Week 3 Section IV Chapter 9: del Cep Exercise

In week 1 we did an exercise creating a smoothed model for delta Cephei. However we didn't discover the frequencies and harmonics for ourselves. We used the value in VSX and simply tried to duplicate the "smoothed" curve in section 1.4 in the text. We also didn't test for the goodness of fit or signal remaining in the residuals. Let's do some analysis from scratch. Don't look up del Cep in VSX or create a phase plot in VStar (since that uses information from VSX). You know from the name what kind of a variable this is, but if you are like me, you probably don't remember specific details about its period and amplitude.

Load del Cep data from JD246000 - JD2457900 using the "New Star From AAVSO Database" selection in the file menu. We will use the visual series. This gives us a lot of data to work with.

Let's not make any assumptions about the light curve or frequencies it may contain. We first should get a feel for the nature of the variable by looking at the light curve data. It is hard to see any pattern when looking at all the data except that the points seem to lie within a range and there is a pattern of seasonal density variations in observations. If you zoom in on the data between two seasonal gaps, say one in the middle of the data range, potential periodicity starts to become visible and if you zoom in further to say around 100 days it looks very likely that the data is periodic. If you zoom in even further to something that would seem to be around 10 periods, the periodic nature is apparent and you start to get an idea of the non-sinusoidal shape of the repeating pattern. You can get a rough idea of the period by taking an average of the intervals between, say 10 maxima or minima. The mean time between selections tool is a convenient way to do this but it is also easy to do manually with pen and paper or a spreadsheet. The mean period should come out to something just over 5 days for an equivalent frequency of around 0.18 or 0.19 cycles per day. You could have started by doing a DCDFT standard scan on the visual data but looking at light curve data at different time scales gives you a better idea of the variability you are dealing with.

By visual examination there doesn't seem to be an overall long term trend in the data that might affect the results of periodic analysis. To confirm that, run a polynomial fit of degree 1 to see if it is actually flat, or nearly so. Examining the model on the R tab of show models reveals that the slope is only about 2.23×10^{-5} based on an epoch at the midpoint of the data. Over the 1899 days of data the effect would only be about 0.04 magnitudes change compared to apparent amplitude of approximately 1.1 magnitudes based on visual examination, if there is actually a trend at all. The slope is sufficiently small that it could simply result from differences in density of data near minima vs. maxima in different parts of the curve. However, since trends can add frequencies to periodic analysis let's be cautious and de-trend the data. Take a screenshot of the linear model on the R tab and save it as **<yourlogon>_delCep_Vis_R_LinModel.<usual suffixes>**

Now we have a very rough idea of the principal frequency in the light curve and have de-trended residuals on which to perform DCDFT analysis. Before working with the de-trended residuals, perform a DCDFT on the visual data with minimum frequency of 0.0011, max Frequency of 4 and the default scan resolution of 0.000135. These are the same parameters we will use for DCDFT on the de-trended data. Save this Periodogram as **<yourlogon>_del Cep_Vis_0.0011-4-def_Full.<usual suffixes>** in case you want to compare it to the corresponding periodogram for the de-trended data. There should only be small differences between them in this case, but it is always good to see for yourself. You probably know that you can have multiple instances of VStar open at the same time in case you want to compare Zooms of periodogram details of the visual data with the de-trended data.

Before doing anything with the de-trended residuals ("Res 0" from now on), go to the residuals tab and save them using save from the file menu. Call the file <**yourlogon>_delCep_Vis_Res0.txt**. This will save the file as a tab delimited text file that you can load directly into VStar. You can add explanatory comment lines (such as column headers) to the file as long as you start those lines with a pound sign (#). Let's give the residual data a name that will be displayed in the plot window and DCDFT windows by adding a comment line

#NAME=del Cep Visual Res 0: De-Trended Visual Data **(don't use any commas on this line)**

It is always a good idea to save successive residuals so that you can go back to a previous iteration rather than starting over from scratch. You Can simply use consecutive numbers in the file names.

If you want to perform a standard scan just to get an idea of what the periodogram looks like feel free to do so. We are going to do our exercise with some different settings (the same settings used for the initial DCDFT on the original data) for reasons outlined in the discussion below.

Let's do our exercise DCDFT on the Res0 data with a frequency range 0.0011 for the low frequency. That corresponds to a period slightly less than half the duration of the data and therefore the data should have at least two full cycles of any signal in the scan range. Since the data clearly has a random component to the time interval between observations, we should be able to differentiate signals above the (sort of) Nyquist frequency (≈1.72 cycles per day, half of the average frequency of observation $= \frac{1}{2\tau}$. Lets pick a round number a bit more than 2x the Nyquist frequency, say, 4 cycles per day. We know from our rough estimate of the primary period that this cut off should allow us to capture up through something slightly above the 20th harmonic. We also know from our examination of the original light curve that the wave shape is somewhat triangular but without the discontinuous jump of the saw tooth waveform in the book. Therefore we should expect that a fair number of harmonics will be required to model the signal but probably nothing near the number needed to get a good model of the saw tooth waveform. We don't know at this point from our examination of the light curve whether it contains multiple fundamental frequencies. However, since del Cep is the prototype DCEP variable, we suspect that if the light curve does contain other fundamentals, they aren't very prominent. Hopefully, our analysis will show this. So we have good reasons other than general thoroughness to look at frequencies above the Nyquist frequency. The choice of 4 cycles per day is somewhat arbitrary but it seems reasonable based on the discussion above. If needed, we can always expand it to higher frequencies. There are approximately 350 primary periods and substantial "randomization of observations in our data so we should be able to determine signal frequencies much higher than 4X the Nyquist frequency, if necessary. Leave the frequency scan resolution at the default, 0.000135. That should be ample to make sure we don't miss any peaks. Make sure that "show top hits " is checked and logarithmic is off. Also apply CLEANest to the top hit frequency. That will add a line at the very top of the top hit list and a black line on the periodogram. The CLEANest output is more precise and you will use it to create your model. Save the periodogram **as <yourlogon>_del Cep_Vis_Res0_per_0.0011-4FScan_Full.<usual suffixes>** and the Amplitude spectrum as **<yourlogon>_del Cep_Vis_Res0_amp_0.0011-4FScan_Full.<usual suffixes>**. Explore the Periodogram by zooming in and out on various features.. If you want to save and present any of these zooms use the same file name as for the complete plot but replace "Full" with the approximate frequency range of the zoom. The top hits list is too long to see in the VStar window. Therefore, save the table of data to a tab delimited txt or spreadsheet file (click on the top entry, scroll to the bottom and shift-click on the bottom entry copy and paste into the file). Name the file **<yourlogon>_del Cep_Vis_Res0_tophits_0.0011-4Fscan_Full. < xls or txt >**.

Try to identify the cause of a bunch of the top hits as fundamental frequencies, harmonics and aliases or side lobes. You don't have to do every one but you should be able to identify at least 8 to 10. It is probably easiest to do this in a table format that you can construct by adding some columns to the TopHits file and saving it as a spreadsheet file. That will also allow you to sort the table in different ways, which can be very convenient. **Use the template accompanying this post**. Instructions for filling in the template are included in footnotes on the template. Common Aliasing sources are the observation (sampling) frequency (this may not be exactly the 1/τ calculated by our formula since observations are randomized), daily gaps, annual gaps, full moon, and perhaps other seasonal variations. Remember the gaps are also randomized so, for example, the frequency of annual gap is unlikely to be exactly 1/365 and the daily gap may not be exactly at a frequency of 1.0 cycles per day. Also because of noise and, to a smaller extent, the resolution of the frequency scan, aliases and harmonics may not appear at exactly the frequencies you get by calculation. Something you may recognize in doing this exercise is that density variations and gap in our observations create real signals in the data that are not real signals in the light from the star.

When creating the table you may want to take screen shots of zoomed subsections of the full plots for reference.

Once you have created the table with sources of signal identified, let's subtract a model that uses the CLEANest frequency output from this DC DFT highest power top hit and harmonics through the highest significant one you discovered. For example if you found a significant 6^{th} harmonic but not the 5^{th} select 6 from the harmonic list in the model DCDFT model dialog. Please save a screenshot of the model from the model function tab as **<yourlogon>_del Cep_Vis_Res0_model.<usual suffixes>** and the residuals in a file as **<yourlogon>_delCep_Vis_Res1.txt**. You may also want to add a comment Header to the file #NAME=<yourlogon>_delCep_Vis_Res1 at the top of the file which will be displayed in plots and DCDFT results if you use this file for the next part of the exercise. Before moving on to the Res1 residuals make a phase plot displaying the Res0 data and the 6 harmonic model using the default epoch of the phase plot dialog and the period corresponding to the CLEANest output frequency you used to make the model. It should be a very good fit by visual inspection.

Moving to the Res1 residuals, Select them as the active plot in plot control and select a reasonable bins size for means. You don't want the bin size to be close to an integer multiple or small integer divisor of the main period, but you would like to have 5 or more points per bin in almost all bins (you can't help the gaps). I picked 4 days. Try several if you like. Check the info dialog for the F statistic and p value. It should not be surprising that the p value is so small that we must reject the null hypothesis that the data results strictly from white noise if for no other reason than the annual gaps are evident in the residuals. Those gaps will certainly have an effect on bin means. It isn't clear that the ANOVA result indicates there is more real signal to find. If you overlay the means plot on the Res0 data plot it appears possible that there may be some remaining signal. Let's do one more thing with the binned **means** of residuals before leaving them. Let's do a DCDFT on them using the same DCDFT parameters as for Res0 and take a screenshot as <yourlogon>_delCep_Vis_Res1_<x>dMeans_Per.<Usual suffixes>. "<x>"is whatever bin size you used for the means. Does one familiar frequency stand out as significant? In my results the background level seemed to be about 4 and none of the other top hits were more than about 10 and, therefore, weren't significant.

Now let's perform a DCDFT on the Res1 residuals themselves using the same settings as used for Res0. Our model frequencies and aliases of those frequencies should now be removed. What are the first things you notice in looking at the periodogram. Are there any significant signals? What does the highest power signal represent? Does there still seem to be what appear to be daily gap aliases of the fundamental frequency and some of its harmonics in the res1 data? I saw that in my results and yet there is no signal remaining at the fundamental and harmonics themselves. You would expect those frequencies to be present as well if the model did not completely remove them. The alias powers are greatly reduced but not completely eliminated. We will learn about a possible explanation for this in the chapter 12 material on interacting frequencies. They may be aliases of interacting frequencies from modulation of the real signal that may be real or only the result of uncertainty in observations.

If you look at the amplitude spectrum, do any of the peaks seem significant? Remember the criterion is a bit different. It is based on 4X the average background level (flat water level) rather than 3X the "forest of peaks" level. In my results no peaks except one at a familiar gap frequency are significant in the amplitude spectrum and in private discussion with Michel Breger, originator of Period04, the amplitude spectrum criterion is the more reliable. Save periodogram, and amplitude spectrum screen shots **as <yourlogon>_delCep_Vis_Res1_per_0.0011-4FScan_Full. <usual suffixes>** and **<yourlogon>_del Cep_Vis_Res1_amp_0.0011-4FScan_Full.<usual suffixes>**. Save your top hits as a file **Cep_Vis_Res0_tophits_0.0011-4Fscan_Full. < xls or txt >**.

You can continue on with more analysis on successive residuals if you like but this has already been a lengthy, but I hope instructive, exercise and I think it should have located the most significant frequencies.

Pleas post your question answers, comments screen shots and files in replies to this topic.