

Using A Linear Regression Algorithm In Python To Quickly Find Variable Stars For Ensemble Photometry

Caleb Derochea, Department of Physics, Photonics, and Optical Engineering (Mentor: Jamie Kern)
Bridgewater State University (BSU), Bridgewater, MA

Abstract

To measure the magnitudes of celestial objects, reference stars with stable magnitudes are used to account for fluctuations in Earth's atmosphere. If no such well-studied reference stars are within the field, other stars must be used so long as they are not variable. We created a Microsoft Excel spreadsheet which harbored a linear regression algorithm to detect star variability and then transferred it to Python to make the process of finding variable stars faster. Via Python packages such as sklearn, numpy, scipy, etc., we were able to produce a quicker and simpler way to analyze the star variability from their raw CSV file data and their accepted magnitudes from an online database. The code was successfully tested on variable stars TZ Aurigae and BX Pegasi.

Introduction

The magnitude scale is how astronomers catalog the “brightness” of objects in space and is based on a logarithmic scale as defined in the equation

$$(m_1 - m_2) = 2.5 \log \left(\frac{I_2}{I_1} \right), \quad \text{Eq. 1}$$

where m_1 and I_1 are the magnitude and intensity of one star and m_2 and I_2 are the magnitude and intensity of another star, respectively.

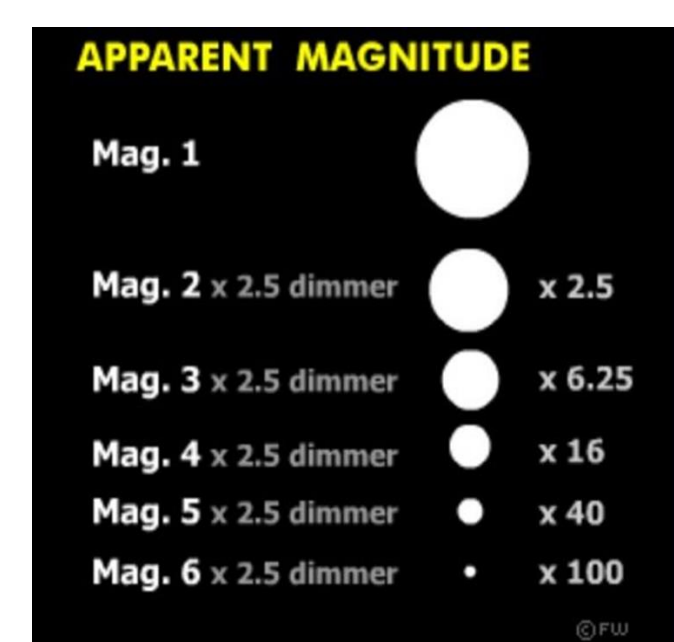


Fig. 1: In 1865, N. R. Pogson placed the six magnitudes that were visible to the naked eye on a logarithmic scale so that a magnitude 1 star was a hundred times brighter than a magnitude 6 star, as seen in the image on the right. The five separate intervals could be evenly be split up using the expression $\sqrt[5]{100} \approx 2.5$.

The images of candidate exoplanet EPIC 210401157-b from the BSU Observatory did not contain any known reference stars. We used ensemble photometry, but this meant a higher margin of uncertainty in our measurement. Using stars that are the least variable in their magnitudes would help reduce the margin of uncertainty. We originally did this via Microsoft Excel but recreated the algorithm in Python to expedite the process.

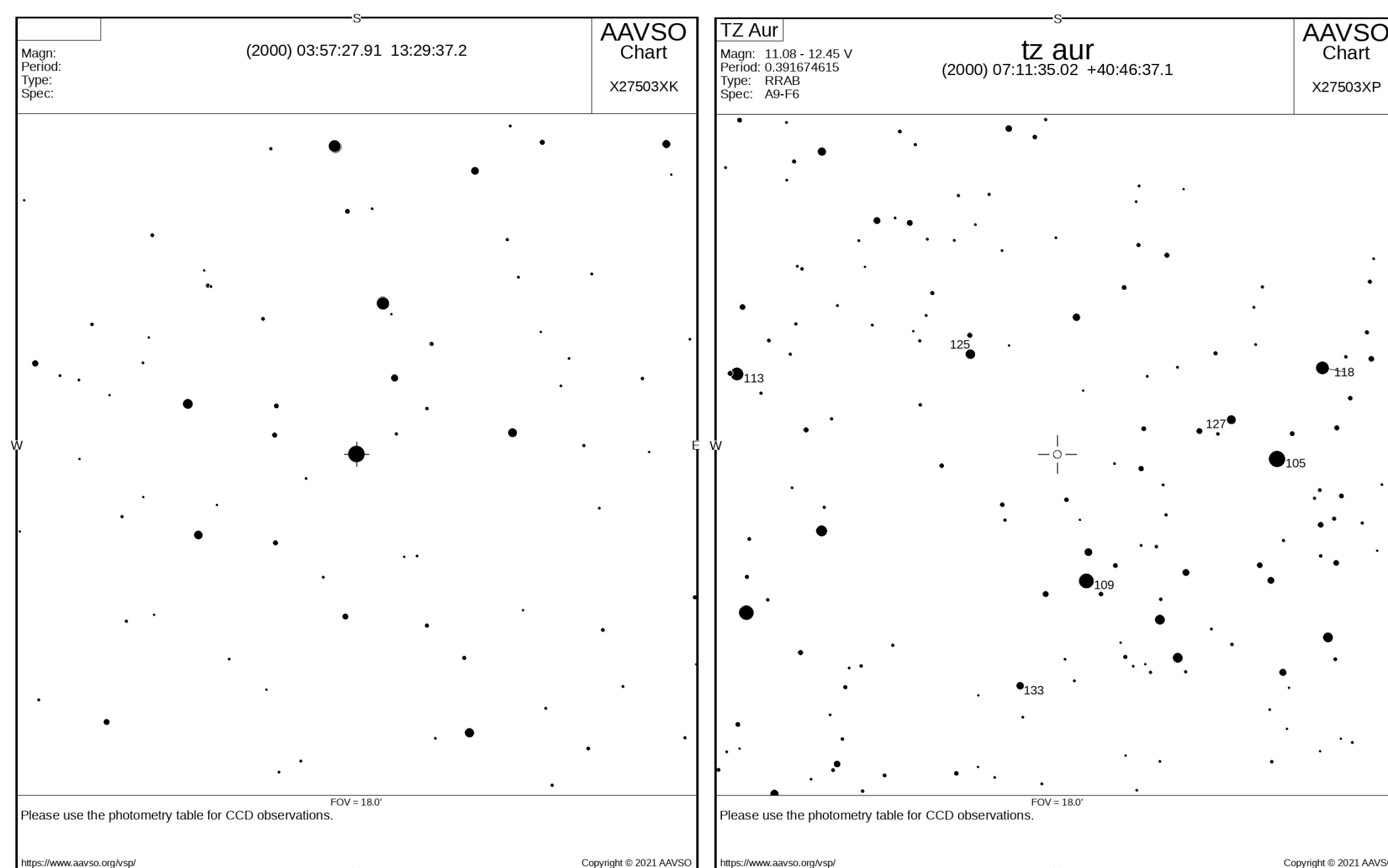


Fig. 2: A candidate exoplanet's field of view (left) and the field of view for the variable star TZ Aurigae (right). The numbers next to a star indicate the magnitude of the star and that said star has a consistent magnitude.

Differential photometry is a process of measuring the magnitude of objects in space (e.g., stars) against a star with a continuous magnitude. Ensemble photometry is similar but uses the averaged magnitudes of multiple stars. So, we used linear regression to find the stars least likely to be variable in a field of view to reduce the error in our measurements.

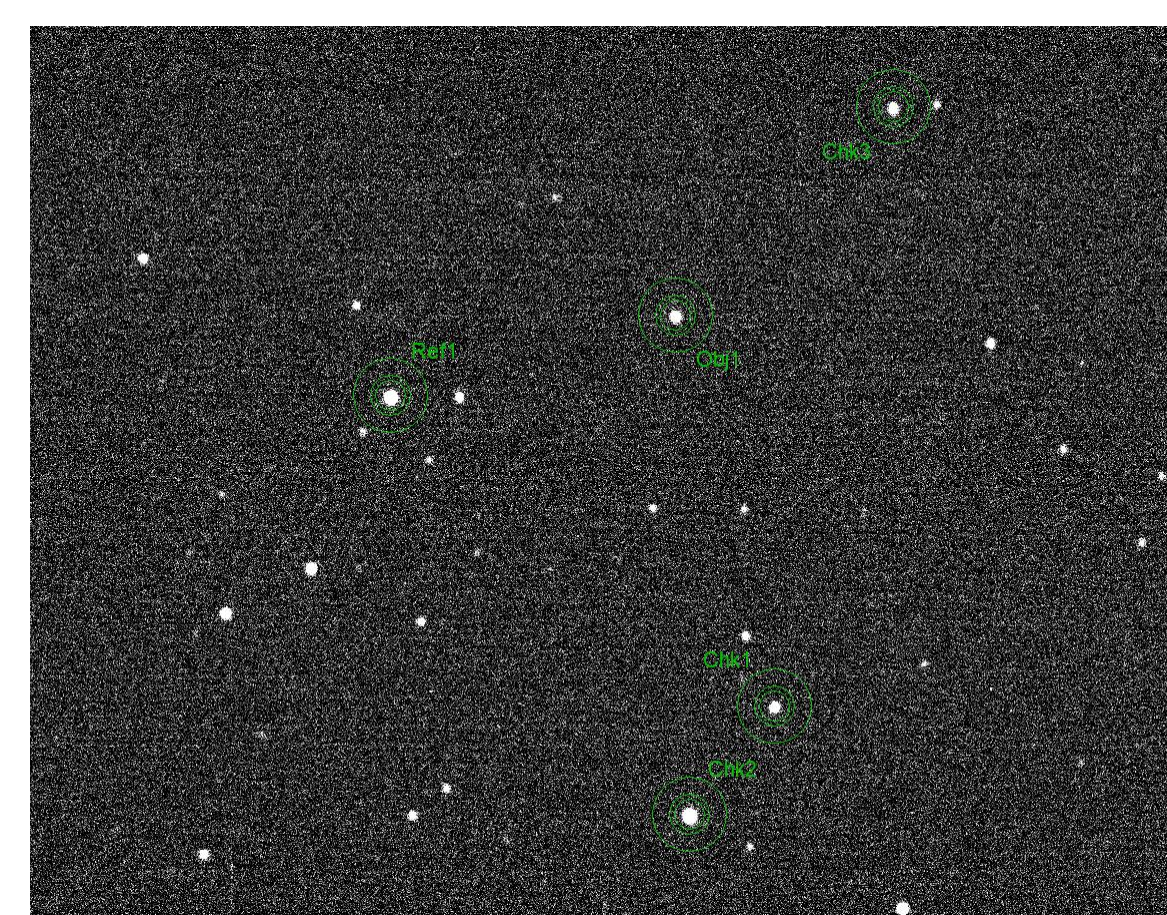


Fig. 3: Left - image demonstrating the photometry process. The main object of interest is selected along with a reference star as well as multiple check stars to check for variability in the reference star. In ensemble photometry, the check stars would be averaged with the reference star to form one mathematical reference star.

Method: Linear Regression

Linear regression is the process of taking multiple data points and finding the line of best fit. Although the magnitudes of each star detected by our camera in the BSU Observatory will not be the same as the magnitudes that have been cataloged, the relation between them is linear, so stars that do not vary in magnitude should almost perfectly match with the line of best fit. Equation 2 is the main basis for the algorithm we used to find such variability:

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad \text{Eq. 2}$$

where Y is the predicted y-values, β_0 is the y-intercept, β_1 is the slope coefficient, X is the independent variable, and ε is the random error of the calculation. Using this, it is possible to create a line of best fit and find any outliers in the data. If one were to take the mean magnitudes of each star within the field of view and test it against their expected magnitudes for the same filter, then, even if the measurements are offset from the actual magnitudes, the relationship between the two should be linear assuming that none of the stars are variable. This makes linear regression a reliable way to find out if a star's magnitude varies not just from its expected magnitude but also over the span of time the data is being collected.

Results Of Code Using TZ Aur Data

To test the Python code, we used data from our observations of well-known variable star TZ Aurigae. Since we had multiple nights' worth of data on it, combining the data from all the nights provided a perfect stress test for the code. The code can take the average of over 900 data points of 16 different stars within about 2 to 5 seconds, which would normally take an hour or more for someone to set up in Microsoft Excel. This substantially decreases the amount of time and energy one needs to spend to find out the variability of stars within a field of view.

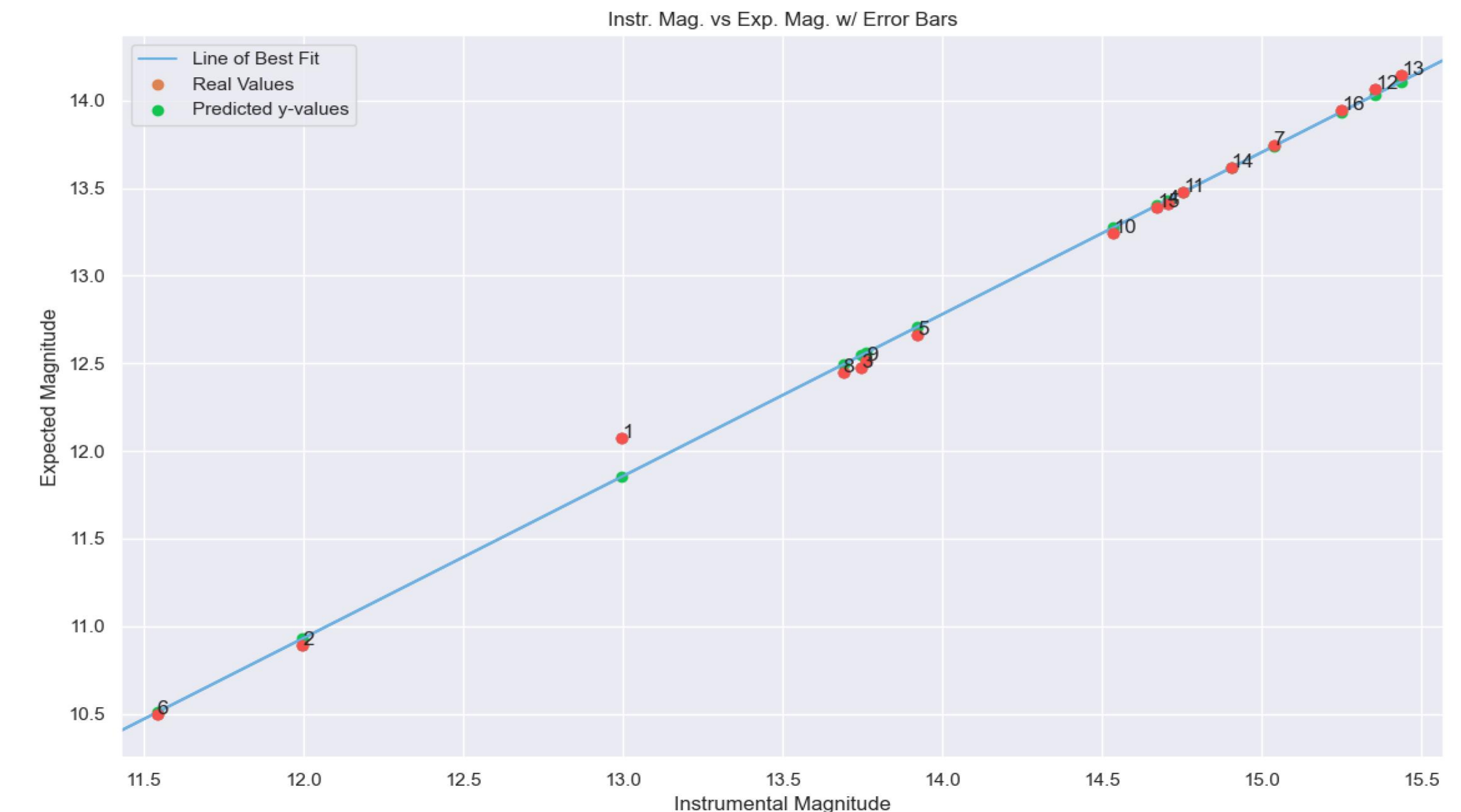


Fig. 4: Above—result of code analyzing the data from the 900 TZ Aur field data points. This plots the instrumental magnitude from the observation against the expected magnitude. Point 1—which varies the most from its expected magnitude—is TZ Aurigae. The error bars are smaller than the data points.

Future Work

This code will be posted on GitHub—along with a CSV file of the TZ Aurigae data and a ReadMe file with instructions—in order to allow amateur astronomers to have free access and help them easily identify possible variable stars. In terms of the future of the code here at BSU, it will be used for future projects involving ensemble photometry, such as variable stars, exoplanets, asteroids, comets, quasars, etc. With this code, it'll be easier to exclude variable stars from ensembles and discover them for future research. We may even be able to contribute to finding exoplanet candidates. This code has the potential to help both students here at BSU and amateur astronomers around the world with astronomical research.

Acknowledgments

I would like to thank Jamie Kern for being my mentor for this project. Her guidance has helped me develop the research and astronomy skills that I possess today.

I would like to thank the BSU Experimental Astrophysics Research (BEAR) Team for the images of TZ Aurigae. They have been a major help in developing and testing my code.

References

1. “Variable Star Plotter.” AAVSO, <https://app.aavso.org/vsp/chart/?star=TZ%2BAur&orientation=visual&type=chart&fov=18.0&maglimit=18.0&resolution=150&north=down&east=right>.
2. “Variable Star Plotter.” AAVSO, <https://app.aavso.org/vsp/chart/?ra=03%3A57%3A27.91&dec=13%3A29%3A37.2&orientation=visual&type=chart&fov=18.0&maglimit=18.0&resolution=150&north=down&east=right>.
3. Berry, Richard, and James Burnell. *The Handbook of Astronomical Image Processing*. Willmann-Bell, 2011.
4. Paxson, K. B., “Differential Ensemble Photometry by Linear Regression”, Journal of the American Association of Variable Star Observers (JAAVSO), vol. 38, no. 2, p. 202, 2010.