Hydrogen Line Galaxy Drift Scan Data Analysis Spreadsheet User Guide

by Jason Burnfield   25 September 2024

Introduction

This document is meant to be a companion guide to my Hydrogen Line Galaxy Drift Scan Data Analysis Spreadsheet. It includes an in-depth description of each tab on the sheet, what the various columns do, and how to use them. I hope that this document will make the use of the spreadsheet easier to understand. Each tab in the spreadsheet will be covered in the order in which they appear referred to by the title on the tab.

Purpose of the Spreadsheet

The purpose of the Hydrogen Line Galaxy Drift Scan Data Analysis Spreadsheet is to analyze the data from one or more drift scan observation sessions to generate a plot that hopefully reveals a clear spectral profile that matches the spectral profile obtained from professional sources.

Reference Profile Tab

The first tab is entitled "<*galaxy designation*> Reference Profile" (e.g. "M31 Reference Profile"). This tab contains a reference spectral profile in tablature form derived from professional survey data and then plotted using the x-y plot tool. Above the plot is a citation of the source including a URL where the paper can be found. This reference profile plot is reflected on the "compare" tab as a reference guide against which to compare the data from your own observations. I the template I have included reference profiles for M31, M33, M101, M81, NGC2403, M74, Holmberg II, and IC10. I suggest you delete all reference profile tabs except the one you wish to use and rename the spreadsheet with the name of the galaxy you are observing. If you wish to make observations of galaxies not included in the template, you can modify the "reference profile" tab to include the reference profile for a galaxy of your choosing. Simply find a professional survey containing the integrated spectral profile of the galaxy you want, convert the plot in the survey to a table of velocity and antenna temperature values. Then adjust the horizontal and vertical scale of the plot to accommodate the profile you are using. You must make sure that the reference profile plot in the "compare" tab points to the reference profile tab that you are using.

Raw Data Tabs

Next are the raw data tabs where the data from the observation files generated by the AirSpy SDR#/IF Averaging software is copied for analysis. There may be multiple raw data tabs in this section of the spreadsheet depending on how many observation sessions have been included. Each of these tabs should be named with the date of the observation session in *ddMmmyy* format (e.g. "6Sep24"). If you are adding a new raw data tab, select the last raw data tab already in the sheet and insert a new sheet after it. Then rename the new tab based on the date of the new observation session. Copy the contents of the previous raw data sheet into the new one and then delete all cells to the left and below and including cell “B2” to clear it as a template in which to paste the new data files.

Column “A” in raw data tab is labeled “Velocity” and contains the VLSR corrected doppler shift velocities calculated from the frequency values in column “B” which contains the frequency values from the first raw data file. The formula used for this conversion is the non-relativistic approximation doppler formula:

$$V = 300{,}000*(1420.406/Fo - 1)$$

The VLSR correction value contained in cell “A2” is subtracted from this value to get the final corrected velocity.

You must type into cell “A2” the VLSR correction value obtained from the following URL:

https://www.gb.nrao.edu/cgi-bin/radvelcalc.py

Go to the site and enter the center coordinates of the galaxy you are observing in RA and Dec format as well as the date of the observation. You don’t need to change the time because the contribution to the correction value from the rotation of the earth is negligible on the velocity resolution scale of these observations. This also means that it doesn’t matter that this tool calculates the correction based on the GPS coordinates of the GBT in Green-Bank, the value will work for you anywhere on Earth. The correction value you want to use is labelled “V_LSRK” on the site. From day to day, this value will only change by fractions to single digits of km/s. If you get two vastly different correction values for sequential dates, something is wrong with the calculation (double check your data entry for the coordinates and date).

Column “B” is labeled “Frequency” and is the place where you must copy the frequency values from the first data file. Leave row “1” for column labels and copy the data starting at cell “B2”. Select the “Data” menu and then the “Text to columns” option. Select “Fixed width” and click through the dialogue boxes until the text has been expanded into two adjacent columns. You will need to move the time/date stamp from cell “B2” to “C2”. All subsequent data files will use the same frequencies so you will need to delete the frequency column from those files as you copy them into the sheet.

Columns “C” and following will contain the linear amplitude values copied in from each of the data files starting with the pre-transit data, followed by the main-transit data, and finally the post-transit data. Begin with copying the next data file into column “D” starting at cell “D2”. Use the same method as before to expand the text into two adjacent columns. Select only the frequency values in column “D” and delete them selecting the “Shift cells left” option. Repeat this process until all of the data files for each category (“pre-transit”, “main-transit”, and “post-transit”) have been copied into the sheet. You may need to move the labels for the categories to mark the appropriate columns according to your observation plan and the time stamp on each data file. The number of columns in each category will vary depending on how many data files are generated for each.

Analysis Tab

The next tab is named “Analysis” and it is where the bulk of the analysis calculations take place. This tab separately averages all of the pre-transit, main-transit, and post-transit values together for each observing session at each frequency point. Then the main-transit average value for each frequency point is divided by the average of the pre and post-transit average values and the resulting ratio is converted to dB using 20Log. Finally, the dB values are smoothed using a 5-point running average.

Cell “B1” contains a center frequency value that must be adjusted to correct for VLSR. You must adjust this value until the velocity value in cell “B3” matches (within 2 or 3 km/s) the value in cell “A3” on each of the raw data tabs. These should not vary by more than 2 or 3 km/s from each other. If you observe for multiple days, the cumulative change in VLSR correction may make it necessary to delete a row in one or more of the raw data tabs to get them all to start at the same velocity value. Over the course of many days, you might have to delete as many as two or three rows from some of the raw data tabs to keep everything aligned properly.

Column “A” contains frequency values generated from the center frequency typed into cell “B1”.

Column “B” contains VLSR corrected velocity values which will be the actual values used in the spectral plots.

Column “C” averages the ratios contained in the “Main vs Ref” section starting with column “E” and these values are converted to dB using 20Log.

Column “D” computes a 5-point running average to smooth the values in column “C”. This column is used to generate the spectral plots.

The section starting with column “E” is where we compute the main-transit vs pre and post-transit ratios for each observation session. These columns are labelled “Main vs Ref” with an index number to indicate which observation session that column refers to (e.g. “Main vs Ref2”). The cells in these columns compute the ratio of main-transit vs the average of pre and post-transit values from the “Pre”, “Main”, and “Post” columns to the right which will be described in the next section.

If you are adding a new day of observation data to an existing spreadsheet, you will need to insert a new copy the previous “Main vs Ref” column right after it, re-index the new column, and change the formulas in the cells to point to the new “Pre”, “Main”, and “Post” columns that you will add to the next section of this tab.

At the end of the ratio calculation section is a blank column that aids in quick navigation on the sheet when adding new observations.

The next section of columns has three columns for each observation session labelled “Pre”, “Main”, and “Post” followed by the index number to indicate which tab the data is pulled from (e.g. “Pre3”). These columns average together all the pre-transit, main-transit, and post-transit values respectively for each frequency value in the observation data from that observation session's raw data tab.

If you are adding a new day of observation data to an existing spreadsheet, you must insert a new copy the previous "Pre", "Main", and "Post" columns after last set of columns already in the sheet, re-index the new columns, and change the formulas in the cells to point to the "pre-transit", "main-transit", and "post-transit" cells in the new raw data tab.

Plot Tab

The "plot" tab contains an x-y plot of the data pulled from columns "B", "C", and "D" in the "Analysis" tab. The x-axis is VLSR corrected doppler velocity in km/s. The y-axis shows both the Combined and Smoothed data from the "Analysis" tab in dB.

Flattened Tab

The "flattened" tab contains a polynomial flattening function that can be customized to correct for the curvature of the receiver response over the operating bandwidth. This allows the background to be flattened and set to zero so the final spectral profile can be seen clearly and will be displayed in dB above background.

The "velocity" and "smoothed" data values are pulled from the "Analysis" tab and placed in columns "A" and "B" respectively.

Column "C" contains the flattening function (called "func") generated by the polynomial coefficients which are typed into cells "E2", "E4", "E6", "E8", "E10", and "E12. My suggestion is to start will all coefficients set to zero and then go to a linear function (using 1^{st} and 0^{th} order coefficients only). Only go up to higher order coefficients if the background is curved enough to warrant it. The goal is to make the flattening function match the general background curvature of the uncorrected blue plot in the area around the expected spectral profile (leaving some room on either side). The result should be that the corrected orange plot will be mostly flat with its background close to zero.

Compare Tab

The "compare" tab contains a plot of the reference profile in blue at the top of the page followed by the corrected plot from the observations in orange. The horizontal scale of these two plots should match exactly so a direct comparison can be made visually between the two.

Final Plot Tab

The "final plot" tab contains an optimized version of the corrected (orange) plot on the "compare" tab. This has the vertical and horizontal scales optimized for the best presentation of the spectral profile. I also chose a color scheme to look like an oscilloscope screen (a purely aesthetic choice).