Week 2 Section II: Chapter 5 Polynomial fit Exercises

Exercise 1

Load Eps **Aur data** for the period **JD2455050 to JD2455150** from the AAVSO database and select only the Johnson V series to be plotted.

Fit a first order polynomial model and display it in the raw plot window with the Johnson V series and also display the polynomial fit metrics (from the model selection on the Analysis menu) so that the metrics do not obscure the data and model. Take a screenshot and save it as **<yourlogon>_EpsAur_1stO_JV-Model-fitstats.<usual_suffixes>**. Also take a screenshot of the R model tab of the linear fit and save it as **<yourlogon>_EpsAur_1stO_JV_Rmodel.< usual_suffixes>**. Just from visual inspection of the model fit to the data, does the model look reasonable? Clear the fit metrics and set the plot window to display just the residuals. Does it look like signal may be still be present in the residuals?

Now let's get a quantitative evaluation whether the residuals contain only white noise. Display only the residuals and means of residuals using 10 day means (with error bars) in the plot window. As always don't forget to click on "Apply" after changing the means bin size. There isn't any particular magic about the selection of bin size. Ten days simply gives us a reasonable number of bins and a reasonable number of points per bin to make an evaluation about whether the residuals result only from white noise. Give the plot a new name that describes what is being displayed and save a screenshot as **<yourlogon>_EpsAur_1stO-resids-Means10d.<usual_suffixes >**. **Does visual inspection of the residuals and their means suggest whether or not the residuals are strictly the result of white noise?**

Apply current mode ANOVA. What does the p value of the ANOVA calculated on the residuals tell you? Save a screenshot of the current mode anova as **<yourlogon>_EpsAur 1stO-resids-CMA_Means10d.<usual_suffixes>**. Try other bin sizes to get an idea how sensitive the p value is to bin size and **specifically include in your comments whether the p value crossed from one side of the 0.05 threshold to the other**.

Display the Raw residuals only in the plot window. **Do they have a clear horizontal "S" shape as opposed to a "bowl" shape?** That is an indication that the residuals are likely to be dominated by a 3rd order polynomial. That means a 3rd order polynomial is a good starting place for a model to fit to the residuals left after subtraction of the linear trend (data "pre-whitened" by a linear model). However, it doesn't mean that a higher order polynomial might not be a better fit. I tried up through the 7th order and the third order was the best fit according to both the AIC and BIC statistics. As you would expect the RMS of residuals got smaller as the order of the polynomial was increased, but by small amounts. If you want to try fitting several different orders of polynomial to the residuals after subtraction of the linear trend, I suggest you save the first set of residuals as a new data file from the residuals tab. Save it with either a ".txt" suffix since a save of the residuals tab is a tab delimited text format. You can add a chart title to the file as well, with a first line starting #NAME=plot name

Fit a third order polynomial to the residuals from the 1st order polynomial and display the model along with the residuals and 10 day means of residuals. The residuals you see are the **residuals from this new**

model fit not the residuals to which you fit the model. This is a little different than when you fit the first model with the Johnson V data displayed. Since we selected residuals as the series to be displayed, you see the residuals from the most recent model fit. Display the fit metrics for the latest model in a similar manner as for the first order fit. Save a screenshot as **<yourlogon>_EpsAur_res1_3rdO-model-fitstats.<usual suffixes selections>**. How do these metrics compare to the metrics for the 1st order fit to the data? If the first set of residuals contained signal that could be modeled by a 3rd order polynomial it is reasonable that the metrics would be better - smaller the RMS of residuals and smaller AIC and perhaps also the smaller BIC, which in this case means more negative.

Display the R model info window for the 3rd order fit to the residuals and save a screenshot as **<yourlogon>_EpsAur_res1_3rdO_Rmodel.<usual suffixes selections>**.

Now let's quantitatively evaluate residuals of the 3rd order fit to residuals. Close the model info and model windows use the current mode ANOVA tool on the residuals from the 3rd order fit. Make sure the 10 day bin means with error bars are also displayed and save a screenshot as **<your logon>_EpsAur_res1_3rdO-resids-CMA_Means10d.<usual suffixes selection>**. **What does the p Value indicate? Specifically, even if you can't exclude the null hypothesis that the data results only from white noise, does that prove the residuals result only from white noise or only indicate that there is a significant probability that the data are only the result of white noise?**

Now you have two models the first order polynomial fit to the data and a third order polynomial from the fit to the residuals after subtracting the first model.
$f_1(t - t_0) = \alpha_o = \alpha_1(t - t_0)$, and
$f_2(t - t_0) = \beta_o = \beta_1(t - t_0) + \beta_2(t - t_0)^2 + \beta_3(t - t_0)^3$. Since both models use the same observation reference epoch, t_0 is the same for both models, and the two models can be added together to form our "total" model of the signal:
$f_{1+2}(t - t_0) = \gamma_o + \gamma_1(t - t_0) + \gamma_2(t - t_0)^2 + \gamma_3(t - t_0)^3$. It gets a lot complicated if t_0 is different for different models, for example, fitting different "bias adjusting models to different subsets of the data, say to different observers, and then are combining those bias adjustments together with an overall model fit to all of the data. The t_0 values would be different for each subset model since VStar uses the average of the data point JDs of the series, (including a filtered series), as t_0. Then you can't simply add the models together.

It is usually better, however rather than adding the two models together to run a single fit on the original data using the highest order polynomial you used in your sequential fits, a third order polynomial in this case.
$f_3(t - t_0) = \delta_o = \delta_1(t - t_0) + \delta_2(t - t_0)^2 + \delta_3(t - t_0)^3$

The results in most cases will be the same or very close to the same if the noise is white or close to white. For example the sum of coefficients I got for the sequential fits was identical to the coefficients from the single third order fit to the original data out to at least 13 significant figures (as far as the display of the numbers was expanded). However, there may be some small disagreement if the noise isn't white, for example, if there is auto-correlation of the noise.

Try doing a third order fit to the original data. Save a screenshot of the R model info dialog as **<yourlogon>_EpsAur_3rdO_JV_Rmodel.< usual_suffixes>** and fit statistics as **<yourlogon>_EpsAur_3drO_JV-model-fitstats.<usual_suffixes>. Are the summed coefficients of the sequential fits the same as in the single 3rd order fit to the Johnson V data?** Finally save an image with the Current Mode ANOVA results from residuals with the same 10 day means. **Are the F statistic and their p value the same for the single 3rd order fit as for the two-stage fit?** Take a screen shot of the residuals plot with Current Mode ANOVA results showing as **<yourlogon>_EpsAur_3drO_JV-resids_CMA.<usual_suffixes>.**

Post your responses as a reply to this post with image attachments.

Exercise 2

In this exercise we are going to locate a time of minimum of a star using two different sets of sparse data. Then we locate a minimum of another star that has a much denser set of data and compare the results. Let's locate a minima of Beta Lyrae (**Bet Lyr**). Looking at the JV light curve of the star in the new AAVSO enhanced light curve generator (https://www.aavso.org/LCGv2/) there seems to be a nice block of data between 2457500 and 2457700 with a particularly good patch of data between 2457600 and 2457640. Zooming in on that reveals that there is a nice block of data for the minima at about 2457609. If we want to locate the minima we want to include enough data to locate that critical point but not include the adjacent maxima, and if there are enough points, we want to exclude the inflection points on either side of the minima where the graph starts turning from an upright bowl to inverted bowls. If you include the adjacent maxima you might fool the modeling program into returning one of the maxima as the extreme in the model. There aren't a lot of points so we can't really tell where the inflection points are but a visual inspection of the light curve using the Enhanced LCG zoom box seems to indicate that a date range of 2457607.0 - 2457611.0 might be a good choice. We can check this choice since Bet Lyr is a well-studied star, we know from VSX (the Sebastian Otero estimates in remarks not the 2019 epoch and period values) that the period was **12.94061713** days at **epoch 2455434.8072** (about 6 years before our observations). If the system orbit is circular (and that won't be far off for an EB binary system) then the next maxima should appear at minimum +/- approximately ¼ of the Period (between the primary & secondary minima), or about +/- 3.2 days from the minimum. Therefore, the choice of something around +/- 2 days from 2457609 for our data seems reasonable. Without the additional information from VSX we would simply have to rely on visual inspection of the data to select an appropriate data range.

Let's load this date range (**2457607.0 - 2457611.5**) into VStar using New Star from AAVSO database. After loading the data, the minimum seems to be closer to 2457609.5 Since we are trying to locate the time of an extreme point of something that is bowl shaped you might expect that the best fit will be an even order polynomial. That would be true if our data and the "real" signal it contains are symmetrical around the minimum, but they may not be in all cases. Therefore, we will try even and odd polynomials.

Let's try to find the time of minimum for the Johnson V series. Start with order 2 and repeat for increasing orders until both the AIC and BIC tell you the fit is getting worse. Because our series is bowl shaped, if this occurs on an odd order also try the next higher even order. It may still be a better fit.

Make a table (xls, or tab or comma delimited txt, doc, or pdf) with the following column headers, Order, RMS, AIC, BIC, Min_JD and Min_Mag. The table will have one row for each polynomial order fit you make. RMS, AIC and BIC will come from the "Metrics" tab of the model info dialog and the Min_JD and Min_Mag come from the "Extrema" tab. Just make sure that you have the correct model selected when you select "Show Model."

Save your file containing the table as **<yourlogon>_BetLyr_JV_min_table.<xls, txt, pdf, png, doc>.** You can now upload files with xlsx, docx or csv suffixes. If you save the table as a.txt file I don't care if it is comma delimited or tab delimited.

Which fit do you think gives the best estimate of the time of minimum of the signal contained in these data? How did you arrive at your conclusion? Save the plot window (or take a screenshot) with Johnson V data and the model you chose as giving the best estimate of the time of minimum displayed as **<yourlogon>_BetLyr_JV_Best_<order#>.<usual suffix choices>.** Don't forget to change the title of the plot to something that describes its content. Please notice that I did not ask for the model that gives a minimum closest to the JD you might calculate from the epoch and period in VSX. Compare the raw plots of your models superimposed on the Johnson V data. Which one do you think most <u>realistically</u> represents the underlying eclipse signal? By realistic I mean captures the shape of the light curve during the eclipse rather than noise and avoids prominent fit "artifacts" (wiggles in the model particularly at the ends of the data set that don't represent signal). Save the plot (or make a screenshot) as **<yourlogon>_BetLyr_JV_realistic_<order#>.<usual suffixes choices>**. **If you chose different models for best minimum estimate and most realistic, explain why?** Please do not infer from this question either that they should be the same or should be different. If choosing the best model to use seems a bit more difficult you expected the article on Beta Lyra gives some insight into why that may be.

Now repeat the process for the visual data in the same time period including appropriate name changes (vis to replace JV) to your saved files. The same questions apply as for the analysis of the Johnson V data. **How close are the results of your "best" time of minimum estimates for Johnson V and visual**?

If you want to see how things can go wrong if you pick too wide a JD range when fitting a critical point try fitting polynomials to the visual data over a range of JD 2457605.0 - JD 2457613.0 and see what happens when you get up to the 6^{th} and 7^{th} order fits. The metrics are still improving at 7^{th} order so you are fitting more curves, than required with a more restricted date range, but also see what happens to the fit "artifacts" at the ends of the datasets and the Extrema estimates? Egad! We are overcome by nonsensical results.

VStar does not provide error estimates for the time of minima estimate. However, if you were reporting the time of minimum for a newly discovered binary system you should report an error estimate for the time along with your estimate. One way to estimate the error is to calculate the times on each side of the minima at which the polynomial model differs from the minimum or maximum by the standard error of the maximum or minimum and use that as the standard error of the time. unfortunately computing the standard error of the maximum or minimum is complicated and far beyond the scope of this course. There are other methods of estimating the time of minimum and associated error. The most common one is the Kwee-van Woerden Method. See posts in Week X in the section on maximum-minimum detection

How close are your times of minima to the predicted time of minimum based on the VSX epoch (of primary minimum for an eclipsing binary system) and period? **Calculate the differences as Observed minus Calculated (O-C)**. As a first approximation, assume Bet Lyr is strictly periodic with only this one period present and the period is not evolving from the period at the reference epoch. Assume the epoch of our observations should be exactly **168 VSX periods** after the VSX epoch:
Our results are JD but easily convert to HJD using http://britastro.org/computing/applets_dt.html.

I don't want to influence your selections, and therefore, will hold my results until most of you have posted yours and we are in discussion about the exercise. We don't cover O-C diagrams in this course because VStar doesn't create them. They are covered in the last Chapter of the text for the course, however. In my results I will include a calculation of the O-C offset resulting from the change in period mentioned in the text. Try your hand at making the correction if you are so inclined and include it in your results.

For the last part of this exercise let's locate a minimum of **V0711 CrA**. This is another EB star. Load the data for JD **2457597** - JD **2457599** from the AAVSO Database. This is a dense set of Data for a minimum, and if you look at the observations tab in VStar you will see that it is all transformed data from a single observer. Among other things, that means this data should generally conform to our assumption of zero mean white noise except, of course, from the noise-like variability that is caused by the active nature of this system. After inspecting the data we just loaded, it looks like the inflection points where the curvature starts changing from an upright bowl to inverted bowls are at about points JD **2457597.96** and JD **2457598.08**. So reload just the data with those starting and ending JD. **In your post for this exercise include the JDs you would have picked for the inflection points**.

Now let's repeat the exercise we did on the Bet Lyr on the Johnson V data for V0711 CrA Start with a second order polynomial again. Make a table and **save the same images with appropriate name changes. How did the best polynomial order giving the "best" time estimate for this data compare to the orders you used for the sparse Bet Lyr data? If you chose different orders for best minimum time estimate and most realistic order for Bet Lyr did this data allow better agreement between the best minimum estimate order and the most realistic fit order? If there are differences in agreement between the two fit orders or other qualitative aspects of the result, can you think of reasons why these differences might occur? What is the HJD of minimum you estimated from the data for V0711 CrA and what offset (Observed - Calculated) is that in days and phase from the closest predicted minimum JD using the epoch and period in VSX (E = HJD 2448500.517; P = 0.700892 d).** If the closest calculated time of minimum is later than the observed minimum this will be a negative number. This isn't a reliable O-C comparison since it assumes that the closest integer number of calculated periods is actually the corresponding number of periods to the minima we measured. It may be true but not always. For example, our data for V0711 CrA, is almost 13,000 periods from a VSX epoch that is only given to 3 decimal places, we need intermediate observations to compare to the calculated ephemeris to make sure our assumption isn't off by one or more periods in either direction. In the case of Bet Lyr, where the signal appears to contain multiple frequencies that may not be harmonically related, the interaction of frequencies can shift the maxima and minima away from the Points where the maxima and minima of the primary frequency occur. In some cases, multiple fundamentals can make the signal appear chaotic making it impossible to reliably detect maxima and minima of any fundamental frequency visually, and the times of light curve maxima and minima may vary in an apparently (but not really) random way. The two attached images of a signal synthesized in Excel show two periodic functions comprised of the same two fundamental frequencies with the same phase angle at t_0. The only difference is the ratio of the amplitudes. We will delve into this further when we get into multi-periodic signals.

Post your tables, pictures, answers to questions, observations and questions as replies to this topic. These exercises should provide plenty of discussion to occupy us this week.