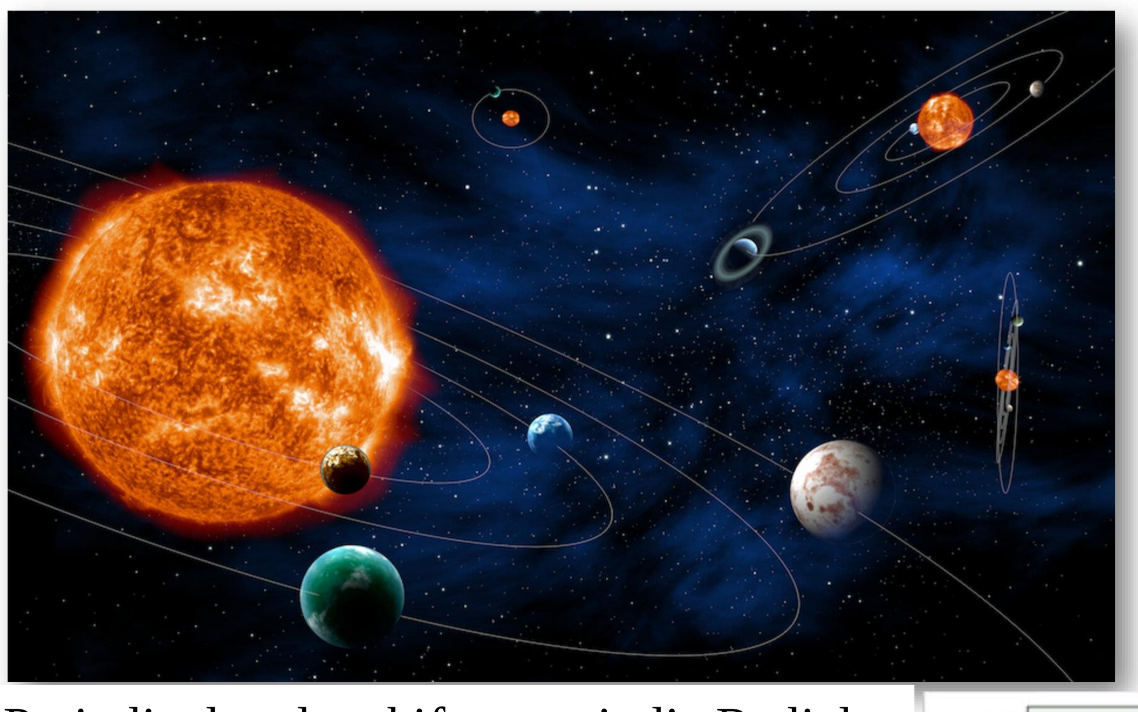
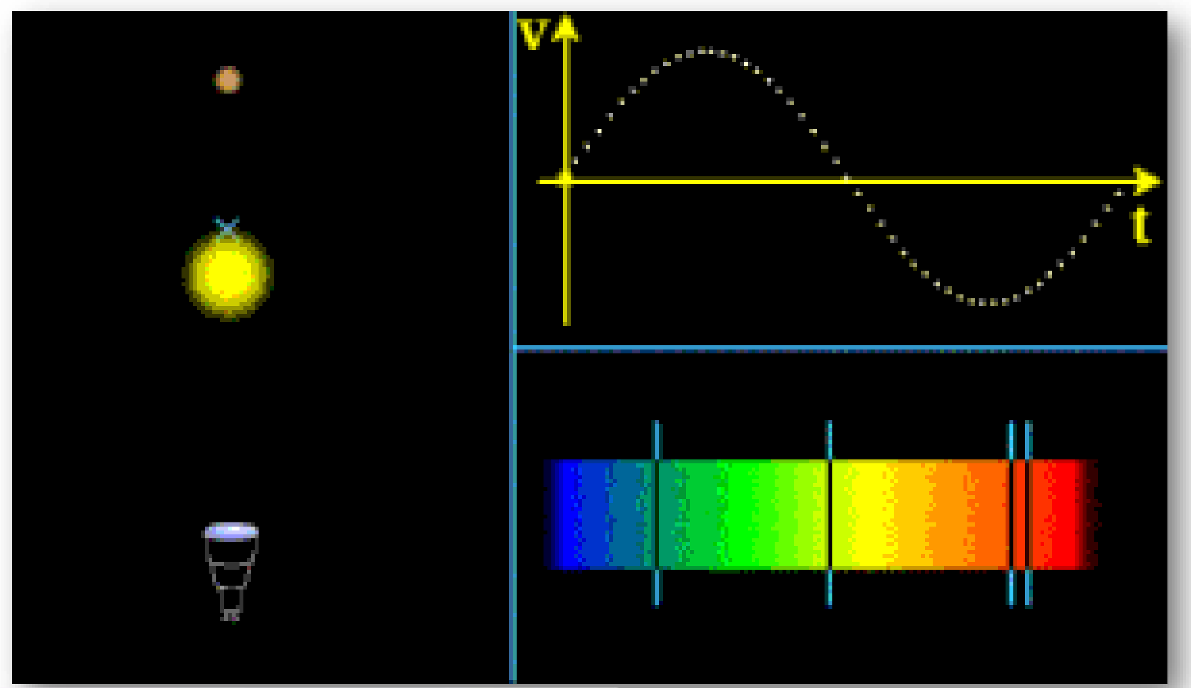
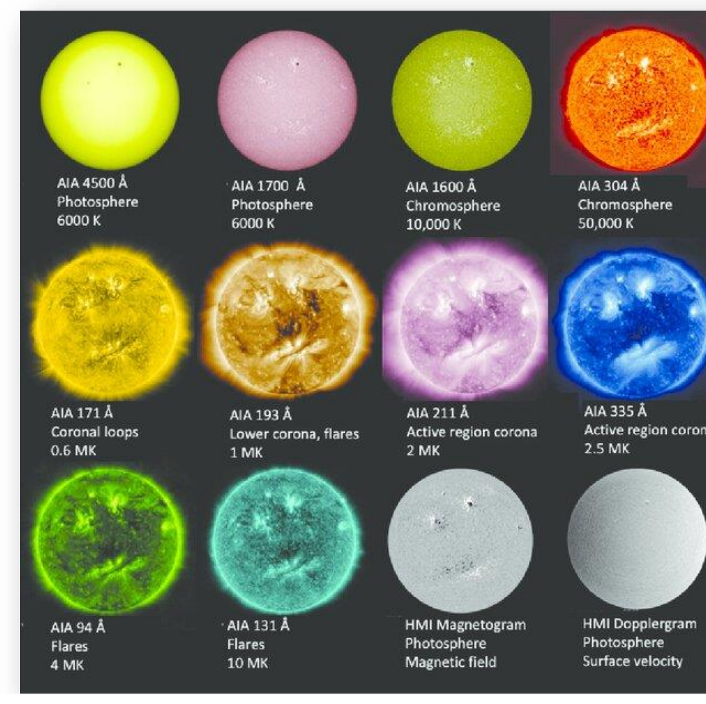
Machine Scientist for Research

Example: Understanding and Mitigating Stellar Activity Problem in Exoplanet Detection

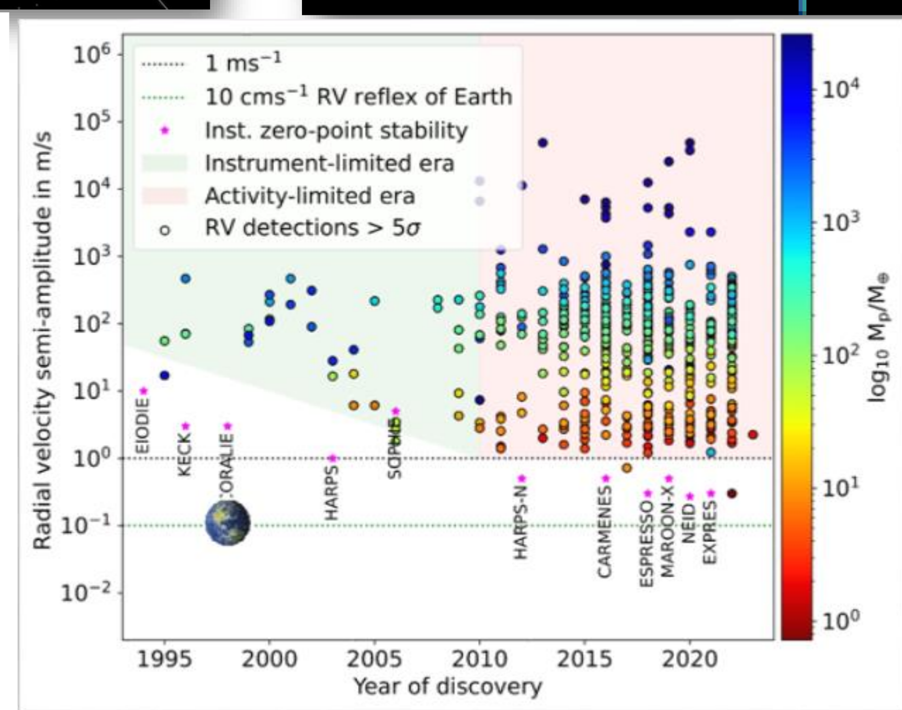
Mangesh Daspute (Ariel University); Supervisor- Lev Tal-Or



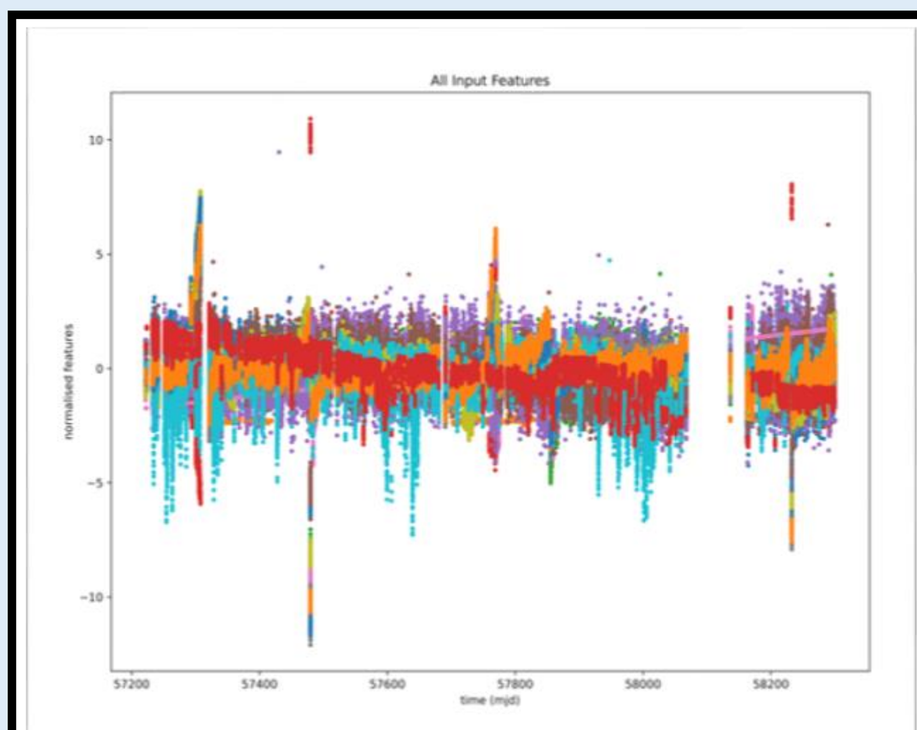


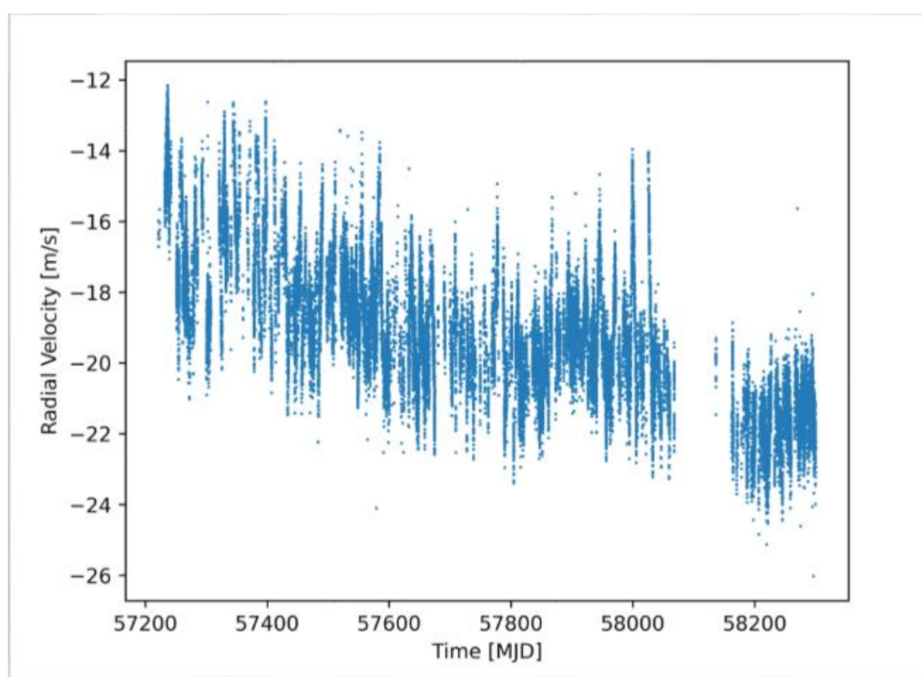


- Periodic doppler shift or periodic Radial Velocity (RV) change of star indicates there is an exoplanet.
- **Problem:** Stellar activity produces noise in the radial velocity data.
- It hinders the detection of earth mass exoplanets because the exoplanet signal is lost in the noise.
- Existing methods are inadequate in determining the noise.



- **Solution:** We use a data-driven approach to determine the effect of stellar activity on radial velocity. We took 3 years of HARPS-N Solar spectrograph data containing 34 thousand examples. We used 24 known stellar activity indicator absorption lines (eg. Calcium, Hydrogen, Iron etc.) from each spectrum.
- **Input Data:** Normalized fluxes of these 24 activity indicators.
- **Output Data:** Radial velocity data of the sun in its rest frame.
- **Neural Network:** *Bayesian Machine Scientist* with 5 parameters and one output.





- **Future Plans:** Use all 5000+ absorption lines and their properties like FWHM, depth, BIS etc. from spectra as input.
- Use PCA before training to identify the most important, related lines and properties
- Use data from the NEID and ESPRESSO solar spectrograph which has a greater number of examples and better precision.
- Validate findings using injection retrieval simulations of exoplanets.

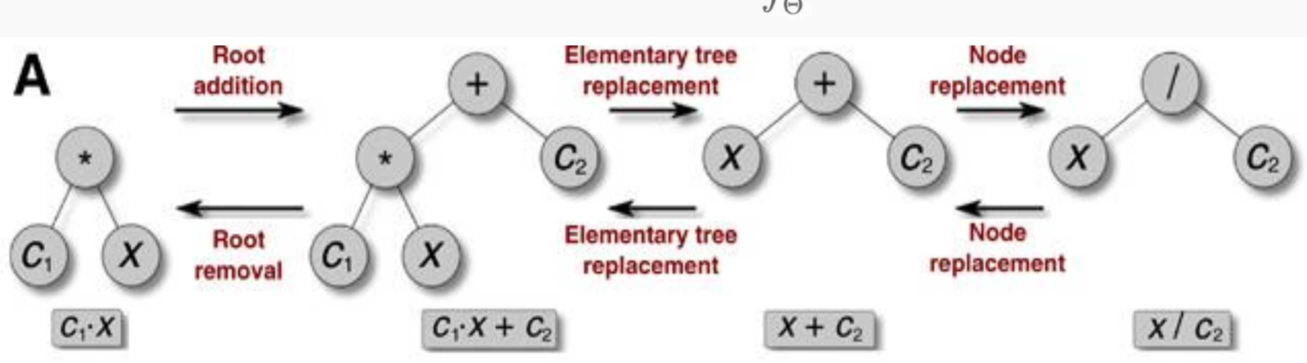
Bayesian Machine Scientist Algorithm[1]

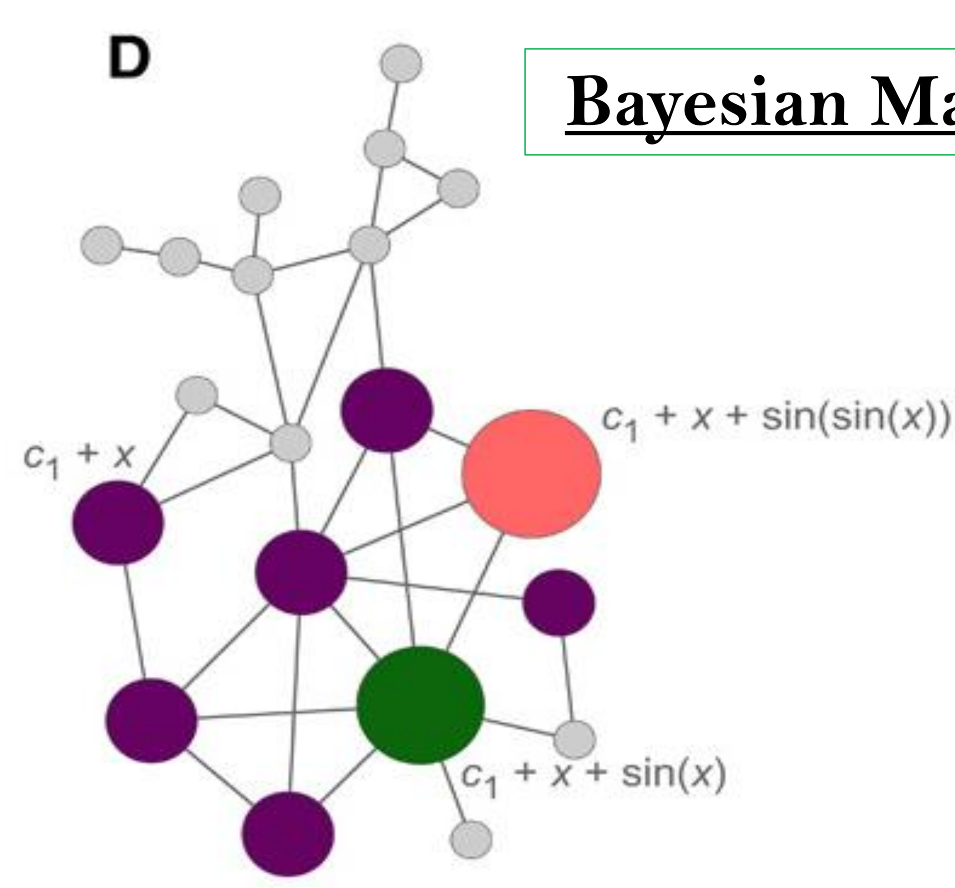
The posterior can always be written as:

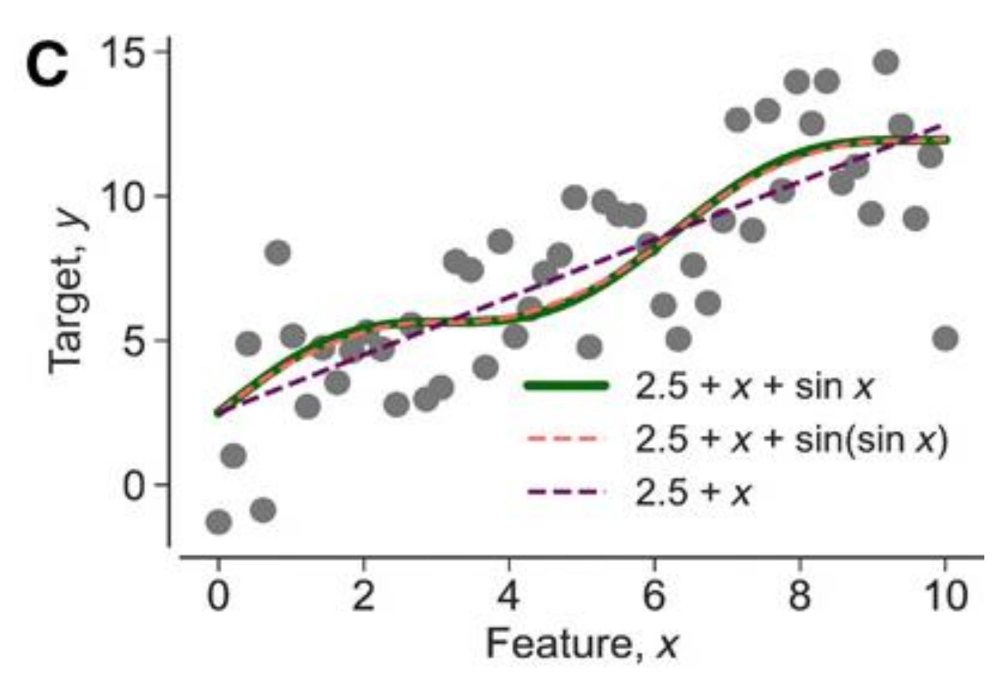
$$p(f|D) = \frac{1}{p(D)} \int_{\Theta} d\theta\, p(D|f,\theta)\, p(\theta|f)\, p(f) = \frac{e^{-\mathcal{L}(f,D)}}{p(D)}$$

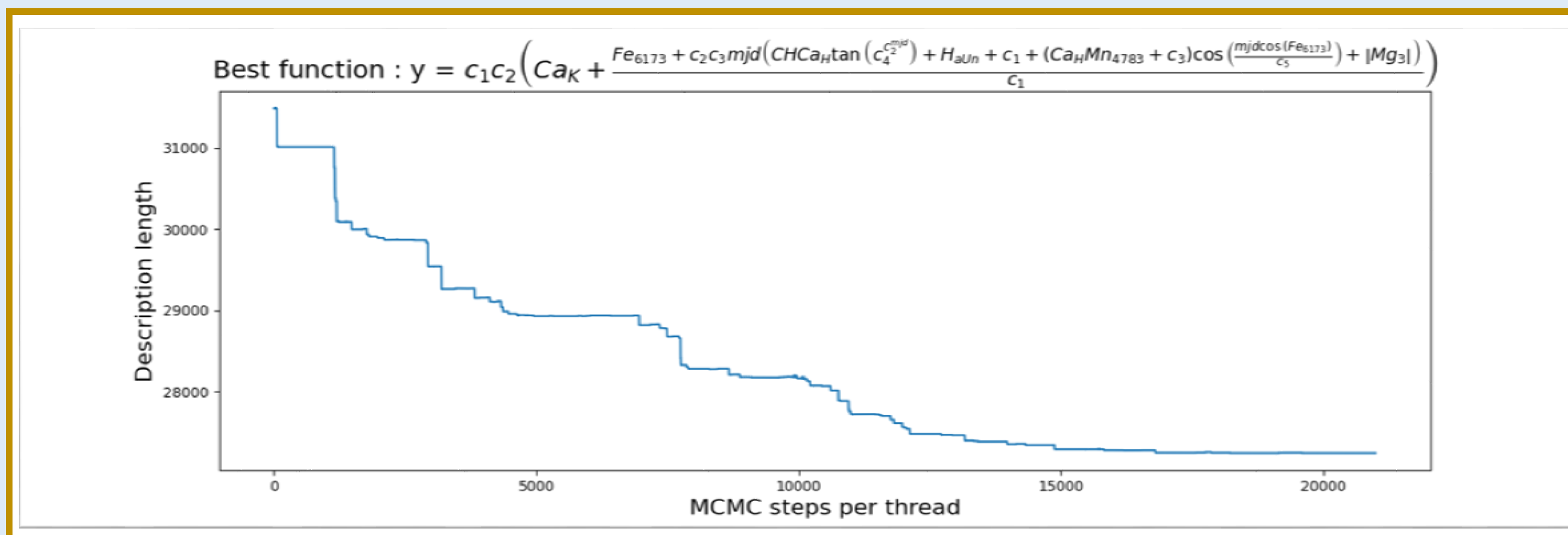
with the ***description length***:

$$\mathcal{L}(f,D) = -\log p(f,D) = -\log \int_{\Theta} d\theta\, p(D|f,\theta)\, p(\theta|f) - \log p(f)$$

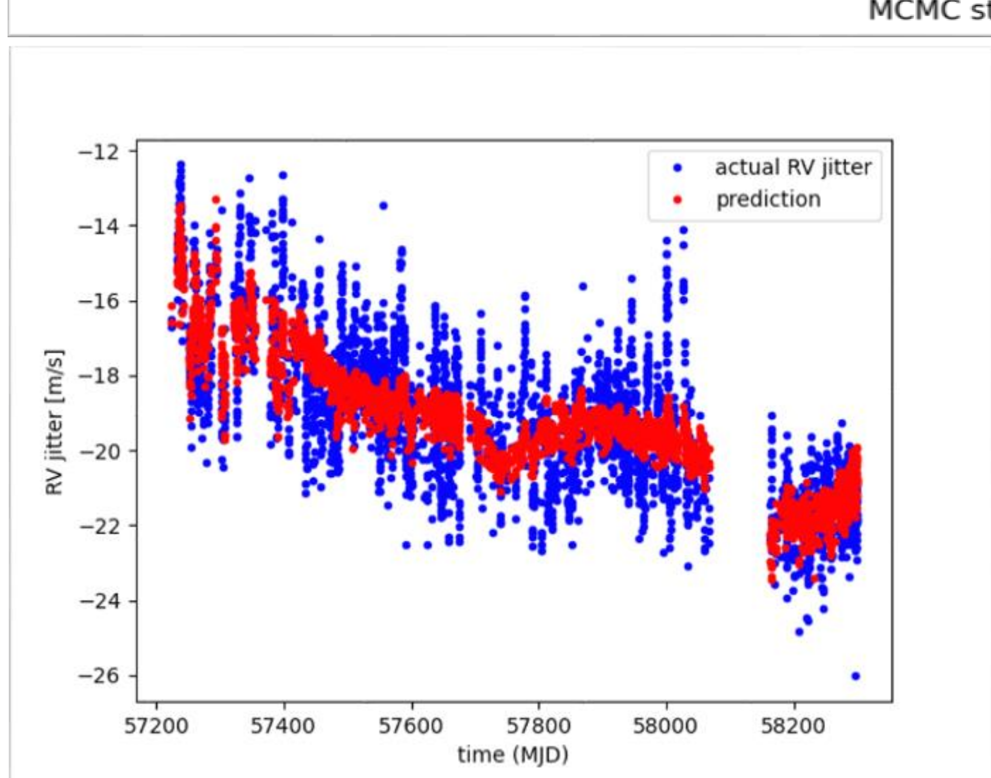


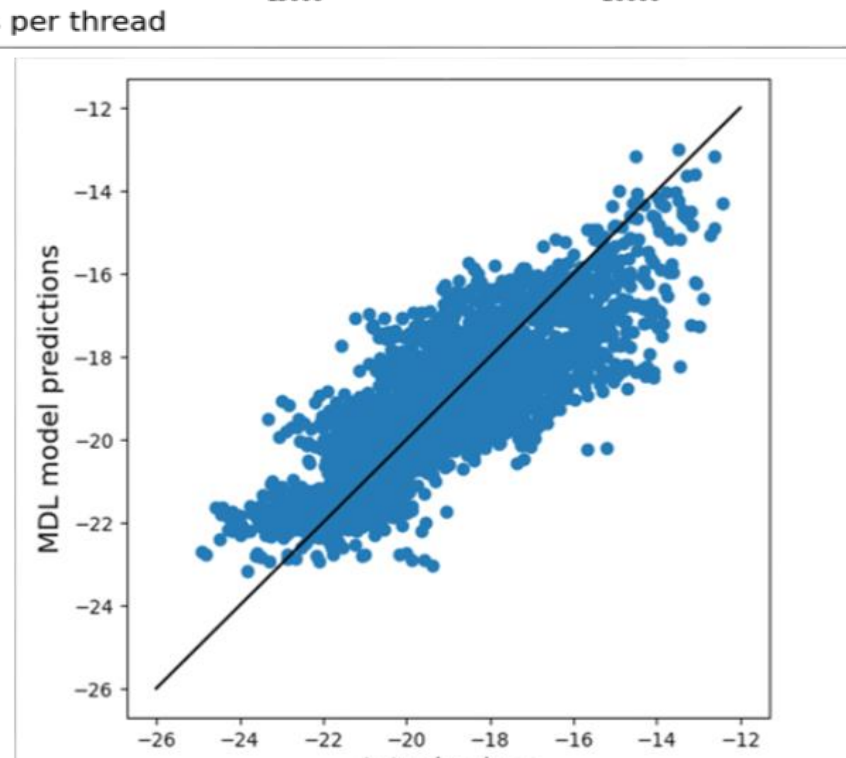






- **Results:** We were able to determine the effect of stellar activity on radial velocity.
- Accuracy = 41%
- Reducing the prediction from target output data reduced the scatter from 2.04 to 1.22 .
- 1.6 m/s effect is modelled out from the noise, Which is significant compared to the measurement error of about 0.3 m/s.





Scan QR code to try Bayesian Machine Scientist

References: [1] Guimerà et al. (2020). A Bayesian machine scientist to aid in the solution of challenging scientific problems. Science Advances, 6(16), eaav6971. https://doi.org/10.1126/sciadv.aav6971