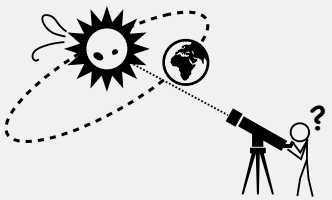


ÆSTRA: A Deep Learning Tool for Stellar Activity Mitigation in EPRV Measurements

Yan Liang[1], Joshua N. Winn[1]
[1]Department of Astrophysical Sciences, Princeton University

PRINCETON UNIVERSITY

Introduction

Detection of **Earth-like exoplanets** within habitable zones around Sun-like stars is very difficult due to the **inherent variability of host stars**.

Doppler radial velocity (RV) for an **Earth-like planet: 0.1 m/s**
Apparent RV induced by **stellar activity: ≥ 1 m/s**

Many methods address activity signals at the RV level, but a more refined approach is needed at the **spectral level.**

We present ÆSTRA*, an **unsupervised deep-learning approach** that exploits the rich information in stellar spectra to mitigate stellar activity and enhance RV measurements **without relying on detailed physical models**.

*ÆSTRA: Auto-Encoding STellar Radial-velocity and Activity

Spectrum Auto-encoder & RV Estimator

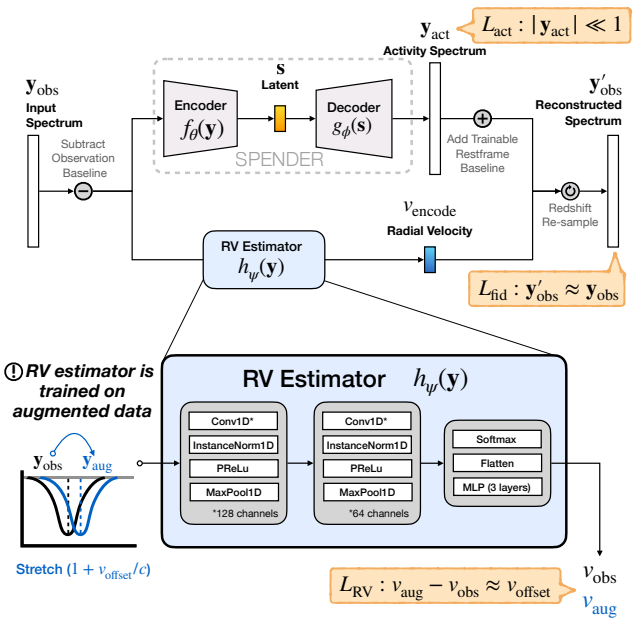
Network Input: Observed spectrum $\mathbf{y}_{\rm obs}$, without prior RV information

Network Outputs:

Doppler-invariant latent vectors $\mathbf{s}$: encodes intrinsic line variations

Rest-frame spectrum $\mathbf{y}_{\rm rest}$: comprehensive model of stellar activity

Radial velocity $v_{\rm encode}$: combined with $\mathbf{y}_{\rm rest}$ to compute $\mathbf{y}'_{\rm obs}$



Training via Data Augmentation

The loss function is designed to impose structure on the latent space by explicitly disentangling the Doppler shift from stellar activity.

$$L_{\rm total} \begin{cases} L_{\rm fid} = \dfrac{1}{NL}\sum_i^N \mathbf{w}_i \odot (\mathbf{y}_{{\rm obs},i} - \mathbf{y}'_{{\rm obs},i})^2 & \text{"Good reconstruction"} \\ L_{\rm RV} = \dfrac{1}{N}\sum_i^N \left[\dfrac{1}{\sigma_v^2}(v_{{\rm aug},i} - v_{{\rm obs},i} - v_{{\rm offset},i})^2\right] & \text{"Meaning of a Doppler shift"} \\ L_{\rm act} = \dfrac{k}{NL}\sum_i^N \dfrac{\mathbf{y}^2_{{\rm act},i}}{\sigma_y^2} & \text{"Activity-induced perturbations are small"} \end{cases}$$

The RV Estimator can be trained independently using $L_{\rm RV}$

Perfectly Reconstructs Simulated Data

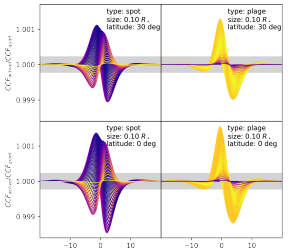
We created a set of simulated spectra using the SOAP 2.0 code, which models the effects of stellar activity on CCFs.

To construct full spectra, we:

Build a quiet spectrum:
- Generate a list of line locations
- Randomly draw width and depth

Perturb each line using CCF:
- Calculate the CCF with four random active regions drawn from a prior
- Perturb each line by multiplying the CCF ratio:

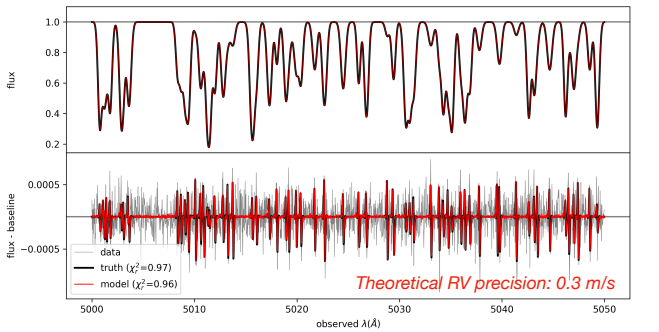
$$f_{\rm CCF}(v) = \frac{\rm CCF_{active}}{\rm CCF_{quiet}}$$



CCF ratios for a solar-type star color-coded by brightness during varying phases of stellar rotation.

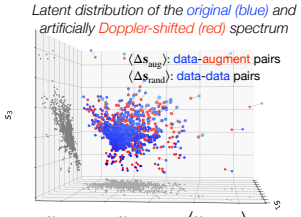
SOAP Parameter Priors ($N_{\rm active} = 4$)

Activity type	-	$P(\text{Spot}) = P(\text{Plage}) = 0.5$
Size [R_*]	Gaussian	$\mu = 0.10,\ \sigma = 0.05$
Longitude [°]	Uniform	$0° \le \phi < 360°$
Latitude [°]	Uniform	$-30° \le \theta \le 30°$



Grey: the observed spectrum with four active regions and photon noise
Red: the reconstruction *Black: noise-free underlying spectrum*

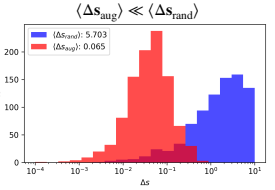
Recover 0.1 m/s Earth-like signals from random stellar activity



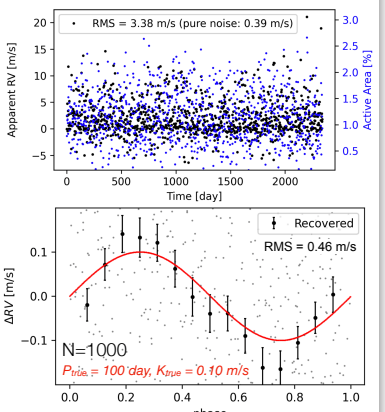
$v_{\rm recover} = v_{\rm encode} - \langle v_{\rm encode}\rangle_{D_{10}}$

Averaged over a ***Gaussian kernel***

Doppler-shift-invariant encoding: $\langle \Delta s_{\rm aug}\rangle \ll \langle \Delta s_{\rm rand}\rangle$



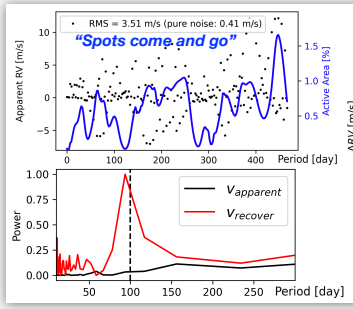
RV scatter: 3.38 m/s → 0.46 m/s



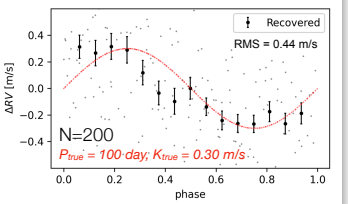
Parameter Constraints:

$P = 100.0^{+0.27}_{-0.26}$ [day] $K = 0.12^{+0.01}_{-0.01}$ [m/s]

Recover 0.3 m/s signals from starspot evolution



RV scatter: 3.51 m/s → 0.44 m/s



Parameter Constraints:

$P = 101.9^{+1.3}_{-1.3}$ [day] $K = 0.29^{+0.03}_{-0.03}$ [m/s]

Conclusion & Outlook

Conclusion

- ÆSTRA effectively detects 0.1 m/s signals in simulated data amidst 3.4 m/s stellar activity and 0.3 m/s photon noise.
- Successfully captures stellar activity with strong time structures.

Limitations

- Tested only on simulated data, without telluric/instrumental noise.
- Requires ~200 spectra.
- Not utilizing time-domain information.

Future Plans

- Apply ÆSTRA to solar data.
- Incorporate time-domain analysis,
- Deal with telluric features.
- Optimize for fewer spectra.
- Apply to exoplanet hosts, improving characterization and validating low S/N candidates.