National Aeronautics and Space Administration

Measuring Extreme Precision Radial Velocities (EPRVs) Using Deep Learning

Principal Investigator: Virisha Timmaraju (174); Co-Investigators: Ian Colwell (174), Samuel Halverson (383), Alexander Wise (Pennsylvania State University), Alexander Wise (Pennsylvania State University)

Program: FY22 R&TD Topics
Strategic Focus Area: Extra-solar planets and star and planetary formation

Objectives

1. Characterizing stellar noise components in terms of their pixel-by-pixel effects on the spectrum, using DL.

2. Quantifying the contribution of each stellar RV noise component to the RV error in each spectrum down to, or below, instrumental noise levels, using DL.

3. Determining the data requirements of neural networks in terms of constraints on SNR, resolution, cadence, and number of spectra to effectively train a neural network to characterize and/or quantify each component of stellar RV jitter.

Background

- State-of-the art methods use cross-correlation function (CCF) or search for activity-sensitive lines [1],[2],[3],[4].
- CCF studies [1],[2] cannot account for differences in responses of individual to stellar activity.
- Activity-sensitive line searches focus on line depth changes [3],[4], rather than asymmetric line-shape changes which are more highly correlated with stellar RV jitter [2].
- Our method aims to globally characterize all such changes in the spectrum, by use of a large quantity of high quality input data (34450 HARPS-N spectra over 3 years) and by harnessing the power of DL methods to probe the effects of stellar activity on the spectrum at unprecedented detail.

Approach and Results

Datasets and Preprocessing (Figure-1-B.1): Training data is 3 years of HARPS-N sun-as-a-star spectra (34450 spectra) from 2015 to 2018. The RV corrections are provided by the HARPS-N team. Alpha-shape Fitting to Spectrum (AFS) algorithm [2] implementation in the RvSpectML package is used for continuum normalizations. Interpolation uses a sinc kernel, preventing the introduction of noise due to intra-pixel sensitivity.

Ancillary Datasets (Figure-1-B.1)**: (1)** Helioseismic and Magnetic Imager (HMI) onboard the Solar Dynamics Observatory (SDO) provides near single-granule spatial resolution photometric maps of the solar surface. These data were reduced using the SolAster Package to quantify solar activity conditions. **(2)** Observing conditions for each observation were provided by the HARPS-N team. In tests of our neural network, we use data such as seeing, humidity, and air column density to isolate telluric noise component(s) and validate network performance.

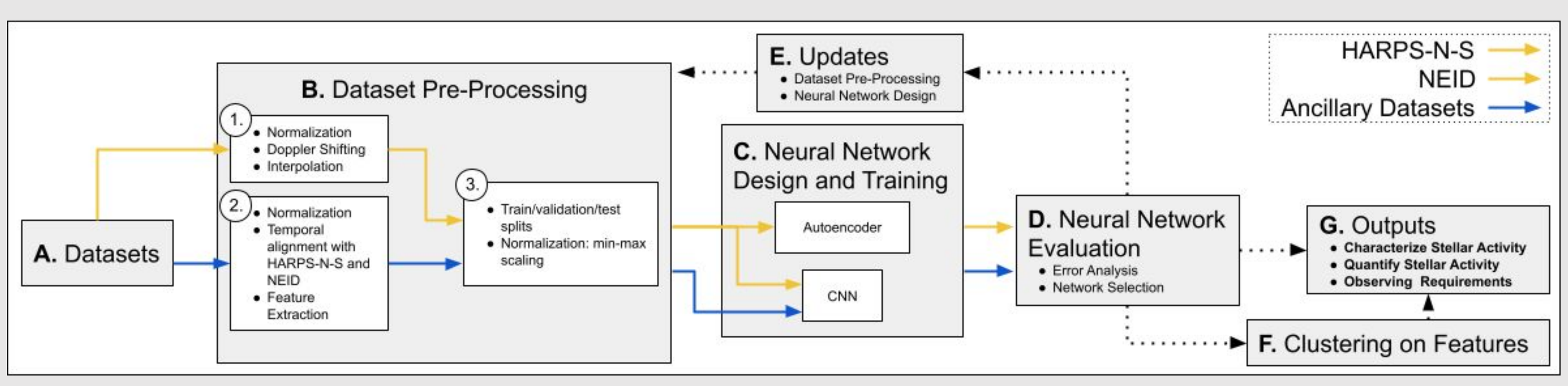

Figure-1: Technical Approach Flow Diagram

Clustering (Figure-1-F): HARPS-N spectral lines are empirically generated using the public RvLineList package. In order to measure the change in each line over time and identify lines with similar changes, a feature extraction step followed by clustering is performed. The feature extraction step involves, for each line, deriving the correlations between the pixels associated with adjacent time steps. Then, a few statistics like mean, standard deviation, skew, kurtosis are all derived as features for each line to quantify the change in each line over time. In order to characterize and group together similar manifestations of stellar noise, DBSCAN clustering algorithm was used. The clustering results are validated by correlating the clustered lines with stellar activity signals from the ancillary data. The validation results are depicted in figures 2 and 3 and indicate that the features affect the quality of clustering. So, using this as a baseline, we wish to explore physically derived features in the coming year.

Deep Learning (Figure-1-C): HARPS-N spectra are subsetted to extract lines associated with a spectral line mask (G2.Espresso). Individual CNNs are trained using a single spectral line as input and ancillary value as the target. This approach results in ~5k x N models created, where we have ~5k spectral lines and N ancillary features. This was demonstrated by targeting injected RVs, an optional output of the preprocessing pipeline. This helps understand and highlight lines that are sensitive and insensitive to stellar variability; lines sensitive to stellar variability will have a high RMSE/MAE while lines insensitive to stellar variability have a low(er) RMSE/MAE. Ancillary variables representative of stellar activity are swapped out as the target variable to understand what lines are more or less sensitive to a given type of stellar activity (below)

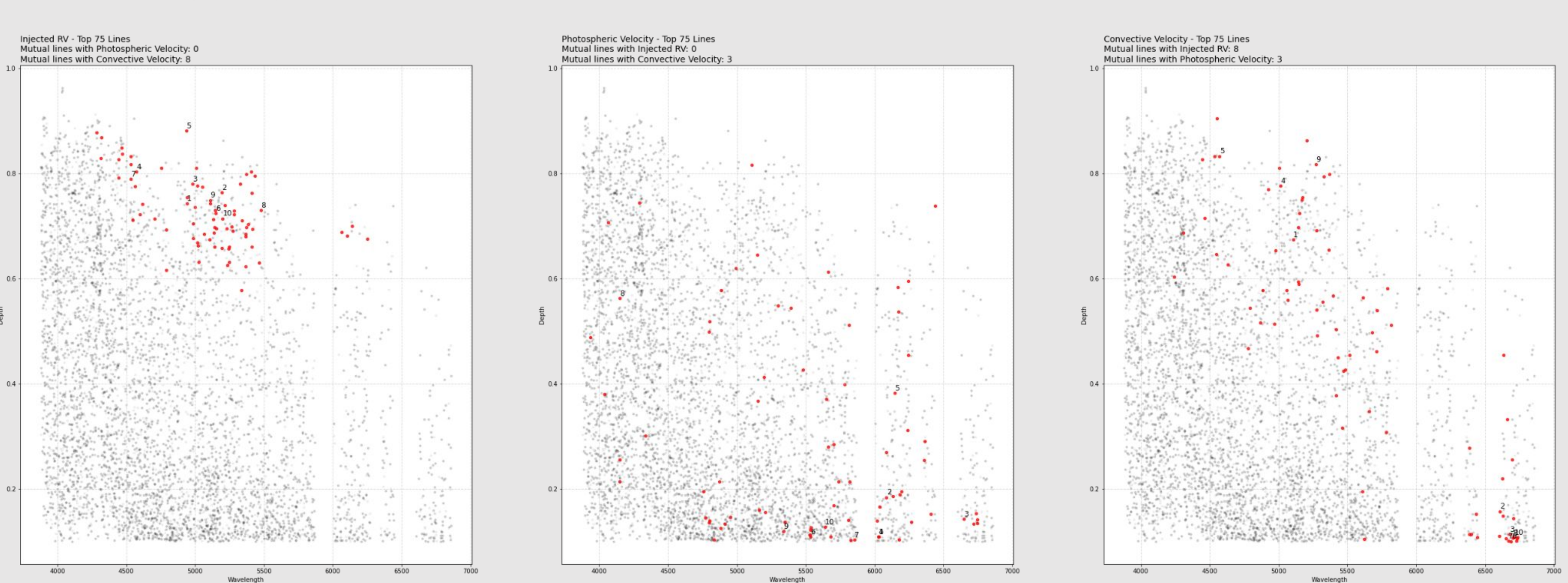

Figure-4. The three charts correspond to different targets the CNNs were trained on - an injected planetary RV, photospheric velocity, convective velocity (left to right). The red highlighted points correspond to the top-75 best performing lines relative to their target values (columns) and show distinct patterns across the three different sets.

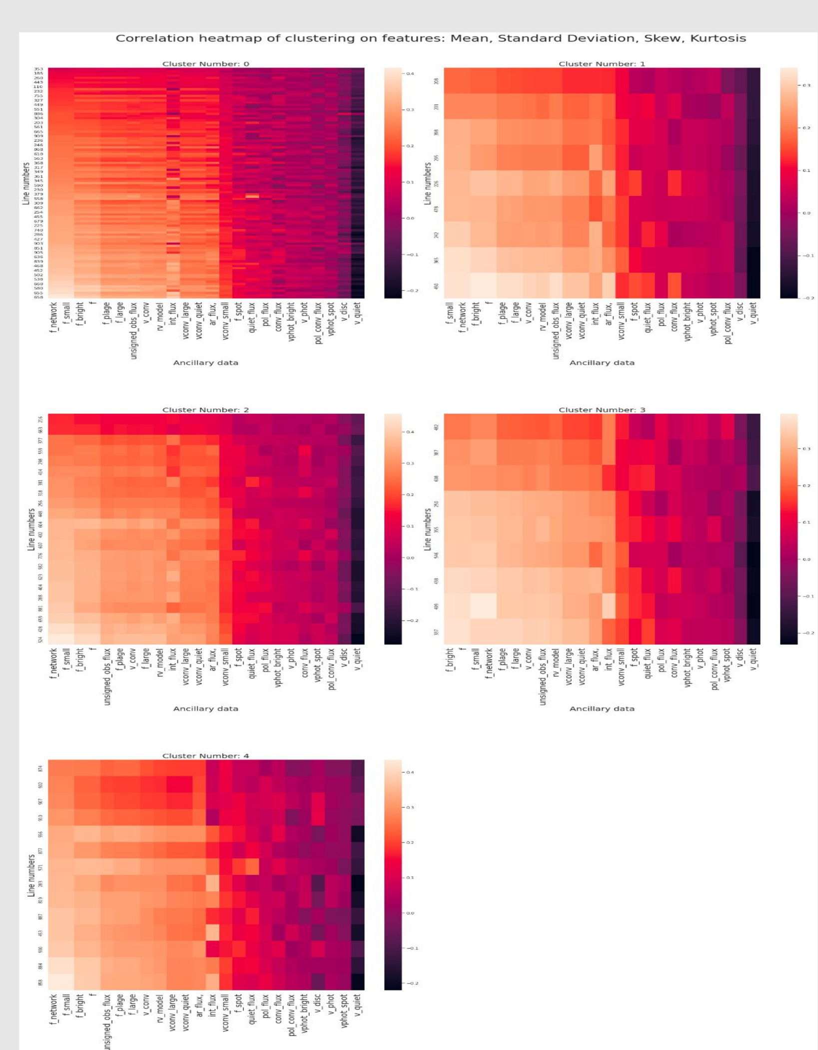

Figure-2: Various stellar activity components (from the ancillary dataset) are correlated with the spectral lines in each cluster. The clusters in this figure are derived by applying dbscan on four statistical features. The correlations of the clustered lines with the stellar activity components seem weak.

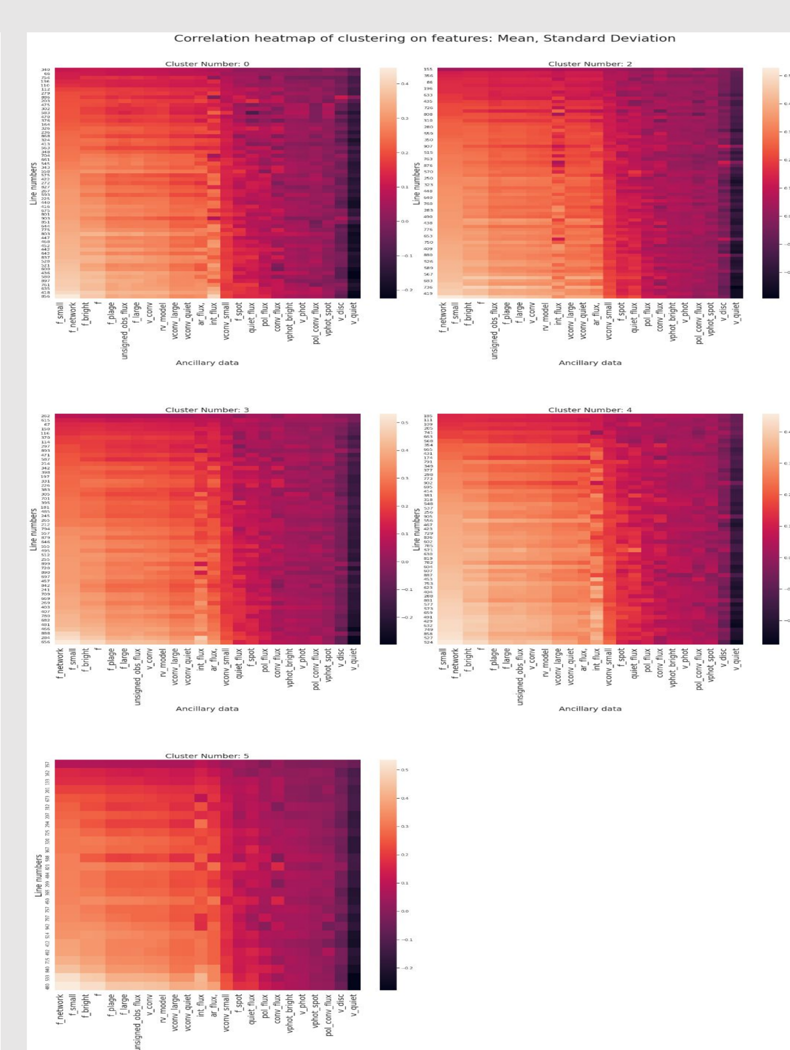

Figure-3: The clusters in this figure are derived by applying dbscan on two statistical features. The correlations of the clustered lines with the stellar activity components seem relatively stronger in a few clusters.

References

[1] Xu, Xin et al. 2019, The Astronomical Journal 157.6: 243. doi:10.3847/1538-3881/ab1b47
[2] de Beurs, Z. L., Vanderburg, A., Shallue, C. J., et al. 2020, arXiv:2011.00003
[3] Lanza, A. F., Malavolta, L., Benatti, S., et al. 2018, Astronomy & Astrophysics, 616, A155. doi:10.1051/0004-6361/201731010
[4] Wise, A., Dodson-Robinson, S. E., Bevenour, K., et al. 2018, The Astronomical Journal, 156, 180. doi:10.3847/1538-3881/aadd94
[5] Ning, B., Wise, A., Cisewski-Kehe, J., et al. 2019, The Astronomical Journal, 158, 210. doi:10.3847/1538-3881/ab441c

PI/Task Mgr. Contact Information:

Email: virisha.timmaraju@jpl.nasa.gov

Significance/Benefits to JPL and NASA

- NASA is beginning to invest resources in ground-based radial velocity (RV) surveys to support its space-based search for habitable exoplanets (NASA ExEP Science Gap List).
- Stellar activity is considered to be the largest source of noise in EPRV instrument teams' RV error budgets.
- Improvements the characterization and quantification of stellar RV jitter in EPRVs will enhance stellar activity mitigation algorithms, boosting the efficiency of upcoming missions like JPL's HabEx, which directly image habitable exoplanets, by ~50% (R. Morgan, EPRV working group report), improving our chance of detecting biosignatures.
- Establishing the data requirements for neural network approaches to disentangle stellar noise sources will help inform future EPRV survey designs on the best way to allocate limited telescope time, and help NASA decide the amount of observing time to be purchased on telescopes in order to meet their EPRV goals.

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

www.nasa.gov

Clearance Number: CL#
Poster Number: RPC#024