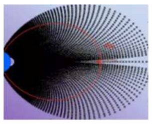
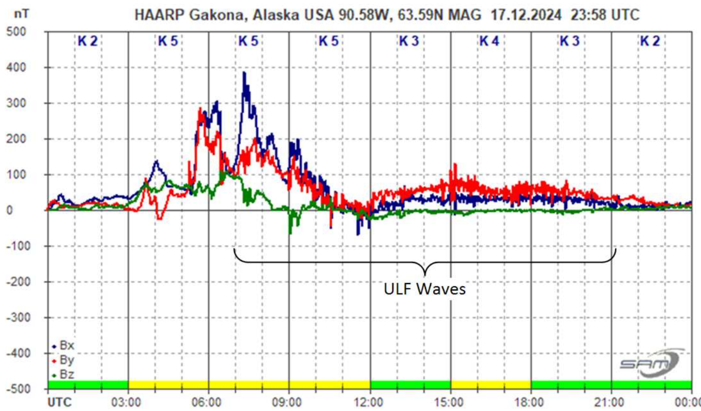
Calculating ULF Waves Spectra from SAM-III Magnetometer Data

Whitham D. Reeve

1. Introduction





The SAM-III magnetometer collects discrete time series data and displays it on a magnetogram similar to the one shown right. ULF Waves are seen as rapidly varying, low amplitude signals riding on the X-component (blue) and Y-component (red) traces. In this example, ULF Waves are visible from about 0700 to 2100.

The time domain data may be converted to the frequency domain by the *Discrete Fast Fourier Transform* (abbreviated FFT) for analysis of the ULF Waves frequency components. Software applications are available that perform the FFT calculations including Matlab, Mathematica, Mathcad, Python and Excel. This document describes using the *FFT Analysis* tool in Excel. FFT Analysis is part of the *Analysis Tookpak* and is installed in Excel as an *Add-in*. The tool is poorly documented, so this article provides detailed procedures. Some procedural information in this document related to using this Add-in was adapted from online sources.

The FFT Analysis returns a table that contains the two-sided power spectrum (amplitude and frequency) corresponding to the SAM-III time-domain data (amplitude and time). The table has both positive and negative frequency components and a dc (zero frequency) term.

2. Considerations

<u>Data Use</u>: The *original* data produced by the SAM-III magnetometer should never be altered. The data should be copied to a new file or Sheet in Excel before it is analyzed or processed. The SAM-III data file formats are described in more detail in the *SAM-III Software Setup Guide*.

<u>Windows internal processes</u>: The Microsoft Windows operating system internal processes are known to occasionally corrupt the incoming serial bit stream from the SAM-III processor as it is being collected by the SAM_VIEW software. Data corruption is rare and occurs most often around a HH:00:00 time stamp but can occur at any random time. Corruption manifests as errored data on a single line that is easily repaired. The errored data represent a very small fraction of the total data acquired in a 24 hour period.

<u>Data selection</u>: Generally, data is selected for the time range of interest, in particular, the time range of the ULF Waves to be analyzed. The entire 24 hour file is not selected unless 24 hours of data are to be analyzed. This avoids scrolling through large amounts of unneeded data when setting up and plotting the FFT for ULF Waves that last only a few hours.

<u>Number of datapoints</u>: The number of datapoints *N* to be used in the Excel FFT Analysis tool and related calculations must be a power of 2 (for example, ..., 256, 512, 1024, ...). The higher the number of datapoints for

a given sample interval, the better the frequency resolution of the transform. However, larger datasets require longer calculation times, and Excel becomes bogged down by large datasets. Datasets as large as 4096 samples have been analyzed with the Excel FFT Analysis tool while developing these procedures. See next item.

Sample interval: The SAM-III sample interval (or sample period) may be set from 1 to 120 seconds (sample rates from 1 Hz to 8.3 mHz) in 1 second steps. Common settings are 1, 10 and 60 seconds, corresponding to sample rates of 1 Hz, 100 mHz and 16.7 mHz. If the SAM-III is set to a 10 second sample interval, each hour of data contains 360 samples. If the analysis period is, say, 4 hours, there will be 1440 data samples produced during that period. In this case, the number of datapoints to be used for analysis in the Excel FFT is limited to 1024, not 1440, because the FFT requires the number of datapoints to be a power of 2. For reference, in a 24 hour period, a 1 second sample interval results in 86 400 datapoints, a 10 second sample interval results in 8640 datapoints, and a 60 second sample interval results in 1440 datapoints.

Windowing: Windowing, or tapering, is often used when transforming data from the time to frequency domain. In practical datasets – datasets of limited size – windowing is used to reduce spectra leakage (or sidelobes), which is where the energy of a signal is spread across more than one frequency bin. Without windowing, the FFT spectrum output does not perfectly match the actual input signal. However, windowing introduces amplitude errors. These errors depend on the type of window and type of input signal (sine, pulse, noise). When windowing is applied, the data is multiplied by a window function. A common window function is the Hamming Window, which is approximated by the coefficients $h(k) = 0.54 - 0.46 \cos(2\pi k/N)$, where k is the data index and N is the total number of datapoints. In this case, the window has a cosine shape. A Rectangular Window (also called Uniform Window), in which the data is multiplied by 1, does not alter the data and is the same as not applying a window at all. Many other window functions exist, such as Blackman, Hann (also called Hanning), Flattop and Sine, and may be used depending on specific requirements or preference.

Aliasing: Aliasing is the folding of higher frequencies to lower frequencies during data processing, thus contaminating the output spectra. Aliasing is avoided by filtering the input signal so that it contains no frequency components higher than one-half the sample rate. For example, if the SAM-III is set to a 10 second sample interval (100 mHz sample rate), the ULF Wave frequency measurements must be limited to periods longer than 20 seconds (frequencies < 50 mHz), corresponding to Pc 3, Pc 4, Pc 5, Pi 1 and Pi 2 ULF Waves pulsations. The frequency response of the SAM-III fluxgate sensors is not known but it is believed to be below 500 mHz. If the sample interval is, say, 10 seconds, frequency components higher than 50 mHz may contaminate the FFT. It is possible to apply a lowpass filter to the SAM-III data during post processing but prior to applying the FFT. Lowpass filtering of the SAM-III data is not covered here.

Frequency resolution: Frequency bin (or FFT bin) spacing is equivalent to frequency resolution and is fs/N, where fs is sample rate and N is the number of datapoints. Equivalently, the resolution is $1/(N * ts)$, where ts is the sample interval. For example, if the dataset has 512 samples and a sample rate of 100 mHz, the resolution is (100 mHz/512 =) 0.2 mHz. In this example, frequency components separated by < 0.2 mHz cannot be resolved. For a given sample rate, the resolution cannot be improved without longer observation times (and more datapoints); however, ULF Waves do not necessarily occur for long time periods, thus limiting the frequency resolution in many practical situations.

Other FFT processing errors: In addition to the errors due to leakage (spectra) and windowing (amplitude), other errors creep into FFT processing including scalloping loss and processing loss. These affect the power amplitude of the transformed signal and are ignored in this document because the presence of the spectra is more important than its exact amplitude.

Plotting: The FFT analysis produces the power amplitude of each frequency component, including a dc component. The relative power amplitudes are sufficient for the purposes of analysis. The two FFT frequency components are identical, so only the positive component is usually plotted. The dc component represents low frequency components below the resolution of the analysis, typically lower than a fraction of a mHz, so it does not need to be plotted. The amplitudes and frequencies are plotted on linear scales in this document. Either of these quantities may be plotted on a logarithmic scale but that is not covered here.

---

3. Procedures

The procedures described in this section consist of four basic steps: Prepare the SAM-III data files; Calculate the FFT of the data using the Excel FFT Analysis tool; Calculate the frequency scale associated with the transform; and Plot the results.

A. Prepare the SAM-III data files

1) The SAM_VIEW application software produces several files during normal operation. The filename to be used for the transform is the *yyyymmdd.txt*, where *yyyy* is the year, *mm* is the month and *dd* is the day of the file. This file contains comma and space delimited, columnar, time-stamped, raw data for each of the three magnetic components, X (north-south), Y (east-west) and Z (vertical).

2) Start Excel, navigate to the SAM-III data folder and then Open the desired .txt file (in the File Explorer window that opens, select Text Files in the dropdown to the right of the File name field to see the .txt files in the folder). Step 1 of 3 of the Text Import Wizard appears. Select the Delimited radio button. The .txt file does not have headers. Click Next.

3) Step 2 of 3 appears. Select Comma and Space and click Next.

4) Step 3 of 3 appears. The first data column is automatically selected. Select the *Date* radio button and the format *D/M/Y*. All other columns are *General format*. Click Finish.

5) Immediately after the file opens in Excel, Save As a .xls or .xlsx file. Do not skip this step because the original text file data may be accidentally overwritten. The data to be analyzed will be in Sheet1.

6) In the new Excel file, delete the columns containing only the text X, Y and Z. Each of these columns are simply indicators for the magnetic component data in the column to its right. The actual data columns are labeled in the next step.

7) Insert a row at the top of Sheet1 and label the associated columns, in order, Date, Time, X-data, Y-data and Z-data. At this point, there should be columns A (Date), B (Time), C (X-data), D (Y-data), and E (Z-data).

8) Scroll through the data and repair any corrupted lines and missing time-stamps. There are no ill-effects of replicating data or inserting a missing time-stamp on occasional, single lines of data.

9) Calculate the horizontal magnetic component, H, from the X- and Y-components. Label a new column F as *H*. Insert the formula = SQRT (*C-cell*^2 + *D-cell*^2), where *C-cell* and *D-cell* are the data cells in the X-data and Y-data columns, respectively. Example: For the X-data in cell C2 and Y-data in cell D2, the formula in cell F2 is =SQRT(C2^2+D2^2). Copy the formula to all cells in the column corresponding to the data. For convenience, Format all cells in column F as Number with 0 Decimal Places.

10) Normalize the horizontal component data. Label a new column G as *H-Norm* for normalized data. Normalize the data by subtracting the first *H* value (cell F2) from itself and from all other *H*-data cells. If only part of the dataset is being used in the transform, normalize only from the beginning to the end of the time period to be used. Example: The data to be used starts at cell F2. The formula in cell G2 is =F2-F$2 and in G3 is =F3-F$2, and so on. Alternatively, the actual value of cell F2 may be subtracted, as in =F2-10360, = F3-10360, and so on, where the constant 10360 represents the actual value of the data in cell F2 used for normalizing.

11) Select and copy the *H-Norm* data in column G of Sheet1. Paste (as Values) the data into column B of new Sheet2. If more than one time period is to be analyzed, the data for each time period may be placed in new Sheet3, Sheet4, and so on. This allows different time periods to be analyzed on different Sheets. Note that the time-stamps are not needed in the new Sheet2 so it is helpful to Rename the new Sheet2 tab to the time period being analyzed. Example: If the time period is 15:00 to 18:00, Rename Sheet2 to *H 1500-1800*, where H indicates Horizontal component data.

12) Label column A in the new Sheet2 as *Index k*. Enter the *index k* values starting with the value 0 as the first index value. Increment the index by 1 for each subsequent value. The index should count from 0 to N – 1. Example: Cell A2 contains 0 and cell A3 contains =A2+1, cell A4 contains =A3+1, and so on. If N = 4096 datapoints, then the last *index k* cell value is 4095.

13) The new Sheet2 should now have two occupied columns: Column A labeled *Index k* and column B labeled *H-Norm*.

B. Calculate the Fast Fourier Transform

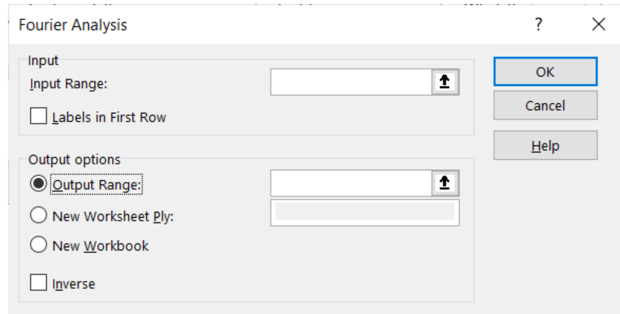
1) Enter the window function formula into a new column C in new Sheet2, and label it *Win*. Generally, the function will include a reference to the *Index k* in column A and the total number of datapoints *N* to be used. Copy the function to the cells in column C. Example: For the Hamming Window and N = 4096 datapoints, in cell C2 enter = 0.54-0.46*cos(2*pi()*A2/4096). If the Rectangular Window is used, enter the value 1 in the cells in column C.

2) Label a new column D as *H-Win*. Multiply the *H-Norm* cells in column B by the *Win* cells in column C. Example: in cell D2 enter = B2*C2. Copy the multiplication to all cells in column D corresponding to the data to be used. Column D now contains windowed data. If the Rectangular Window is used, the values in columns B and D are identical. For other window types, the data will vary with position in the column. For convenience, Format the cells in column D as Number with 3 or 4 Decimal Places.

3) Label four new columns E as *FFTk*, F as *xk (Hz)*, G as *xk (mHz)*, and H as *|FFTk|*.

4) Select the *Data Analysis…* option from the Tools menu or the Data ribbon. The *Data Analysis* window will open. If the *Data Analysis* option is not available, use online resources to learn how to install the *Analysis Toolpak Add-in* (it is free).

5) The *Data Analysis* window will show a list of *Analysis Tools*. Select *Fourier Analysis* and click OK. The *Fourier Analysis* window will open (right).



6) Click in the field to the right of *Input Range:* and select the cell in column D that is on the same row as the first datapoint used. Drag the selection in the input data column D to select the desired data. Be sure this data is the same as used in the previous steps and the number of datapoints is a power of 2.

7) Click on the *Output Range* radio button under the *Output Options*. Click in the field to the right of *Output Range:* and select the cell in column E that is on the same row as the first datapoint to be used. Drag the selection along column E the same length as the input data in column D.

8) Click OK to calculate the complex FFT coefficients and place them in column E. Excel provides a warning and will not do the calculation if the data range is not a power of 2. Note that Excel always displays 15 digits for both the real and imaginary parts of the FFT coefficients, and the number of displayed digits cannot be changed. If, after the analysis is run and the coefficients have been calculated, any changes are made to the data being transformed, the FFT Analysis tool does not automatically update the coefficients and the analysis must be run again.

9) Calculate the magnitude of the complex FFT coefficient. Select the cell in column H that is on the same row as the first datapoint used and enter the formula =IMABS(*FFTk-cell*) * (2/*N*) where *FFTk-cell* is the cell of the first generated FFT value in column E and *N* is the number of datapoints. <u>Example</u>: For cell H2 and N = 4096 datapoints, enter =IMABS(E2)*(2/4096). Copy this cell formula to all cells in column H corresponding to the data to be used. For convenience, Format the cells in column H as Number with 1 or 2 Decimal Places.

C. <u>Calculate the Frequency Scale</u>

1) Populate column F with the formula = *k-cell* / (*N* * *ts*) where *k-cell* is the corresponding value for the *Index k* from column A, *N* is the number of datapoints, and *ts* is the sample interval in seconds. <u>Example</u>: For cell F2, N = 4096 datapoints and 10 second sample interval, enter =A2/(4096*10).

2) Populate the column G cells with the formula = (*xk-cell*) * 1000, where *xk-cell* is the corresponding value in column F. <u>Example</u>: In cell G2 enter =F2*1000. Columns F and G now contain frequencies in Hz and mHz, respectively. Either may be used as the frequency scale in the next step, but mHz is more convenient for ULF Waves.

Upon completion of Steps A, B and C, the first three rows of Sheet2 should look similar to below, but the values will depend on *N* and the sample interval.

	A	B	C	D	E	F	G	H
1	Index K	H-Norm	Win	H-Win	FFTk	xk (Hz)	xk (mHz)	|FFTk|
2	0	0	= 0.54-0.46*COS(2*PI()*A2/4096)	=B2*C2	1154....	=A2/(4096*10)	=F2*1000	=IMABS(E2)*(2/4096)
3	1	3	= 0.54-0.46*COS(2*PI()*A3/4096)	=B3*C2	2140....+138...i	=A3/(4096*10)	=F3*1000	=IMABS(E3)*(2/4096)

Columns:

A Index k = 0 ... N–1 in increments of 1, where N is the number of datapoints (4096 in this example);
B Normalized H-component data from Sheet1, a total of N rows not including header row
C Window function, the example shows the Hamming Window
D Windowed data from Window function in column C applied to the H-Norm data in column B
E Resulting FFT data, real + imaginary of the form r + i
F Frequency scale in Hz based on the index, sample period (10 s) and N (4096)
G Frequency scale in column F converted to mHz (to be plotted on horizontal scale)
H Magnitude (absolute value) of the FFT data in column E with N = 4096 (to be plotted on vertical scale)

D. Plot the Spectra Relative Magnitudes with Respect to Frequency

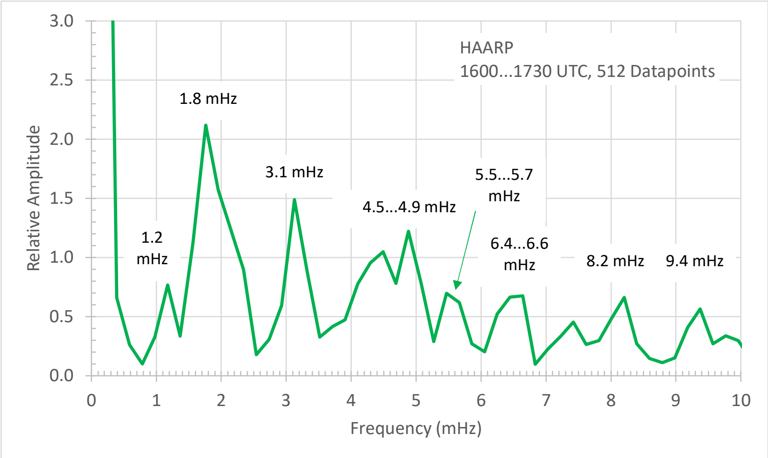
1) The plot X-axis uses the *xk* values in mHz from column G and the Y-axis uses the *|FFTk|* relative magnitude values (no units) from column H. Both axes are plotted on a linear scale.

2) The FFT is two-sided and both sides are identical, so only one side needs to be plotted. Select the upper one-half of the data in columns G and H and then Insert Chart Type *X-Y (Scatter)* with *Straight Lines*. When selecting the first-half of the FFT data, note that the first data cell contains only a real number (dc or zero frequency coefficient) and the cell immediately after one-half of the FFT data ends contains another real number (folding frequency coefficient); all cells in between contain real + imaginary numbers. The folding frequency is one-half the sample rate, or fs/2. The dc and folding frequency components need not be plotted. Example: The FFT contains N = 4096 datapoints and the corresponding index values range from 0 to 4095. Select the FFT (*|FFTk|*) and frequency [*xk (mHz)*] data corresponding to index values from 0 to 2047. This range includes the dc component but not the folding frequency component. If the dc component is not to be plotted, select the FFT and frequency data corresponding to index values of 1 to 2047.

3) In general, the frequency axis will have values from 0 to fs/2. Example: If the sample rate of the SAM-III data is 100 mHz (10 second sample interval), the frequency axis of the resulting chart will have values from 0 to 50 mHz. Similarly, if the sample rate is 1 Hz, the frequency axis will have values from 0 to 500 mHz.

4) Format and label the axes, traces and chart area as desired. Generally, the range of values displayed by both the x- and y-axis will require reduction to effectively zoom into the frequencies of interest, typically frequencies < 10 mHz. Suggested chart details: Aspect ratio: 1.67, Size: 3”x 5” (or 4” x 6.667”), Axis font: 9 or 10 point Calibri, Axis title and chart title font: 10 or 11 point Calibri, Plot line width: 2.0 pt, Arrow callouts width: 1.0 pt.

The annotated plot below shows an example of the results from Steps A, B, C and D above using 512 datapoints. This plot corresponds to the time period 1600 to 1730 on the magnetogram at the beginning of this document. Both the vertical and horizontal scales have been truncated from the full FFT Analysis to better show the individual spectrum components from 0 to 10 mHz.


HAARP
1600...1730 UTC, 512 Datapoints
1.8 mHz
3.1 mHz
4.5...4.9 mHz
5.5...5.7 mHz
1.2 mHz
6.4...6.6 mHz
8.2 mHz
9.4 mHz
Relative Amplitude
Frequency (mHz)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
1
2
3
4
5
6
7
8
9
10

Document Information

Author: Whitham D. Reeve

Revisions: 0.0 (Original draft started, 18 Nov 2022)
0.1 (Revised procedures, 19 Nov 2022)
0.2 (Added Introduction and Considerations, 20 Nov 2022)
0.3 (Updated text and added examples, 24 Dec 2024)
0.4 (Additional text updates, 25 Dec 2024)
0.5 (Added plotting info, 28 Dec 2024)
0.6 (Added magnetogram, 29 Dec 2024)
0.7 (Minor text edits, 04 Jan 2025)
0.8 (Refinement based on add'l data, 05 Jan 2025)
0.9 (Add'l considerations, 07 Jan 2025)
1.0 (Revised Introduction, 14 Jan 2025)
1.1 (Minor text edits, 08 Feb 2025)

Word count: 3516
File size (bytes): 147376