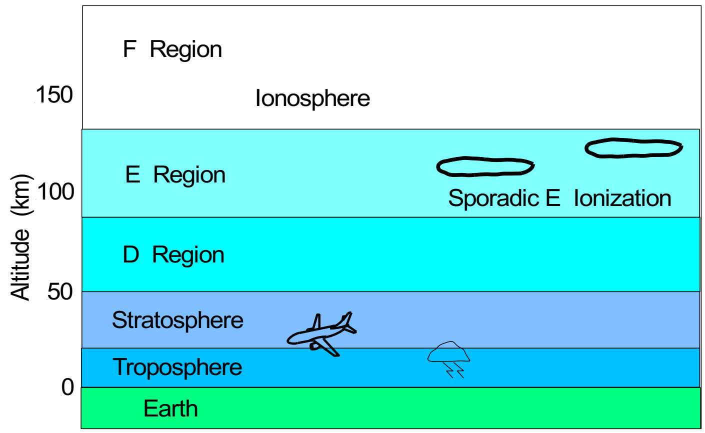

Figure 46 Sporadic E ionization (source: author)

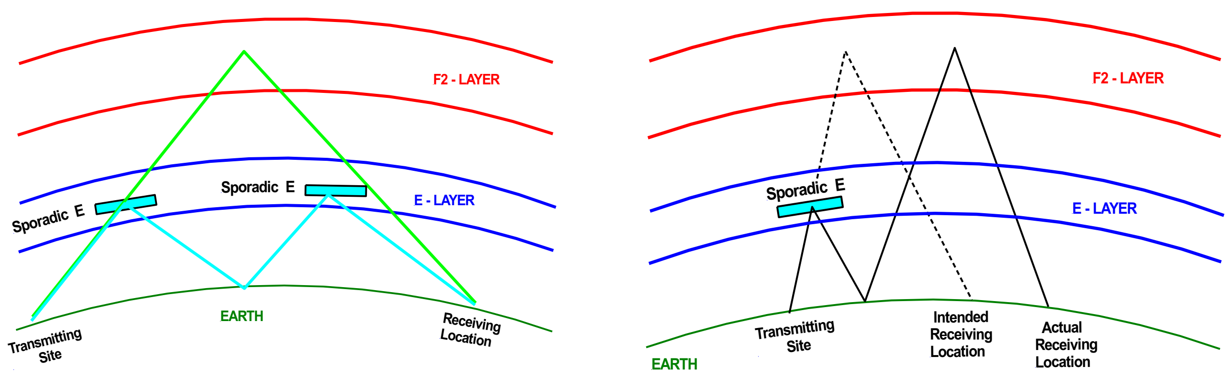

Figure 47 Disruptions in propagations paths due to sporadic E (source: author)

Figure 48 illustrates the important fact that sporadic E reflections can occur on the top side of the E region as well as on the bottom side shown in Figure 47.

Sporadic E zones are particularly strong in the low latitude equatorial region where they are essentially a daytime phenomenon with little seasonal variation. It is believed that they are formed in this part of the world by instabilities in the equatorial electrojet.

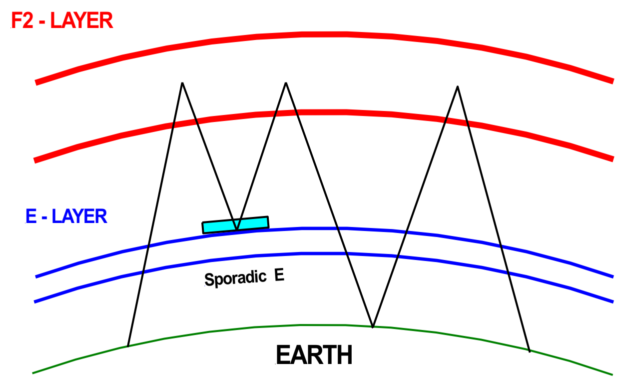

Figure 48 Topside Sporadic E reflection (source: author)

At mid latitudes sporadic E zones tend to be weaker than elsewhere. Their occurrence is subject to diurnal and seasonal variations. They tend to be more prevalent during the summer than in winter and during the day than at night, particularly in mid-morning and near sunset. It is believed that sporadic E patches at mid latitudes form as the result of wind shear in the upper atmosphere in combination with meteoric debris. Enormous numbers of meteors burn up in the E region of the atmosphere (Figure 49). The meteoric debris is largely monatomic metallic ions consisting of iron, sodium, magnesium, and other similar elements. These monatomic ions are relatively small compared to the much larger molecular ions which comprise the E region. Because of their small size, the rate of electron-ion recombination is lower than for molecular ions.

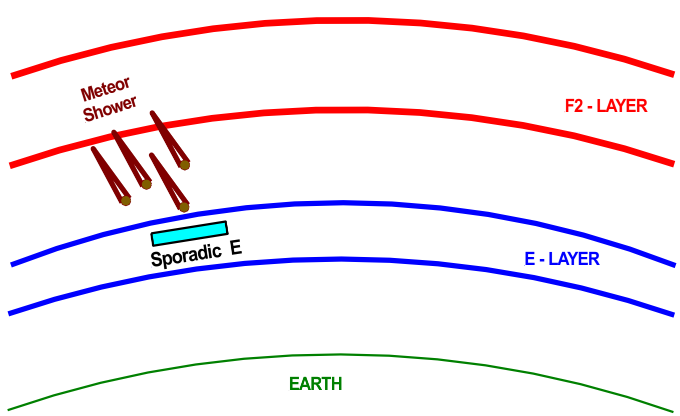

Figure 49 Meteor showers creating sporadic E patches (source: author)

Atmospheric gravity waves, associated with traveling ionospheric disturbances (TIDs), create upper atmosphere high velocity winds that travel in opposite directions at slightly different altitudes (Figure 50). This set of conditions produce what is known as vertical wind shear. Meteoric debris becomes trapped between the wind reversals at locations where the wind velocity tends to be low. Within the pockets of trapped debris, the rate of electron-ion recombination (the rate of electron loss) is lower than elsewhere in the E region. Consequently, relatively high electron concentrations develop in these pockets producing sporadic E patches.

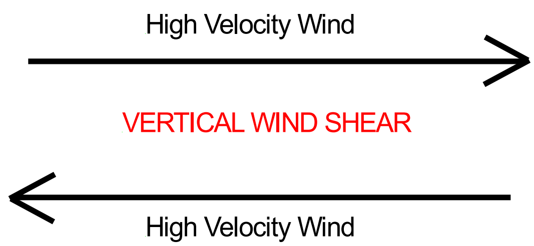

Figure 50 Vertical wind shear (source: author)

At high latitudes sporadic E zones occur mainly at night with little seasonal variance. They are attributed to ionization by incoming high energy charged particles entering the auroral region from the magnetosphere. Clouds of auroral E_s drift westward in the evening and eastward in the early morning at speeds between 200 and 3,000 meters per second, much like the auroral itself. Sporadic E zones within the polar caps have a different characteristic. They are weaker and extend across the polar caps in the form of ribbons in a roughly sunward direction.

17.9.3 Multi-path Propagation

Communications between Los Angeles and Denver, a distance of 800 miles, can occur on 40 meters in either one hop or in two hops of 400 miles each as illustrated in Figure 51.

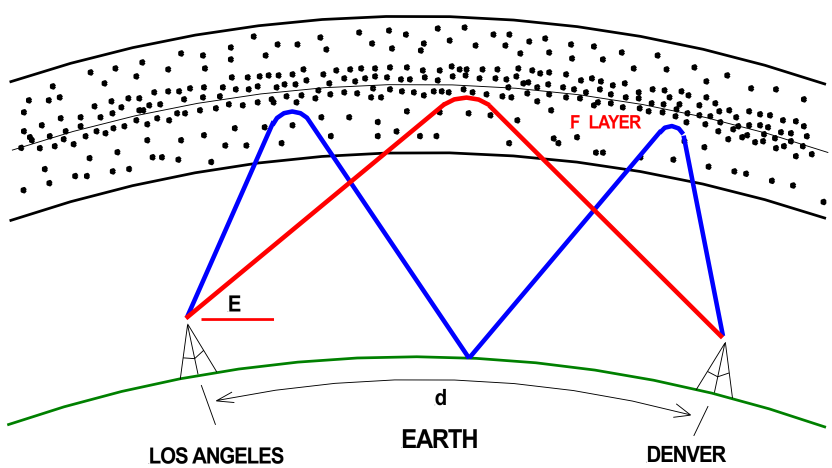

Figure 51 Communications between Los Angeles and Denver (source: author)

The 40 meter skip distance chart in Figure 52 shows that an elevation angle of roughly 20° is needed to reach Denver in one hop while an angle of 38° is required for the double hop path. A critical frequency of 6 MHz supports both propagation paths.

In general, we have relatively little control over the elevation angle at which our antenna radiates. The 40 meter half wavelength dipole shown in Figure 53 radiates at all angles from 20 to 90°. Radiation from this antenna will reach Denver following both the single and double hop propagation paths. There is nothing that we can do about this!

The single and double hop signals will be out of phase when they arrive in Denver since the double hop signal has to travel a longer distance to reach Denver. Consequently, the two signals will interfere with each other reducing the strength of the received signal, in addition to producing distortion and fading.

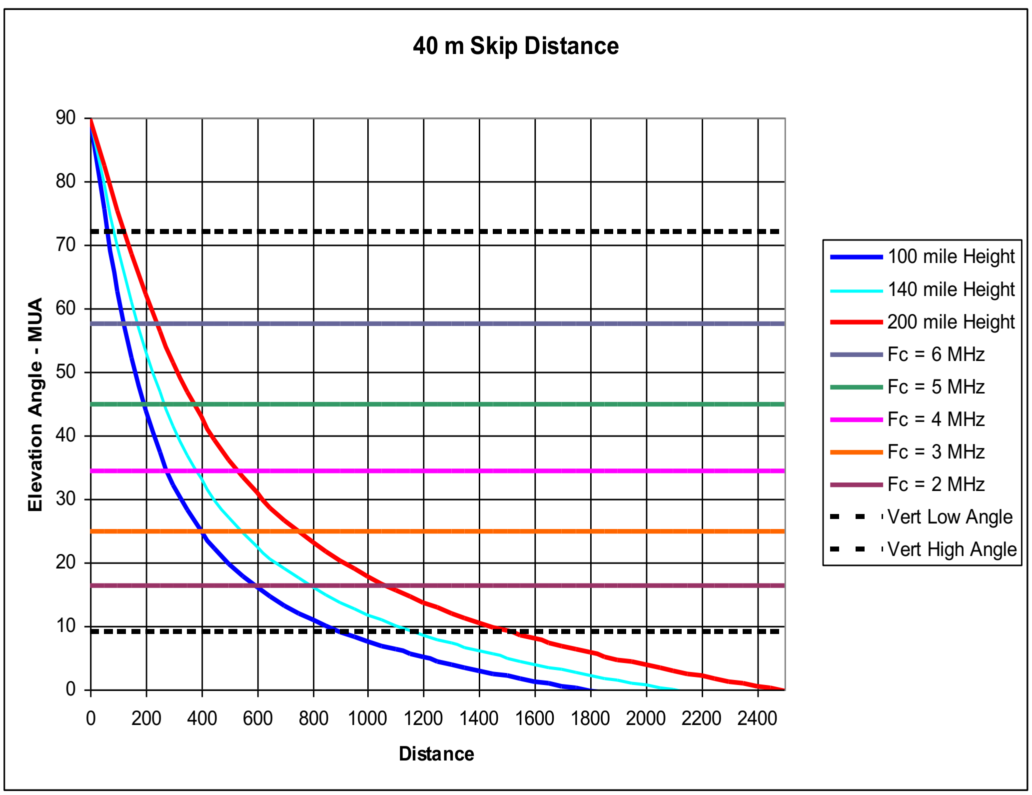

Figure 52 40 meter skip distance chart (source: author)

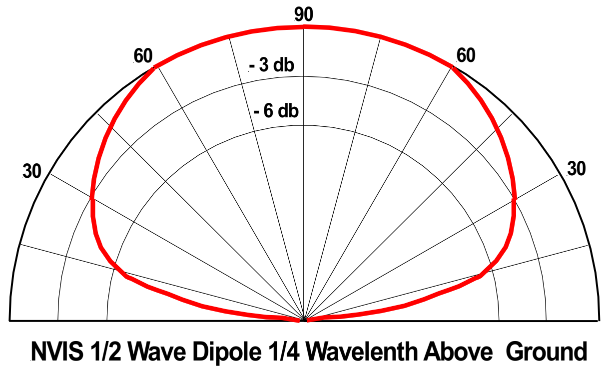

Figure 53 Elevation pattern for an NVIS 40 meter antenna (source: author)

The maximum usable frequency at a critical frequency of 6 MHz and an elevation angle of 20° is

$$[MUF]_{20^\circ} = \frac{f_{cF2}}{\sin E} = \frac{6\ MHz}{\sin 20} = 17.5\ MHz$$

At an elevation angle of 38° the maximum usable frequency is

$$[MUF]_{38^\circ} = \frac{f_{cF2}}{\sin E} = \frac{6\ MHz}{\sin 38} = 9.7\ MHz$$

Consequently, the double and single hop propagation modes are supported on 40 meters since a 7.2 MHz signal is below the MUF for both a 20° and 38° elevation angle.

The 20° elevation angle MUF of 17.5 MHz supports communications with Denver on 20 meters (14.2 MHz), but the 38° elevation angle (MUF = 9.7 MHz) does not. This is illustrated by the 20 meter skip distance chart in Figure 54. In this chart an elevation angle of 38° is above the elevation angles permitted by the 6 MHz critical frequency. Transmitting at an elevation angle of 38° on 20 meters will result in the transmitted signal penetrating the ionosphere and being lost to outer space. Consequently, transmitting at an elevation angle of 20° is the only possible path to Denver on 20 meters. Moving from the 40 meter to the 20 meter frequency band eliminates the Los Angeles to Denver multi-path interference problem.

Finally, the half wavelength 20 meter dipole antenna shown in Figure 55 supports the 20° elevation angle single hop path to Denver better than the 40 meter dipole in Figure 53.

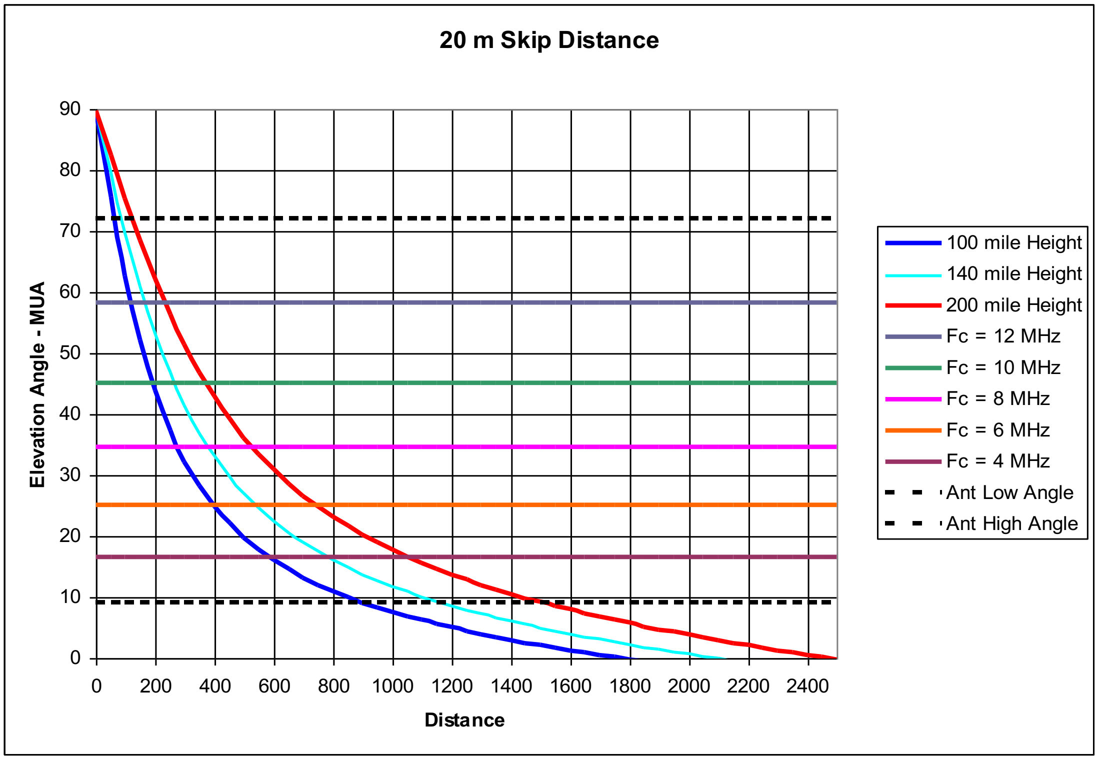

Figure 54 20 meter skip distance chart (source: author)

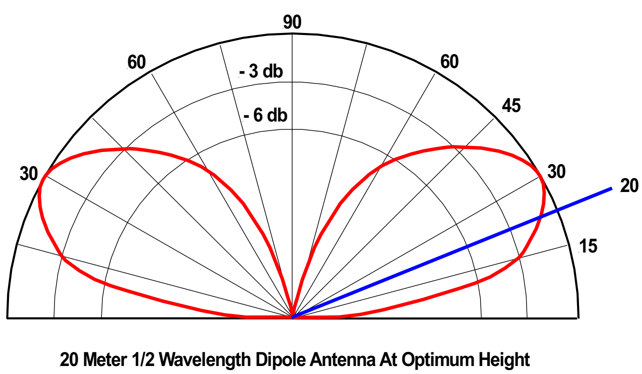

Figure 55 Elevation pattern for a 20 meter 1/2 wavelength dipole antenna (source: author)

This brings us back to a common theme. Optimum communications through the ionosphere is achieved by operating at the highest possible frequency.

17.9.4 Backscatter

Figure 56 shows the first hop of a transmitted signal scattering when it returns to Earth. In this case, part of the scattered signal travels back through the ionosphere ending up in the skip zone. Consequently, some of the stations in the skip zone will hear the transmitting station when normally they would not.

This example brings up another situation. We assume in our simple propagation diagrams that the elevation angle of our transmitted signal remains the same hop after hop as our signal makes multiple reflections from a flat, smooth, horizontal Earth. This over simplification is definitely not the case, particularly for reflections in mountainous areas. The reflection elevation angle may frequently change. This affects hop distances and even the possibility that a high reflection angle could cause the signal to penetrate the ionosphere and be lost to outer space.

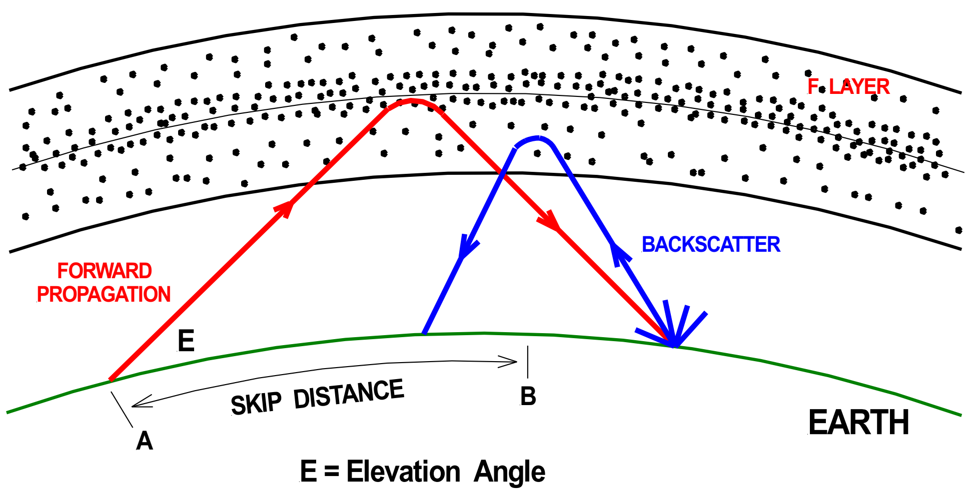

Figure 56 Backscatter propagation (source: author)

17.9.5 Great Circle Propagation

A great circle path is the shortest distance between any two locations on Earth's surface. A great circle is formed by a plane passing through the two points of interest and the center of the Earth as shown in Figure 57. A great circle always divides the Earth in half. Thus, the equator and lines of longitude are great circles. However, lines of latitude are not great circles since they do not cut the Earth in equal halves.

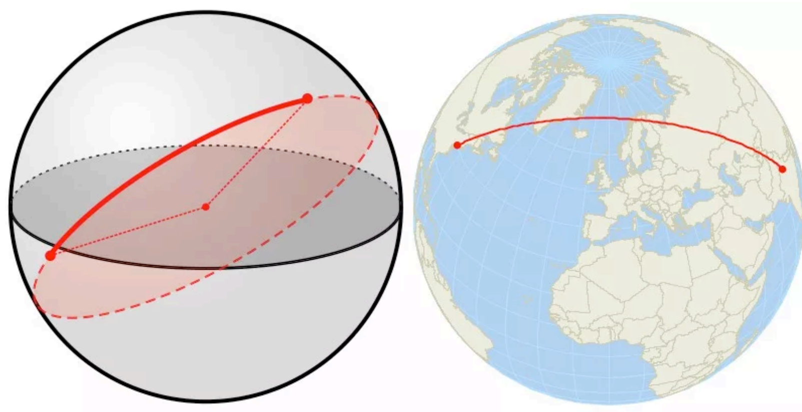

Figure 57 Great Circle path (source: Caliper Corporation)

Normally, radio waves follow great circle paths since a great circle is the shortest distance between the transmitting and receiving stations. However, high latitude ionospheric irregularities including

- Ionospheric troughs,
- High latitude spread F,
- Patches,
- Blobs, and
- Traveling Ionospheric Disturbances,

can seriously alter signal propagation paths. (See the "High Latitude Ionosphere" chapter for an in-depth discussion of these irregularities.) In addition, diverging ordinary and extra ordinary waves in the polar regions complicate the problem further. For example, a signal intended to travel along the great circle path over the polar region from eastern United States to India (Figure 57) could instead end up in Egypt or perhaps Korea. In general, great circle paths become less meaningful for signals propagating through Earth's polar zones.

17.9.6 Gray Line Propagation

In Figure 58 the line dividing night and day is called the terminator. It is also referred to as the gray line and twilight zone. It is actually a fuzzy line due to the bending of sunlight in Earth's atmosphere. Thus, a more appropriate term would be gray zone.

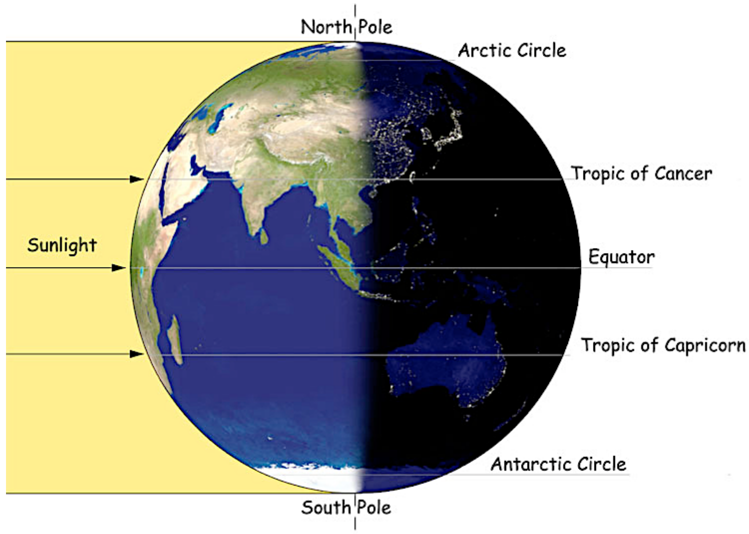

Figure 58 Earth's terminator at equinoxes (source: cseligman.com)

The position of the gray line is constantly changing as the Earth rotates throughout the day and orbits around the Sun during the year. As illustrated in Figure 59, the Earth's axis is always tilted 23.5° with respect to its orbit. In Figure 58, the gray line runs north and south passing through the Earth's geographic poles during the March 21 and September 21 solar equinoxes. During the December and June solstices (December 21 and June 21), the gray line is tilted 23.5° with respect to Earth's axis, as shown in Figures 59 and 60. During the course of a year, the gray line traverses a 47° sector of the Earth north and south of the equator as the Earth orbits the Sun. The width of the gray zone also varies. The transition between night and day occurs quickly near the equator while in the polar zones it occurs more slowly. Consequently, the gray zone is wider at high latitudes than at the equator.

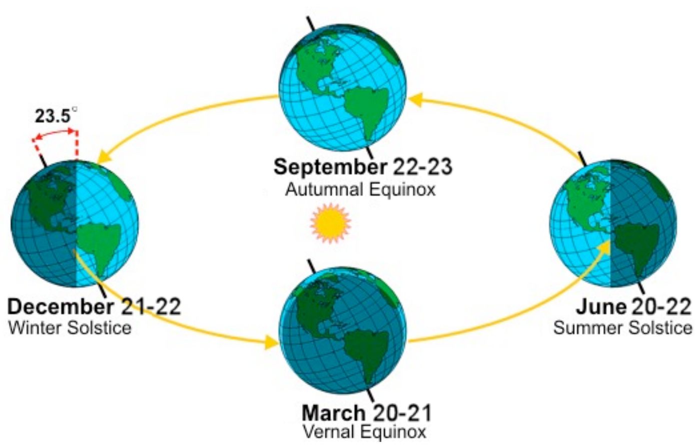

Figure 59 Earth's orbit around the Sun (source: NOAA National Weather Service)

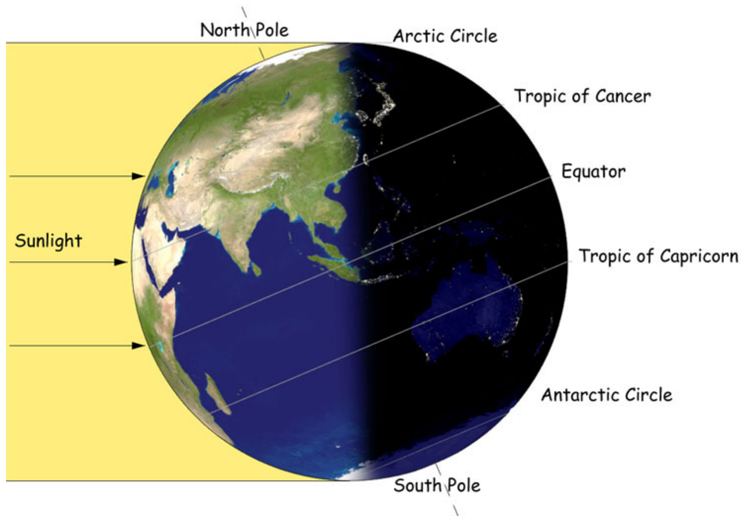

Figure 60 Earth's terminator at solisis (source: cseligman.com)

The ionosphere changes significantly during sunrise and sunset. At sunrise the F region builds rapidly as solar ionization resumes. The D layer also reappears but more slowly. The reverse occurs in the evening. The D layer disappears soon after sunset while the F region slowly declines throughout the night. The delay between the appearance and disappearance of the D region relative to the F layer, illustrated in Figure 61, provides a period of enhanced communications.

In the morning a strong F region permits excellent communications along the gray line before the signal absorbing D layer has a chance to develop. In the evening the D layer disappears quickly again permitting a period of excellent communications while the F region is still strongly ionized.

In Figure 62 excellent communications from northern Europe to eastern Africa occurs at sunrise. In western Canada good communications into the south Pacific occurs at sunset. Seasonal variations in gray line orientation, illustrated by Figures 58 through 60, provide optimum paths to slightly different parts of the world each day. These periods of enhanced communications last for about 45 minutes to an hour.

Since D level absorption is inversely proportional to frequency squared, gray line propagation is very important for 80 meter and 40 meter DX, less so for 20 meters, and usually not relevant for 15 meters.

$$Absorption \propto \frac{1}{f^2}$$

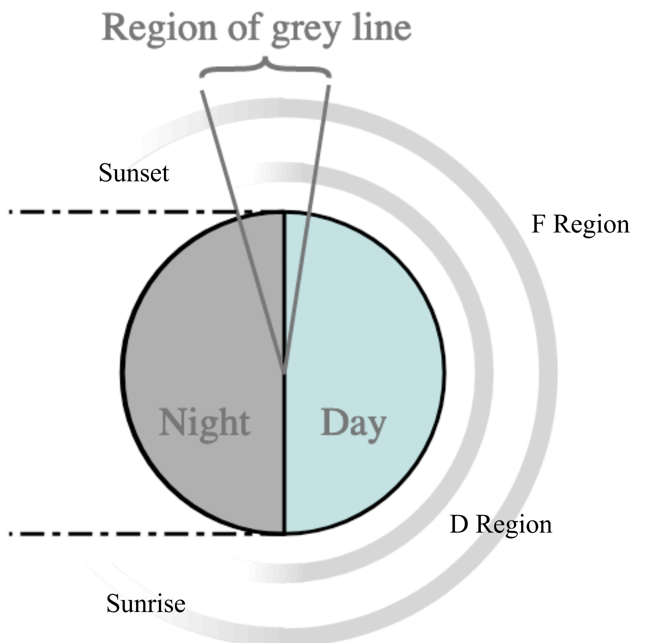

Figure 61 Gray line formation (source: Electronics Notes)

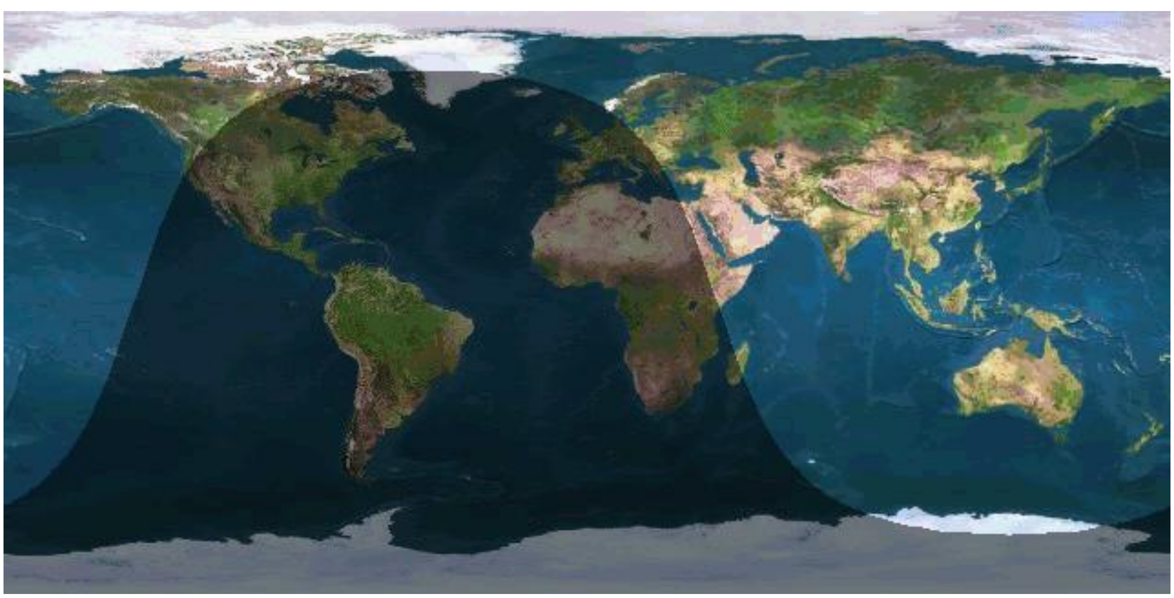

Figure 62 Gray line propagation

17.9.7 Equatorial Sporadic E Propagation

At low latitudes, ionization irregularities resulting from the equatorial electrojet are believed to be responsible for creating sporadic E patches. The patches appear daily in a band about 5° either side of the magnetic equator.

A low elevation angle signal can travel a considerable distance by reflecting from an equatorial sporadic E patch followed by a subsequent F layer refraction, as illustrated in Figure 63.

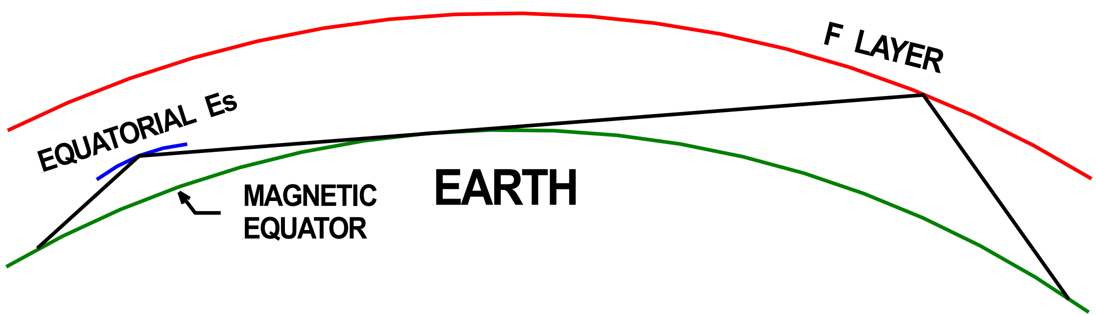

Figure 63 Equatorial Sporadic E propagation (source: author)

17.9.8 Transequatorial Propagation (TEP)

Transequatorial propagation is primarily a long distance 6 meter propagation mode between:

- Central America and South America,
- Japan and northern Australia,
- The Mediterranean and South Africa

as illustrated in Figure 64.

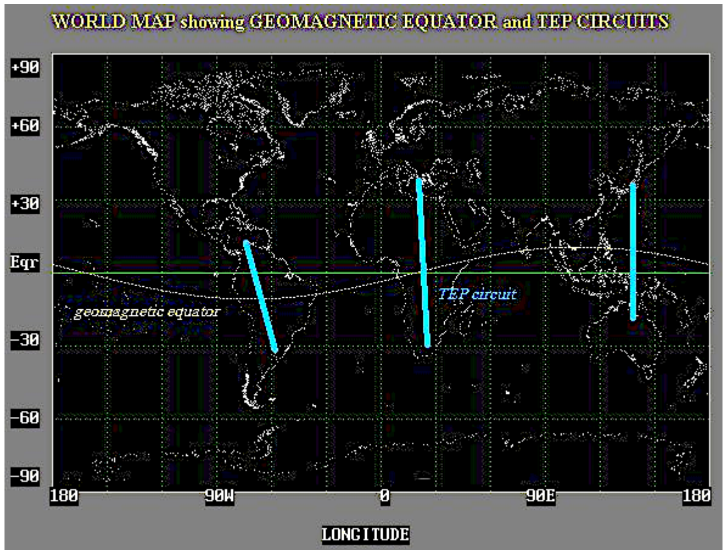

Figure 64 Transequatorial propagation paths (source: Australian Space Weather Services)

Transequatorial propagation is caused by the equatorial fountain effect illustrated in Figure 65. During the day ions and electrons drift upward in the ionosphere due to the force exerted by perpendicular electric and magnetic fields along the geomagnetic equator. The electric field gradually weakens and finally disappears at an altitude of around 800 km. With the electric field no longer present, charged particles (ions and electrons) travel under the influence of gravity and pressure gradients along magnetic field lines that curve back to Earth as illustrated in Figure 65. The charged particles combine with those already in the region creating a peak or crest in electron concentrations 15 to 20º north and south of the equator while a trough, or deficiency of electrons, develops over the equator.

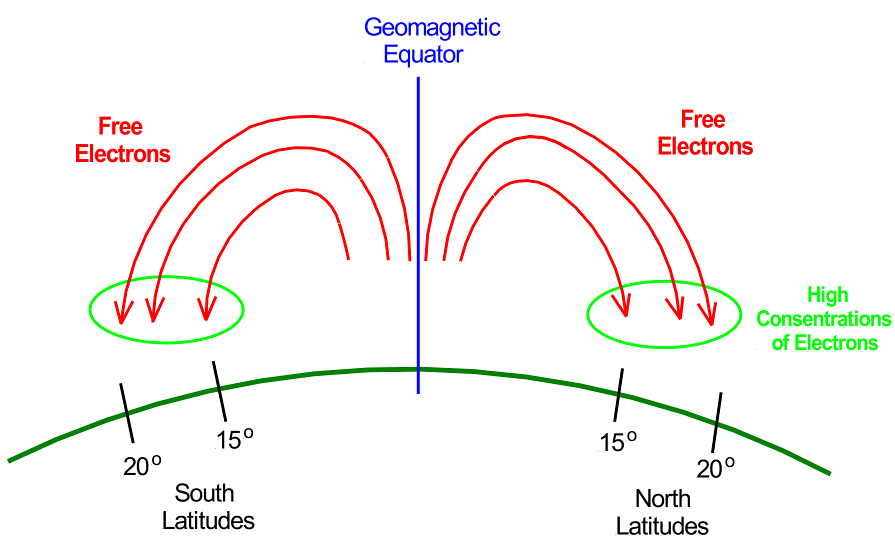

Figure 65 Equatorial plasma fountain (source: author)

Initially, before the fountain effect was understood, the high concentration of electrons appearing at latitudes between 15 to 20 degrees was known as the equatorial or Appleton anomaly. It was believed to be an anomaly since the highest electron concentrations were expected to occur over the equator.

The fountain effect is clearly visible in the winter (December) ionospheric map shown in Figure 66. The black line sloping through South America is the magnetic dip equator. The crests in electron concentrations are the bright pink zones on both sides of the magnetic equator.

The crests usually form in the late afternoon and early evening. While the crests vary from day-to-day and seasonally, they are most pronounced during solar maximum. Notice in Figure 66 that the largest electron peaks (two of them) occur in the winter hemisphere.

F2 critical frequencies within the crests can often reach 18 MHz or higher during solar maximum. The maximum usable frequency at f_{cF2} = 18 MHz and an elevation angle E = 17° is

$$[MUF]_{17^\circ} = \frac{f_{cF2}}{\sin E} = \frac{18\ MHz}{\sin 17} = 61\ MHz$$

with the frequency of optimum transmission

$$FOT = 0.85\ MUF = 0.85(61\ MHz) = 52.3\ MHz$$

occurring in the 6 meter frequency band. Because of the fountain effect and associated equatorial trough, critical frequencies along the magnetic equator are typically several MHz less than in the crests.

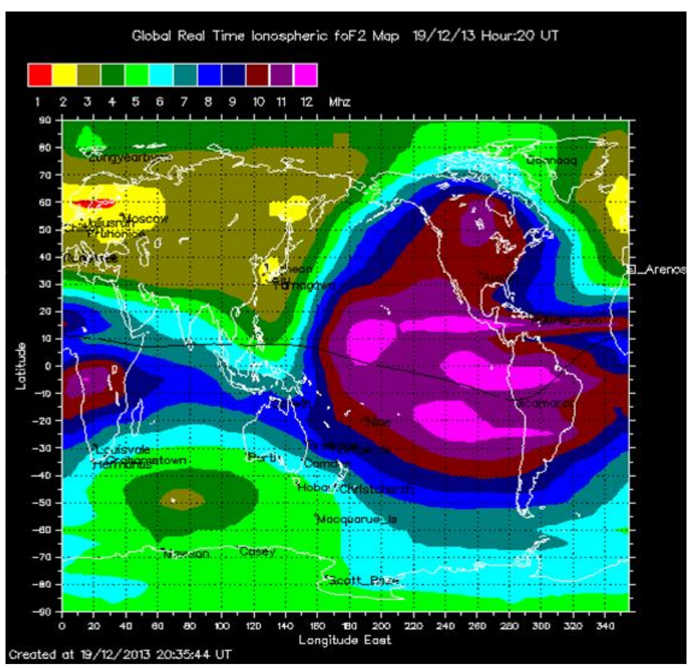

Figure 66 Ionospheric map showing fountain crests (source: Australian Space Weather Services)

The fountain effect distorts the general form of the ionosphere throughout the low latitude equatorial region leading to the Transequatorial 6 meter propagation phenomena, as illustrated in Figure 67. Typically, a radio signal transmitted from one hemisphere to the other requires multiple hops through the ionosphere to reach its destination, with signal attenuations occurring with each hop. Instead, a radio signal experiencing Transequatorial propagation reflects off one fountain crest, travels across the equator to the second crest, and then reflects back to Earth. Transequatorial propagation allows a radio signal to traverse long distances from the transmitter to receiver in one hop, greatly minimizing signal loss.

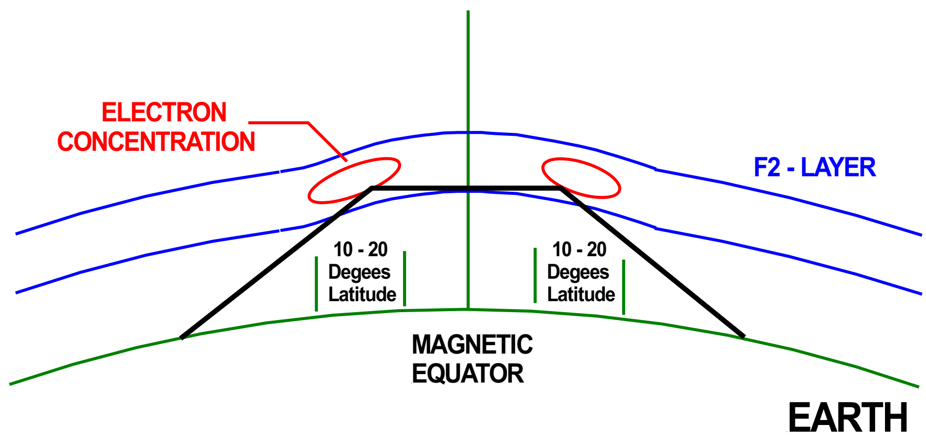

Figure 67 Transequatorial Propagation (source: author)

17.9.9 Ionospheric Ducting

If a signal is injected into the ionosphere at just the right angle, and under the right conditions, it can become trapped between the F and E layers, reflecting off the bottom of the F and the top of the E layer, as illustrated in Figure 68. This injection can sometimes occur when a signal reflects from ionospheric irregularities. A signal traveling in an ionospheric duct avoids multiple passes through the energy absorbing D layer and reflections from the ground. Consequently, the signal can travel a long distance with little signal loss. Experiencing ionospheric ducting is weird. When ducting is present, communications half way around the world can occur with the same signal strength, clarity, and stability as communicating locally on a 2 meter repeater.

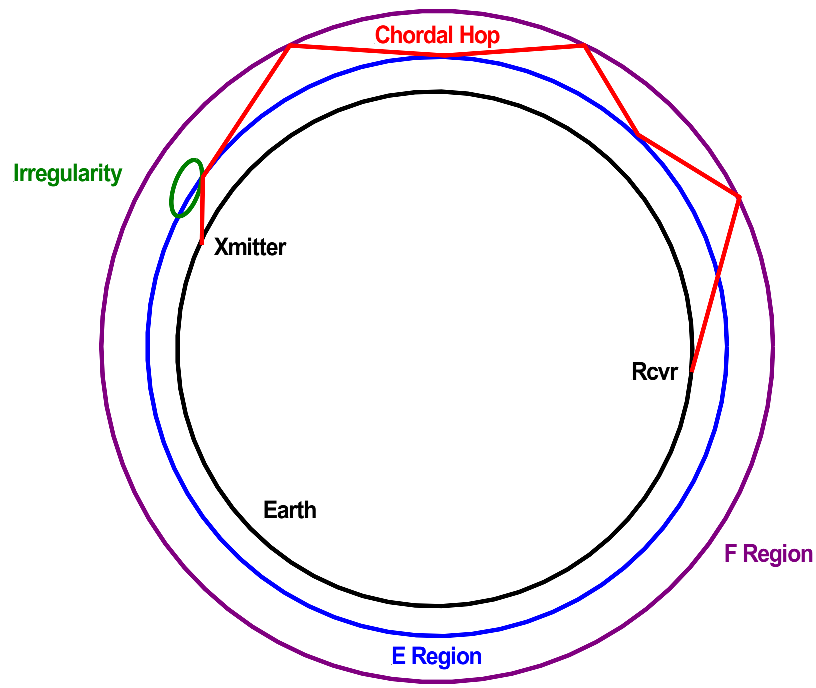

Figure 68 Ionospheric ducting (source: author)

17.10 HF Non-Skywave Propagation

In concluding this chapter, it is important to point out that there are two non-skywave HF propagation modes, specifically:

- Line of sight, and
- Ground wave propagation.

Historically these two propagation modes were very important.

Prior to the mid 1960's the primary amateur radio bands were 160 through 10 meters. While the VHF/UHF 2 meter, 220 MHz, and 440 MHz bands were assigned for amateur radio use, they were in general used only by experimenters (technicians, engineers, and scientists). The reason for this was that the vacuum tube transmitters and receivers of the day were physically too large to support communications at these VHF and UHF frequencies. Figure 69 shows the author's 1962 amateur radio rig. The wiring in these big radios was sufficiently long that it radiated at VHF frequencies making the technology impractical for VHF/UHF work. For comparison, the small radio on the far left in Figure 69 is the author's 1990 Kenwood TS-440 transceiver. The TS-440 was a 100 watt semiconductor HF radio that far out performed the 100 watt vacuum tube rig.

Figure 69 Author’s 1962 ham radio station (source: author)

Opening up the VHF and UHF bands for general amateur radio use had to wait for development of semiconductor technology and the subsequent tiny hand held and desk top VHF/UHF transceivers. Today 2 meters is the most popular amateur radio band, being used primarily for local communications.

The question is, what frequency band was used for local communications prior to the introduction of semiconductor technology? The answer is 80 meters, which was used both day and night. How can that be? During the day 80 meters is dead due to intense D layer absorption. Yes, 80 meter skywave communications is dead during the day, but not 80 meter line of sight and ground wave propagation.

The distance covered by line of sight on 80 meters is essentially the same as line of sight 2 meter coverage. The range of ground wave propagation increases as the electrical conductivity of the ground increases. It is greatest over sea water and smooth flat fertile ground. The range of ground wave propagation is also proportional to signal wavelength as illustrated in the following table. Long wavelength signals travel further along the ground than short wavelength signals.

Frequency Band	**Typical Ground Wave Distance**
80 meters	68 miles
40 meters	50 miles
20 meters	30 miles

Consequently, the range of 80 meter ground wave is similar to that of 2 meter non-link repeaters. 80 meters is a good local communications band with one exception, the long 80 meter antennas. Because of the long antennas, 80 meters has largely been abandoned with local communications switching to the highly desirable VHF and UHF bands. However, the physics has not changed. 80 meters is a good back up option should VHF and UHF repeaters fail during a large scale disaster.

One could ask why 40 meters, with its shorter antenna, was not use for location communications? The answer is relatively simple. The amateur radio bands were fairly congested, particularly with wide bandwidth (6 KHz) AM phone traffic. While invented, single side band (SSB) had not yet come to the amateur radio bands. At the time, amateur radio operators wanted to relieve 40 meter congestion and so moved all location communications to 80 meters.

References

Davies, Kenneth; "Ionospheric Radio"; Peter Peregrinus Ltd.,1990

McNamara, Leo F.; "The Ionosphere: Communications, Surveillance, and Direction Finding"; Krieger Publishing Company, 1991

Hunsucker R. D.; Hargreaves, J. K.; "The High-Latitude Ionosphere and its Effects on Radio Propagation"; Cambridge University Press 2003

Nichols, Eric P.; "Propagation and Radio Science"; The American Radio Relay League, Inc. 2015

Yeang, Chen-Pang; "Probing The Sky With Radio Waves"; The University of Chicago Press, 2013

Devoldere, John; "Low-Band DXing" fourth edition; ARRL, 2005

Levis, Curt A. ; Johnson, Joel T.; and Teixeira, Fernando L.; "Radiowave Propagation Physics and Applications"; John Wiley & Sons, Inc., 2010

Goodman, John M.; "Space Weather & Telecommunications"; Springer Science+Business Media Inc. 2005

Ahrens, C. Donald; "Essentials of Meteorology"; Wadsworth Publishing Company, 1998

UCAR Center for Science Education (UCAR SciEd); https://scied.ucar.edu/learning-zone/atmosphere/

Khazanov, George V.; "Space Weather Fundamentals"; CRC Press, 2016

Foukal, Peter; "Solar Astrophysics third edition"; Whiley-VCH Publishing Company, 2013

Ganushkina, N. Yu.; Liemohn, M. W.; Dubyagin. S.; "Current Systems in the Earth's Magnetosphere"; AGU, March 8, 2018

Gallagher, Dr. D.L.; "The Earth's Plasmasphere"; Space Plasma Physics, Marshall Space Flight Center, Huntsville, Al, September 05, 2018

Moore, T. E., Morwitz, J. L.; "Stellar Ablation of Planetary Atmospheres"; August 9, 2007

Yau, Andrew W.; Abe, Takumi; Peterson, W. K.; "The Polar Wind: recent Observations"; Department of Physics and Astronomy, University of Calgary

Carroll, Bradley W. and Ostlie, Dale A.; “An Introduction to Modern Astrophysics”; Addison-Wesley Publishing Company Inc., 1996

Cander, Ljiljana R.; “Ionospheric Space Weather”; Springer Nature Switzerland AG 2019

Moldwin, Mark; “An Introduction to Space Weather”; Cambridge University Press, 2008

Campbell, Wallace H.; “Introduction to Geomagnetic Fields”; Cambridge University Press, 2003

Golub, Leon and Pasachoff, Jay M.; “Nearest Star The Surprising Science of Our Sun second edition”; Cambridge University Press, 2014

Loff, Sarah: ”Explorer and Early Satellites”; National Aeronautics and Space Administration, Aug 3, 2017

Minzner, R. A.; “The 1976 Standard Atmosphere Above 86 km Altitude” NASA Goddard Space Flight Center, 1976