National Aeronautics and Space Administration

FY24 Topic Areas Research and Technology Development (TRTD)

Characaterizing and Quantifying Stellar Noise Sources

Principal Investigator: Virisha Timmaraju (174)**; Co-Investigators:** Hamsa Shwetha Venkataram (174), Samuel Halverson (383)

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.

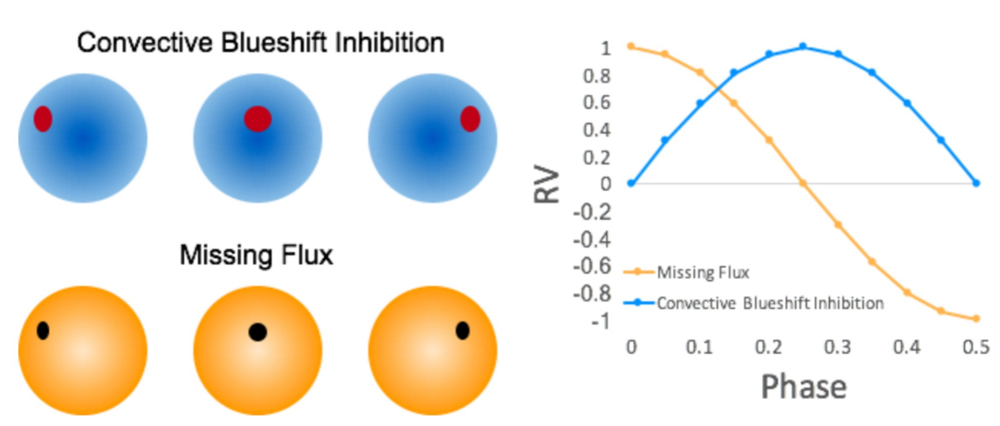

Figure 1. Left: Illustration of two spatially correlated sources of stellar RV jitter: convective blueshift inhibition from strong magnetic fields and rotational RV imbalance caused by missing flux in star spots. **Right:** Plot of the RV signals versus rotational phase corresponding to convective blueshift inhibition and missing flux.

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.

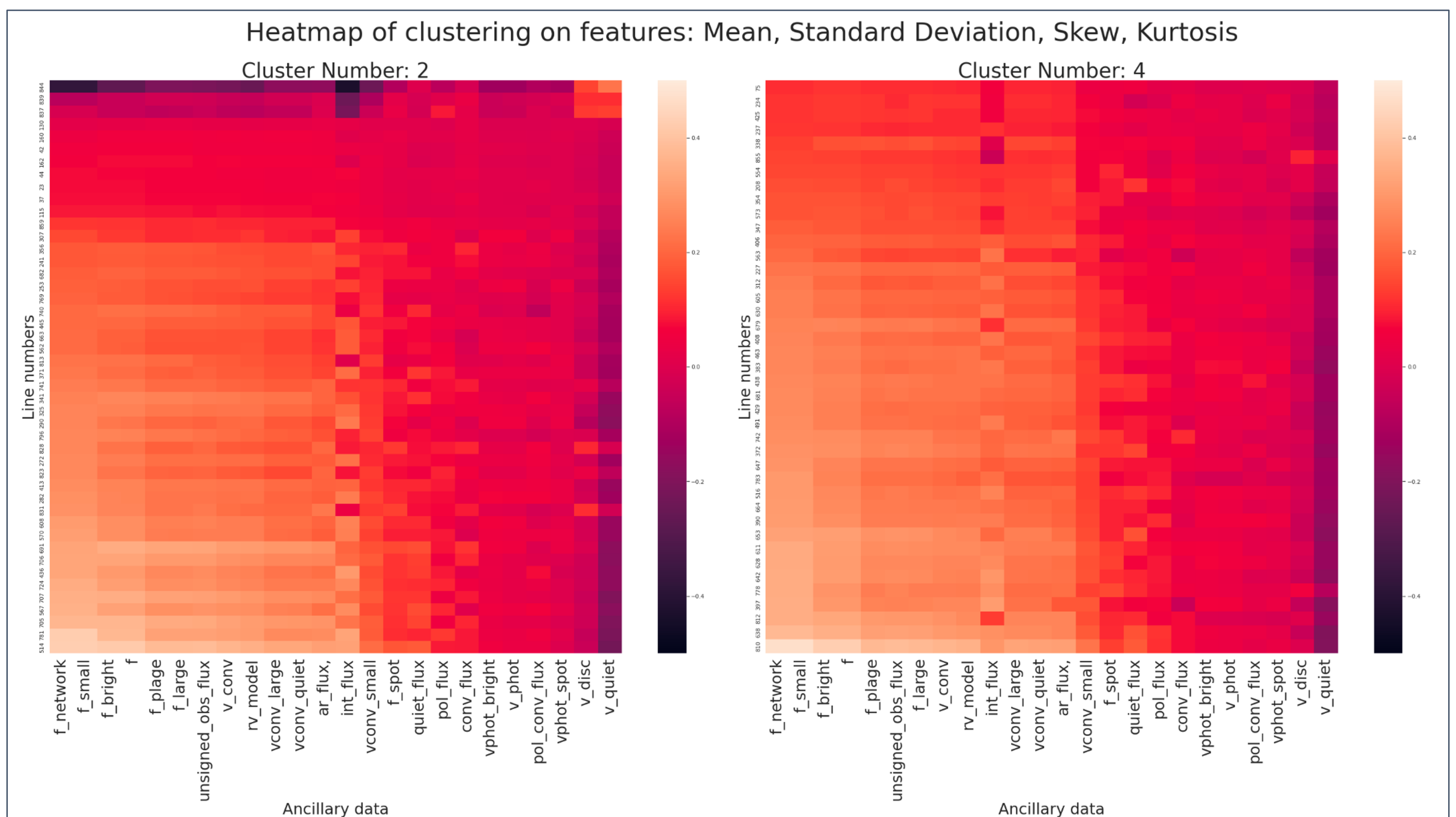

Figure 3: Various stellar activity components (from the ancillary dataset) are correlated with the spectral lines in each cluster. The clusters in this figure are derived by applying dbscan on four statistical features. The correlations of the clustered lines with the stellar activity signals seem relatively strong in some clusters.

Objectives

Datasets and Preprocessing (Figure 2-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 2-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 the "sun-as-a-start" RV model, the convective and photometric velocity components, and more as targets for the network during training.

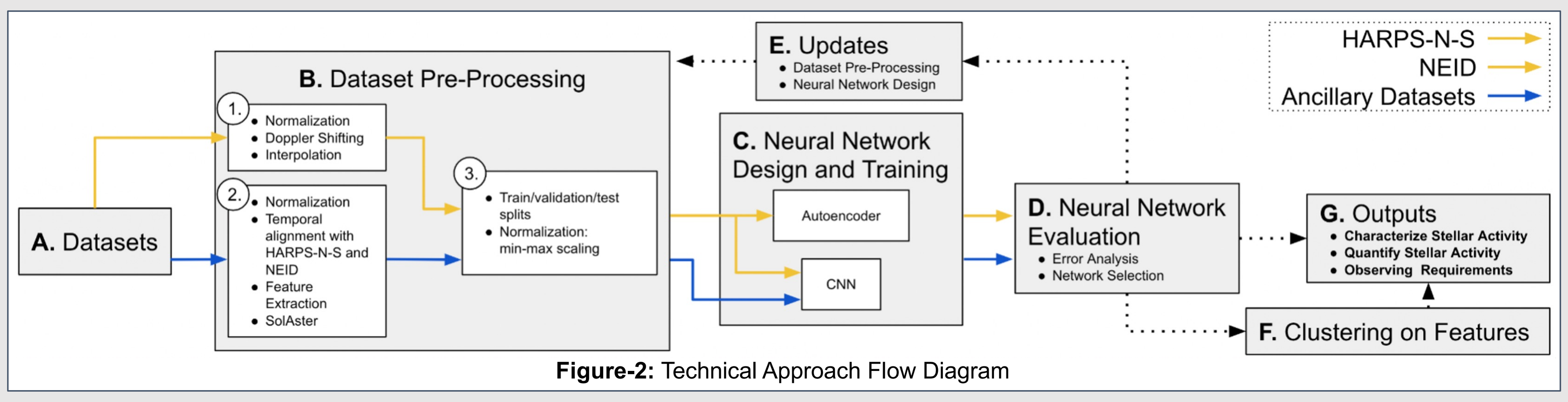

Figure-2: Technical Approach Flow Diagram

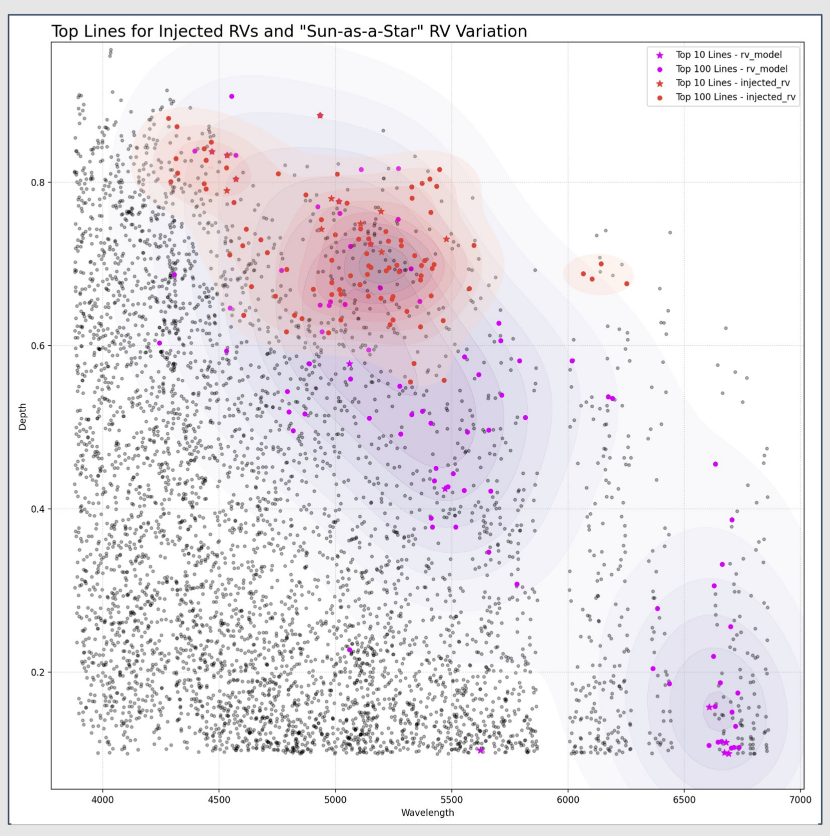

Figure 4: Each spectral line from HARPS-N spectra was input to a CNN targeting an injected planetary RV (red) and separately the "sun-as-a-star" RV variation (purple). The top 100 top performing lines for each trial are highlighted.

Deep Learning (Figure 1-C): Individual spectral lines are subsetted from HARPS-N using a line mask (G2.Espresso). CNNs are trained using a single spectral line as input (15 x ~35K) and the time aligned ancillary (photometric velocity, convective, etc.) value as the target. This approach results in ~5k x N trained CNNs (~5k spectral lines; N=23 SolAster features). We first demonstrated this line-by-line CNN concept by targeting injected RVs, an output of the preprocessing pipeline; this highlights lines that are insensitive to stellar variability (low RMSE) and lines that are sensitive to stellar variability (high RMSE). The CNNs' targets are changed to "sun-as-a-star" RV components, providing a more direct way to understand what lines are more, or less, sensitive to a given type of stellar activity (Figure 4). Larger CNNs are trained on different subsets of multi-line inputs to help understand how the removal of lines sensitive to stellar activity affect our ability to identify planetary RVs (Figure 5).

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 mean, standard deviation, skew and kurtosis for the pixels associated with each line. Then, the mean and standard deviation are derived across time, 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, depicted in Figure 3, indicate that the features affect the quality of clustering. So,

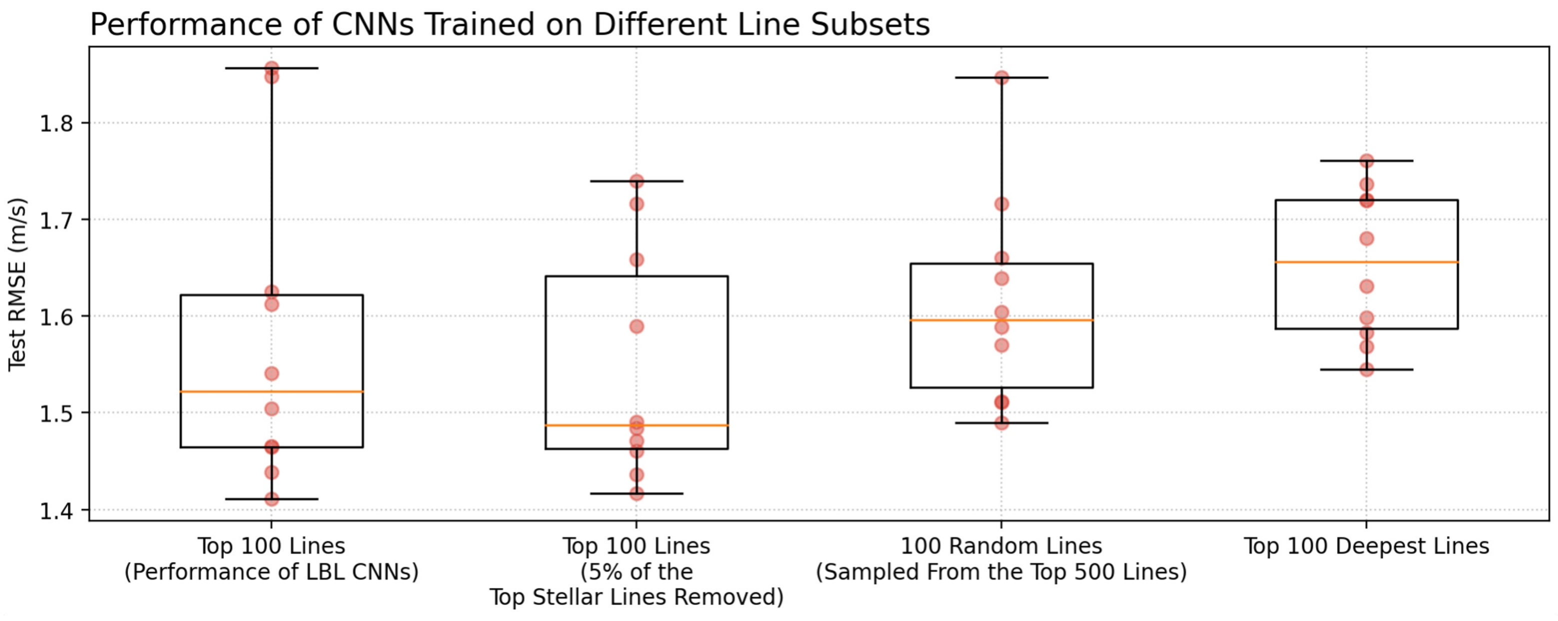

Figure-5: Performance results of the same CNN architecture trained on different subsets of lines (inputs) targeting injected planetary RVs. The removal of lines sensitive to stellar activity (second boxplot from the left) results in ~25% lines dropped from the top 100 lines (left most boxplot). This removal of lines appears to improve the average performance of the CNNs.

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.

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., Ciswski-Kehe, J., et al. 2019, The Astronomical Journal, 158, 210. doi:10.3847/1538-3881/ab441c

PI/Task Mgr. Contact Information: ***virisha.timmaraju@jpl.nasa.gov***

National Aeronautics and Space Administration
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California
www.nasa.gov

Clearance Number: CL#00-0000
Poster Number: RPC#