Errors In Incoherent Scatter Measurements CEDAR Workshop 2003

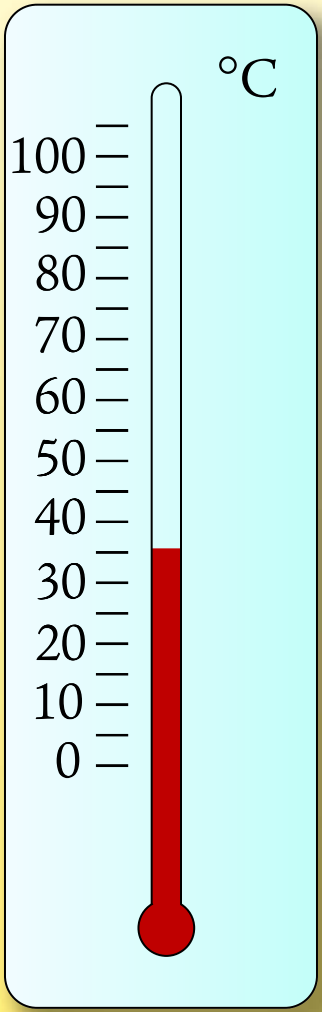
Michael Sulzer

Arecibo Observatory/NAIC

Consider this simple thermometer and some of its associated measurement *errors.*

There are random errors. Example: each person who looks at the level might associate it with a different position.

There are systematic errors. Examples: the tube might be mounted at the wrong level. The tube might contain an incorrect amount of alcohol.



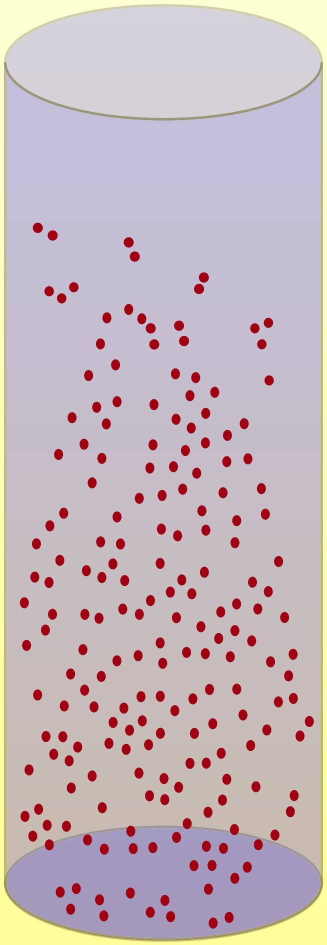
Nevertheless, there is something simple about this measurement and deterministic in nature at the accuracy that we would normally associate with it.

Many measurements are much more complicated in nature, and they are usually the ones that are usually of interest in science.

Consider this tube containing a gas at a very low density in a very strong gravitational field.

Suppose that we have a *Magic Molecular Marker* which allows us to know the position of each molecule at any time.

There are two ways to measure the temperature:
1. Since the G field is strong there is a gradient in molecular density with height.



2. Since we know position with time, we know velocity, and we can associate the motion of the molecules with a Maxwellian velocity distribution.

Both of these measurements are inherently statistical in nature: errors should decrease if we "average" over time.

To summarize some things about measurements:

Simple

Some measurements consist of evaluating a quantity that is, for all practical purposes, deterministic. (Example: the level of fluid in the simple therometer)

There are both systematic and random errors in such a measurement. (Possible example of the first: readings by many different people; of the second: poor design and manufacture of the thermometer.

More Complicated

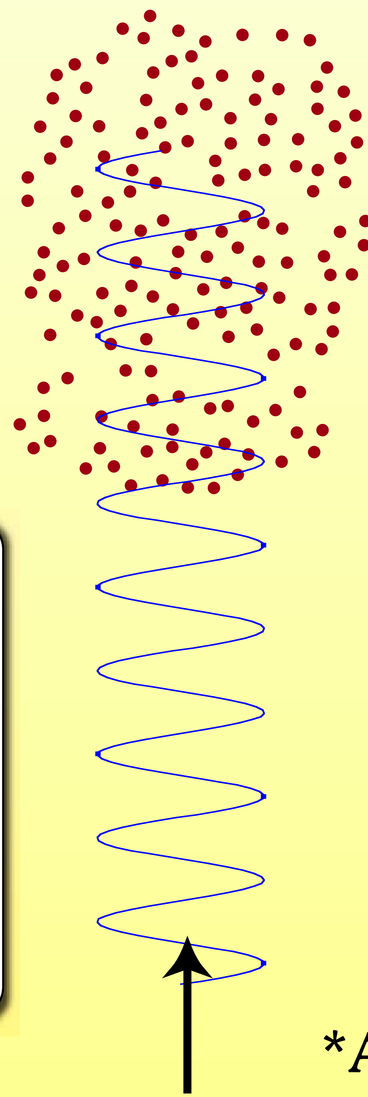
The quantity we want to measure might be statistically related to a set of "simple" measurements. (example: temperature of a gas)

Furthermore, there might be no easy way to go from the "simple" measurements (each with its own set of errors!) to the required quantity. How does one define and obtain the "best" answer?

Incoherent Scatter measurements are inherently statistical in nature: we measure approximations of "expected values".

Incoherent Scatter is the sum of the scatter from many electrons. The sum depends on the phases of the individual electrons.

It is better to think in an equivalent way. IS measures the spatial Fourier component of the electron density fluctuation that matches the radar k vector.



We have N electrons, labeled by i:

$$S(t) = \sum_{i=0}^{N-1} a e^{i(\omega_0 t - k r_i(t))}$$

Signal

Amplitude

Phase

Location

The evolution of this Fourier component in time gives the power and spectral (or ACF*) measurements.

*AutoCorrelation Function

It is also natural to develop the theory of IS using the spatial Fourier transform of the electron density fluctuation

This provides a way to compare measurements with the theory and thus to measure useful parameters about the ionospheric plasma.

The power in the scattered component is closely related to the electron density.

The time variation of the scattered component (through the spectrum or ACF) gives temperatures, velocities and ion composition.

The theory, interesting though it is, is not the topic of this talk. We have what we need with the above explanation.

ISR is older than "convenient" computing!

The HP-35

1972-1975

Courtesy of the HP Museum

The first easy way to calculate a sine!
Marketing studies said it would not sell. It sold out everywhere in the world. People always underestimate the value of ease of computation. It cost $395, a lot of money then.

Incoherent Scatter is 14 years older than the HP-35. We are still assimilating improvements in computing into our experiments and analysis. Computing used to be very expensive and difficult.

In this talk we will:

1. Look at the errors in techniques, both old and tested, and new and under development.
2. First we will look at the power profile, ISR's most basic way of getting information about the electron density, but also a good way of leading into the lag profile, the method for getting spectral information.
3. We will consider the "long pulse technique" only, the simplest way to get spectral information, but we are still developing techniques for its analysis.

Here is an interesting problem:

1. The data we put in the database consists of values with errors.
2. The errors have two parts: random and systematic.
3. The data we put in the database not distinguish between the two.
4. This matters: if the user wants to do further averaging, the statistical error decreases, but the systematic does not in general.
5. How can we eliminate this problem, and thus do a better job of providing data for the community?

A solution:

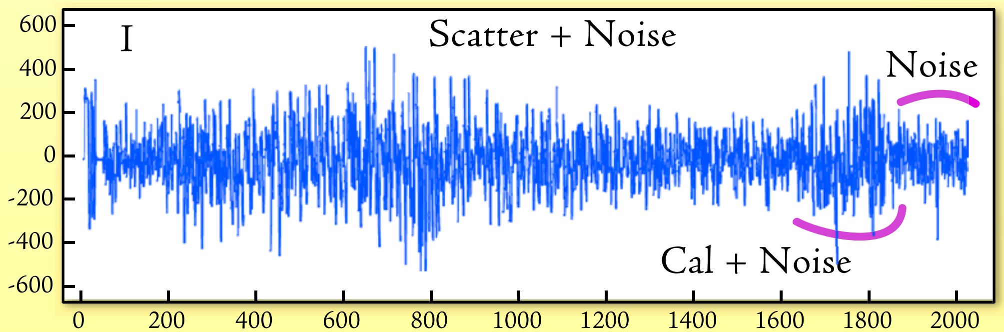
Most of the systematic errors in the parameters derived from IS data can be greatly reduced by the use of good numerical processing.

In particular, the application of inverse techniques can potentially reduce these problems to very small levels. Therefore: make the job of the data user easier by greatly reducing the systematic errors!

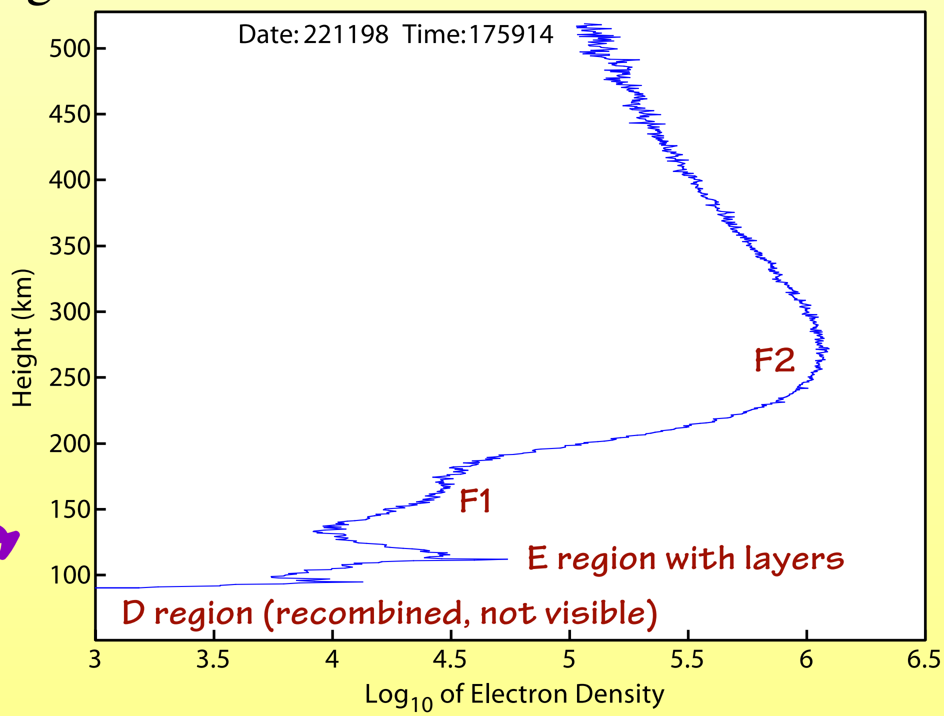
We can do this, using convenient, powerful computing and new processing correctly applied.

The power profile: a simple application of IS for measuring electron density

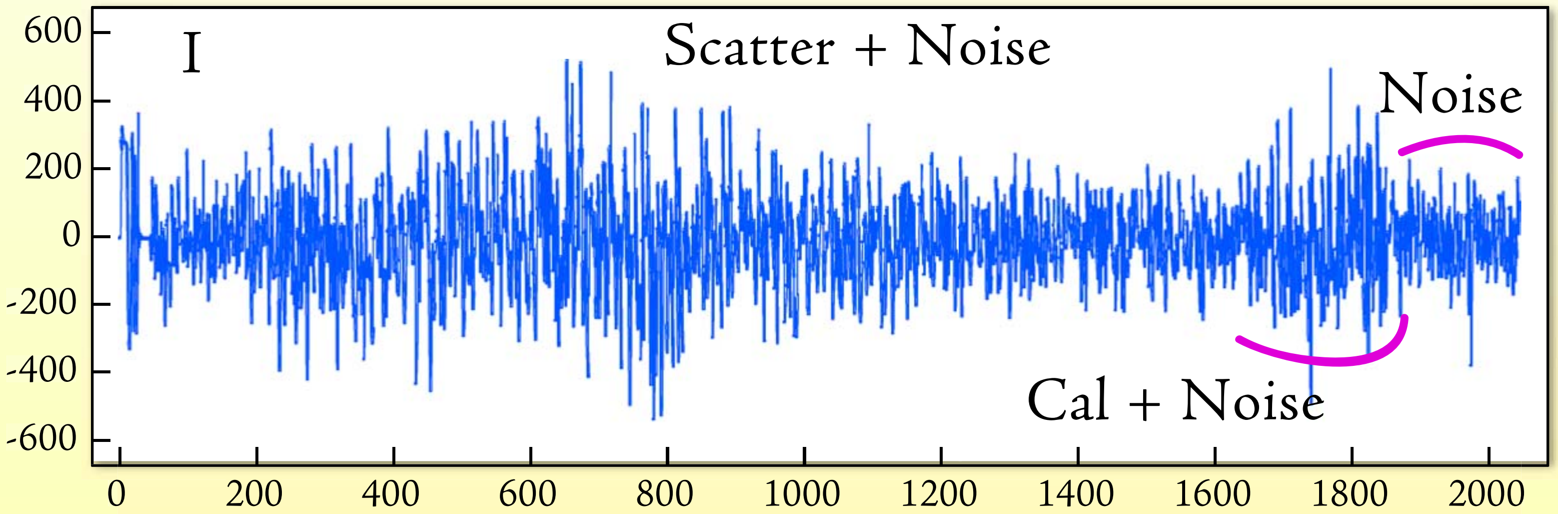
1. A power profile program can use a short pulse or a coded pulse intended for pulse compression. Why? It does not need to measure correlation from one time to the next.
2. The errors in a power profile measurement are fairly simple to study:
   a. First, there are statistical errors relating the noise-like nature of IS.
   b. There are errors in obtaining an absolute calibration.
   c. There are more subtle errors like changes with altitude of the ratio of electron to ion temperature.



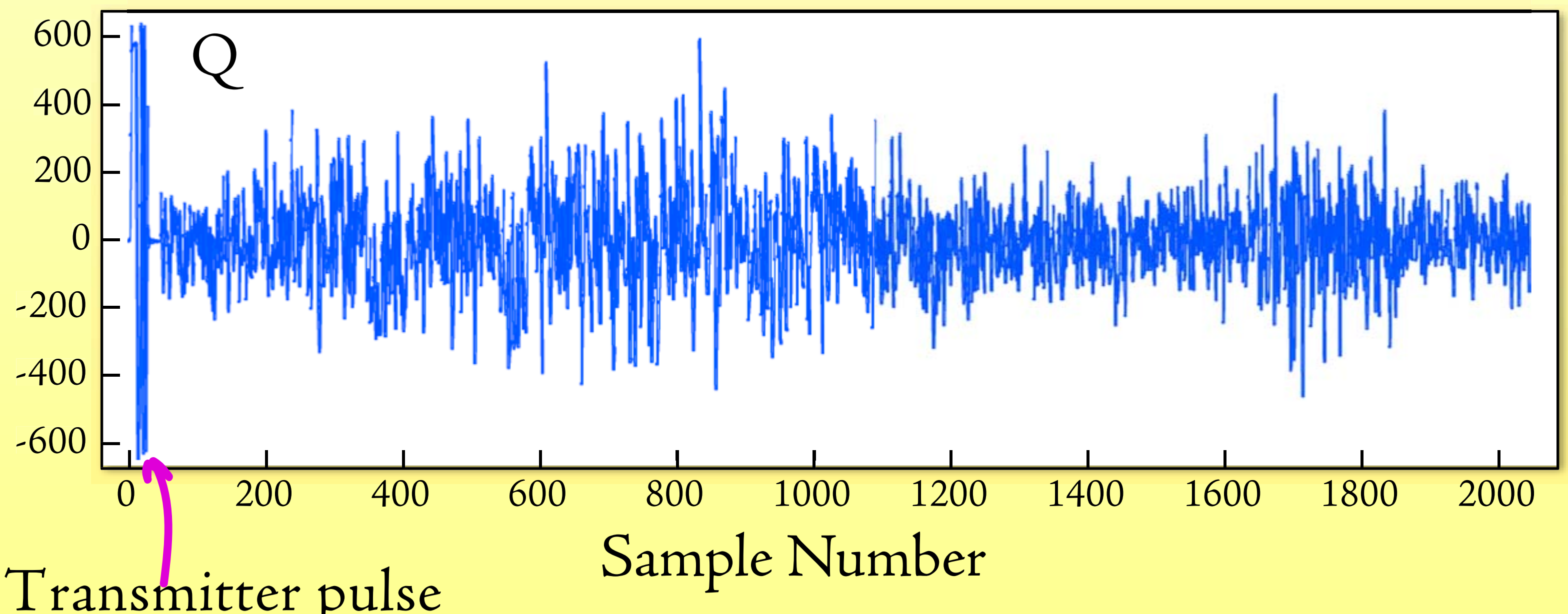
How do we get from raw samples to the electron density profile?
How do these errors affect the result?

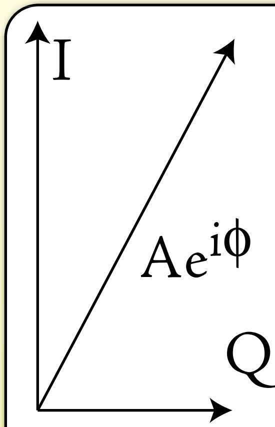


The Raw Samples of Single Radar Pulse and the IS signal, Calibration and Noise



Errors from one pulse are huge. We must process and repeat.

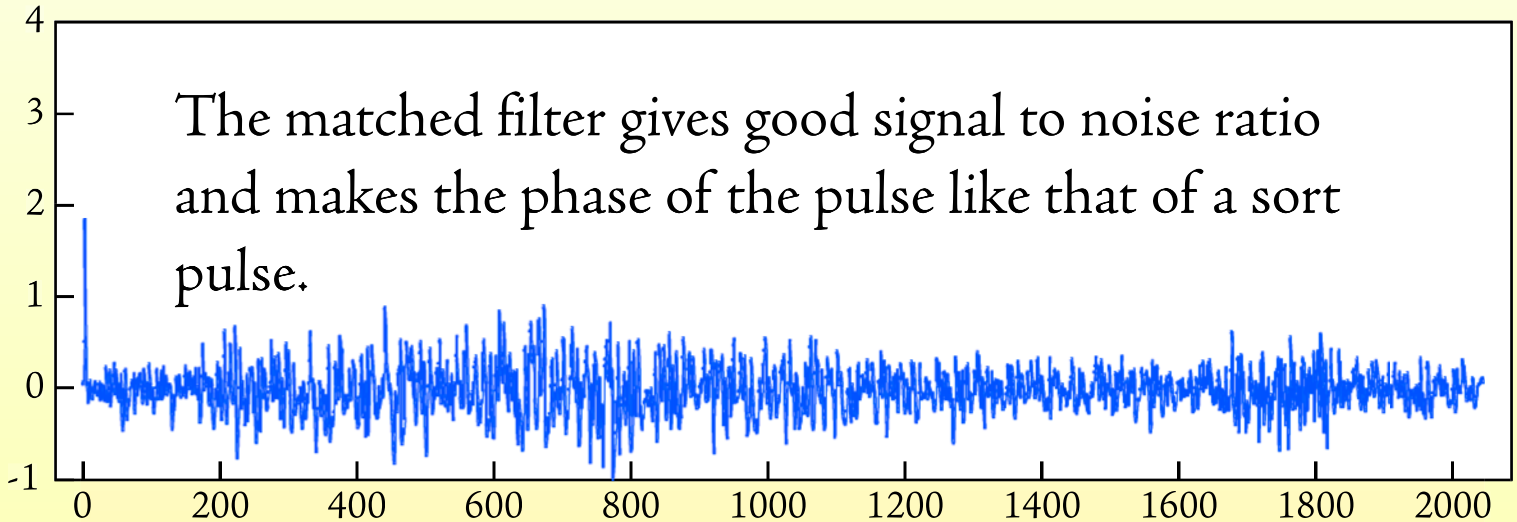


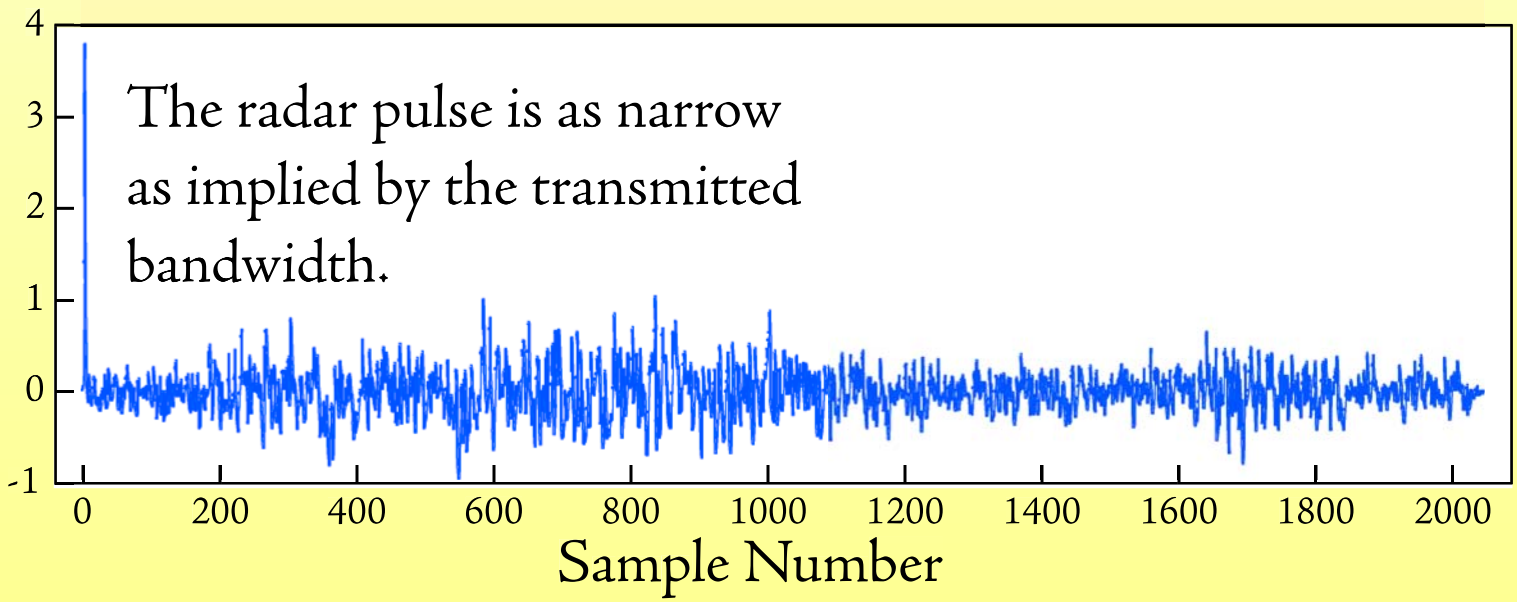


The Signal has amplitude and phase. We use the projections.

Old way to make I and Q: analog baseband mixer. New way: IF sampling, digital processing. Why? Smaller errors.

The same, but after "decoding", passage through a matched filter

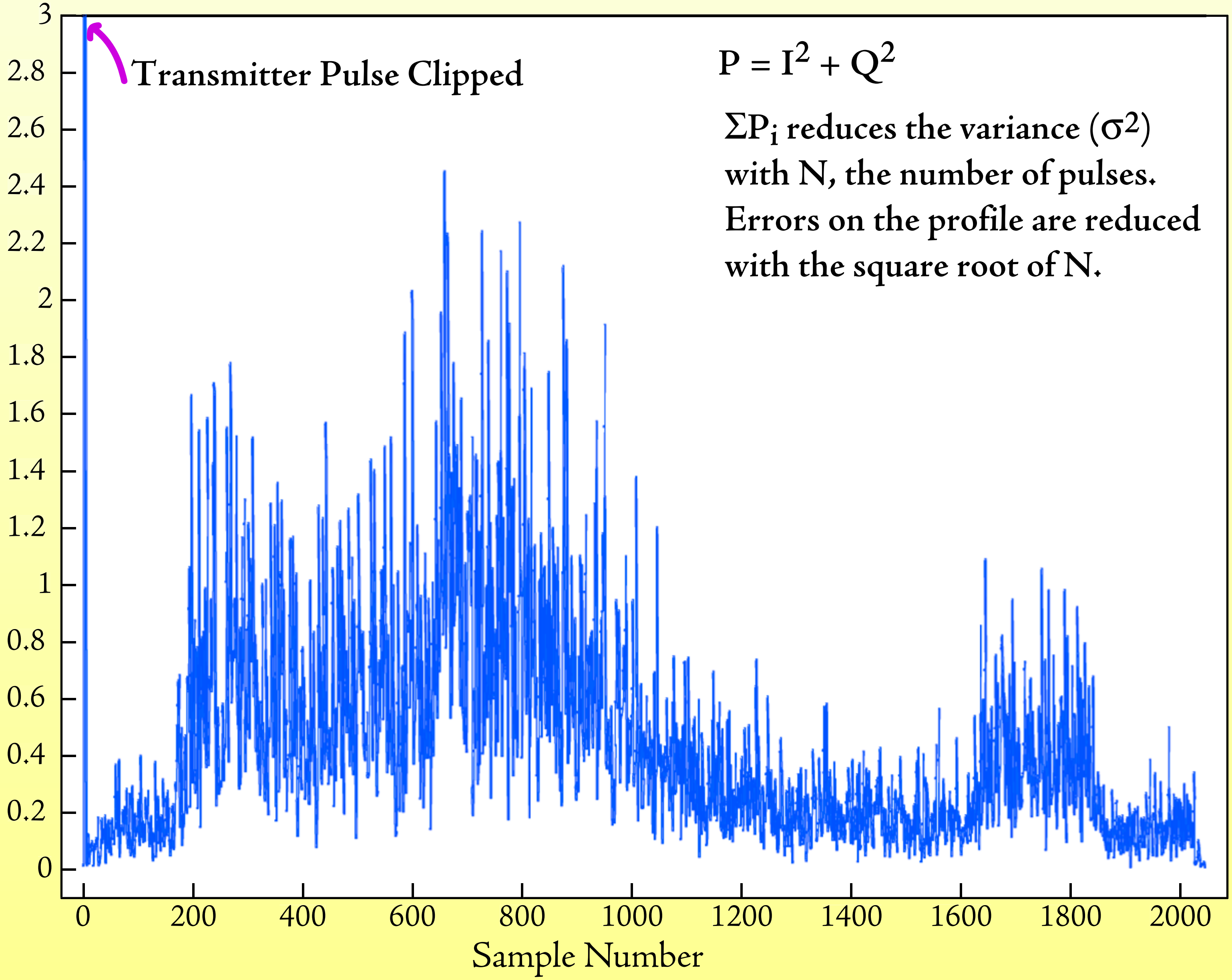




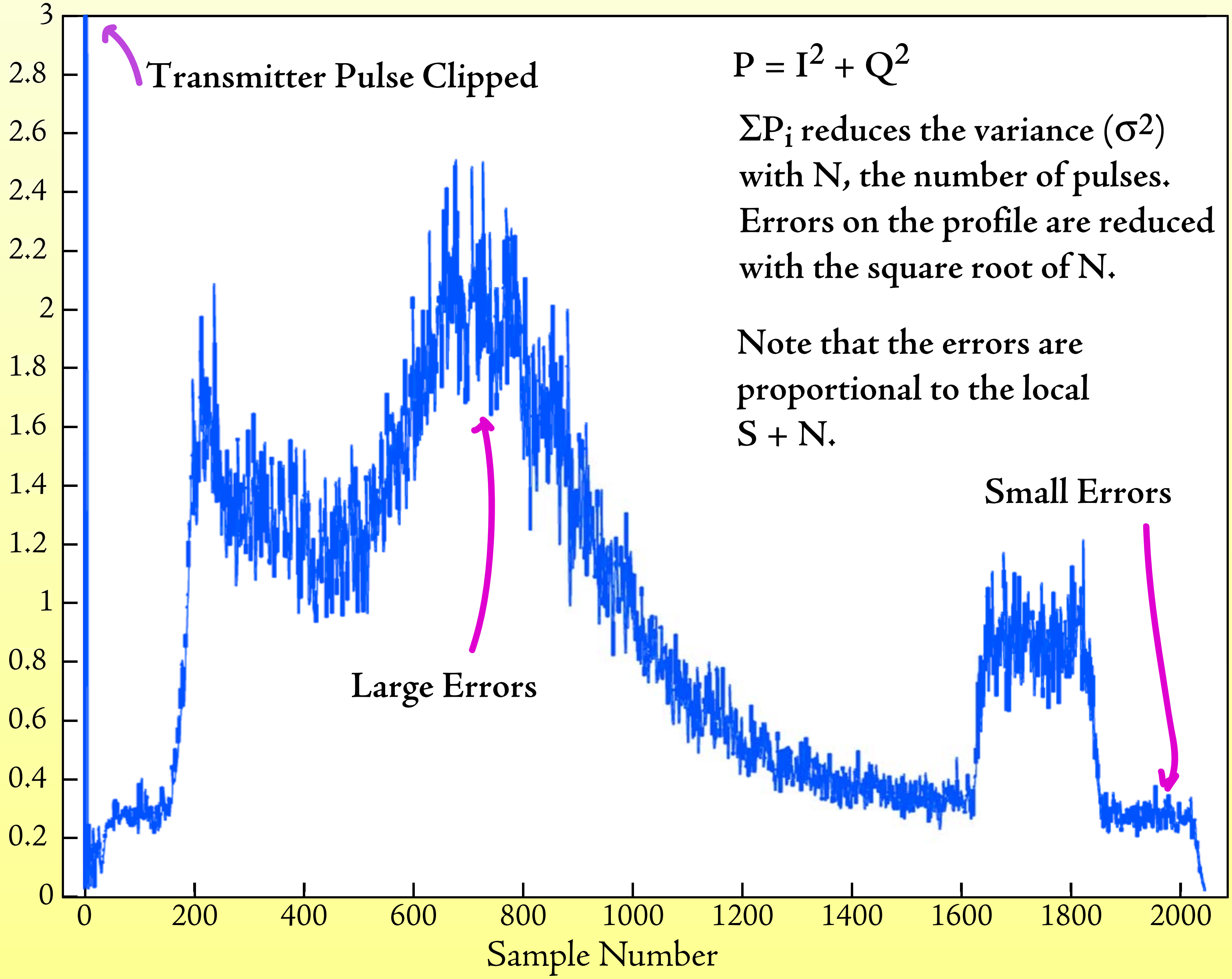
The analog bandpass is "flat"; the matched filtering is done digitally in the computer. Other processing could allow other uses for the data. (meteors, for example)

The use of a "coded" pulse and processing incorporating "decoding" reduces errors due to smearing, and keeps the signal to noise ratio high.

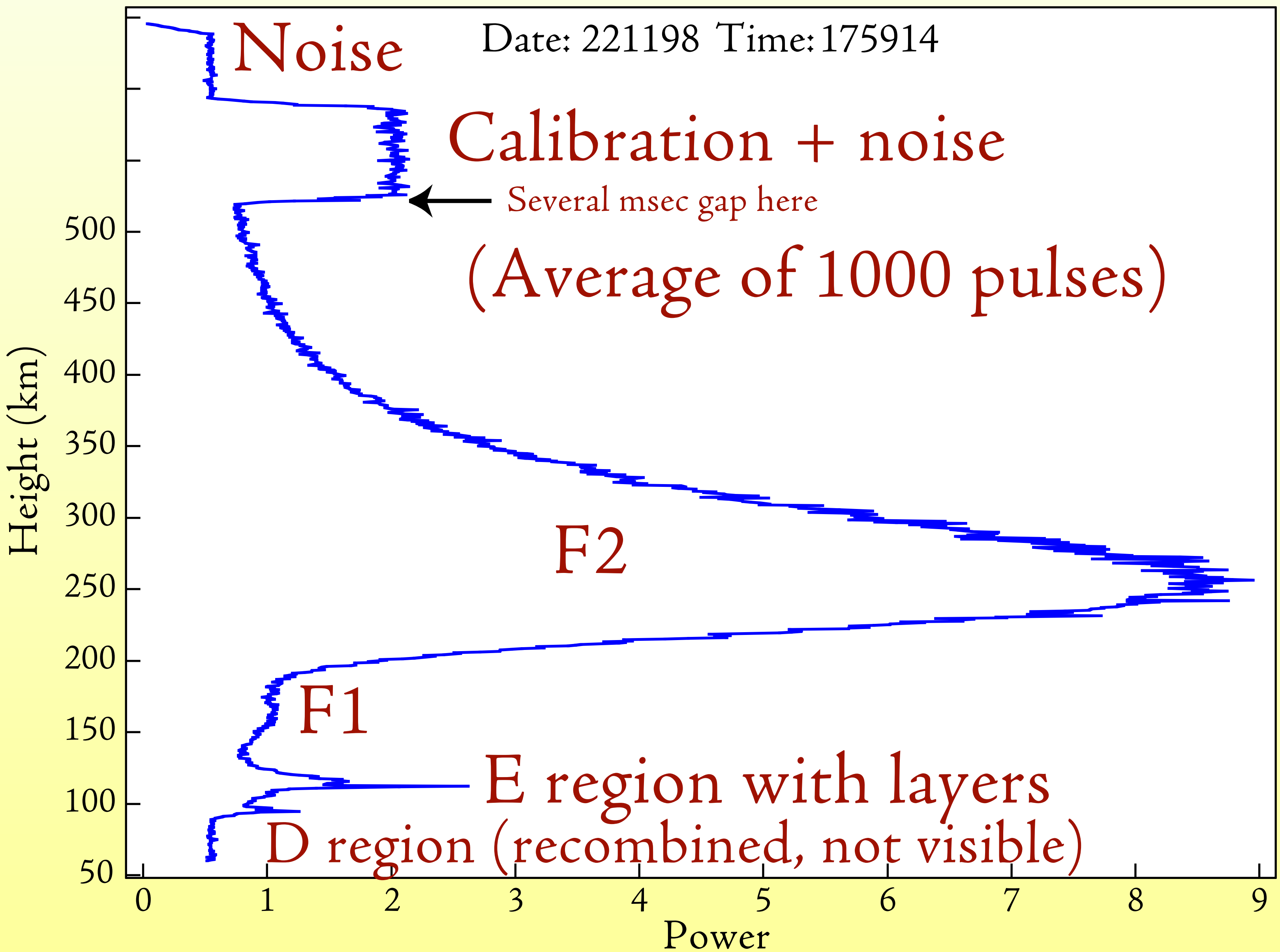
Compute the power, and accumulate for four pulses.



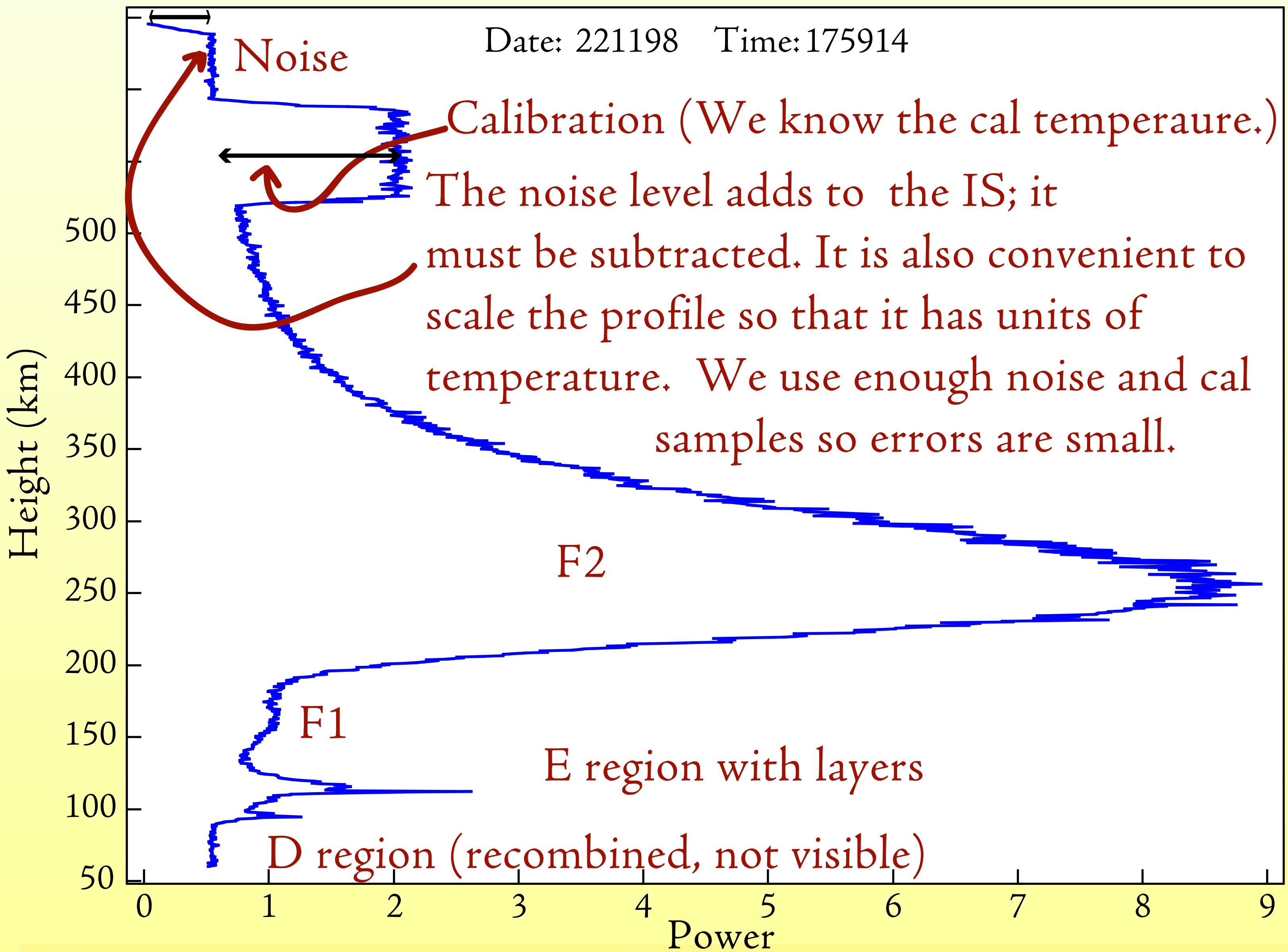
Compute the power, and accumulate for 75 pulses.



A Power Profile from the Arecibo 430 MHz Radar

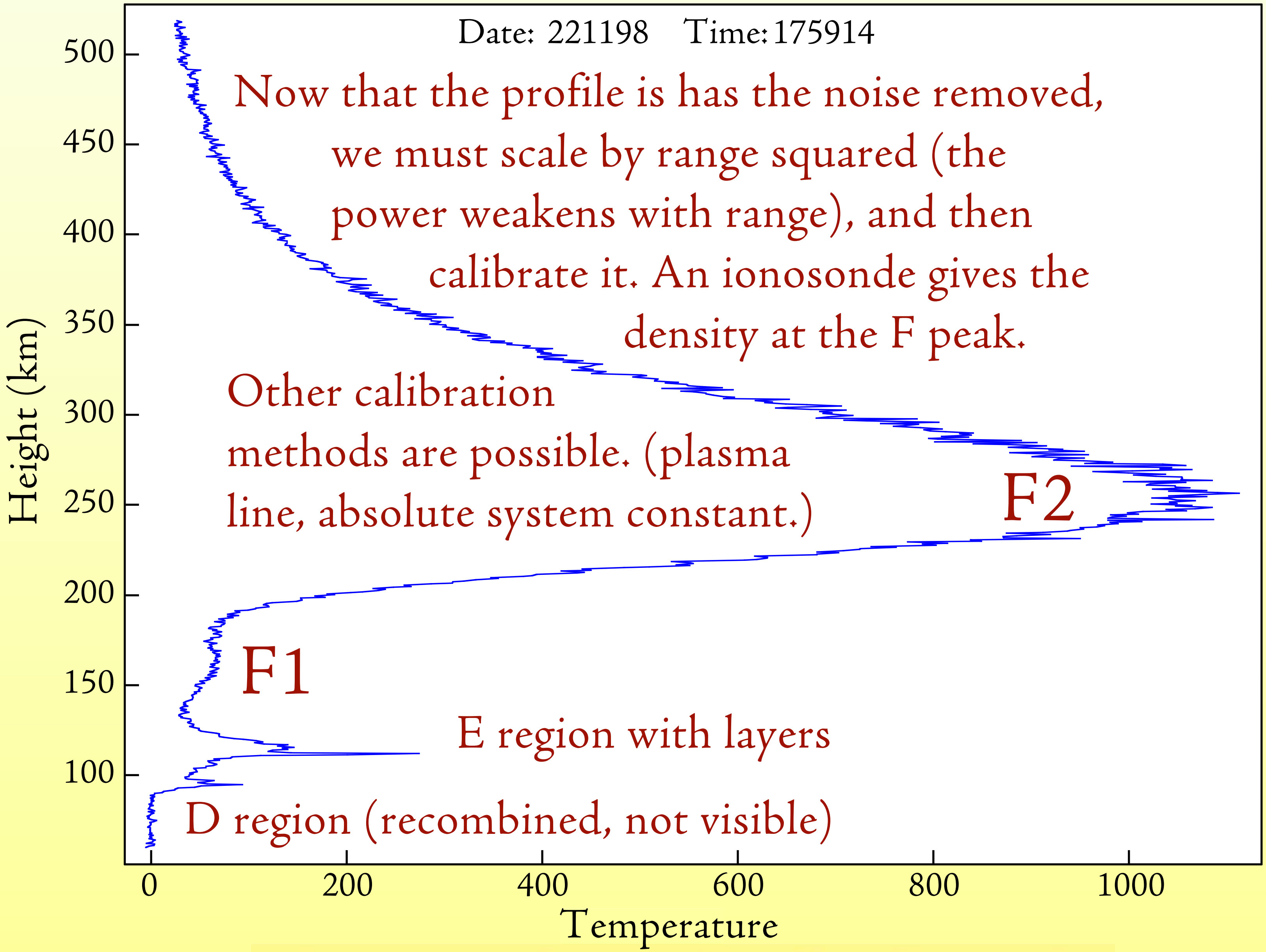


Processing the Power Profile: Step 1



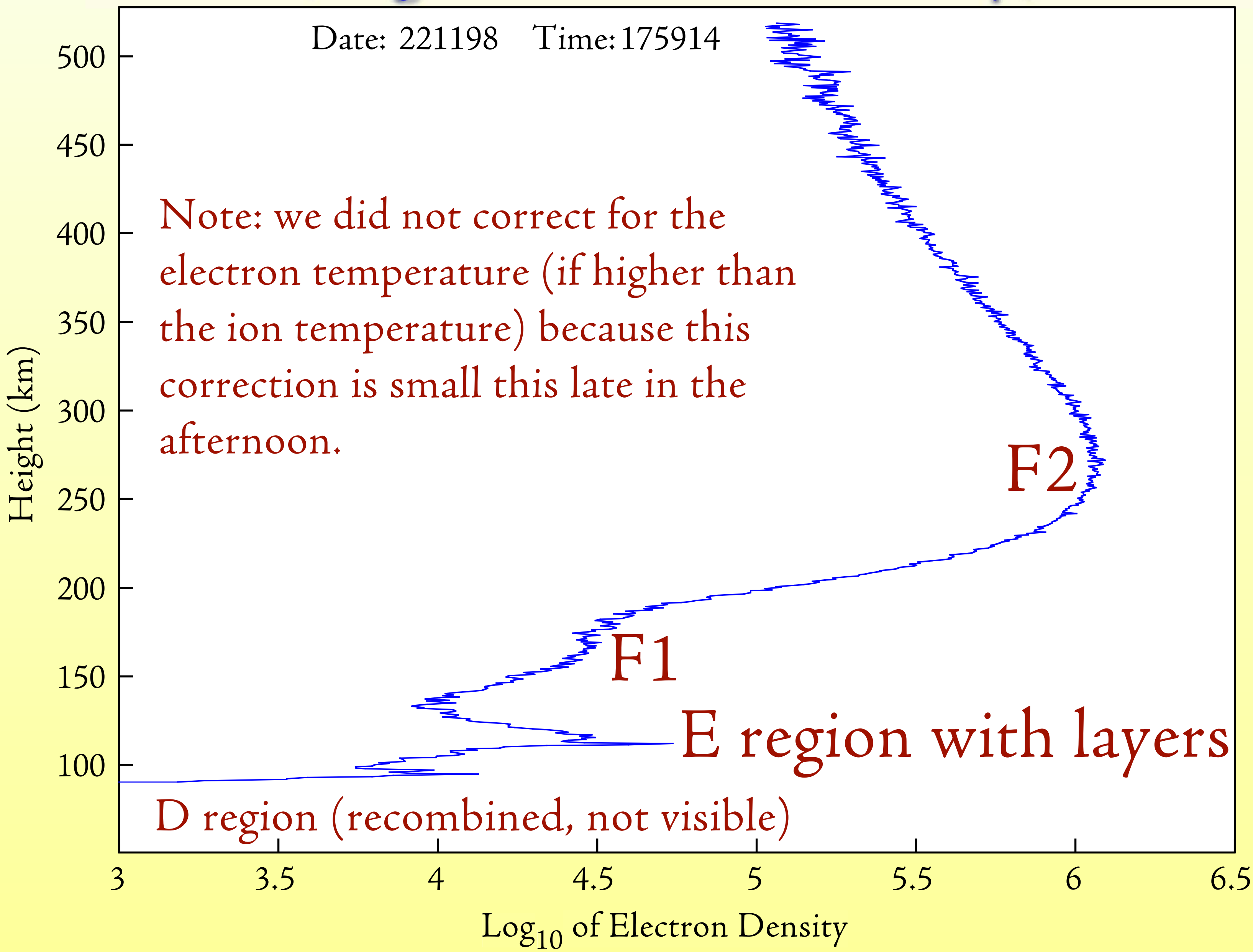
Next we show the result of these operations and continue the process.

Processing the Power Profile: Step 2



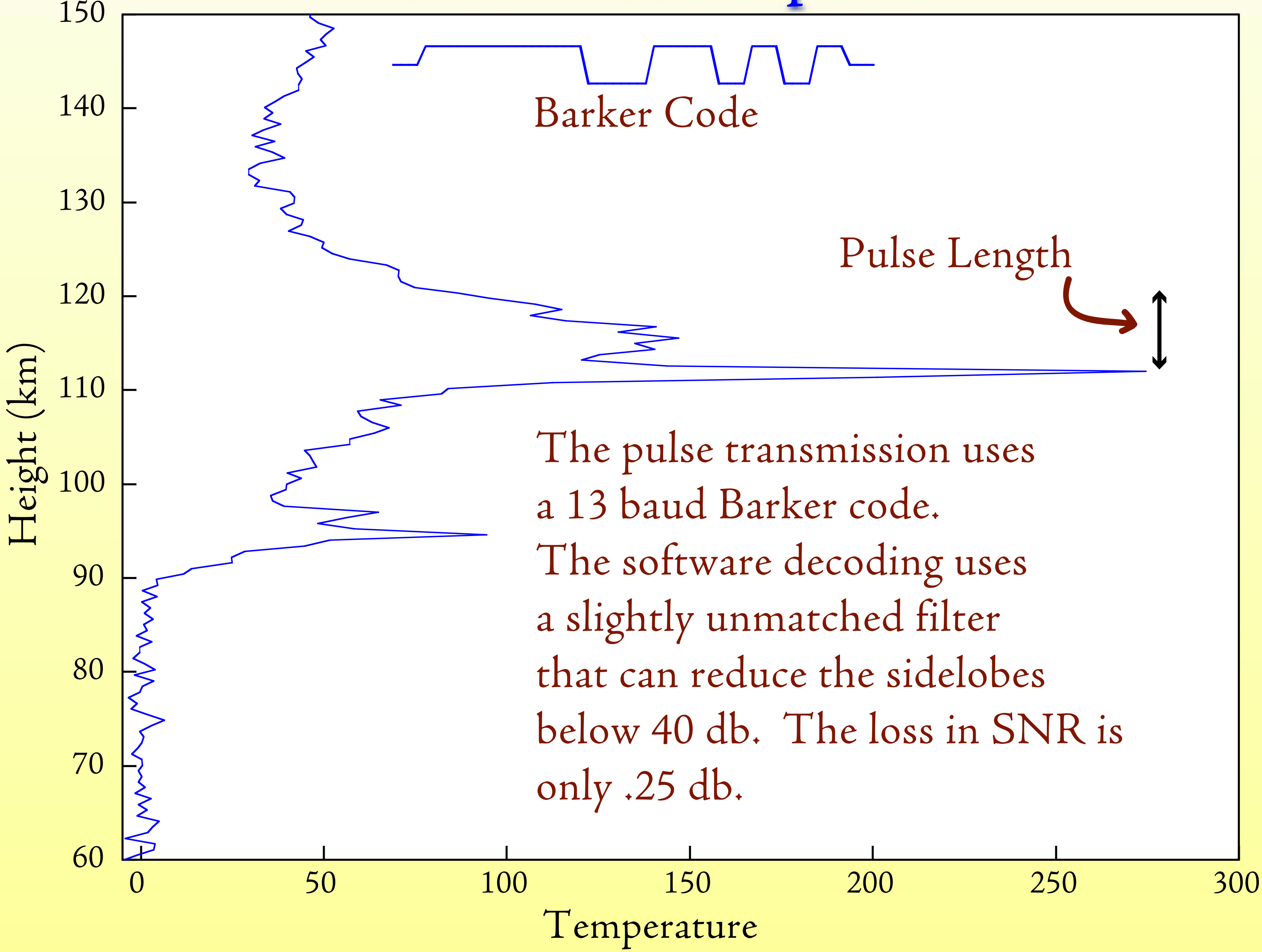
Next we show the final result.

Prcoessing the Power Profile: The profile (Log10 of electron density)



Next we look the range resolution.

How are we able to get such good range resolution with a 7.8 km pulse?



The pulse transmission uses
a 13 baud Barker code.
The software decoding uses
a slightly unmatched filter
that can reduce the sidelobes
below 40 db. The loss in SNR is
only .25 db.

Power Profiles: Summary

1. A power profile has statistical errors due to the noise like nature of IS. The errors are proportional to to S + N, the sum of the IS signal and the system noise.

2. Errors can occur from the noise noise subtraction if inadequate numbers of noise samples are used in determining the noise level.

3. Calibration errors:
   a. Ionosonde: Could be as high as 10%.
   b. Absolute system constant: Opinions differ as to how stable such an approach is.

Are there better ways to measure electron density than power profiles?

1. Faraday rotation is a really good method at low frequency radars, especially Jicamarca, but also at Kharkov and Irkutsk.
2. High resolution plasma line measurements can give spectacular results (daytime only), but this is not yet a standard technique that is easy to apply. Probably only Arecibo can do this.

More about signal and noise

1. A power profile has statistical errors due to the noise like nature of IS. The errors are proportional to to S + N, the sum of the IS signal and the system noise.

Although this is an elementary concept, it has important consequences:

First, maximizing signal to noise ratio is relevant only when the signal is small compared to the noise.

When the signal is large it is the source of the errors.

To improve the measurement, then it is necessary to increase the number of independent samples, even at the expense of SNR.

We will look at an example of this later.

Spectral Analysis of ISR Data

- The ISR frequency spectrum is very useful.
- It is possible to measure several ionospheric parameters simultaneously:
    - one, two, or three (sometimes) plasma temperatures;
    - one, two, or three (often) plasma densities and electron density;
    - one, or occasionally two velocities of plasma constituents.

- Measuring the IS spectrum can be quite complicated.
- The radars need different techniques due to their operating frequencies, etc.
- Different regions of the ionosphere require different techniques.
- This talk will concentrate on one technique used:
    - in the F region or topside ionosphere • by:
    - most of the IS radars (but not all).

- It is convenient to put the errors into three categories:
    - First, statistical errors (the noise-like nature of IS, and system noise)
    - Second, errors due to the length of the radar pulse
    - Third, systematic errors in the non-linear least fitting process.

From raw samples to geophysical parameters: The steps in spectral analysis of IS data

1. Correlation or FFT Analysis

(Accumulation over many radar pulses)

2. Correction for Pulse length effects

(Simple or sophisticated techniques could be used.)

3. Non-linear least squares fitting

(Additional corrections might be necessary.) (Multiple ranges can be fit simultaneously.)

Note: Steps 2 and 3 can be combined.

From raw samples to geophysical parameters: The steps in spectral analysis of IS data

1. Correction or FFT Analysis

(Accumulation over many radar pulses)

We begin by looking at the raw samples used in step one. How do we design the technique?

2. Correction for Pulse length effects

(Simple or sophisticated techniques could be used.)

3. Non-linear least squares fitting

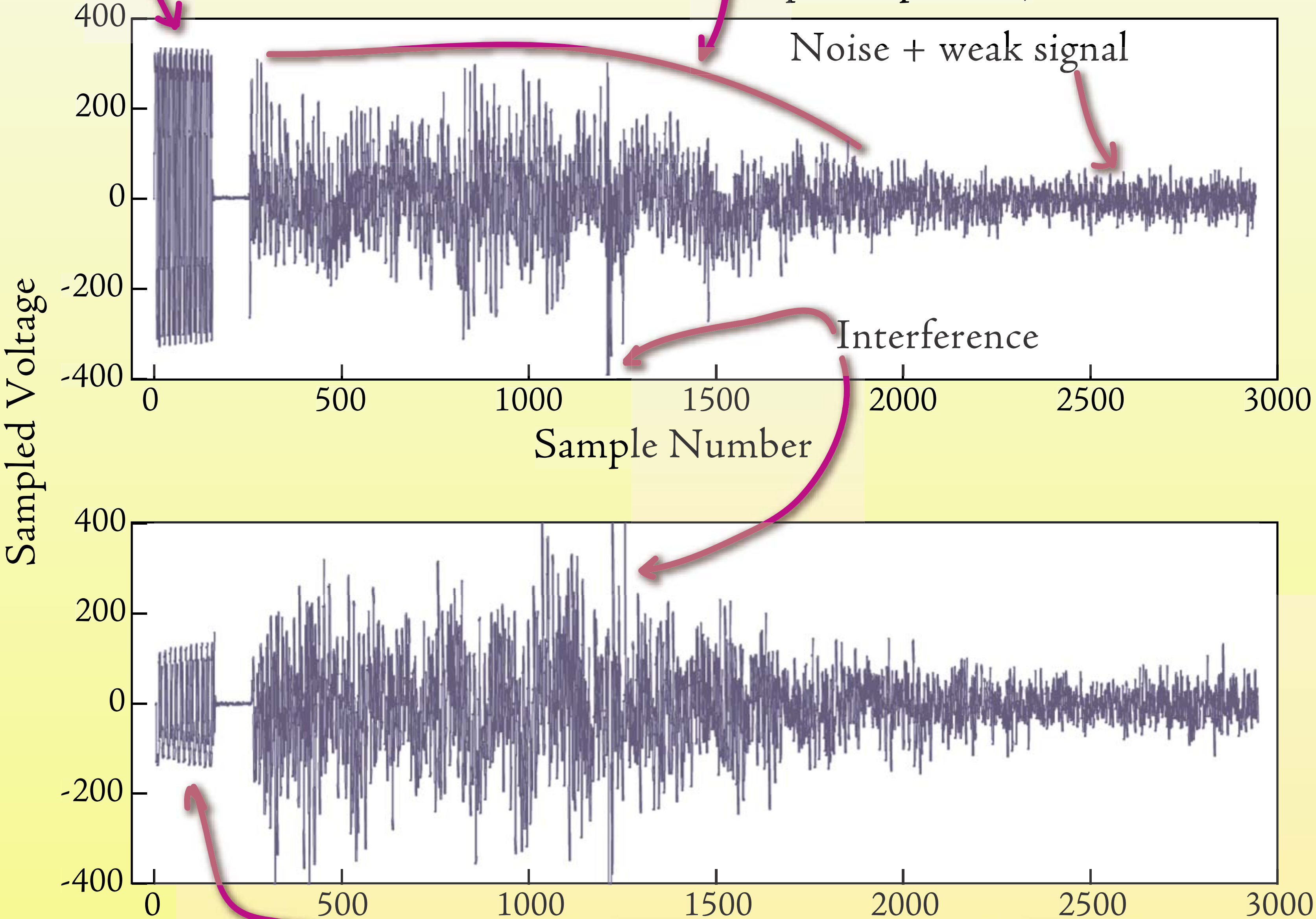
(Additional corrections might be necessary.)
(Multiple ranges can be fit simultaneously.)

We will see that how we use the radar depends a lot on what we expect to see.

Spectral Analysis: raw samples

Transmitter Samples
(Why is this a coded pulse rather than an uncoded pulse?)

Scatter + noise
(Why does this appear correlated over time unlike the samples of the power profile?)



Other Questions:

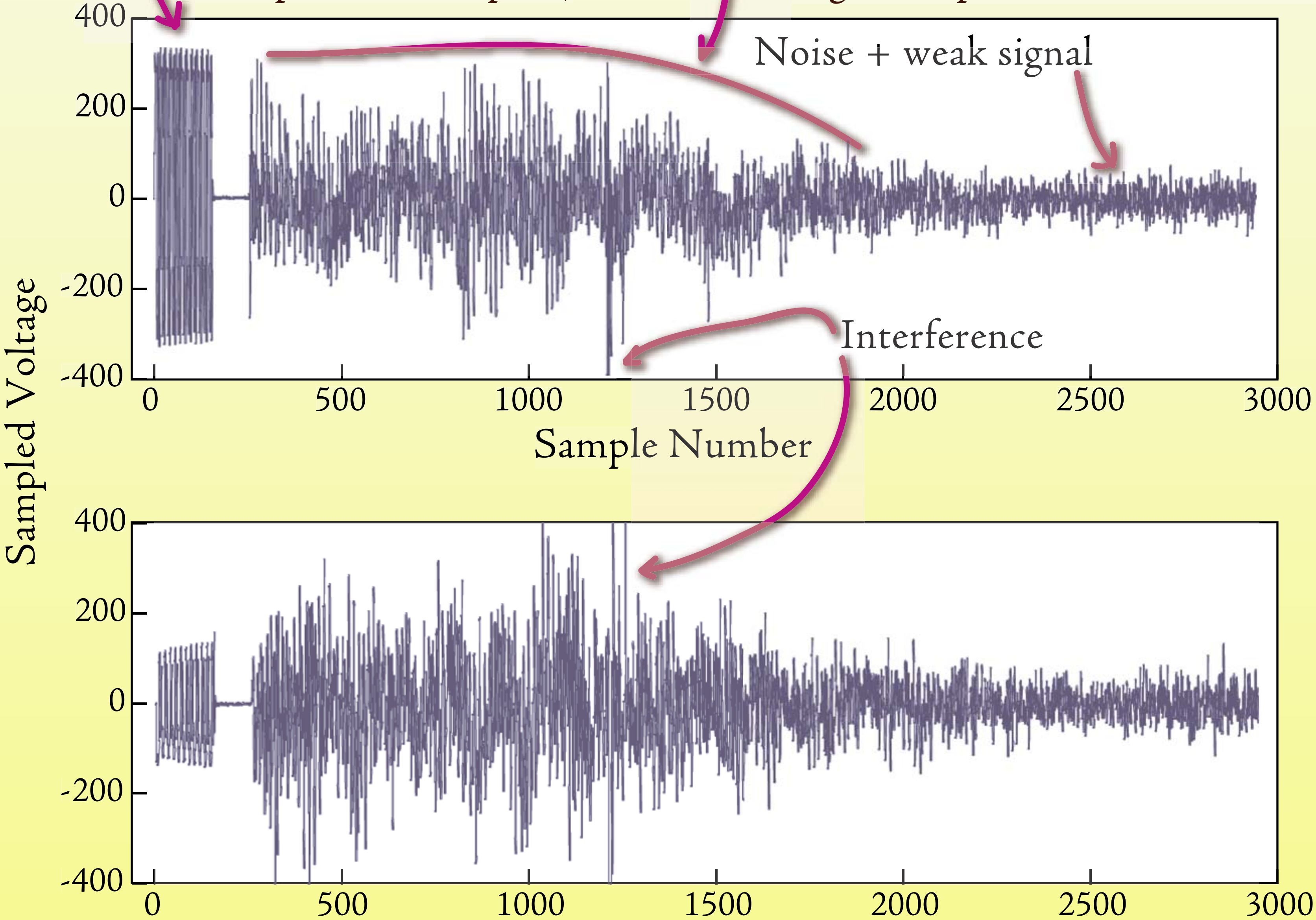
How does one make use of the high signal to noise ratio to reduce the errors in the measurements?

What is the significance of the slope on the envelope of the transmitter sample?

Spectral Analysis: raw samples 2

Transmitter Samples
(When signal is large it is best to generate more independent samples.)

Scatter + noise
(It is necessary to use a pulse at least as long as the correlation time. When the signal is large, it is possible to see how fast it changes.)



A phase code is used to generate seven frequencies at once, much like having seven smaller radars transmitting and receiving at once.

Although there is some droop in the amplitude of the transmitter pulse, the major effect of the falling high voltage is a phase change with time. This "chirp" must be accounted for in velocity measurements.

From raw samples to geophysical parameters: The steps in spectral analysis of IS data

1. Correlation or FFT Analysis

(Accumulation over many radar pulses)

How does the radar pulse smear information across range? How do we keep as much as possible?

2. Correction for Pulse length effects

(Simple or sophisticated techniques could be used.)

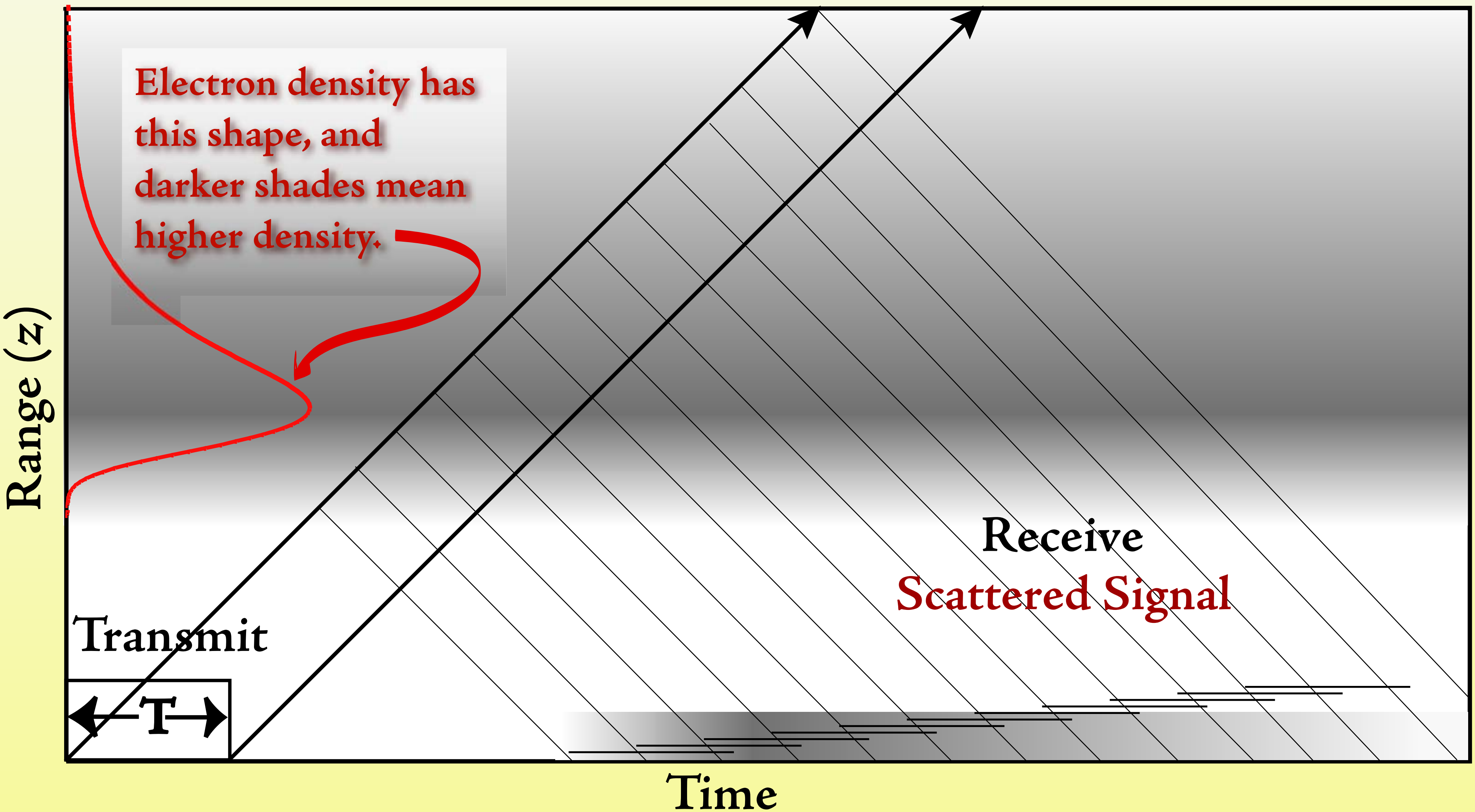
3. Non-linear least squares fitting

(Additional corrections might be necessary.) (Multiple ranges can be fit simultaneously.)

We will look at the computation of lag profiles, averages over range and delay.

Range-time diagram of a long pulse experiment

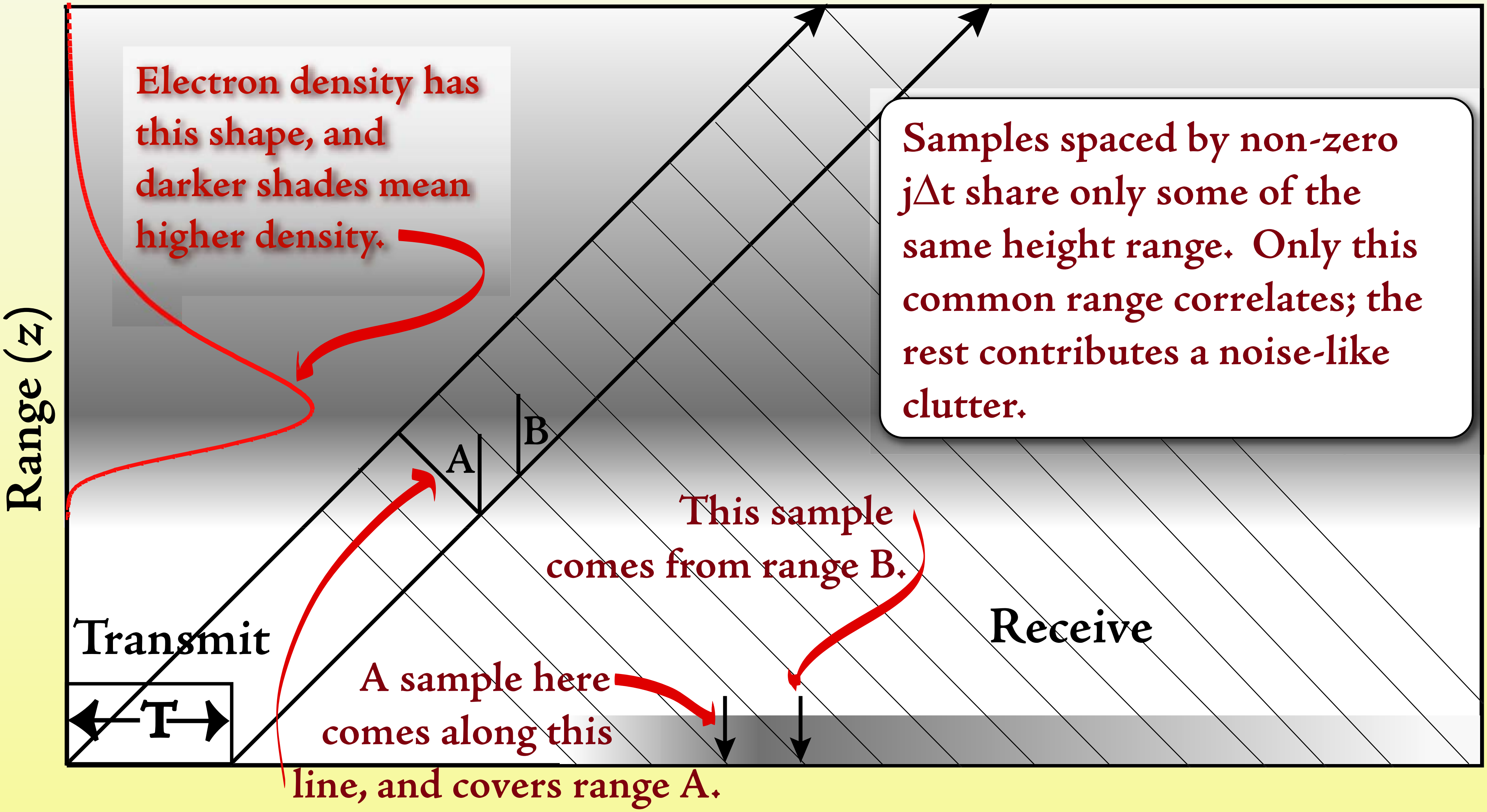
A radar converts range into time. If we think of the ionosphere as composed of many narrow slabs, the return from each slab is approximately as long as the pulse.



Since the transmitted pulse is long, the returns from neighboring slabs overlap. This can cause problems if the parameters change significantly from slab to slab. Note that the magnitude squared of the received signal is power profile.

The spectral information is in the set of lag profiles

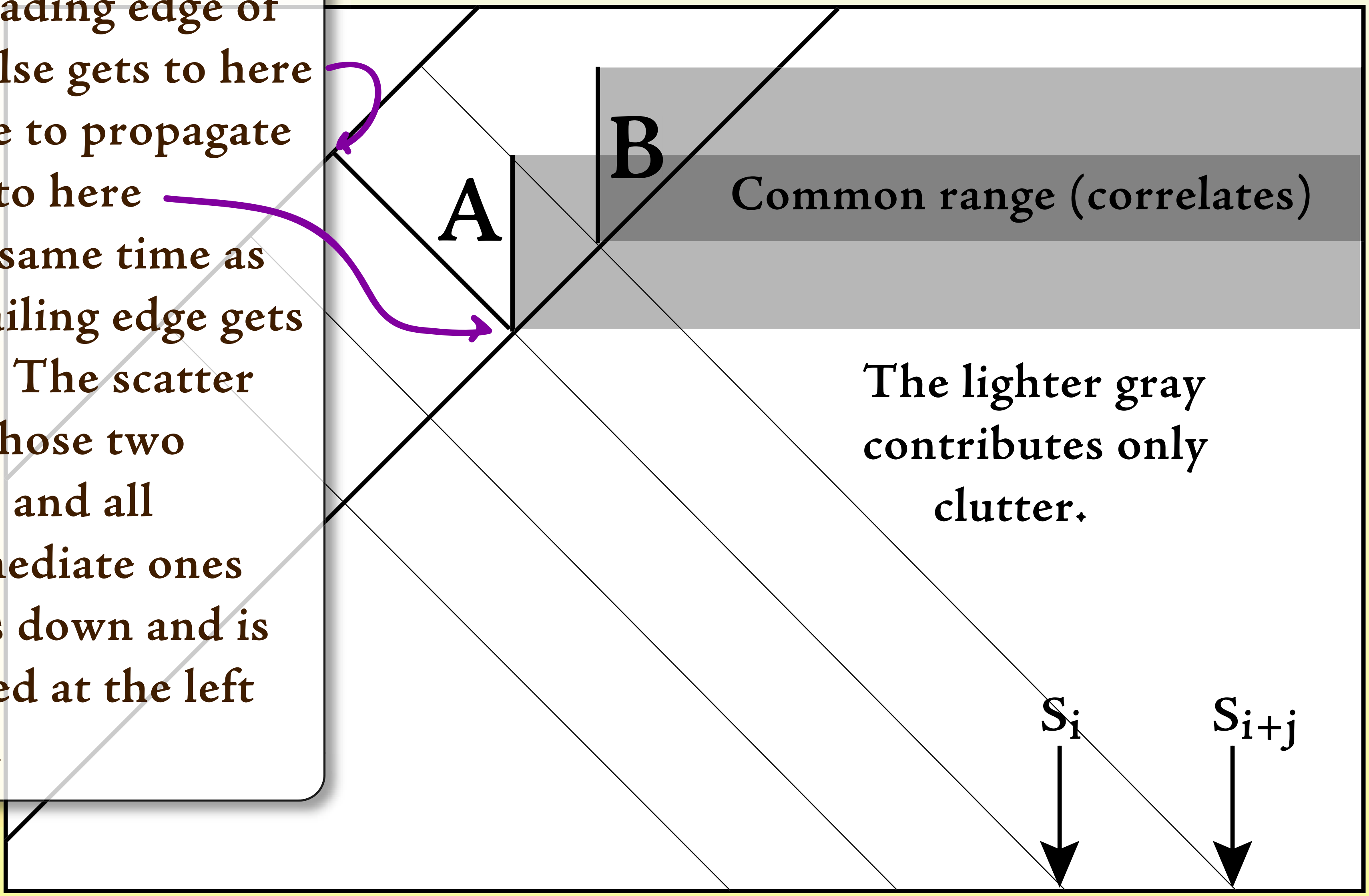
If the samples are S_i, then the lag profile for delay $j\Delta t$ are $S_i{}^*S_{i+j}$ for all i giving non-zero products. $j\Delta t$ starts at zero, and is limited by pulse "overlap".



The overlapping parts of ranges A and B can correlate. Lag profiles of different delay have a different range smearing and a different weight owing to the different common ranges. These effects must be accounted for.

Look at a blow-up of the last graph:

The leading edge of the pulse gets to here in time to propagate down to here at the same time as the trailing edge gets there. The scatter from those two ranges and all intermediate ones travels down and is sampled at the left arrow.



Range B is defined the same way at a later time, and so it is larger (higher) and sampled at the right arrow.

More on lag profiles

If this represents the samples from a radar pulse,

then here is how to make a lag profile for this delay.

We need lag profiles for each possible delay with a limit of the pulse width.

An ACF results from taking values in this direction. However, there are three potential sources of errors: 1. alignment in range, 2. range resolution, and 3. variation in range resolution with delay. We will discuss these later.

From raw samples to geophysical parameters: The steps in spectral analysis of IS data

1. Correlation or FFT Analysis

(Accumulation over many radar pulses)

Now we look at a few things about errors in spectral analysis involving spectra or ACFs.

2. Correction for Pulse length effects

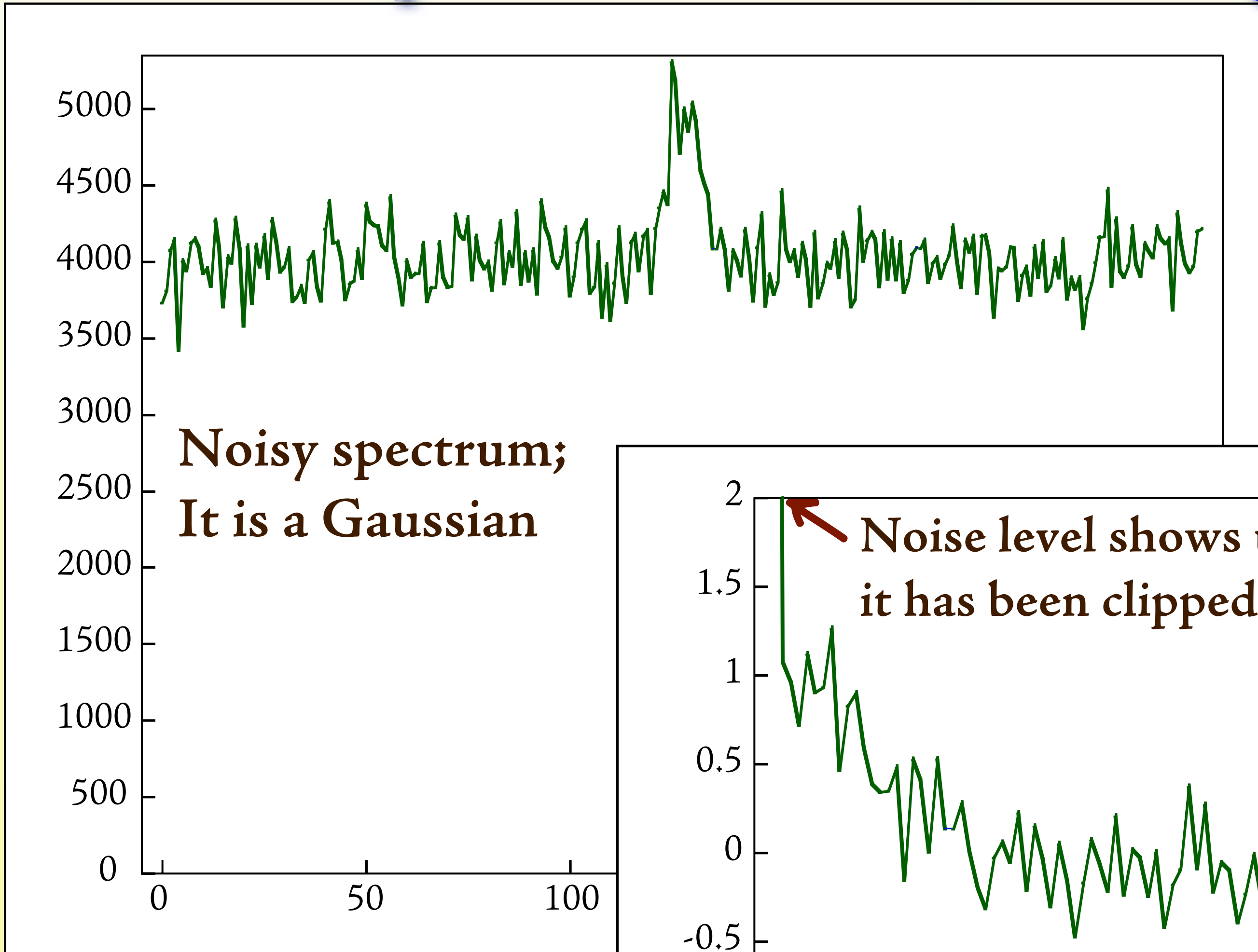
(Simple or sophisticated techniques could be used.)

3. Non-linear least squares fitting

(Additional corrections might be necessary.) (Multiple ranges can be fit simultaneously.)

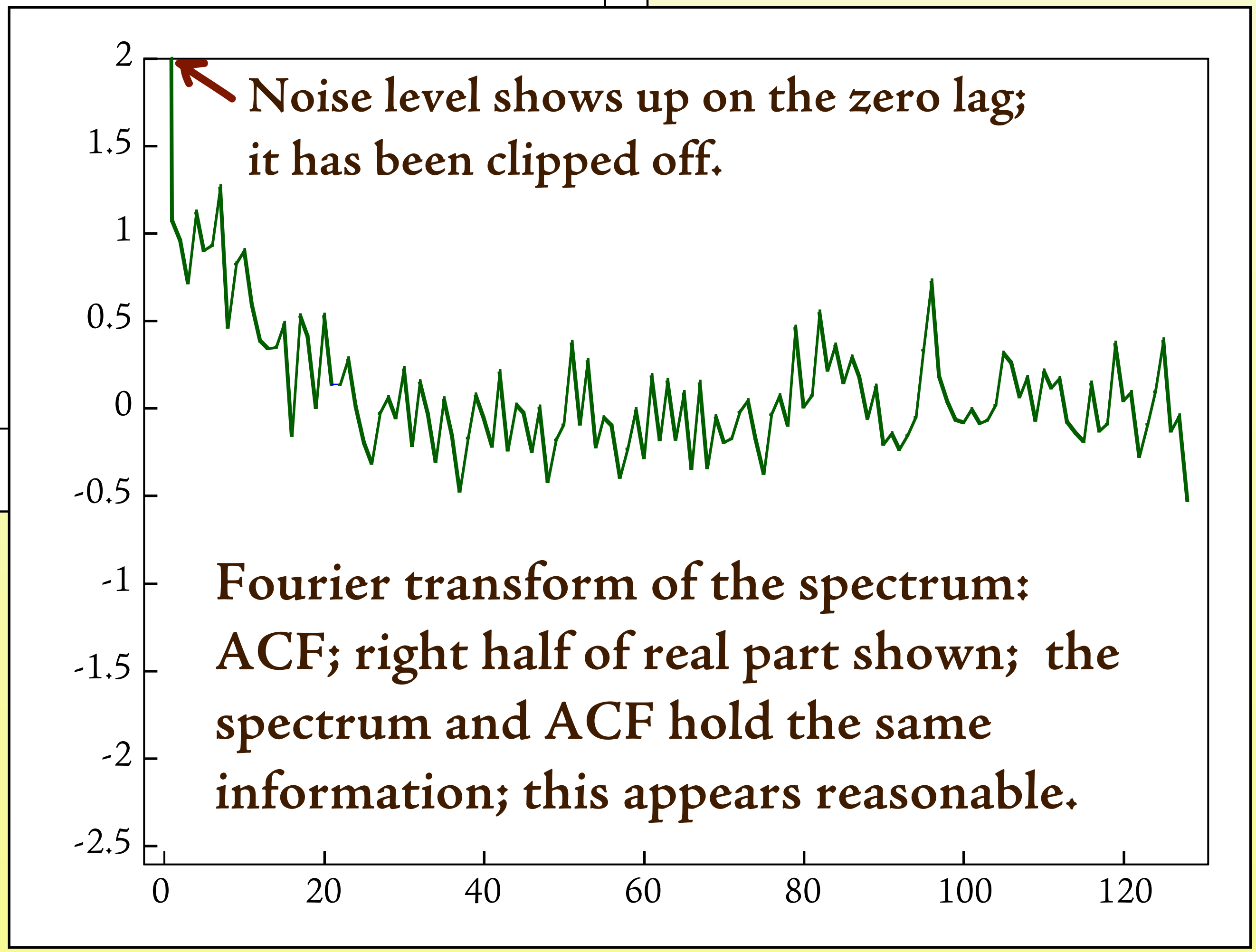
The differences between the high and low SNR cases are interesting.

Errors on Spectra and ACFs, Low Signal to Noise Ratio



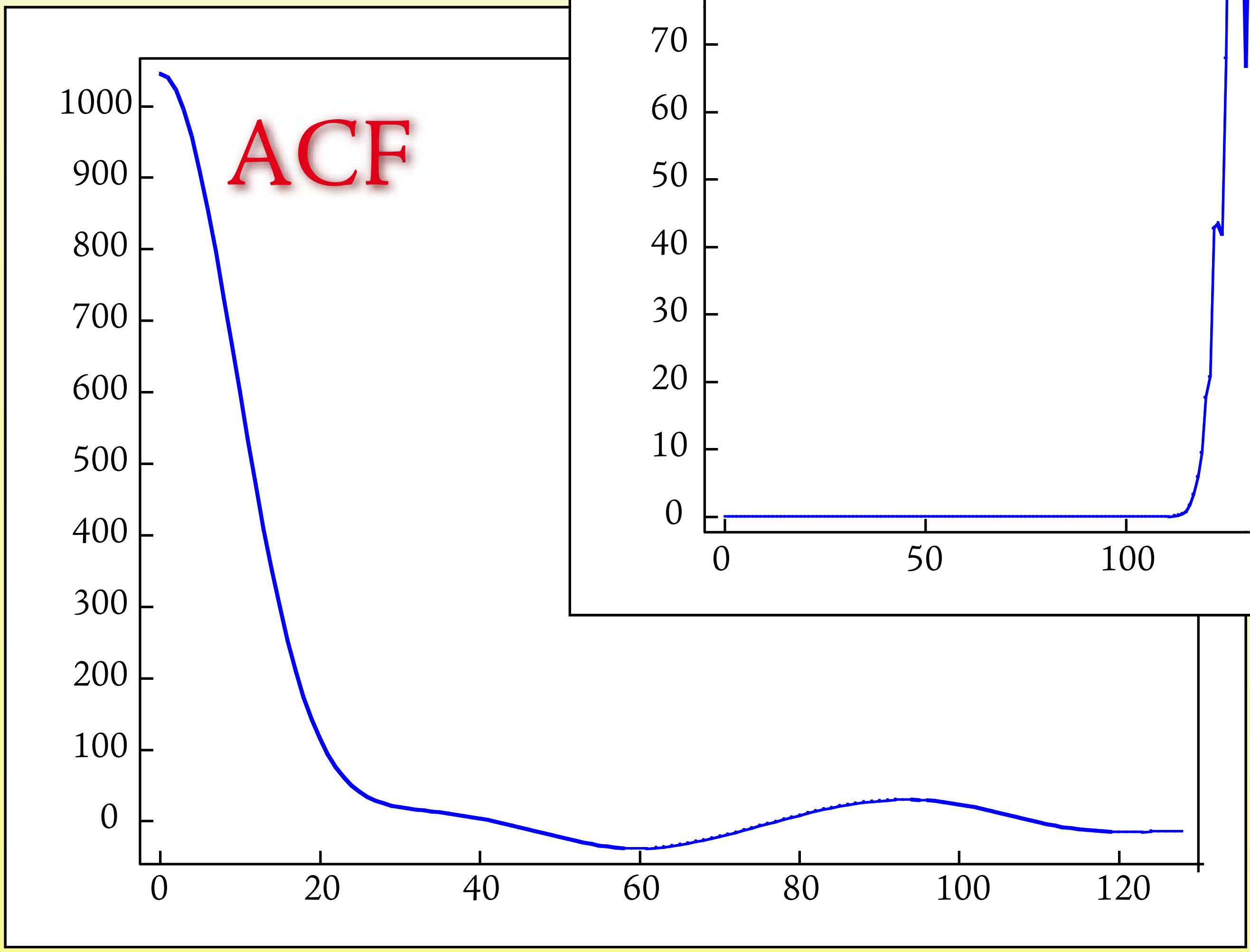
Errors on both are nearly independent from point to point, and nearly the same from point to point.

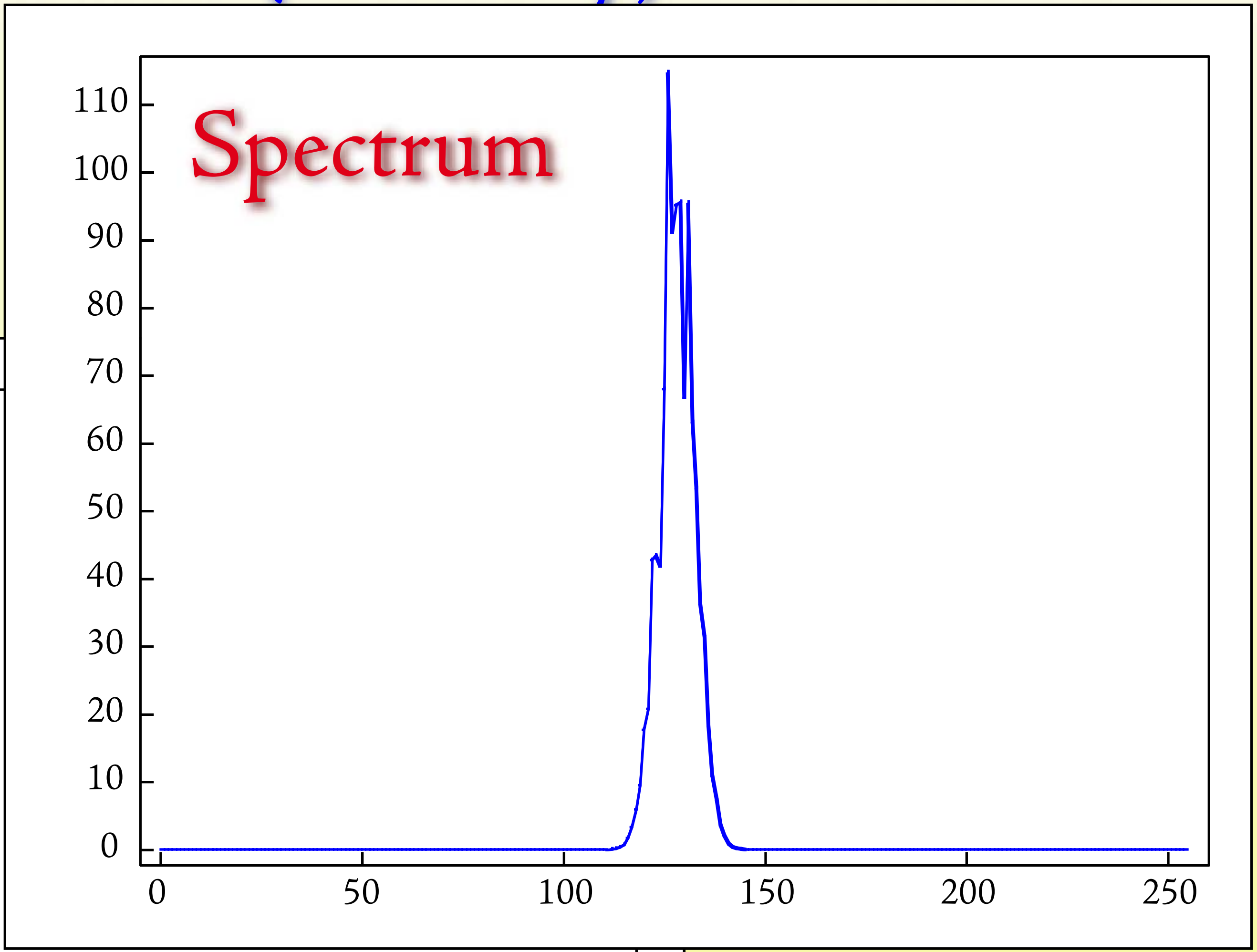
1000 independent spectral estimates accumulated.



Consider these plots, a spectrum and an ACF
Which is a better (less noisy) measurement?

That is, which has more information about the process it represents?

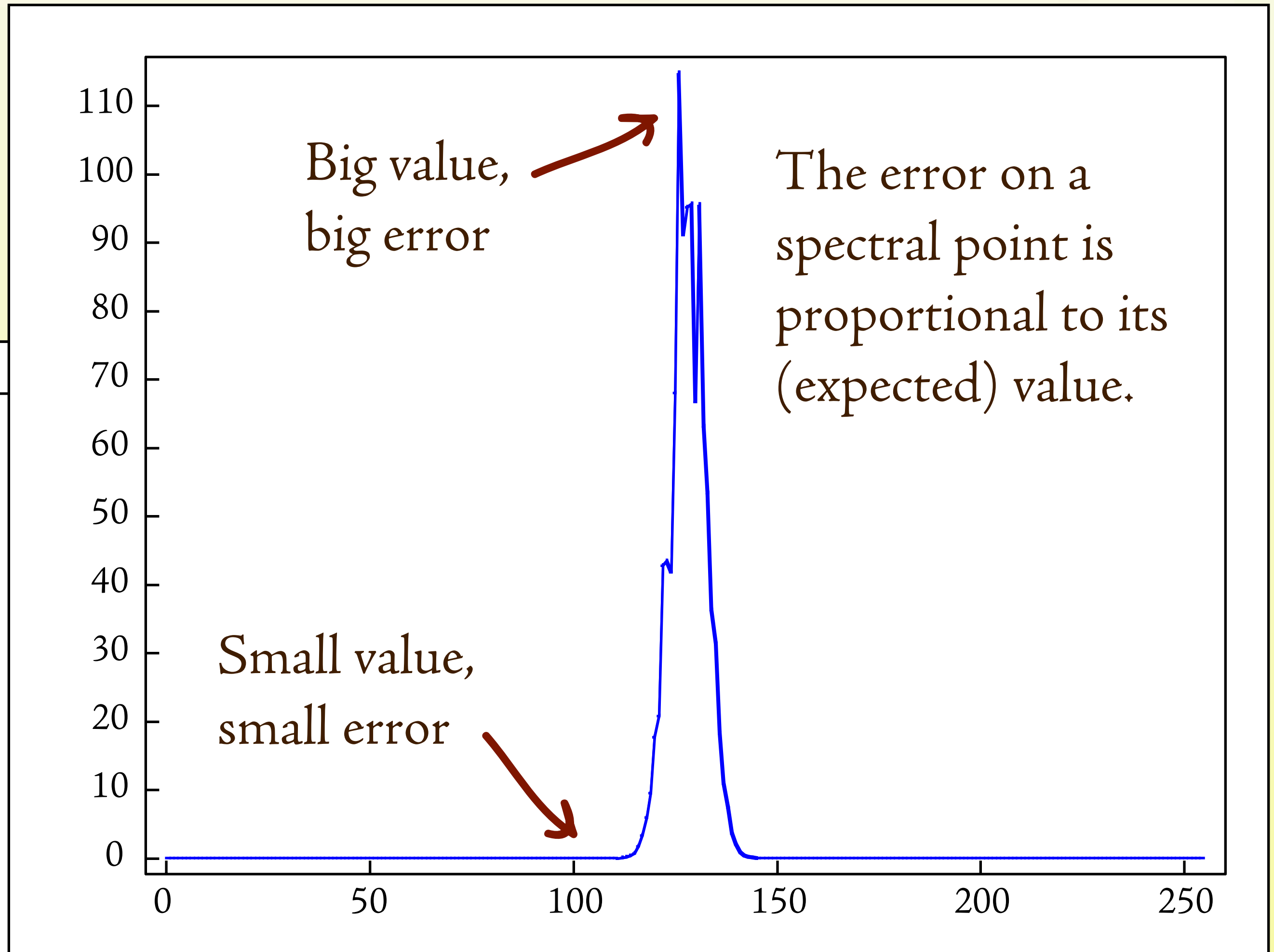




Consider: is this a high or low SNR measurement?

Errors on Spectra and ACFs, Very High SNR, (same data for both) 100 independent samples

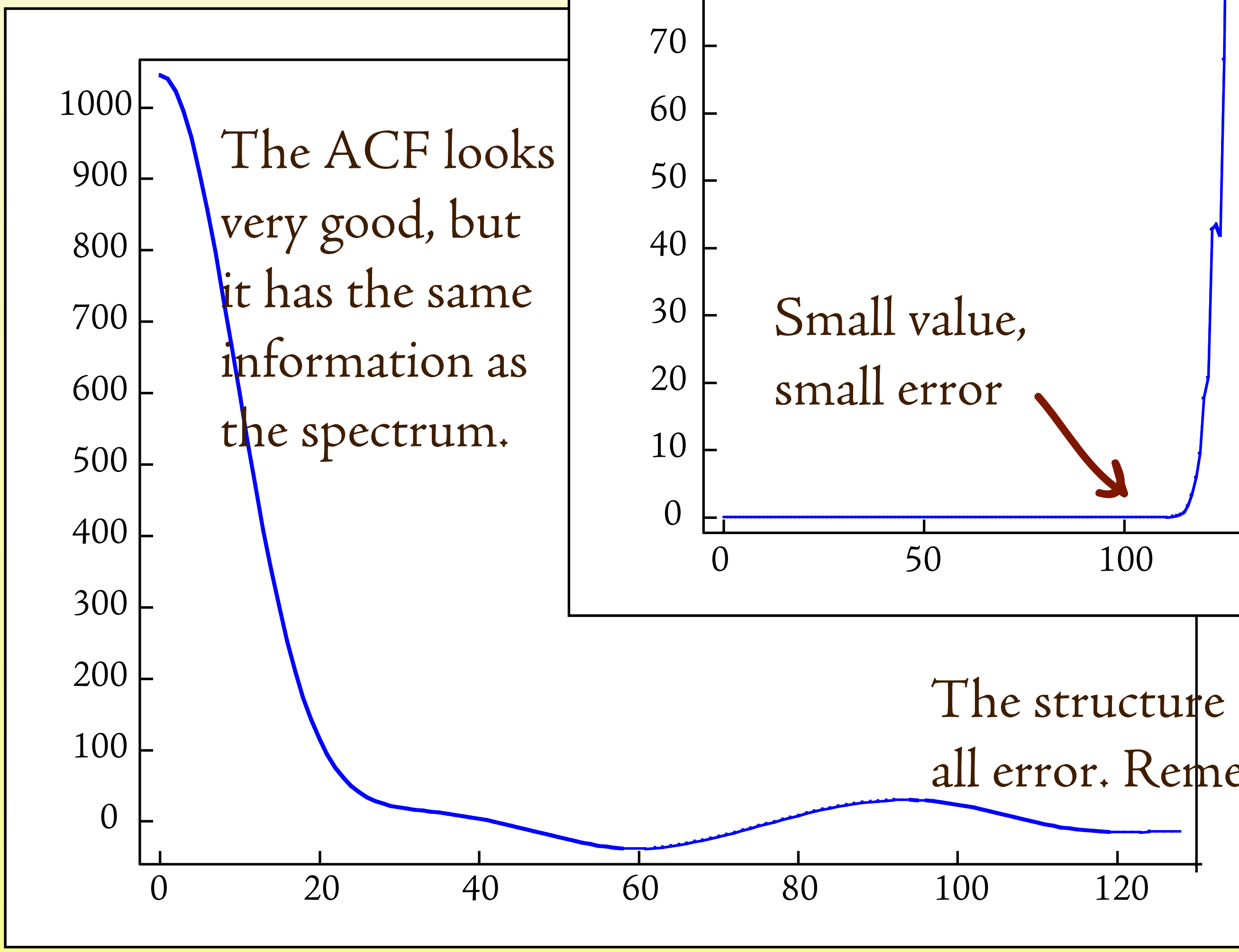
Errors on spectrum are nearly independent from point to point, but highly correlated on the ACF!



Big value, big error

Small value, small error

The error on a spectral point is proportional to its (expected) value.



The ACF looks very good, but it has the same information as the spectrum.

The structure on this part of the curve is all error. Remember, the Fourier transform of a Gaussian is a Gaussian.

From raw samples to geophysical parameters: The steps in spectral analysis of IS data

1. Correlation or FFT Analysis

(Accumulation over many radar pulses)

2. Correction for Pulse length effects

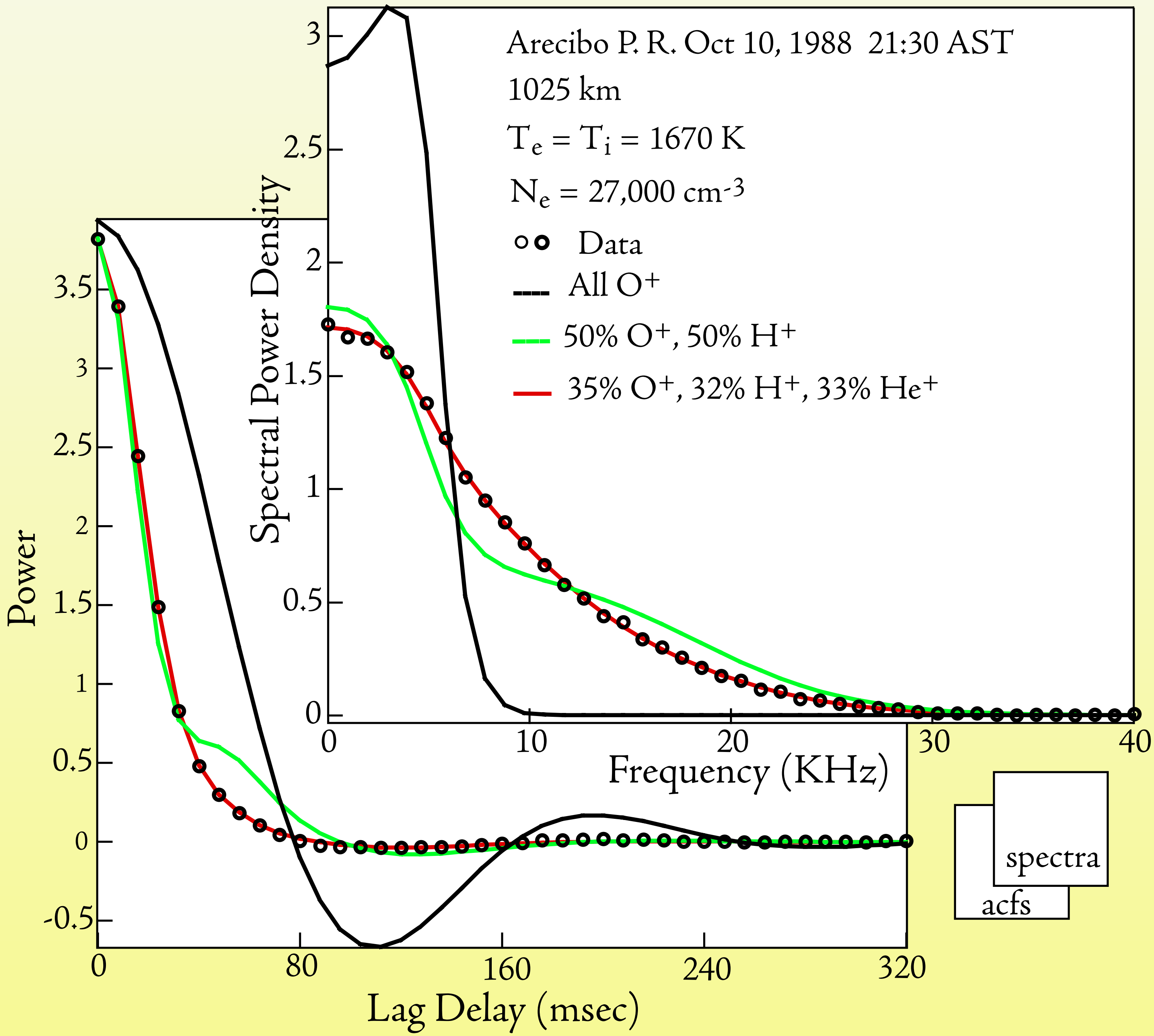
(Simple or sophisticated techniques could be used.)

3. Non-linear least squares fitting

(Additional corrections might be necessary.)
(Multiple ranges can be fit simultaneously.)

Now we look NLLS fitting. This is useful in a situation where we have a parametrized model which describes our function. IS spectra (or ACFs) have such a model.

The Effects of Ion Composition on the Spectrum and ACF



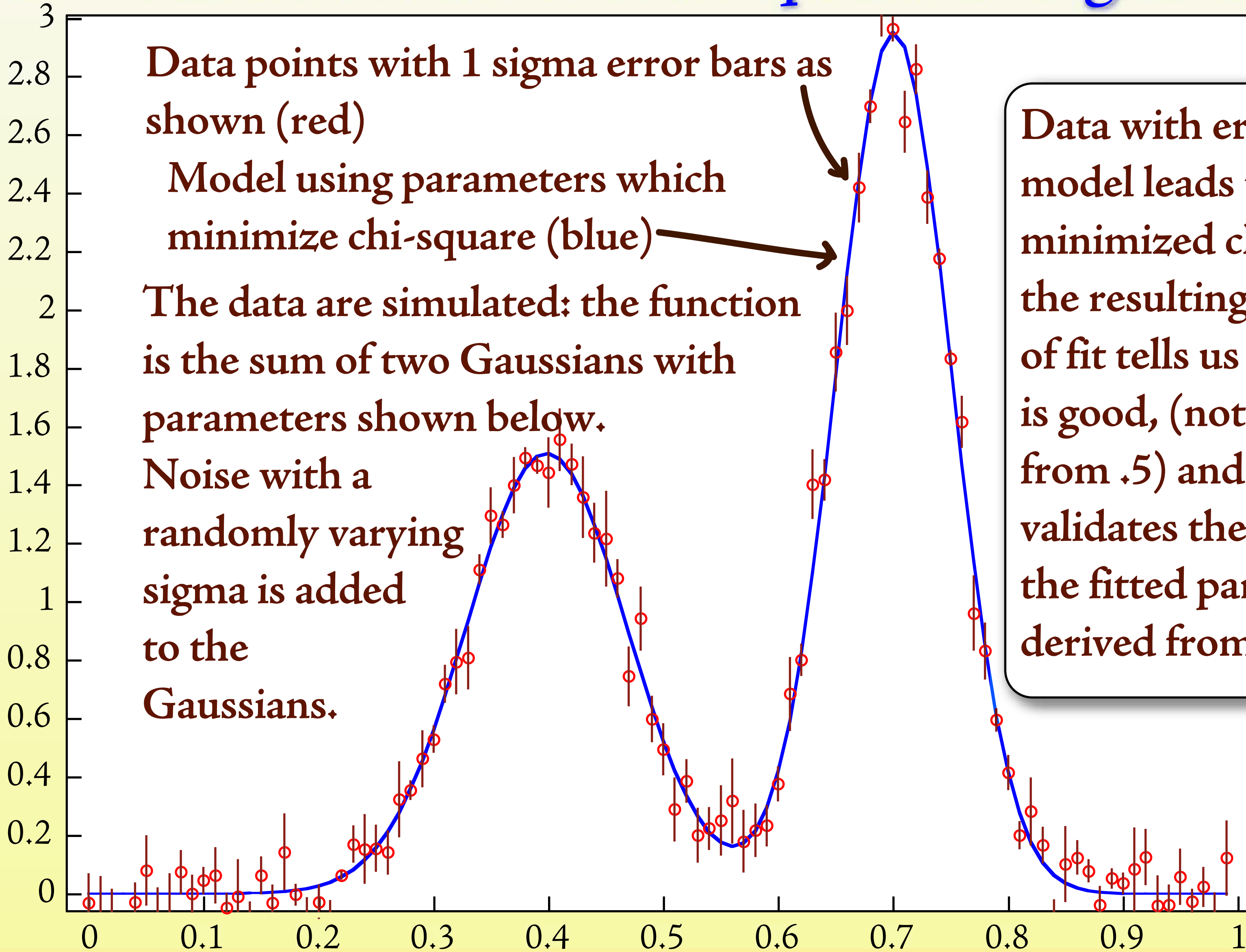
The circles are the data; we show both spectra and autocorrelation functions.

The line closely approximating the data is a fit using three ions.

The other lines show that it is not possible to get a good fit using one or two ions since the spectral shape is incorrect.

There is little noise on this data, and so the fit was not so difficult. There are cases where the fit can go the wrong way and end up with the wrong answers, but usually the Q value (or the temperatures) shows that it is a "bad" fit so that you know what happened.

The essentials of non-linear least squares fitting: a simulation



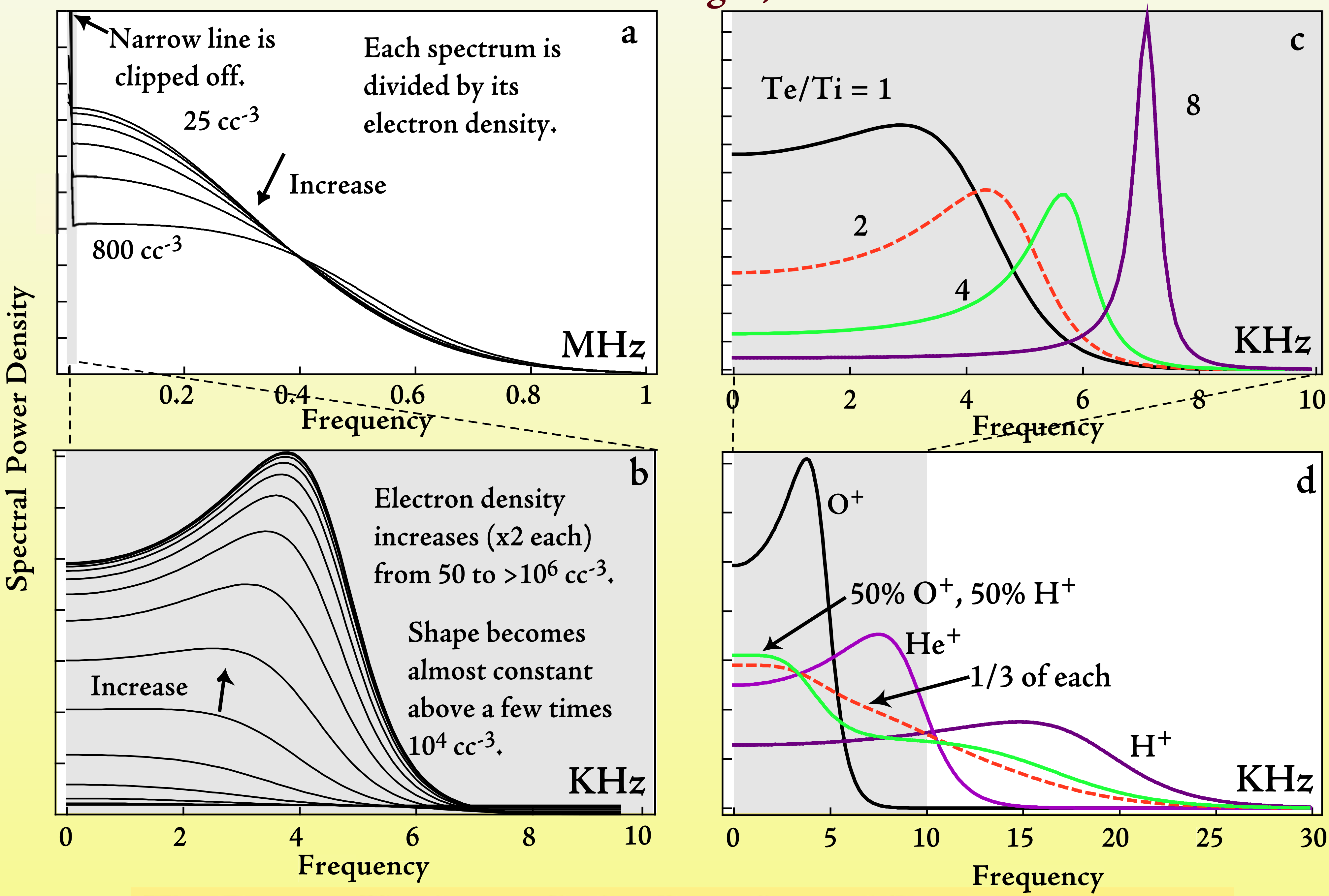
Data with errors and a model leads to a minimized chi-square; the resulting goodness of fit tells us if the model is good, (not too far from .5) and thus validates the errors on the fitted parameters derived from the fit.

	Amplitude	Center	Width	Amplitude	Center	Width
The simulation parameters:	1.5	0.4	0.1	3	0.7	0.07
The "fitted" parameters:	1.5099	0.39837	0.098772	2.9522	0.70073	0.07101

The goodness of fit Q: 0.485931 (The probability of getting a larger chi-square by chance)

A "Zoo" of Incoherent Scatter Spectra, no B or Collisions

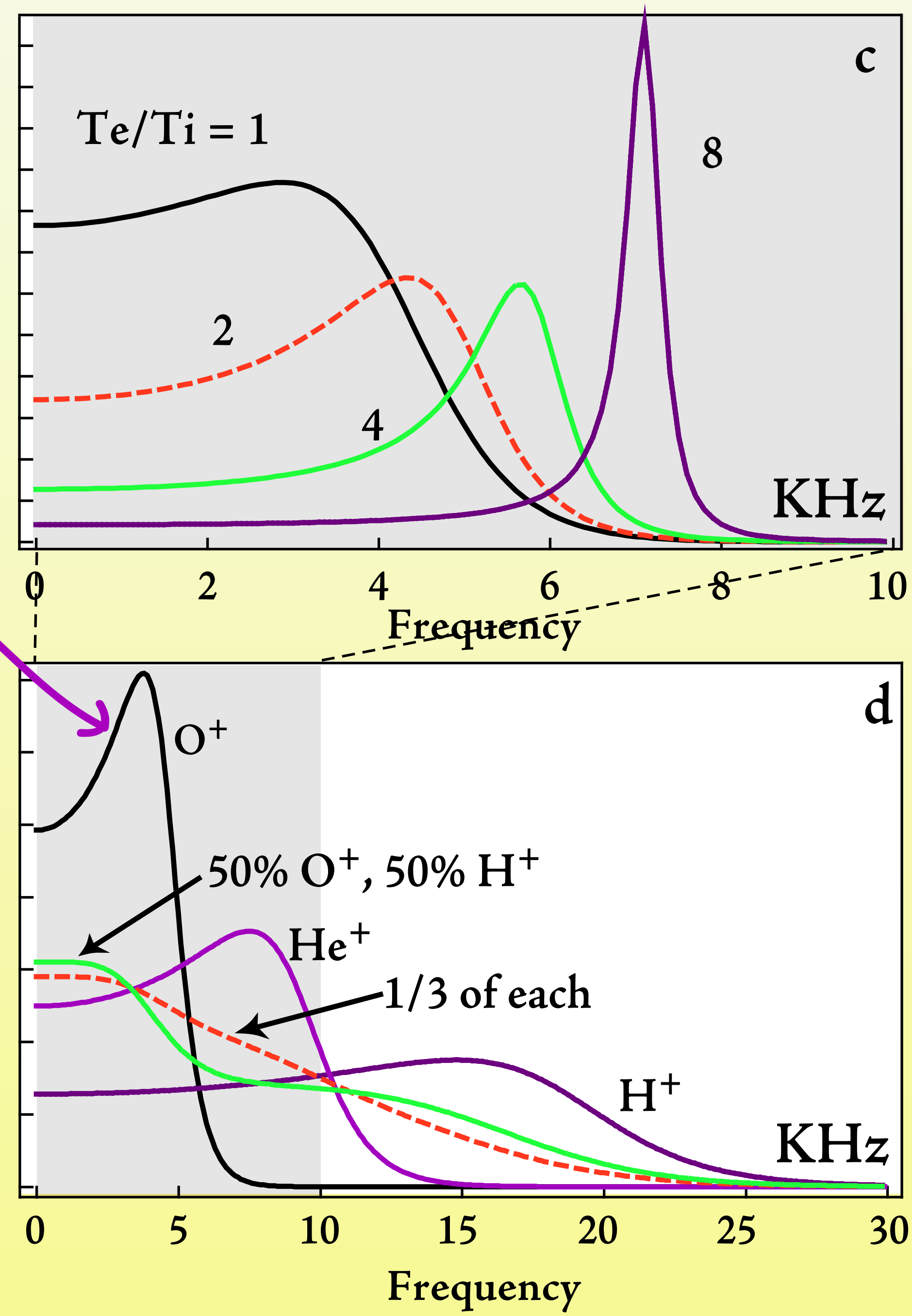
Arecibo Wavelength, 1000K



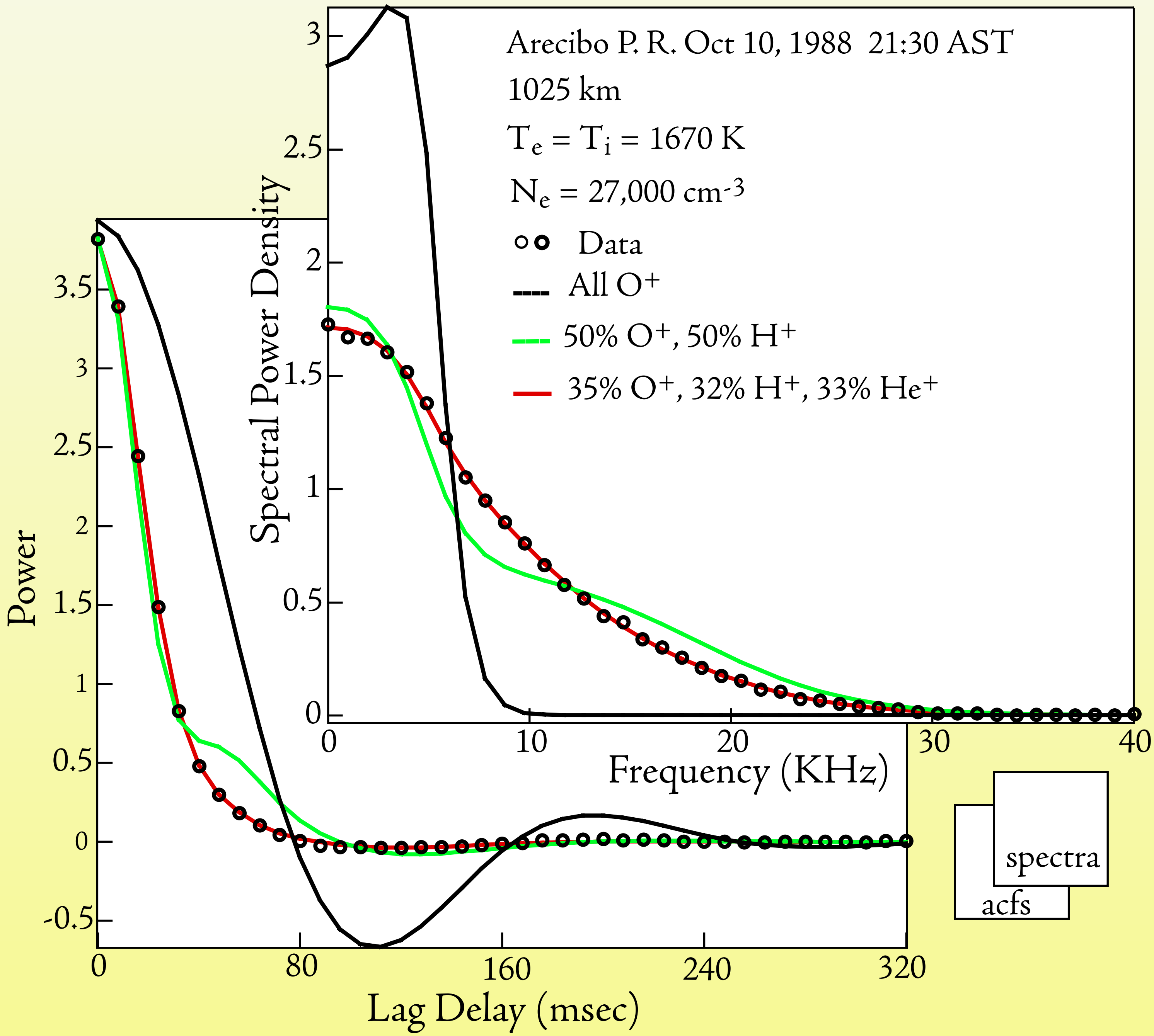
IS radars usually observe the ion line, plot c, with Te/Ti in the range 1-4.

Be careful; solutions are not unique!

The width of the ion line depends upon $T_i^{.5}$. The ion thermal velocity distribution is the key determining factor. However the ion mass controls the velocity as well as T_i, and so this spectrum could be O+ at 1000K, He+ at 250K, or H+ at 62.5K. Only the first is reasonable in the F region, and so it is possible to decide. NLLS fitting is iterative; one establishes where the fitting process goes by where one starts it, as well as by the freedom allowed in the fit. Ion identity cannot be confused when only one is allowed, but when two or three ions are possible, it is possible to get errors.



The Effects of Ion Composition on the Spectrum and ACF



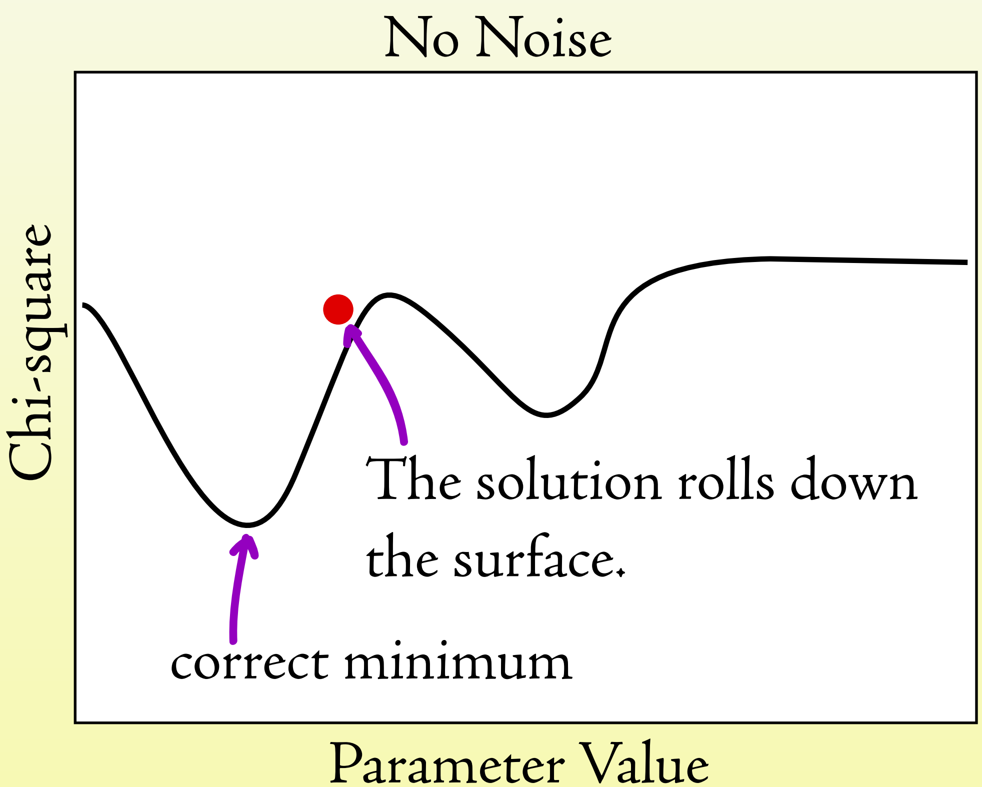
The circles are the data; we show both spectra and autocorrelation functions.

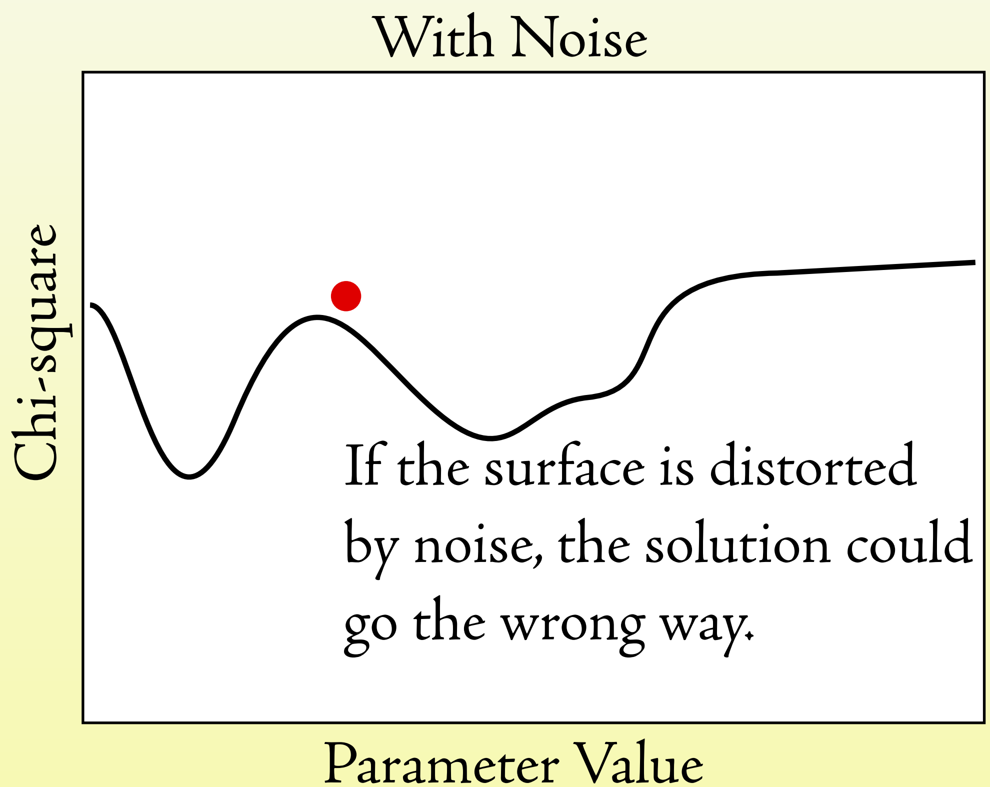
The line closely approximating the data is a fit using three ions.

The other lines show that it is not possible to get a good fit using one or two ions since the spectral shape is incorrect.

There is little noise on this data, and so the fit was not so difficult. There are cases where the fit can go the wrong way and end up with the wrong answers, but usually the Q value (or the temperatures) shows that it is a "bad" fit so that you know what happened.

What can happen with two minima





This problem is more likely to happen if the initial values are chosen badly.

How to fix the "two minima" problem:

One needs to lower the level of noise:

1. Assuming one cannot improve the radar, one can try to reduce the noise by averaging more data.
   a. One can average over time and/or range.
   b. Averaging spectra resulting from different parameters (temperatures, composition) can result in spectra that correspond to no actual set of parameters. It certainly does not yield spectra with the average of the parameters.
   c. Intelligent averaging (fitting data to simple models, Savitsky-Golay filters, etc.) can improve the performance, but is not always good enough.
2. One can recognize that the basic problem is that simple variations of the parameters do not lead to simple variations in the spectra.
   (Example: Increasing the fraction of light ions adds a wider spectral component, but it also "rounds" the shape of the heavier ion component.)
   Therefore, one needs to adopt a method where one assumes a simple variation in the parameters, and computes the spectra. This involves fitting multiple spectra simultaneously with a simple model connecting the spectra.
   - **We do not have physical models for range and time variation as we do for frequency variation. Let us stop and look at pulse width effects now.**

From raw samples to geophysical parameters: The steps in spectral analysis of IS data

1. Correlation or FFT Analysis

(Accumulation over many radar pulses)

We look at a way to correct for pulse length effects by using an inverse technique (linear regularization). This involves finding the "short pulse" profile from the measured profile (deconvolution). Noise goes up!

2. Correction for Pulse length effects

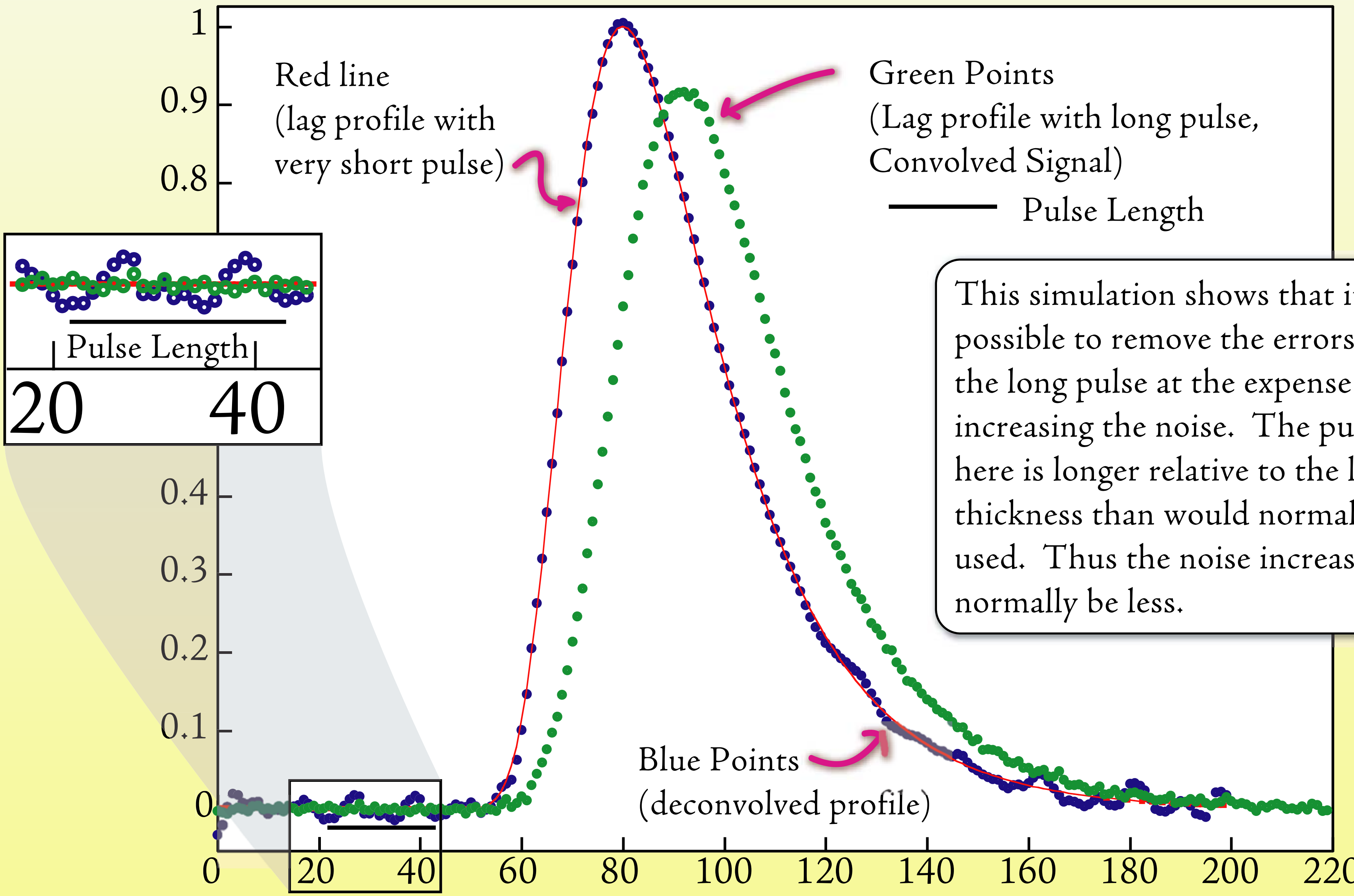
(Simple or sophisticated techniques could be used.)

3. Non-linear least squares fitting

(Additional corrections might be necessary.) (Multiple ranges can be fit simultaneously.)

Linear regularization is an inverse technique suitable for deconvolution when we have some "missing information". It does not involve a model, but can use a smoothing rule. Remember, we have no range models.

Linear Regularization of a Lag Profile (Simulation)



This simulation shows that it is possible to remove the errors due to the long pulse at the expense of increasing the noise. The pulse used here is longer relative to the layer thickness than would normally be used. Thus the noise increase would normally be less.

Defining Errors in the Simulation of Linear Regularization

SUM over many(POWER(FFT(test profile - deconvolved function)))

→ Spatial Spectrum of total errors

Systematic and noise

POWER(FFT(SUM over many(test profile - deconvolved function)))

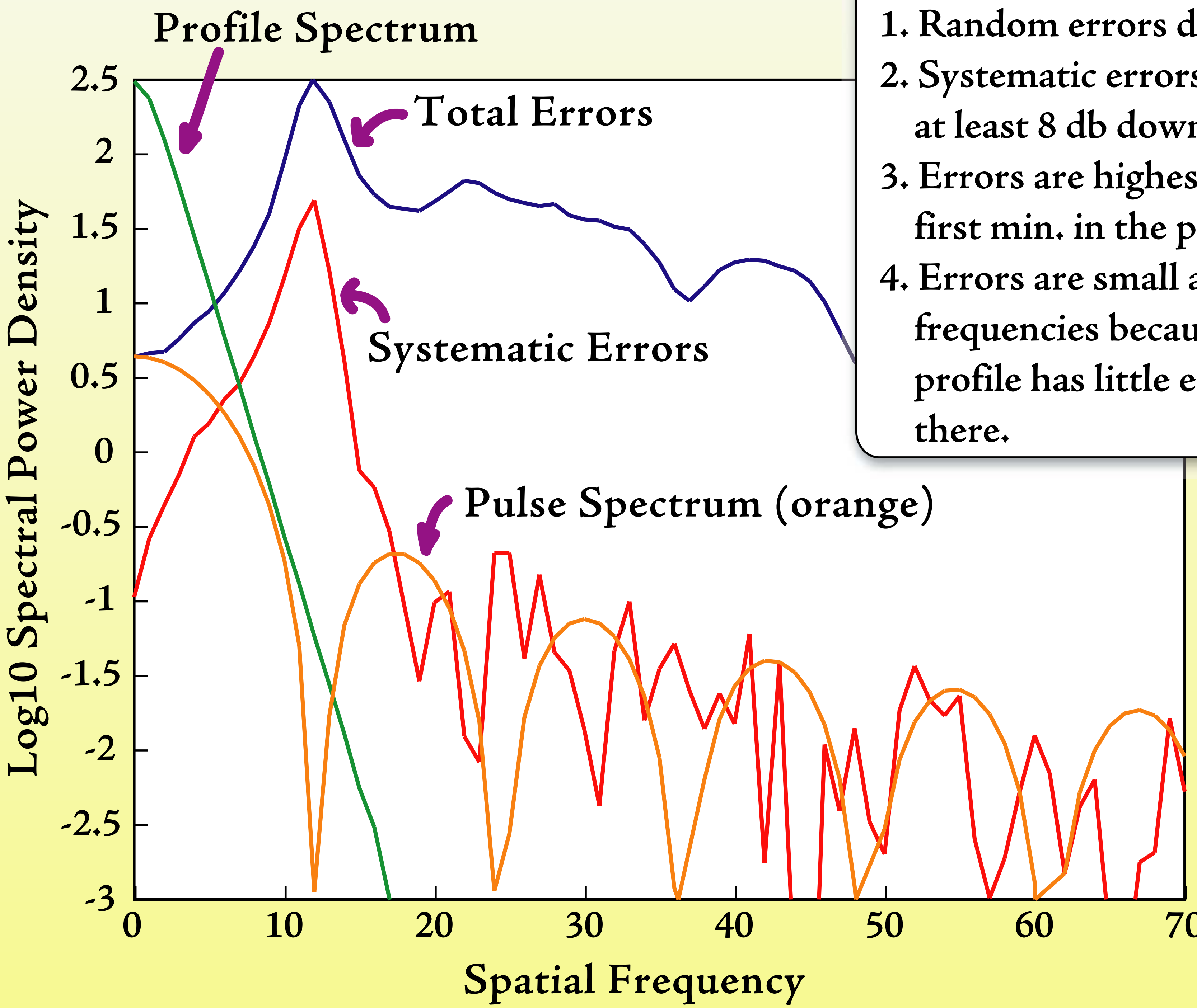
→ Spatial Spectrum of systematic errors alone

At what spatial frequencies (relative to the spectrum of the convolving pulse) do we expect the peaks in the errors?

Hint: what is missing from the convolution of the pulse and the profile?

Another question: How could we modify the radar technique to put back in what is missing?

Spatial Spectral Domain Errors in Linear Regularization

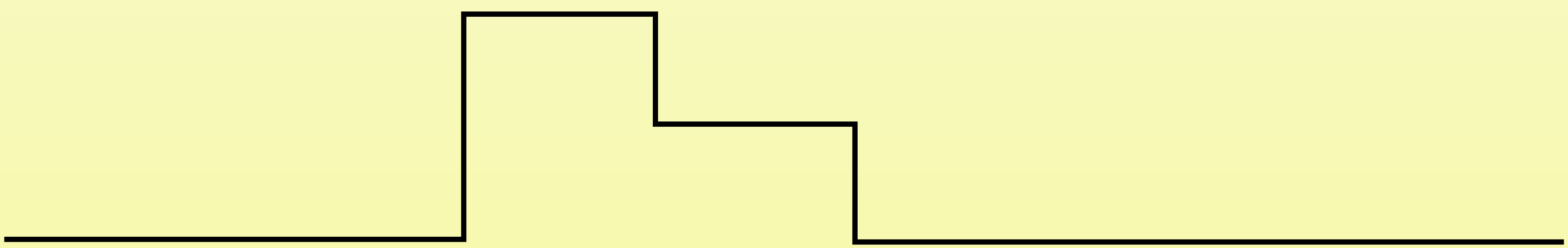


1. Random errors dominate.
2. Systematic errors are at least 8 db down.
3. Errors are highest at the first min. in the pulse.
4. Errors are small at high frequencies because the profile has little energy there.

How to reduce the errors in the Linear Regularization

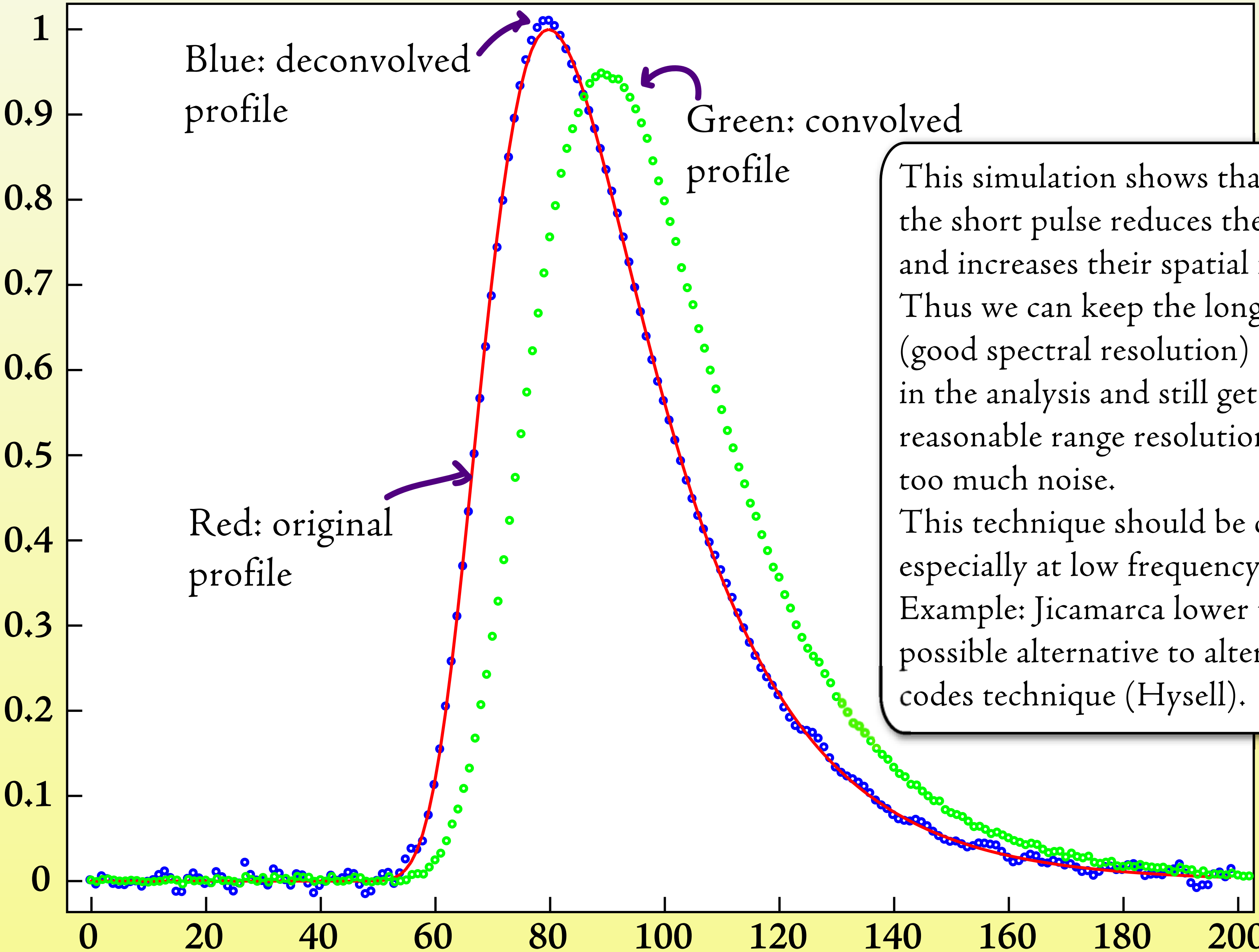
The following technique was first suggested by Lehtinen. This is a modification:

We need a long pulse to the long lags in the ACF.
We need a short pulse to get good range resolution.
Therefore alternate the use of both. If we add the profiles of each (paying attention to correct statistical weighting), we get a convolving function that might look like this:



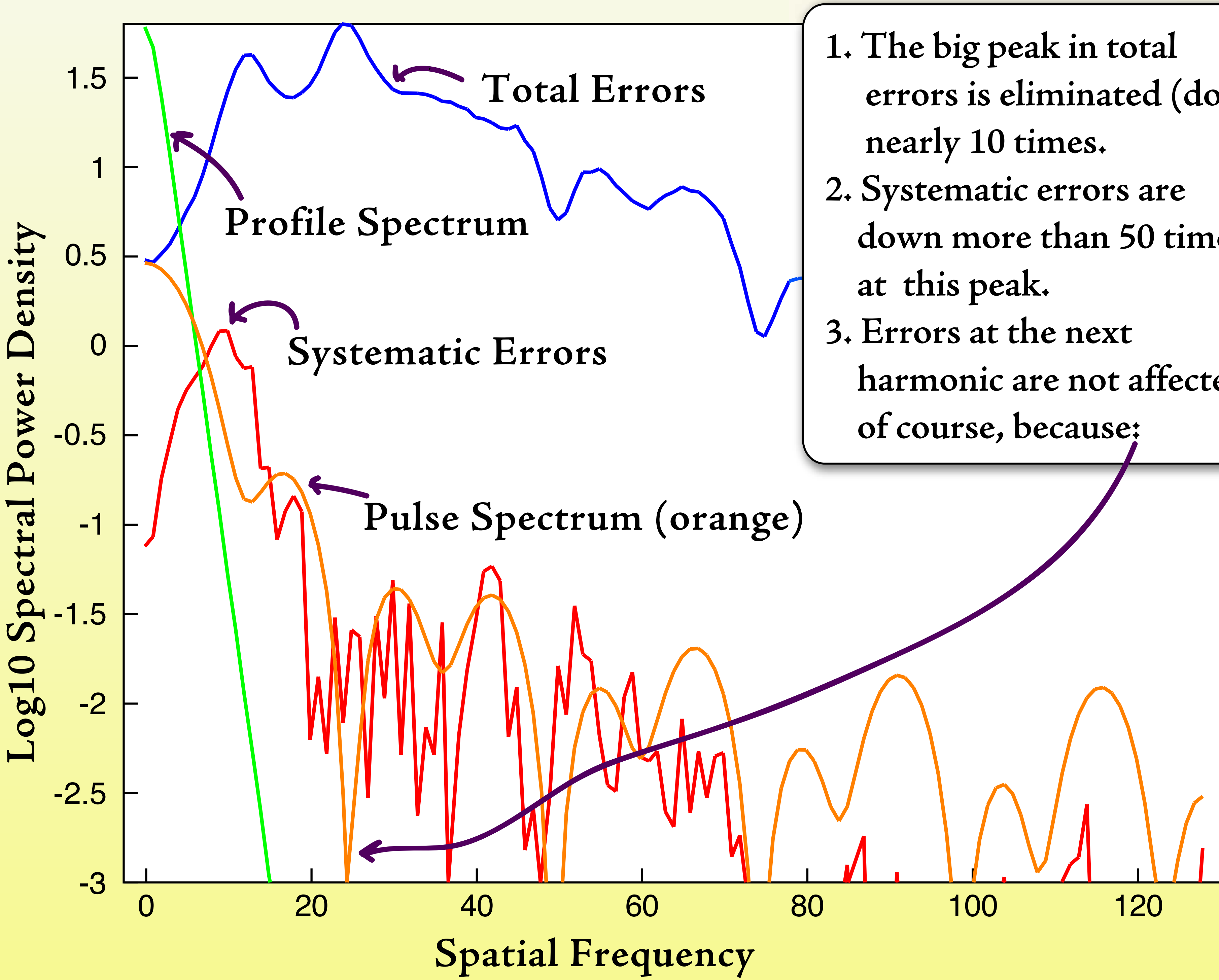
We can apply LR to this just as easily as to the function for a single pulse. This waveform has more higher spatial frequencies than that for a single pulse. Therefore we expect some improvement. How much?

Linear Regularization of a Lag Profile (Simulation with two pulses, different lengths)



This simulation shows that adding the short pulse reduces the errors and increases their spatial frequency. Thus we can keep the longer lags (good spectral resolution) required in the analysis and still get reasonable range resolution without too much noise.
This technique should be considered especially at low frequency IS radars.
Example: Jicamarca lower topside, possible alternative to alternating codes technique (Hysell).

Spatial Spectral Domain Errors With Two Pulse lengths



1. The big peak in total errors is eliminated (down nearly 10 times.
2. Systematic errors are down more than 50 times at this peak.
3. Errors at the next harmonic are not affected, of course, because:

Putting Linear Regularization and Fitting Together

Here is a summary of the data analysis process we have been discussing:

1. Compute and accumulate lag profiles as long as desirable.

2. Use linear regularization to deconvolve the lag profiles. Determine the errors (how big?, how correlated?). The errors are essential for the fitting.
   - •Note: at this stage we have higher range resolution than we need, and more noise than we would like. This might cause difficulties in the fitting, so...

3. Do the non-linear least squares fitting.
   - •Note: apply the technique discussed earlier of fitting across a range of heights at once. How much range? My feeling is that one wants to use a range no larger than necessary to perform the required averaging, using simple models for the range variation of the parameters. Then move to the next range cell, with some overlap for error check.