Using Digital RF data to derive Doppler Shift and Ionospheric Heights: *Steps Along the Way*

Gwyn Griffiths G3ZIL

A presentation to the HamSCI Science Telecon 31 October 2024

Slide 1

I welcome this opportunity to present work that is very much in progress, hence the *Steps Along the Way*. Preparing this talk has already helped me better structure my thoughts, but it's also left me questioning my initial assumption on high and low rays.

This up-front list of acknowledgements is there for the record, and it is a true tribute to HamSCI connecting and enabling communities to work together.

Slide 2

"Steps along the way" are at the heart of this talk – the HamSCI Personal Space Weather Station database at the University of Alabama and the quick-look plots on its website are absolutely brilliant. But, as we'll see, the data set is far, far richer than can be seen in the summary plots. Getting to the underlying data and working with it weaves throughout this talk.

I'll cover reading in Grape data from digital_RF files, how to calculate Doppler shift using a ridiculously simple algorithm, how we can separate out Doppler shift, amplitude and frequency spread from different propagation modes into separate columns, and by calculating height of reflection perform a reality-check on the Doppler measurements and assumptions we have made.

However, the incredible richness of the Grape data sets brings complications. In this case study we'll meet bimodal Doppler spectra, where simple algorithms are not appropriate. Many weeks of work have gone into trying different methods of separating propagation modes, and a robust, general solution still eludes me. I suspect this is also a major challenge for HamSCI researchers. Nevertheless, in this case study I've made sufficient progress to share what appear to be credible results for bimodal Doppler shifts, but pose a conundrum when converted to reflection height and compared with predictions from PyLap modelling.

Slide 3

Let's start with the quick-look spectrogram for today's test case. I'd like you to sit well back from the screen, and burn into your memory the traces between 14:00 and 20:00 UTC.

And I'll forgive you your quizzical look as you ponder, why has it taken him weeks to examine this plot?

Slide 4

As befits a test case this data has several nice attributes: it is clean signal and looks to have a high signal to noise ratio. Those attributes make it amenable to the ridiculously simple algorithm for finding Doppler shift. There is some vertical banding from an unknown cause, but it should not worry us, neither should the ghostly traces. I'm pretty sure that those are from the propagation mode two-hop side scatter.

Slide 5

Our first step is to read in the data from the digital_RF file. Having found the site and day of interest on the PSWS website we click to download the data, unzip the folder, which will be automatically named ch0.

There's get-you-started code on the digital_RF Github site, here are the lines of Python code needed to get the data into a numpy array. Nathaniel's RX888 was set up to record nine time standard stations, and with the array index starting at zero in Python, array index 8 for frequency is 25 MHz, which is what we want. The data rate is 10 Hz, and so I'm going to read in 720,000 samples to take me to 20:00 UTC. We set data folder location, use two digital_RF functions to get details on the channels in our folder and the start time, before the simple read function into the array input that will have all frequencies, from which we extract 25 MHz into the data_25MHz array as a series of complex numbers representing the magnitude and phase of the baseband signal centred at 0 Hz.

Slide 6

I like to get a feel for the actual data I'm working with, and this simple time series plot showing the in-phase and quadrature amplitudes serves that purpose.

It also reinforces my comment about a high SNR.

Slide 7

Which means that this ridiculously simple algorithm to estimate Doppler shift can be used. Don't let its full title: Argument of the Complex Autocorrelation at One Lag put you off.

Expanding our time series, lets take two samples, one at time t and the next one, at time t plus dt, where dt is the sample interval, here 0.1 seconds.

We'll form the complex conjugate of the second sample, simply by changing the sign of its imaginary part.

We can plot the first sample and the complex conjugate of the second sample on the complex plane, on the right.

Next we multiply our complex signal at time t by the complex conjugate at time t plus dt, here plotted as the magenta diamond. Taking the phase angle of the result, where zero is along the real X axis, gives us -36˚, which we divide by the number of degrees in one cycle, 360˚, and by the sample interval dt to give us 1 Hz: The frequency of our input signal – from just two samples. Of course our Grape Doppler will not be a pure sinusoid, and so we'll average the correlation function over one minute, 600 samples, before taking its argument, or phase angle.

Slide 8

Also in the time domain I estimated signal level by adding the mean and root mean square values in each one-minute interval.

Recalling the spectrogram from the PSWS website you will not be surprise to see two regions, these are signal levels from two quite different propagation modes.

They have distinctive, and non-overlapping, signal levels. The high signal level is from one-hop propagation, in orange. Present before 25 MHz opened for one-hop and as the eclipse temporarily closed the band, and with a signal level some 40 dB lower, in brown, the propagation mode was most likely two-hop sidescatter. PyLap modelling, bottom right, suggests that the scatter was from the Gulf of Mexico. That is the region where transmit rays from WWV and simulated transmit rays from W2NAF overlapped.

Having identified the two propagation modes from their non-overlapping amplitudes we can easily separate their Doppler shifts, hence the bottom graph shows one-hop propagation Doppler in blue, and two-hop sidescatter in cyan. There is no confusion between the two propagation modes.

Slide 9

There are cases where we can separate propagation modes based on their frequency spread – their spectral width. Frequency spread can be estimated in the time domain via the magnitude of the autocorrelation function. The tricky part is estimating the noise N, which I have assumed to be zero in this high SNR example. I also cannot find what this spread metric represents, is it 50% 90% of the signal power, or some other value. Insights welcome!

In this time series we again have a clear and non-overlapping distinction between one-hop and two-hop sidescatter signals. My experience is that there is normally rather little variation in two-hop sidescatter frequency spread, as here before the band opened. In contrast, as the band closed due to the eclipse, the very high initial value, followed by a decay, appears to be a specific eclipse signature. The outgoing and the return rays cross regions affected differently. I won't pursue this further today – my main message is that there are situations where frequency spread can be used to separate propagation modes.

Slide 10

I'm still building up my intuition for interpreting Doppler shift, not helped by its dependence on frequency. Converting Doppler shift to reflection height removes the frequency-dependence and enables me to compare with expected diurnal changes and ray-trace models such as PyLap. Here I have essentially followed the method in this paper by Kristina Collins and colleagues. My initial height is from PyLap, with some concern, as I had to use a smoothed sunspot number of 170 to get the path to open at 14:20 rather than the accepted SSN for April of 103.

Nevertheless, the expected features are present, the morning descent, bottoming-out, then a rise at the start of the eclipse, with a height rise of at least 38 km, the band closing before the expected turn-around.

And so we have a credible picture of the change in reflection height from Doppler shift calculated by the very simple autocorrelation algorithm.

Slide 11

Before we leave the autocorrelation algorithm let's look at its results on a 7 km true line of sight path between WWV and two co-located Grape receivers. Through the commitment and effort of Dave Swartz W0DAS and colleagues we have results from a Grape V1.2 DRF reporting as W0DAS 1 on the left, and a Grape RX888 WsprDaemon reporting as WW0WWV/0 on the right.

On the left, the frequency shift from exact 10 MHz for the Grape 1.2 at about -11 milliHertz is most likely from an uncorrected sound card oscillator. Dave uses a horizontal dipole for this receiver, and I suspect the greater variation in frequency shift, and bump in frequency spread were likely from multipath near vertical incidence skywave – despite being only 7 km distant.

On the right we have the Grape RX888 – only dependent on a GPSDO for its frequency accuracy and stability. And we have impressive results: a mean offset of 0.28 milliHertz and an rms variation of 0.16 milliHz, which is 1.6 parts in ten to the power eleven. There is no evidence of NVIS multipath, probably because the antenna here is a vertical.

This reference site close to WWV reconfirms the Grapes as high precision scientific instruments: for the very highest performance – use an RX888 and GPSDO and seek to avoid multipath.

Slide 12

Earlier, I suggested you sat back, now's the time to peer closely at this zoomed-in spectrogram from the PSWS website. There are quite distinct – for want of a better description – 'Loops' - as the band opened and as it closed. There were other times at which the Doppler trace was likely bimodal – for instance … here … and here. And so, you might well ask – might the Doppler shift be bimodal throughout, just that we cannot separate at this resolution?

The implication is that there were two propagation paths – what could they be?

Slide 13

There are clues as to what the two propagation paths might be in this ray trace that HamSCI just happens to use in its publicity material. This is for 15 MHz, but the principles are the same for our test case at 25 MHz. There is a skip zone, where one-hop propagation is not possible. The ray that propagates to the very edge of the skip zone is not the ray with the highest elevation. That would be the case if the reflection height did not vary with take-off angle from the transmitter. But it does. Rays at higher take-off angles are refracted higher up in the F region and consequently reach a greater distance than rays at somewhat lower take-off angles.

The implication is that, for a certain distance beyond the edge of the skip zone there will be two rays reaching a receiver – one from a high angle and the other from a low angle. They will not have been reflected from the same height.

Slide 14

With a few lines of Python code added to PyLap we can extract the distance at with rays for a span of elevations (or take off angles) reach the ground. Here is a plot of that data for 14:20 UTC just as the path from WWV to W2NAF opened at 25 MHz. From the PyLap model it is the X, or extraordinary wave, that was first received, from a single elevation angle. A few minutes later, the 'nose' of this plot would have moved to the left, signals would be received at a shorter distance.

Slide 15

Here's that picture for ten minutes later. We'd expect *four* propagation paths – an O and an X each at the low and the high elevation angles. But … the O and the X, at least in PyLap, cluster together – the difference in elevation angle between them being smaller than the difference between the high and the low rays after a few minutes. So its possible we'd only be able to differentiate between the wider-separated low and high rays and not the closer-separated O and X waves (that is, without a polarimeter – which would be neat).

Slide 16

To separate the closely-space Doppler shifts hinted at in the PSWS spectrogram we need some custom frequency domain analysis. Earlier we saw how to read in an IQ time series at 25 MHz.

These few lines of Python code perform a Fast Fourier Transform on Hann-windowed data in chunks of one minute. FFTshift is used to move the zero frequency to the centre for plotting.

Here are four spectra spanning just ten minutes. As an aside, the continuous time series from the Grapes is so valuable in these studies.

At 14:20 we have a broad peak, at a low signal level, centred at about +0.5 Hz, from two-hop sidescatter propagation. Four minutes later, superimposed on that broad sidescatter peak we have a single narrow peak from one-hop propagation along the great circle, just as the path opened. That single peak is so transient – for two minutes later we have two peaks, and the signal level is higher by about 20 dB. Four minutes later, the separation of the peaks has increased, and the level of one has dropped.

There is no sign of four peaks.

Slide 17

Plotting spectra is all well and good, but what I really need are time series of the Doppler shifts at the two spectral peaks. For me, this has been a non-trivial task, with a few dead-ends along the way.

Let's go through the problem, step by step, with this example spectrum for 18:00. Our eyes see two well-separated Doppler peaks. Having tried a simple inflexion point algorithm, then a method based on kernel density estimation, I've settled on a function in Python's signal processing toolbox – find_peaks_cwt, where CWT is an acronym for Continuous Wavelet Transform. I'll outline the CWT approach in the next slide, but for now, what's key is that the function lists the index of amplitude maxima. From those indices we can get the appropriate frequency bin – we now have Doppler in increments of one sixtieth of a Hertz.

For a higher frequency resolution I can use a three-point interpolation – no wider – as we do not want to contaminate one peak with another if closely spaced.

We now have digital Doppler estimates for two co-propagating paths with a resolution of a few milliHertz.

Slide 18

If we think of the more common FFT as a method of fitting sinusoids to a signal, in a continuous wavelet transform we fit a wavelet with a range of widths. One can choose from a number of wavelets, although the default in the Python toolbox, the Ricker wavelet, is well suited for finding peaks.

The actual python code is straightforward, but its use is not entirely satisfactory. We have to 'tune' the range of widths based on characteristics of the dataset. To be specific, I found that with the minimum width at 1 there would be instances of false peaks, from narrow spikes close in to either the true first or second peak. And if the maximum width was more than 4 the algorithm would ignore the normal narrow peaks – they would not appear in enough of the specified range of widths to be accepted.

Slide 19

Having introduced the CWT as working well for the spectrum at 18:00 here, from the same test case dataset, are two examples where it fails.

On the left, the algorithm got the second peak completely wrong at 17:36. it ignored the higher-level peak nearby, and it also ignored a clear peak just HF of the correctly identified peak. Setting minimum width at 1 enabled the CWT to correctly identify the second peak. Consequently, I added a second pass with minimum width of 1 when the initial second peak was below a signal level threshold or greater than a frequency shift threshold from the previous value obtained with width of 2.

At 14:59, recalling that the minimum width was 2, the algorithm reckoned the second peak was at -0.278 Hz. Completely wrong. It rightly ignored the narrow peaks on the left side of the main peak. My guess is that the true second peak is slightly HF of the correctly identified peak. However, setting minimum width at 1 did not pick it out – it's a shoulder rather than a true peak.

Slide 20

Using the analytical methods I've just described here is what the first hour or so of bimodal Doppler data looks like. The zoomed-in 'Loop' from the PSWS is shown for comparison.

Recall that the methods so far only assign labels first and second peak by their amplitude. I'd expect continuity for the two traces – clearly neither the postulated low ray nor the high ray always had a higher signal level.

Slide 21

While on the topic of signal levels, it is useful to look at a perhaps under appreciated component of the PSWS measurement system – the antennas. Setting aside the WWV vertical quarter wave monopole, quarter wave above ground, with nine radials sloping at 45 degrees, here are some notes on Nathaniel's off centre fed dipole antenna Nathaniel. What's clear is that over the range of elevation angles PyLap suggests we're seeing in this study the amplitude response may vary over a 10 dB range. Furthermore, the antenna analysis suggests WWV is received on a sidelobe, and near the edge of a sidelobe at that. These are only model predictions – it might be instructive, useful and even fun to try and measure the beam pattern of Nathaniel's antenna over elevation and azimuth angles relevant to WWV.

And I should mention the recently formed HamSCI Antenna Project group that will undoubtedly look at antenna issues in detail.

Slide 22

Returning to the problem of assigning the two Doppler shifts to the postulated low and high rays, with some trepidation I explored 'Machine Learning', I stress, very much as an amateur!

First, I needed a 'training set' an initial set of Doppler shifts correctly assigned to the postulated two rays. A set of ten seemed sufficient. I'm after capturing the short-time trend within a highly variable natural phenomenon. For the initial training set I reckoned that automatic simple linear regression of the first peak Doppler values with time, followed by comparing the residuals for the First and Second peaks would suffice.

Having got a training set I tried two machine learning methods for assigning subsequent peaks to set one, and by implication, the second peak got assigned to set two. It appears simple – use the 10-minute training set to predict one-minute ahead, then test whether the first or second peak Doppler at that time most closely matched the prediction, then assign the closest to set one. Increment the training set, and repeat.

Slide 23

Here's what the two methods produce. At left is the initial training. Simple assignment by amplitude of first and second peaks gets three time intervals wrong.

It's obvious visually, but we need an algorithm to run automatically. The black line is the least squares fit to the First peak data points, the open circles. As the algorithm steps through the residuals it identifies the three cases in which a swap of first and second peak values give reduced residuals, and then reassigns.

On the right is an example of two machine learning algorithms predictions for one-step ahead. The open circles form the 10-minute training set. The solid black circle at 0.1 Hz is the prediction from the Support Vector Regression method with a Radial Basis Function kernel – too much jargon, I know. While it is, in this instance, a useful prediction, in that it would correctly lead to us choosing the first peak Doppler at this time interval as the next one for this ray, that is, the thick black open circle, over the second peak Doppler, the blue dot, it is not as good a prediction as the one from Facebook's Prophet method, the magenta circle, with the error bars that Prophet provides.

Slide 24

The outcome of all this data analysis and assignment for the first hour is this graph: Doppler shift time series for two rays with consistent data points except for two – a 97% success rate.

I'm actually very encouraged by this graph – I think it shows the exquisite measurement capability of the Grape receiver, and some progress in data analysis for bimodal Doppler.

The miss-assignment highlighted with the red arrow is most curious. The spectrum for that time, on the right, shows that only one peak was found, which was assigned to the High Ray. The very clear peak at 0.384 Hz was completely missed. It fits nicely in the gap in the Low Ray series. There's a way to go for 100% success…

Slide 25

Just how much more there is to do is clear from this graph over the whole five-hour data set. We did well for the first hour. And for the 'Low Ray' the automated analysis looks sensible. But clearly incorrect Doppler peaks have been found and assigned to the 'High Ray', escaping my existing quality control.

Slide 26

Earlier, I referred to the usefulness of deriving height of reflection from the Doppler shift as a reality check, a diagnostic.

Let's step through what the graph at left shows. The red and the green triangles are the heights of reflection from PyLap for the Extraordinary and Ordinary rays for the low and the high paths. In magenta is the height of reflection based on the PyLap initial value and Doppler shift from the autocorrelation algorithm. Very credibly it tracks the PyLap height of reflection for the low ray. The heights from the bimodal spectrum analysis, postulated as the high and the low rays, fall either side of the autocorrelation-derived values – which is good.

However, what I've called the High ray is nowhere near the High Ray reflection height from PyLap. It is also the wrong shape – the PyLap trace is a magnified mirror image.

And so I am flummoxed.

The change in height for the high ray from PyLap implies a large, negative, and reducing Doppler shift. Going back to the spectrogram – there is no sign whatsoever of a large negative Doppler shift. Why don't we see the high ray?

If not high and low rays, what are the two Doppler peaks that we see? Are they the X and the O? Is my method of extracting or interpreting reflection heights for the X and the O rays at fault?

I end with those questions. I've learnt a great deal from the steps along the way about just how difficult it is to process bimodal, let along multimodal Doppler spectra, and even when one ends up with a decent bimodal Doppler data set, it does not match up with my preconception – but that's science for you.

Slide 27

I'll finish with some graphics from a 2023 report by Carrano and Rino from Boston College on wave-optics modelling of HF propagation. First, it is uncommon to see an arrival angle with range plot, top right, similar to ones I've shown today, second) the high ray trace is 'flat topped', not as divergent as in PyLap and finally, in plot B bottom right in a disturbed ionosphere do we see such a plethora of Doppler peaks in Grape receivers, if so, how might we measure and track them all?

Unused slides

Slide 28

Here's an indication of how well the Prophet algorithm manages to forecast the next Doppler shift in the low ray dataset. There is a definite lag after reversals in slope, such that the actual Doppler shifts were outside the Prophet error bounds (and I do not yet know what they signify). It does pretty well on a consistent slope. We can easily spot the two outliers discussed previously, the red and brown arrows. Perhaps there are tuneable parameters I have yet to find to reduce the lag.

Slide 29

I'm grateful to my friend Peter K6RFT for an introduction to Joe Gorin. Joe suggested that I model a sinusoid at the frequency identified as the first peak, and coherently remove it from the data, followed by reanalysis to look for the close-in second peak. Despite refinements to this basic approach: modelling a time-varying amplitude and finding the optimum phase for the modelled signal, results were mixed. In this example one-minute record we see that the automated method worked pretty well in the central part of the time series, but amplitude and phase errors at start and end led to large residuals and a spectrum that showed a spurious mirror-image peak.

Slide 30

I showed the plot on the left at the March HamSCI workshop, picking out the change from one-hop to two-hop sidescatter propagation during the October 2023 eclipse on this path using a frequency spread and signal level scatterplot. It was only when writing my talk that it occurred to me what the green dots might be – the two ray, that is, low and high ray zone. This middle graphic sets out this hypothesis, published in my article in July RSGB RadCom. But I still lacked sure-fire evidence. The Grape spectrograms provide useful evidence. Bottom right is the spectrum for two-hop sidescatter – low amplitude and a wide. Above it we have the bimodal spectrum with two rays. Signal level is high and although individual peaks are narrow, the effective width is much greater. And, as we've seen, the separation of the peaks can change quickly.