Chapter 14

The HF Doppler technique for monitoring Transient Ionospheric Disturbances.

Introduction.

Many types of transient disturbances in the ionosphere, such as those resulting from solar flare activity or the passage of travelling disturbances (TIDs), distort the iso-ionic contours and so produce variations in the reflection height of an HF signal. This change in reflection height gives rise to a change in the phase path, consequently a Doppler frequency shift will be induced into the reflected signal. The magnitude of the frequency shift will be proportional to the rate of change of the phase path. By monitoring the frequency of the reflected signal, the presence of a disturbance which produces height changes can be detected and its time history recorded. If measurements are made on three or more propagation paths whose reflection points are suitably spaced, then a travelling disturbance will produce Doppler signatures which are time displaced on the signals received over the three propagation paths. If no dispersion occurs as the disturbance propagates, the horizontal component of its velocity can be determined from the three time delays and a knowledge of the reflection point separation by simple triangulation.

The Doppler technique provides a low cost, effective method of monitoring transient changes in the ionosphere. It does not always yield an accurate measure of the TID velocity because the driving internal acoustic gravity wave is generally dispersive. Nevertheless, the technique can produce valuable information as indicated later in this chapter.

Choice of operating frequency.

An HF radiowave will be reflected from the ionosphere provided its frequency is less than the F-layer critical frequency. In the Doppler technique, the receiver and transmitter are usually separated by distances of not more than 200 km and so propagation is at steep (almost vertical) incidence. The magnitude of the Doppler shift Δf is related to the rate of change of phase path P by

$$\Delta f = -\frac{1}{\lambda}\frac{dP}{dt}$$

where λ is the wavelength.

A change in phase path can occur either due to a change in the reflection height or because of a change in the refractive index along the ray path. It has been shown (Davies 1965, 1968) that for changes in reflection height the magnitude of Δf is directly proportional to the wave frequency, thus

$$\Delta f \propto f$$

If the frequency shift is induced by refractive index changes then

$$\Delta f \propto \frac{1}{f}$$

For reflections from the E and F layers of the ionosphere, Δf is found to be proportional to f indicating that changes in the reflection height are dominant in producing the Doppler shift.

Since the magnitude of the Doppler shift is proportional to the wave frequency, it would appear that the operating frequency should be chosen as close to the critical frequency as possible. This is not the case, since this choice results in the wave penetrating the layer as the diurnal changes in critical frequency occur. However, since the amplitude of all wave-like disturbances increase with increasing height, ie. decreasing atmospheric density, the displacement of the reflection level and hence the Doppler shift, increase with height, ie. with sounding wave frequency. For radio waves reflected from the mid-latitude F-region at steep incidence, a typical medium-scale TID will produce Doppler shifts of between -0.5 to +0.5Hz. A spectral resolution of 0.01Hz is more than adequate for studies of such events and a resolution of 0.05Hz is normally employed. If the measuring system is designed to measure frequency shifts of the order of 0.01Hz, it is not necessary to operate very close to the F-region critical frequencies and it is best to work at least 1 or 2 MHz below the critical frequency. The frequency should be selected by reference to an ionogram and chosen so that reflection occurs well away from regions of large group retardation. Since the critical frequency differs markedly during the day and night, it is not normally possible to operate a Doppler system for 24 hours on only a single frequency. Two frequencies are usually required, one for night and one for day-time conditions. It is rather difficult to measure the Doppler signature of E-region reflections since the Doppler shifts are very small due to the small displacements of the reflection height and the lower frequencies necessary to obtain E-region reflections. It should be noted that the sporadic E-layer produces a very stable Doppler signature with little or no variations over considerable periods of time. For general studies of TIDs, F-region reflections should be employed.

The Experimental System.

a) Transmitter.

The transmitter consists of any suitable RF amplifier which is capable of producing a CW signal of about 20 watts of RF power at the required HF frequency. Some amateur radio transmitters are quite suitable for this purpose. The transmitter drive is generated from a frequency synthesizer capable of producing the required frequencies and stable to about 1 part in 10^9. A good quality crystal oscillator can achieve this stability and the internal crystal oscillators of most modern synthesizers meet this stability requirement. For example, if we are to measure shifts of say 0.01Hz in a 5MHz signal a measurement accuracy of 1 part in 5 10^8 is needed.

The transmitting antenna normally consists of a simple half wavelength horizontal dipole suspended a quarter of a wavelength above the ground. A lower height can be employed, the requirement being that there is a major radiation lobe in the vertical direction. This ensures that power is radiated at an angle suitable for the steep incidence path between the transmitter and receiver. The dipole antenna is aligned so that it is end on to the direction of the receiver as this reduces the radiation in that direction and so reduces the ground wave field strength at the receiver. If continuous operation during day and night is required, provision should be made for radiating two frequencies to allow for the changes in critical frequency. For two frequencies it is usually easier, cheaper and more reliable to duplicate a complete transmitter installation than to introduce an automatic frequency changer into the system.

b) <u>Receiver.</u>

A standard HF communications receiver can be employed, provided the local oscillator is sufficiently stable. In general, this will not be the case and arrangements will have to be made to generate this frequency from a synthesizer of similar stability to that employed at the transmitter. When the synthesized frequency is set equal to the transmitted frequency minus the local oscillator frequency, the nominal IF frequency will be obtained from the IF amplifier. This frequency is down converted to base band by mixing with a synthesized IF frequency (see fig 1). This mixing is done in phase and in quadrature with two mixers as indicated in the figure. The real and imaginary channels are fed to two A/D converters which digitise the base band frequency at a rate of between 30 to 50 samples per second. These digitized samples are recorded onto magnetic tape or disc for further processing.

Although amplitude information is not required for a simple Doppler experiment, some form of AGC must be applied to the receiver to overcome the large changes in signal amplitude due to fading. The normal receiver AGC system is generally adequate but an attenuator controlled by the computer can be incorporated. The attenuator is changed in such a way as to keep the receiver output amplitude approximately constant during the course of the measurements.

The frequency of the signal is obtained by applying a FFT to the digitized data. This is done off-line although some capability for viewing the data on site is generally desirable. Samples of the data, representing 20 seconds of observing time are Fourier transformed. This provides a frequency resolution of 0.05 Hz within a spectral range of 30Hz. Clearly the amount of

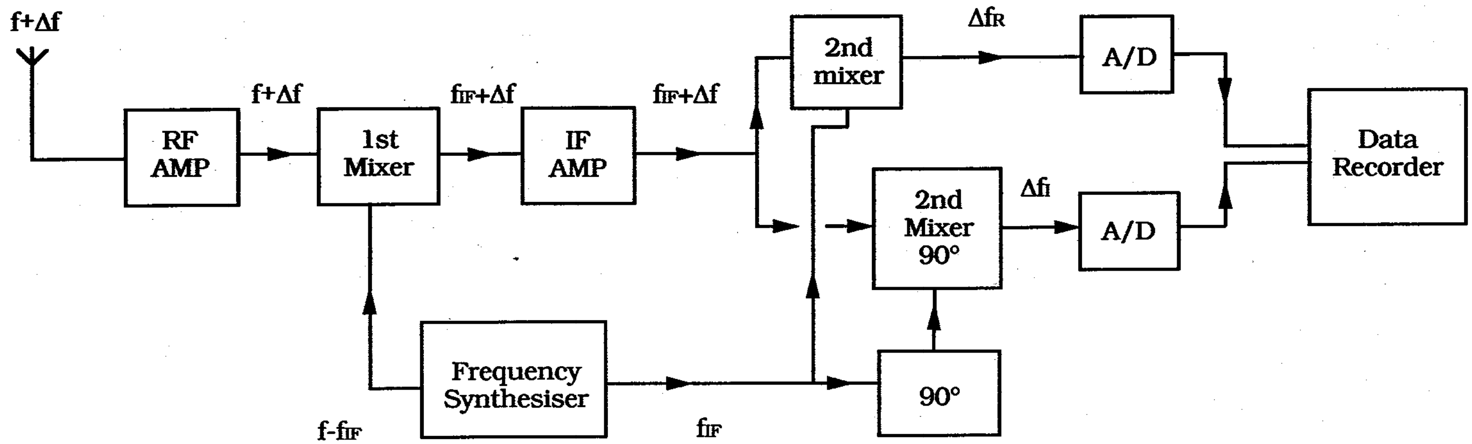

Fig. 1 Block diagram of Receiver elements

data included in each transform can be varied and so the frequency resolution can be changed. A sample of the analysed data is reproduced in fig.2, which contains examples of medium scale TIDs. When 3 or more transmitters are employed for measurements of TID velocities, it is usual to off-set the frequency of each transmitter from its nominal value by a few Hertz. Each transmitter is assigned its own unique off-set so that it can be easily recognized in the subsequent Fourier analysis. Four or five transmission off-sets can be accommodated within a 15 Hz bandwidth so they are within the passband of the receiver even at its narrowest setting. The receiver bandwidth should be as narrow as possible, say 100 to 200 Hz, to reduce unwanted signals and noise.

The receiving antenna can be of any design provided it is sensitive to signals arriving at steep angles of incidence. A convenient antenna is a short active dipole mounted horizontally some 5 to 10 meters above the ground. This produces the required polar diagram and responds well to any frequency within the HF band without adjustment.

The transmitter and receiver sites are usually separated by about 60 to 100 km to reduce the ground wave field strength at the receiver. If TID velocities are to be measured, then the spacing must be large enough to produce significant time delays between the signatures of the event on the three or more propagation paths employed. For medium scale TIDs the horizontal component of the velocity is about 300 ms^{-1} and for a reflection point separation of 30km this produces a maximum time delay of 100 sec. For large scale TIDs with periods of approximately 1 hr, the horizontal velocity can be as high as 1000ms^{-1} For studies of these disturbances, larger spacings of not less than 60 km between reflection points should be adopted. This corresponds to a maximum time delay of about 1 min. Since the TIDs are dispersive, the accuracy of the velocity measurement depends critically on how well the Doppler signatures on the various propagation paths can be correlated.

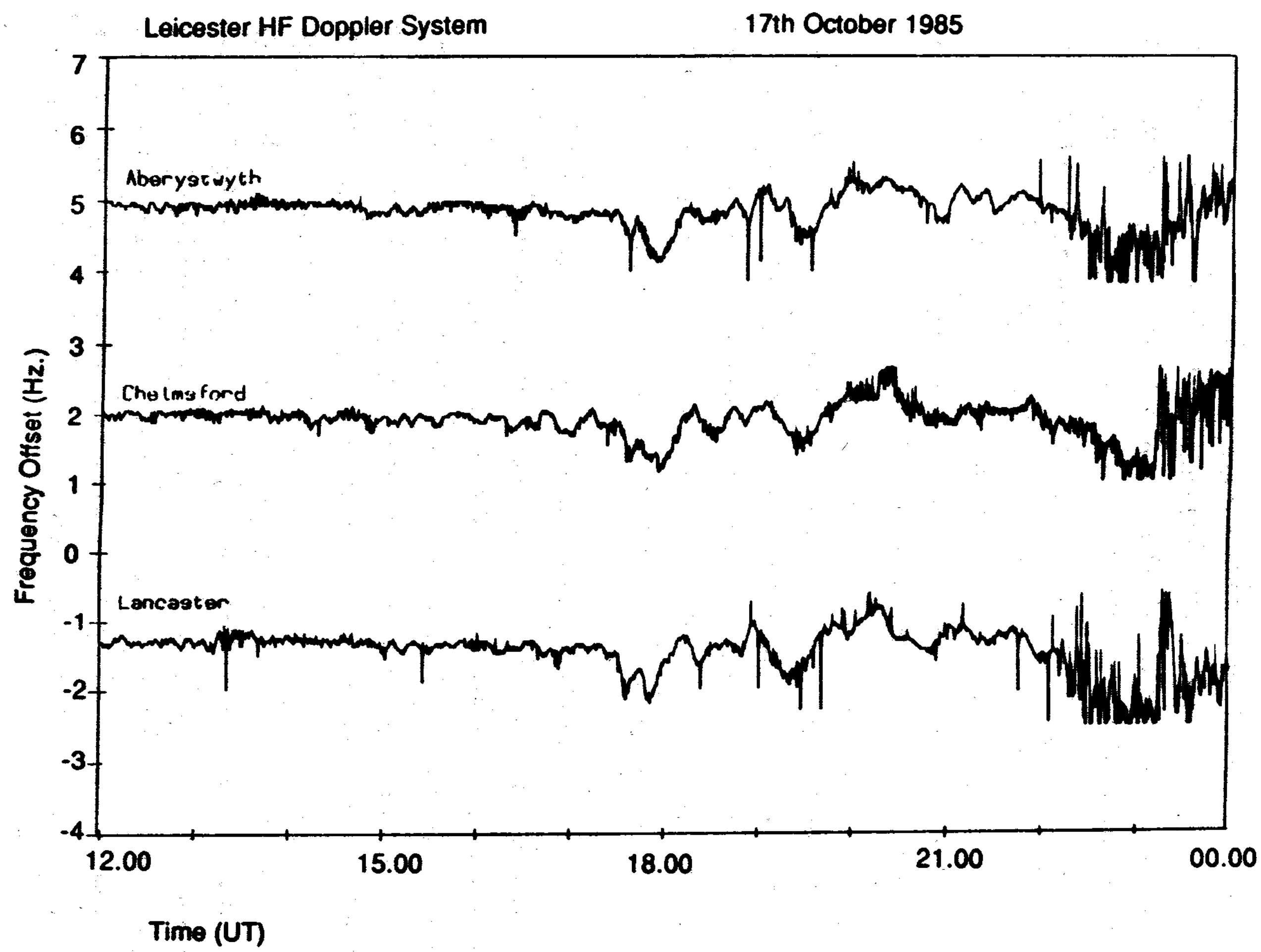

Fig 2. Example of three station Doppler variations from digital system.
Note wave activity from 18.00 to 21.00 UT.

c) Control computer and data logger.

Almost any small computer can be adapted as a controller and data logger for the receiver system. The following are the main considerations when choosing and/or designing a suitable computer:-

i) Interface with the receiver. This is not essential unless the receiver is to be switched between various operating frequencies on an automatic basis (eg. day/night usually have different requirements). On modern receivers this will often be an RS232 interface which can be connected to a serial communications port of the computer. Some receivers may require parallel control interfacing.

ii) A to D conversion. Two analogue to digital converters are required which can be simultaneously sampled at a fixed sample rate (approx. 30-50 samples per second is usual). The conversion time must be such that the input signals do not change substantially during the conversion. With modern converter ICs this is not a problem, since very fast (tens of μs per conversion) conversions can be achieved at little cost.

iii). Data Storage. Large quantities of data can be generated by the receiver system. For example, at 50 samples per second with 8 byte ADCs, and assuming continuous operation, 100 bytes/second $\equiv$ 360 kbytes/hour $\equiv$ 8.64 Mbytes/day are produced. Large storage capacity is therefore needed. If a hard disc is employed (an IBM PC/AT typically has a 30 Mbyte disc), this will need to be periodically off-loaded onto some other medium. Floppy discs are impractical in view of the number which would be required for this amount of data. Cartridge streamer tapes with capacities in the region of 60-150 Mbytes are readily available which fulfill this need. However, if the data is to be transferred to another (large) machine for subsequent processing and analysis then it may be advantageous to employ a conventional $^1/_2$" tape drive (20-30 Mbytes per reel).

iv) Analysis. FFT analysis can be performed on the data collecting/control computer if necessary. In view of the processing times required in such small machines, it is normally preferable to transfer the data to a mainframe or mini computer for subsequent analysis.

The requirements outlined above can be fulfilled by many small desk top computer systems, eg. IBM PC or compatibles. However, a simple microprocessor system is quite adequate and cost effective. A block diagram of one such system, as used at Leicester University, UK, is shown in Fig.3.

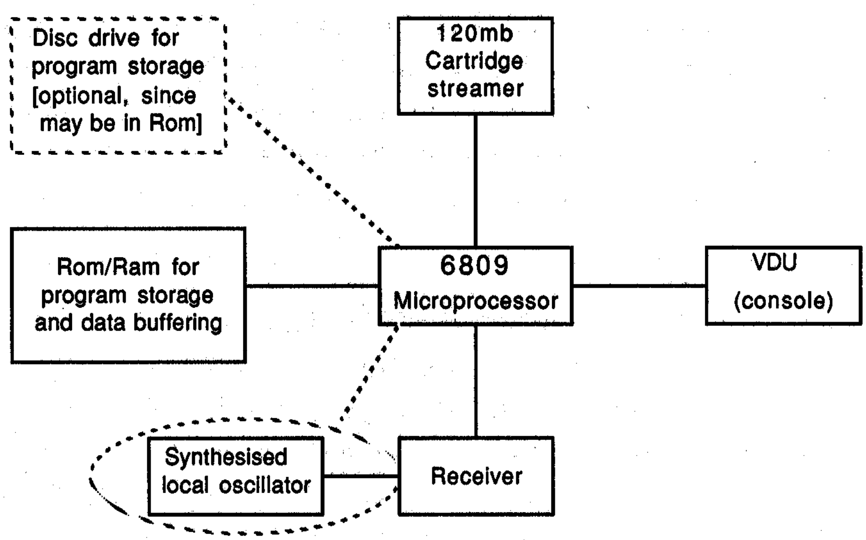

Fig. 3 Block diagram of Receiver control and data loging system.

Data Analysis

a) Determination of TID velocities.

If a TID gives rise to well correlated signatures on 3 or more propagation paths and the locations of the reflection points are known, the horizontal speed and direction of the disturbance across the measuring array can be determined as follows:

Let t_a, t_b and t_c be the relative times at which a characteristic wave of phase φ_o was observed on the frequency-time recordings for three spaced propagation paths. Let the reflection points be at $\underline{r}_a$, $\underline{r}_b$ and $\underline{r}_c$. It is usual to assume that the reflection points are vertically above the mid-points of the transmission paths and that they lie in the same horizontal plane. If the disturbance approximates to an infinite plane wave of the form $\exp[i(\underline{k}.\underline{r}- \omega t)]$ and that it does not change its shape as it propagates, then its characteristic phase φ_o can be written as

$$\varphi_o = \underline{k}.\underline{r}_a - \omega t_a$$
$$\varphi_o = \underline{k}.\underline{r}_b - \omega t_b$$
$$\varphi_o = \underline{k}.\underline{r}_c - \omega t_c$$

Defining $t_o = \frac{\varphi_o}{\omega}$ and $\underline{a} = \frac{\underline{k}}{\omega}$ these equations become

$$t_o = \underline{a}.\underline{r}_a - t_a$$
$$t_o = \underline{a}.\underline{r}_b - t_b$$
$$t_o = \underline{a}.\underline{r}_c - t_c$$

Since $\underline{k}$ is horizontal, $\underline{a}$ is horizontal and there are only 3 unknowns (a_x, a_y and t_o) in the equations. Solving for a_x and a_y

$$a_x = \frac{y_{ac}\, t_{bc} - y_{bc}\, t_{ac}}{y_{ac}\, x_{bc} - y_{bc}\, x_{ac}} \qquad \text{where} \quad x_{ac} = x_a - x_c, \text{ etc}$$

$$a_y = \frac{x_{cb}\, t_{ac} - x_{ac}\, t_{bc}}{y_{ac}\, x_{bc} - y_{bc}\, x_{ac}}$$

Since $\underline{a}.\underline{v} = 1$ and $\underline{a}$ is parallel to $\underline{v}$

$$v_x = \frac{a_x}{a^2} \quad \text{and} \quad v_y = \frac{a_y}{a^2}$$

The horizontal trace speed is then given by

$$v_h = \frac{1}{|\underline{a}_h|}$$

and the azimuth angle of propagation Θ by

$$\Theta = \mathrm{Tan}^{-1}\left(\frac{a_x}{a_y}\right)$$

Since the delay times are measured and the location of the reflection points are known, v_h and Θ can be determined.

It is questionable whether the phase or group trace velocity is measured by the Doppler experiment (Hines, 1974) and many authors avoid the problem by simply referring to a 'horizontal velocity'. Experimental configurations with reflection points separated by tens of kilometres derive velocities which are close to the phase trace velocity, but this is not quite the same as observing the progression of a single crest or trough of one frequency component. Experiments in which the observation points are separated by several hundred kilometres obtain velocities close to the group trace velocity.

b) Spectral Analysis.

The time delays between the various Doppler traces have to be determined and several techniques are available for this task. The simplest method is to produce hardcopy of the Doppler-time variations for each propagation path (similar to that reproduced in Fig.2) and to identify features which appear on all three (or more) by eye. The time delays can then be determined directly from the recordings. This very simple procedure can be adopted only when there are very well defined features with large relative time delays and even then, high accuracy is difficult to achieve. It is also possible to calculate the cross correlation coefficient between the variations on the various propagation paths but again it is difficult to obtain an accurate result, particularly if the disturbances are dispersive which will generally be the case. The method commonly adopted for determining the time delays is that of cross spectral analysis in which a given interval of data is Fourier analysed to obtain the major frequency components present. The phase differences between the same frequency component in the data for each propagation path is calculated. From this the time differences for each path and for each spectral component are obtained.

c) Cross Spectral Analysis.

Cross spectral techniques have been applied to measurements of ionospheric motions with spaced receivers by several authors [see for example, Jones and Maude (1968, 1972)]. The cross spectrum of two signals $x_1(t)$ and $x_2(t)$ is defined as

$$S_{12}(f) = x_1^*(f) . x_2(f)$$

where $x_1(f)$ and $x_2(f)$ are Fourier transforms of $x_1(t)$ and $x_2(t)$, and * denotes the complex conjugate. $S_{12}(f)$ is a complex quantity whose real and imaginary parts are called the cospectrum and quadrature spectrum

$$S_{12}(f) = L_{12}(f) - iQ_{12}(f)$$

This may also be expressed as the product of amplitude and phase functions called the cross-amplitude and phase-spectra.

$$S_{12}(f) = A_{12}(f) \quad e^{iF_{12}(f)}$$

where $$A_{12}^2(f) = L_{12}^2(f) + Q_{12}^2(f)$$

and $$F_{12}(f) = \arctan [-Q_{12}(f) / L_{12}(f)]$$

The cross-amplitude spectrum indicates whether frequency components in one time series are associated with large or small amplitudes at the same frequency in the other series. The phase spectrum indicates the phase relationship between the frequency components of the two time series. The phase differences can be converted into time delays. The velocities of disturbances are determined from these time delays and the geometry of the reflection points as described in the previous section.

The squared coherency spectrum, referred to as the "coherency" can be written as

$$K_{12}^2(f) = \frac{L_{12}^2(f) + Q_{12}^2(f)}{|x_1(f)|^2 \quad |x_2(f)|^2}$$

The coherency $K_{12}(f)$ provides a measure of the noise in the two spectra at a given frequency. When there is perfect correlation the noise is zero and the coherency is unity. Similarly, when the two signals are uncorrelated at a given frequency their coherency is zero. The coherency can, therefore, provide an indication of the confidence that can be placed on estimates of the phase spectrum and hence, on the corresponding velocity estimates.

The Fourier analysis is limited in that the variance of the spectral estimates is not decreased by employing longer data sets. Smoothing the estimates in the frequency domain will decrease their variance and produce a more reliable estimate of the spectrum. The discrete nature of the data, due to the digital recording and subsequent Fourier analysis, makes it impossible to recognise harmonic components whose frequencies are not integer multiples of the frequency spacing. The power at such a component is therefore distributed over a range of frequencies. This is known as leakage. The data are analysed in discrete time blocks and these can be effectively operated upon by rectangular windows, since the Fourier transform of a rectangular window is the sinc function. The large side lobes of this function decay slowly, so leakage may be appreciable. Leakage may be suppressed, however, by smoothing in the time or frequency domains. This corresponds to the use of a convergence factor. Sloan (1969) and Yuen (1977) favour quadratic smoothing (ie. spectral smoothing) against linear (ie. data window) smoothing. Both systems reduce leakage, but in the former case the loss of resolution from smoothing is offset by a reduction in variance.

A Hanning window, which incorporates weights of (0.25, 0.5, 0.25) has been found to be effective for smoothing all spectral estimates in the frequency domain. The width of the main lobe of this window is $W_{main} = 4\Delta f$ where Δf is the frequency spacing of the spectral estimates. Its equivalent bandwidth is (Blackman and Tukey, 1958; Jenkins and Watts, 1968) given by

$$W_E = 0.67\ W_{main} = 2.67\ \Delta f$$

Blackman and Tukey (1958) draw attention to the need for rejection of zero frequency components (including linear trends) from the data before spectral analysis. Even small displacements of the mean from zero can contribute large peaks at zero frequency, and at neighbouring frequencies due to leakage. This is a particularly important criterion for cross spectral analysis where tests indicated that the phase spectrum was particularly sensitive to leakage from zero frequency components.

The data can have near-zero phase spectra, in which case a simple replacement of the sample co-spectrum and quadrature spectrum by their smoothed counterparts in the Equations for $A_{12}^2(f)$ and $F_{12}(f)$ produces the most accurate cross spectral results (Tick, 1967). Thus

$$\bar{F}_{12}(f) = \arctan\left[-\bar{Q}_{12}(f) \;/\; \bar{L}_{12}(f)\right]$$

$$\bar{K}_{12}^{2}(f) = \frac{(\bar{L}_{12}(f))^2 + (\bar{Q}_{12}(f))^2}{|x_1(f)|^2 \quad |x_2(f)|^2}$$

The coherency $K_{12}(f)$ may be transformed to the variable $Y_{12}(f)$ = arctanh $(K_{12}(f))$ via the Fisher transformation, then the 95% confidence interval for $Y_{12}(f)$ can be determined which is independent of frequency (Jenkins and Watts, 1968). Approximately 95% confidence intervals ($\pm\Delta\varphi$) for the smoothed phase spectrum with varying degrees of freedom (ν) can be calculated from

$$\tan \bar{F}_{12} \pm 1.96 \frac{\sec^4 \bar{F}_{12}}{\nu} \left(\frac{1}{\bar{K}_{12}^2} - 1\right)$$

The confidence interval is parametric in $\bar{K}_{12}^2$ but approximately independent of $\bar{F}_{12}$ (Jenkins and Watts, 1968).

d) Rejection Criteria for Phase and Velocity Estimates.

3-station Doppler data provide 3 independent phase spectral estimates, any two of which give a measure of the horizontal velocity of the disturbance at a given frequency. The three velocity estimates thus obtained were rejected by Jones and Maude (1972) if their azimuths differed by more than 10 degrees. The coherency spectrum provides another measure of noise, but Jones and Maude (1972) found that coherency criteria rejected different (fewer) velocity estimates. They suggest that this is simply due to the unreliability of coherency estimates when they are low.

The azimuthal effect of small timing errors is largest in situations where the timing errors and time delays between Doppler traces are comparable. When cross spectral methods are applied to disturbances travelling at high speeds, the 10 degree criterion would thus tend to reject much valid data.

Coherency criteria are therefore more appropriate for the general study of TIDs and coherencies of $\overline{K}_{12}^{2} > 0.50$ are found to be significant. If all three values of Δt fulfil the coherency criterion then the two time delays with greatest coherencies are employed to calculate an apparent propagation velocity.

The most important parameter to be determined directly from HF Doppler data is the time delay, Δt, between perturbations on the different radio paths. The magnitude of the time delays is not only employed to determine horizontal velocities, but also to differentiate between travelling and non-travelling disturbances. The time delays can be determined from either the cross correlation or cross spectral analysis. The cross correlation technique has several disadvantages which are not removed even when digital filtering is applied to the data sets and major difficulties occur for long period waves when data sets are short. The cross spectral analysis is a more powerful technique since it examines the properties of different frequency components without the need for filtering. Spectral estimates are improved by smoothing in the frequency domain, which suppresses leakage and reduces variance. The Hanning function can be conveniently adopted for this purpose. Rejection criteria for velocity estimates, based on coherency have been developed and are generally available in the literature.

Summary

The Doppler technique provides a simple and effective method for monitoring transient changes in the ionospheric electron density distribution. It has been extensively employed for studies of TIDs and by comparing the time delays produced by these disturbances on three or more spaced propagation paths, an estimate of their speed and direction can be obtained. A recent study using this technique as part of the Worldwide Atmospheric Gravity-wave Study (WAGS) has been published by Crowley and McCrea (1988). There are inherent difficulties in the measurement due mainly to the dispersive nature of the acoustic-gravity waves responsible for the TIDs, the presence of more than one wave in the observing region at any given time, and the problems involved in determining the time delays and the precise location of the reflection points. In spite of these difficulties, many of the important characteristics of TIDs have been established from Doppler observations. The low cost and the ability to monitor continuously over long time periods, renders the Doppler method an ideal tool for synoptic studies of thermospheric waves during the forthcoming ISTP period.

References

Blackman and Tukey, 1958, "The Measurement of Power Spectra". Dover, New York.

Crowley G and McCrea I W, 1988. Radio Science 23, 905.

Davies K, 1965, "Ionospheric Radio Propagation" 11 US National Bureau of Standards.

Davies K, 1968, AGARD Lecture Series No. 29.

Hines C O, 1974. J. Atmos.Terr.Phys. 36, 1179

Jenkins and Watts, 1968, "Spectral Analysis and its Applications." Holden-Day, London.

Jones D and Maude A D, 1968. J. Atmos.Terr.Phys. 30, 1487

Jones D and Maude A D, 1972, J. Atmos.Terr.Phys. 34, 1241

Sloan, 1969, IEEE. Trans.Audio and Electro acoust. AU-17, 133.

Tick, 1967, "Spectral Analysis of Time Series". Wiley, London.

Yuen C K, 1977. Proc. IEEE. 65, 984.