Link to HW6 code

1 Task 1

I chose to use the J-band as it is the least distorted of the infrared bands due to its (comparatively) short wavelength. While this choice makes no difference for this assignment, I want to make explicitly clear that I don't appreciate how much effort one must go through when dealing with observations in those particular bands. Ranting aside, the design matrix was constructed around the variables in Equation 3 on the homework assignment, which are $\log P$ and $[Fe/H]$. This meant that the design matrix was

$$\mathbf{X} = \begin{bmatrix} 1 & \log P_1 & [Fe/H]_1 \\ 1 & \log P_2 & [Fe/H]_2 \\ & \vdots & \\ 1 & \log P_n & [Fe/H]_n \end{bmatrix}$$

and the model parameter vector was just the coefficients in question, namely

$$\theta = \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix}$$

Using Equation 12, the best linear least-squares fit parameters were found to be $\alpha = \mathbf{-3.018}$, $\beta = \mathbf{-2.638}$, and $\gamma = \mathbf{-0.077}$ with respective errors $\sigma_\alpha = \mathbf{0.143}$, $\sigma_\beta = \mathbf{0.168}$, and $\sigma_\gamma = \mathbf{0.252}$.

2 Task 2

Figure 1 shows the fit (in red) compared to the scatter of the data (in blue). The fit seems to represent the average trend in the data well. There is not much more to be said.

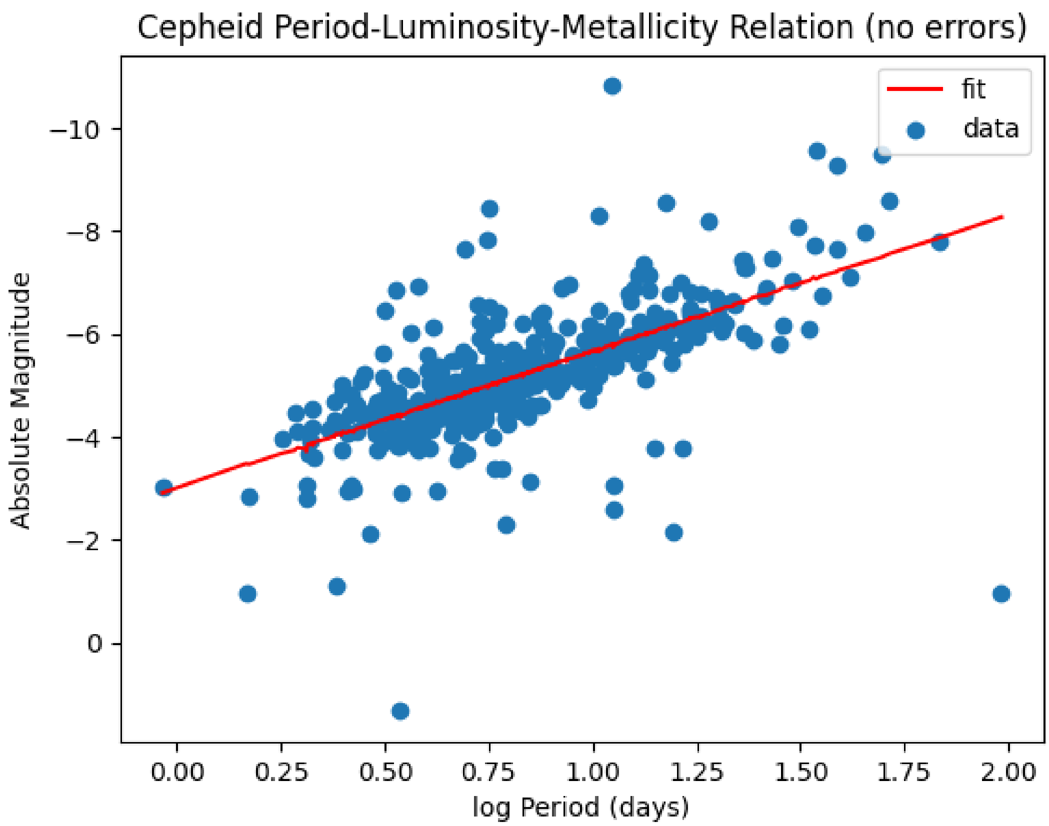

Figure 1: Figure showing the absolute magnitude as a function of the log of the cepheid pulsation period. The red line is made by using the parameters obtained from the linear least-squares model in conjunction with the data to form a line which best represents the average trend in the data.

3 Task 3

When errors on the data were taken into consideration, the best linear least-squares fit parameters were found to be $\alpha = \mathbf{-3.018}$, $\beta = \mathbf{-2.638}$, and $\gamma = \mathbf{-0.077}$ with respective errors $\sigma_\alpha = \mathbf{0.0143}$, $\sigma_\beta = \mathbf{0.0168}$, and $\sigma_\gamma = \mathbf{0.0252}$.

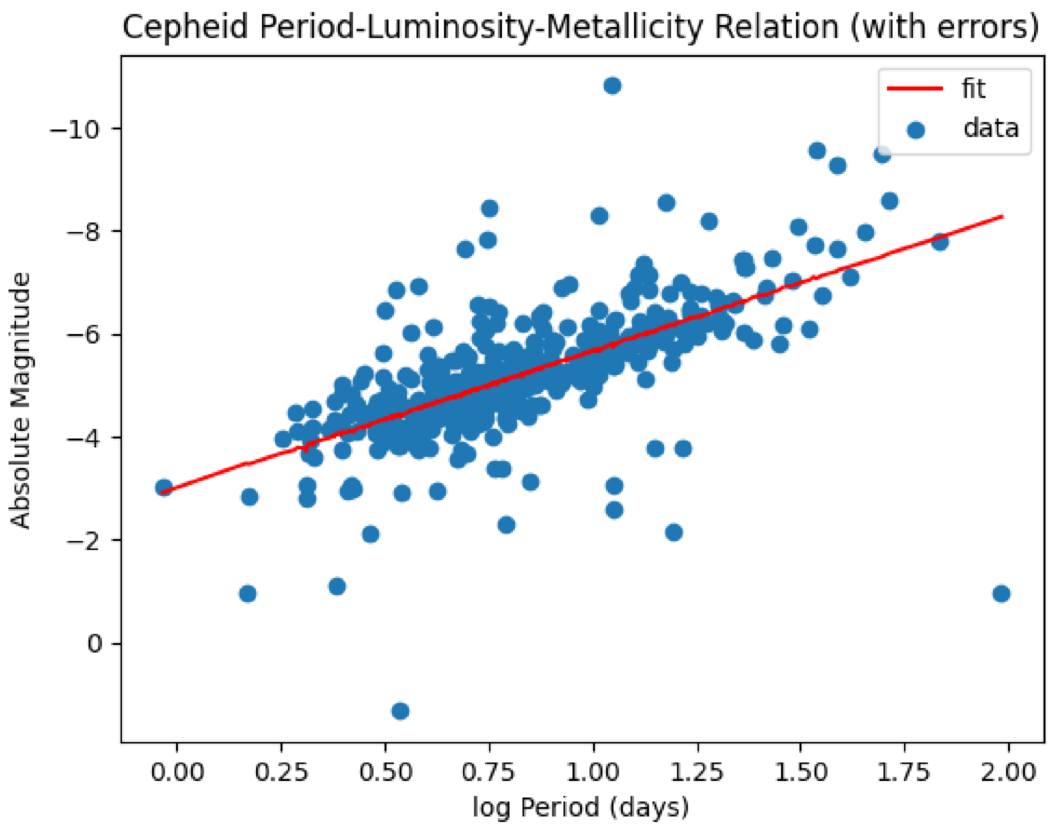

Figure 2: Same plot as Figure 1 with the same parameters but different errors on the parameters stemming from the inclusion of data uncertainty.

4 Bonus Task

The nested model, considering only

$$\theta = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

yielded best linear least-squares fit parameters of $\alpha = -\mathbf{3.007}$ and $\beta = -\mathbf{2.654}$ with respective errors $\sigma_\alpha = \mathbf{0.0139}$ and $\sigma_\beta = \mathbf{0.0159}$. The difference between the nested model plot in Figure 3 and Figures 1 and 2 which consider the metallicity parameter γ is visually negligible. To ensure no considerable difference between the full and nested models, an analysis using the F-test is presented, comparing the χ^2 statistics of the full and nested models along with the free parameters in each. The F-value is given by

$$F \equiv \frac{\left(\frac{\chi^2_{nested} - \chi^2_{full}}{\nu_{full} - \nu_{nested}}\right)}{\left(\frac{\chi^2_{full}}{\nu_{full}}\right)} \tag{1}$$

The F-value was found to be **0.00411** which, when supplied to **scipy.stats.f.cdf** alongside the degrees of freedom in the numerator and denominator of Equation 1 (1 and 449, respectively), returned a CDF value of **0.0511**. Based on the small visual difference between the fits and the paragraph at the bottom of page 2 in lecture notes 14 stating that "a large value of F ... is unlikely to draw by random chance from the F distribution," the conclusion is that the least squares model is not significantly improved through the inclusion of metallicity in the model. This can also be discerned through the CDF value of 5% which effectively states that the inclusion of the γ parameter only improves the model by 5% (this is probably a simplified interpretation of the CDF value).

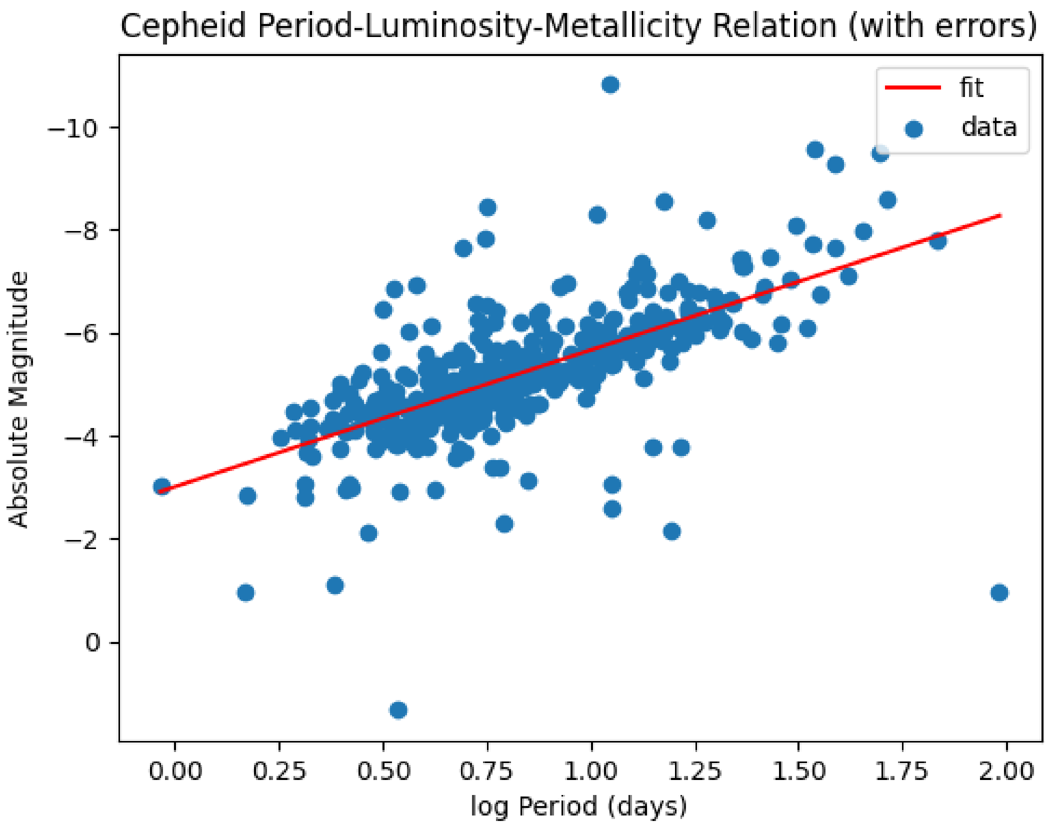

Figure 3: Same plot as Figures 1 and 2, but only considering the parameters α and β with no γ contribution. Clearly, the omission of the metallicity is negligible with regard to the fit's quality.