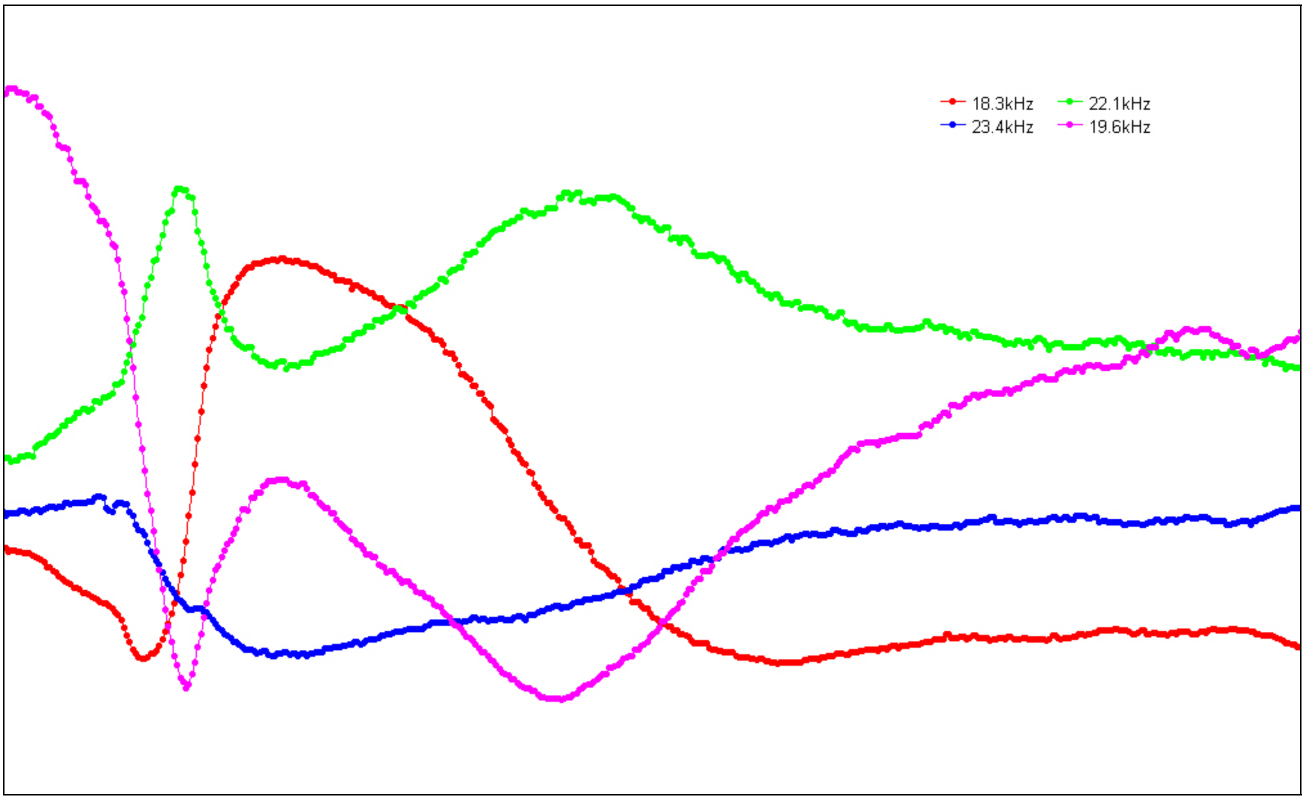
Modelling the Ionosphere

The recent long period of solar inactivity was spectacularly terminated by a series of X-ray flares during January 2010. One of these, an M-class, produced an intense Sudden Ionospheric Disturbance (SID) at 11:22UT on 20th January 2010.

1 A SID observation

The picture below shows the effect it had, at my location near Coventry, on the received signal strength from four Very Low Frequency (VLF) transmitters scattered around Europe.



One thing is immediately obvious; the SID produced remarkably different effects at each of the four frequencies. Of particular note are the totally opposite effects at 19.6 and 22.1 kHz, which are also notable for having two prominent peaks (or troughs).

This causes a problem as a SID report to the BAA is supposed to include not only the start and end times of the SID but also the time of maximum effect. However, where is the maximum when there are two peaks or troughs? It would seem, by comparison with the effects on the other two frequencies, that the maximum should be between the two, but is there any justification for that assumption?

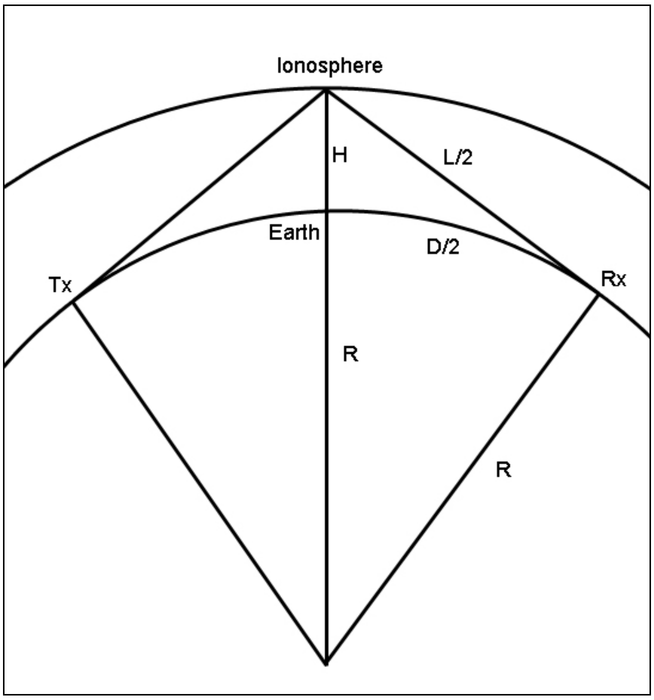
To try to answer these problems I decided to see if a simple empirical model of the changes in the ionosphere during a SID would give the type of effects that had been observed and in particular whether it would produce the mirror image traces.

2 VLF propagation

The simplest model for VLF propagation in the Earth-ionosphere waveguide consists of the addition of two waves – a ground wave that follows the curvature of the Earth and a sky wave which undergoes a specula reflection from the D-layer of the ionosphere. The effective height of the D-layer is then governed by its ionisation level, which in turn is determined by the amount of solar X-rays impinging on it.

During a solar flare the ionisation of the layer increases dramatically and its height rapidly reduces. After the flare, through recombination, its ionisation gradually returns to normal and the layer gradually rises to its original level.

The geometry of the situation is shown in the diagram below:-



Where:-

Tx is the position of the transmitter
Rx is the position of the receiver
D = distance between the transmitter and receiver
R = radius of the Earth
H = height of the ionosphere above the Earth
L = length of the path taken by the sky wave between transmitter and receiver

From which the length of the path taken by the sky wave between transmitter and receiver can be calculated as:-

$$L = 2\sqrt{[R^2 + (H+R)^2 - 2R(H+R)\cos(D/2R)]} \qquad \text{--- (1)}$$

The difference in the path lengths between the sky and ground waves is just (L – D) which translates into a phase difference (in radians) of:-

$P = (L-D)2\pi\nu/c$

Where:-

ν = frequency of transmission
c = speed of light

As there is an inversion of the electric field on reflection from the ionosphere this can be represented by the insertion of an extra half cycle (π) of phase:-

$P = (L-D)2\pi\nu/c + \pi$ --- (2)

The received signal is then the vector sum of the ground wave of amplitude G and phase 0, with the sky wave of amplitude S and phase P. Giving the amplitude of the resulting vector as:-

$A = \sqrt{(G^2 + S^2 + 2GS \cos P)}$ --- (3)

We now have three equations (1) – (3) that relate the received amplitude (A) of a transmitter (frequency ν) at a distance (D) to the height of the ionosphere (H).

These show that as the height of the ionosphere changes so the received amplitude varies between a minimum of (G – S) and a maximum of (G + S) as the phase difference between the sky and ground waves varies between an odd and even number of half cycles, respectively.

The equations can of course be run in reverse to obtain the height of the ionosphere corresponding to any given minima and maxima.

In this way and using distances of 305.98km for the 19.6kHz transmitter (located at Anthorn, Cumbria) and 277.67km for the 22.1kHz transmitter (located at Skelton, Cumbria) allows a comparison to be made between the height of the ionosphere required to receive a minimum amplitude at one frequency at my location and a maximum at the other:-

19.6kHz Amplitude	19.6kHz Height (km)	22.1kHz Height (km)	22.1kHz Amplitude
		29.4	Max
Max	32.6	42.5	Min
Min	47.2	52.6	Max
Max	58.6	61.4	Min
Min	68.3	69.2	Max
Max	77.1	76.4	Min
Min	85.1	83.1	Max
Max	92.6	89.5	Min

This is encouraging as it shows a close match at 68.3 and 69.2km (min at 19.6 kHz and max at 22.1 kHz) and 76.4 and 77.1km (max at 19.6kHz and min at 22.1kHz) which are also consistent with the generally accepted height of the D-layer (60 to 90km).

3 The effect of a SID

To model the change of height of the D-layer during a SID requires a function that matches the effect of X-rays on the ionisation of the D-layer during a solar flare. For this I chose a function that rises rapidly and has a gradual fall:-

$$F_{SID}(e^{(-t / T_F)} - e^{(-t / T_R)})$$

Where:-

t = time
T_F = fall time constant
T_R = rise time constant
F_{SID} = factor dependant upon the strength of the SID, the stronger the SID the larger the value.

Subtracting this function from 1 (to get the opposite effect) and treating it as a modulation of the normal height (H_{NOR}) of the ionosphere gives:-

$$H = H_{NOR}(1 - F_{SID}(e^{(-t / T_F)} - e^{(-t / T_R)}))$$

for the height of the ionosphere at time t.

4 Diurnal effects

The normal height of the ionosphere also varies slowly throughout the day as the altitude of the Sun changes. This diurnal change in height is given by the Chapman equation[1]:-

$$H_{NOR} = H_0 + H_S \log_e(\sec \chi)$$

Where:-

H_0 = height of the ionosphere when $\chi = 0$, ie. when the Sun is overhead
H_S = scale height (which varies from day to day)
χ = Sun zenith angle

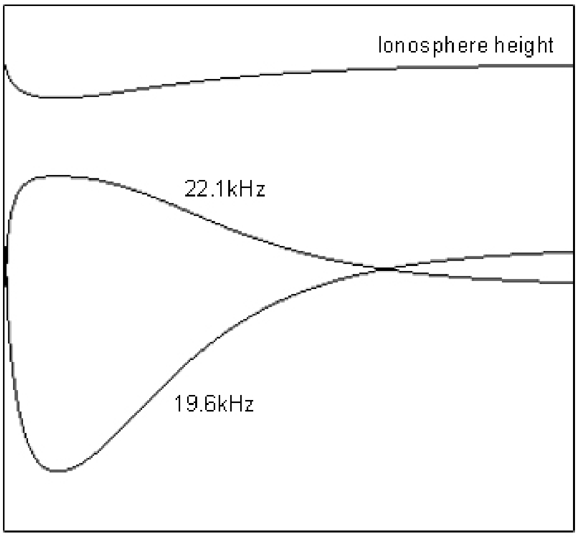
Combining this effect with that of the SID gives:-

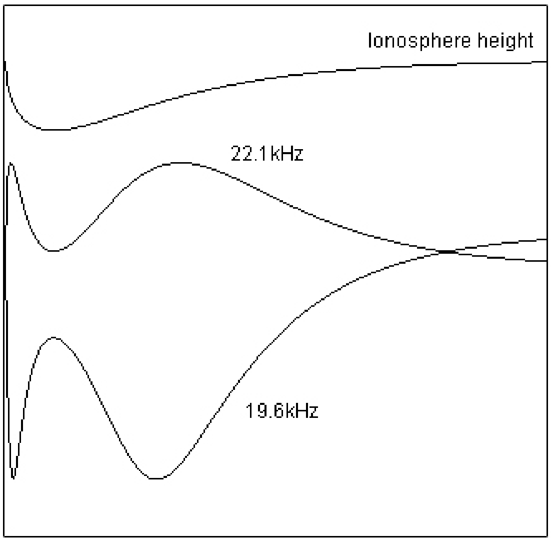
$$H = (H_0 + H_S \log_e(\sec \chi)) (1 – F_{SID}(e^{(-t / T_F)} - e^{(-t / T_R)})) \qquad \text{--- (4)}$$

Equations (1) to (4) then form the basis of the SID model.

5 Output of the SID model

Running the model over the time of a simulated SID, ie. $F_{SID} > 0$, but without any diurnal variation, i.e. $H_S = 0$, produces the following:-



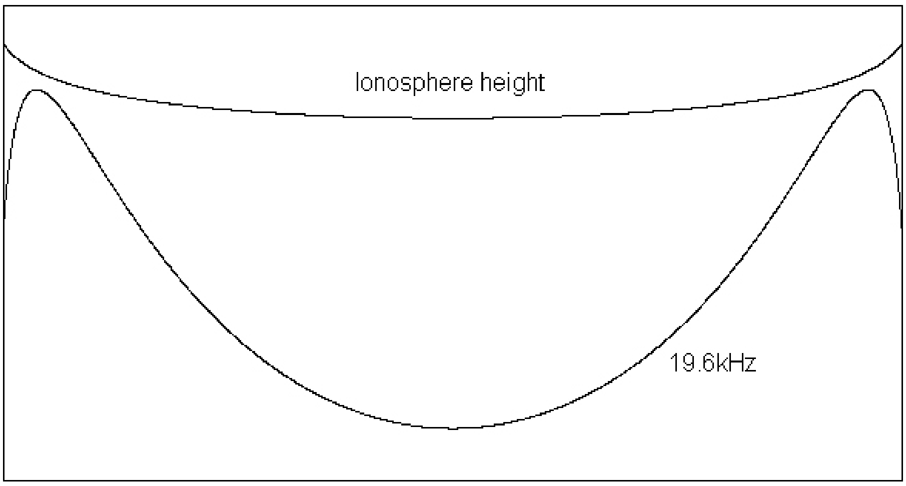


In the left-hand diagram the ionosphere starts at height just above 70km and drops by a small amount that takes it towards the height (69.2km) at which there is a maximum for 22.1kHz, but crucially not through it. This produces the characteristic single peak of a SID.

In the right-hand diagram, the drop in the height of the ionosphere is larger and sufficient to take it through the critical region of peak amplitude and towards the height of the next minimum (61.4km). As the SID declines and the ionosphere returns to its undisturbed height, the critical height is passed through again but more slowly, so producing a broader amplitude peak than the first.

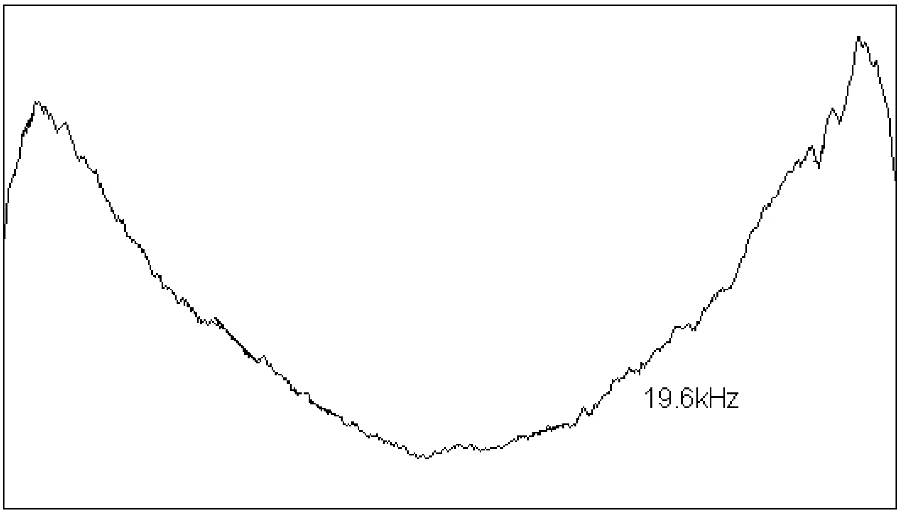
As the critical height for the minimum amplitude at 19.6kHz is only 0.9km lower than that for the maximum at 22.1kHz, the same changes are experienced at nearly the same time but in the reverse direction.

On the other hand, running the model in the absence of a SID, ie. $F_{SID} = 0$ and $Hs > 0$ produces curves of the type shown below for a frequency of 19.6kHz:-



This shows the diurnal variation in the strength of the received signal as the height of the ionosphere drops from 78km during the morning to 71km at local noon and rises again to its original height during the afternoon.

Comparing this with the actual variation seen at 19.6kHz on a SID-less day, shows a striking similarity:-



Finding that the model produced remarkably similar results to the actual observations encouraged me to write a model fitting program that would, given the SID observations as input, produce the values of the variables used in the four equations as output. What is more, if the program could fit more than one frequency at once that would help to constrain the model.

6 Model fitting

The model fitting program is based on an old function minimisation routine[2]. It uses the method of gradient descent and so requires a function to minimise as input. The function to minimise is obviously a measure of the difference between the output of the model and the real data:-

$F = \Sigma(A' - A)^2$

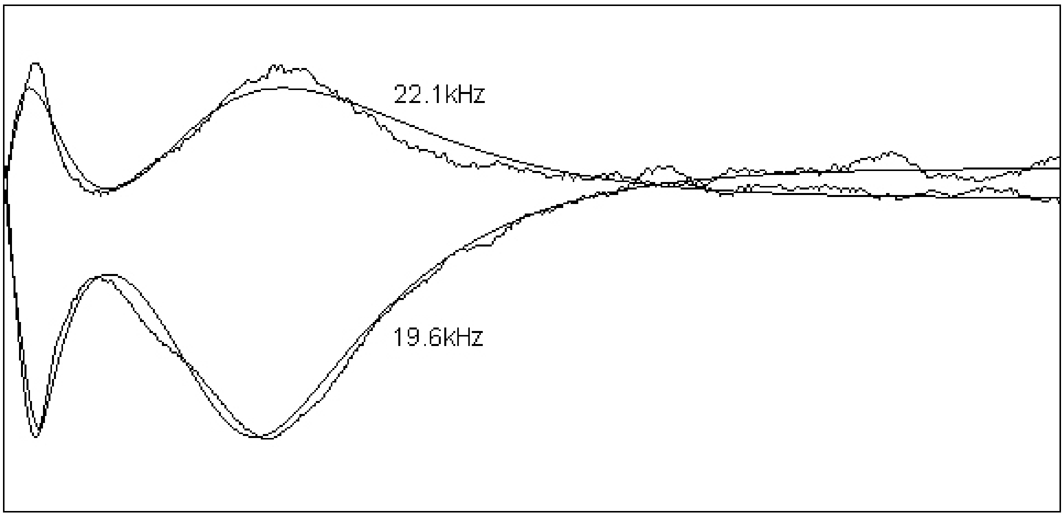
Where:-

A' = amplitude from observation
A = amplitude from model

The sum in the equation is taken over all time and all frequencies of observation and will have a value of zero if the model matches the real data precisely.

7 Fitting a SID

Running the program produces a remarkably good match with the actual 19.6kHz and 22.1kHz SID observations:-



Where the smooth lines are the output of the model and the jagged lines are the observations.

As output, the model gives:-

Height of ionosphere before SID = 72km
Height drop produced by SID = 7km

Although very successful at fitting one or two frequencies at a time, the program is unable to fit all four frequencies affected by the SID at once. A glance at the minimum/maximum table for 18.3kHz shows why:-

18.3kHz Amplitude	18.3kHz Height (km)
Max	44.5
Min	66.1
Max	82.9
Min	97.2

Falling from a height of 72km to 65km the amplitude of the 18.3kHz transmitter should go through its minimum at 66.1km after the 19.6kHz one goes through its minimum at 68.3km, but plainly it actually goes through its minimum before. Running the model with 18.3kHz alone indeed gives a good fit with a starting height just above 67km.

Although at first sight this might appear to be a frequency affect, with the lower frequency penetrating less into the D-layer, it cannot be the case as there is no such effect evident between 19.6 and 22.1kHz. What does appear to be happening though, is a change of effective reflection height with distance. At 659.42km away from my location, the 18.3kHz transmitter is over twice the distance of the other two.

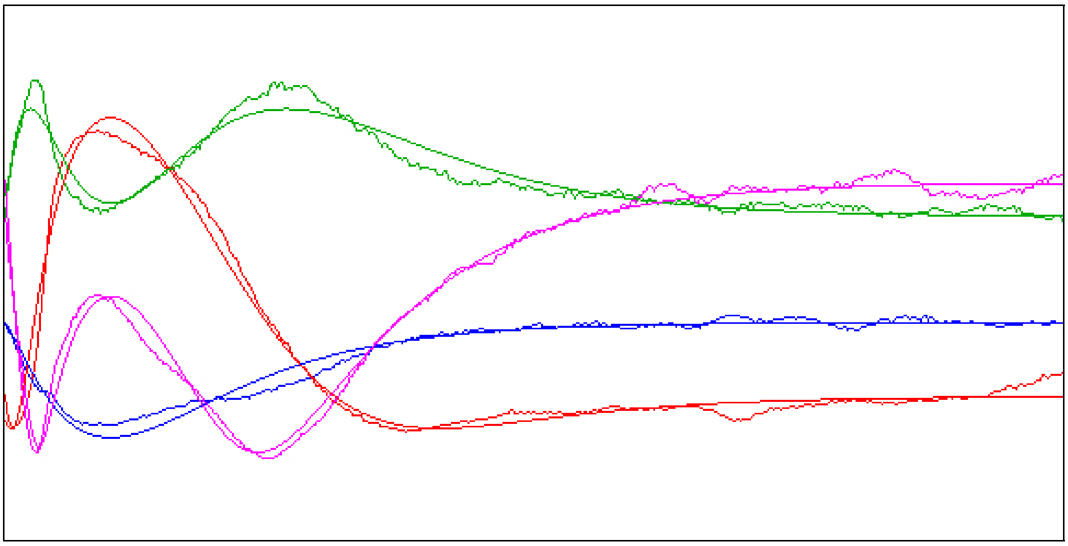
What has to be remembered is that the path of radio waves through the ionosphere is more complicated than the assumption of specula reflection and in fact takes the form of a curve. As the angle of incidence on the ionosphere of the radio waves from a local transmitter is less than that of a remote transmitter they have to penetrate further to be bent back to Earth, so their apparent reflection height is greater.

To take account of this, more rigorous models resort to ray tracing methods, but in my simplified model adding another linear term to equation (4) is sufficient to give it the required flexibility:-

$$H = (H_0 + H_S \log_e(\sec \chi))\,(1 - F_{SID}(e^{(-t/T_F)} - e^{(-t/T_R)}))\,(1 - F_{DIST}(D - D_L)) \quad \text{--- (5)}$$

Where:-
F_{DIST} = distance factor
D = distance to the remote transmitter
D_L = distance to the local transmitter

With this modification the model fits all four frequencies reasonably well:-



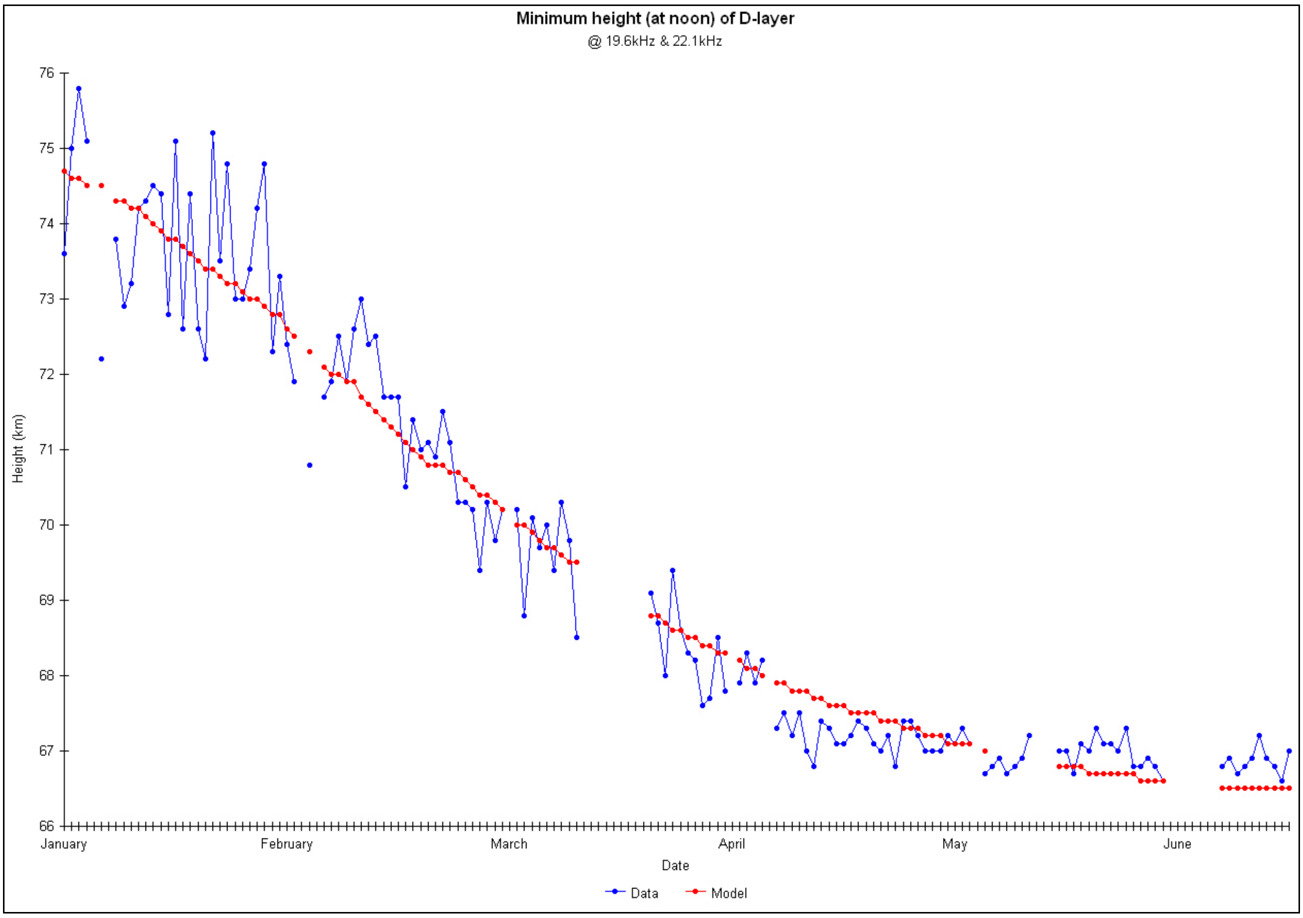
As output, the model gives:-

Transmitter frequency	18.3kHz	19.6kHz	22.1kHz	23.4kHz
Ionosphere height (km)	67.2	72.0	72.4	67.8
Drop during SID (km)	5.7	6.1	6.2	5.8

8 Fitting Diurnal Variation

Even more information about the ionosphere can be obtained by fitting the diurnal variation over a period of time. For not only does the height of the ionosphere vary throughout any given day due to the Sun's changing zenith angle, but also throughout the year as the Sun's declination varies.

The data below (jagged curve) was obtained by model fitting each day's variation at 19.6kHz and 22.1kHz for the first six months of 2010 and taking the resultant minimum height at noon:-



As can be seen, the height of the ionosphere shows a gradual decline as the Sun's declination rises from the winter to the summer solstice.

That seasonal decline follows exactly the same equation as for the diurnal variation:-

$H = (H_0 + H_S \log_e(\sec \chi))$

and can again be model fitted to the data, giving the smooth curve in the figure. That curve shows a good fit with H_0 = 65.6km and H_S = 6.2km.

Now, H_S (the scale height) = kT/mg

Where:-

k = Boltzmann's constant
T = Temperature of the gas in the ionosphere
m = mass of a gas molecule in the ionosphere
g = acceleration due to gravity

This means that given its scale height and temperature, the atomic mass of the ions forming the D-layer in the ionosphere can be determined. Using the fitted scale height of 6.2km and a typical temperature at 70km of 230 ^{o}K gives m = 31.3 u, which is consistent with the known composition being a mixture of NO^+ (30 u) and O_2^+ (32 u) ions.

9 Conclusion

In conclusion, I have shown how a simple model can give a useful insight into the causes of seemingly random changes during a SID and also provide information about the ionosphere itself.

10 References

1 Contributions to the 3D ionospheric sounding with GPS data. PhD Thesis Universitat Politecnica de Catalunya. Miquel Garcia-Fernandez. January 22nd. 2004.

2 Collected Algorithms from CACM. Algorithm 251. M. Wells 13th July 1965 and 5th October 1964. Based on the method of Fletcher and Powell Computer Journal 6, 163-168 1963.