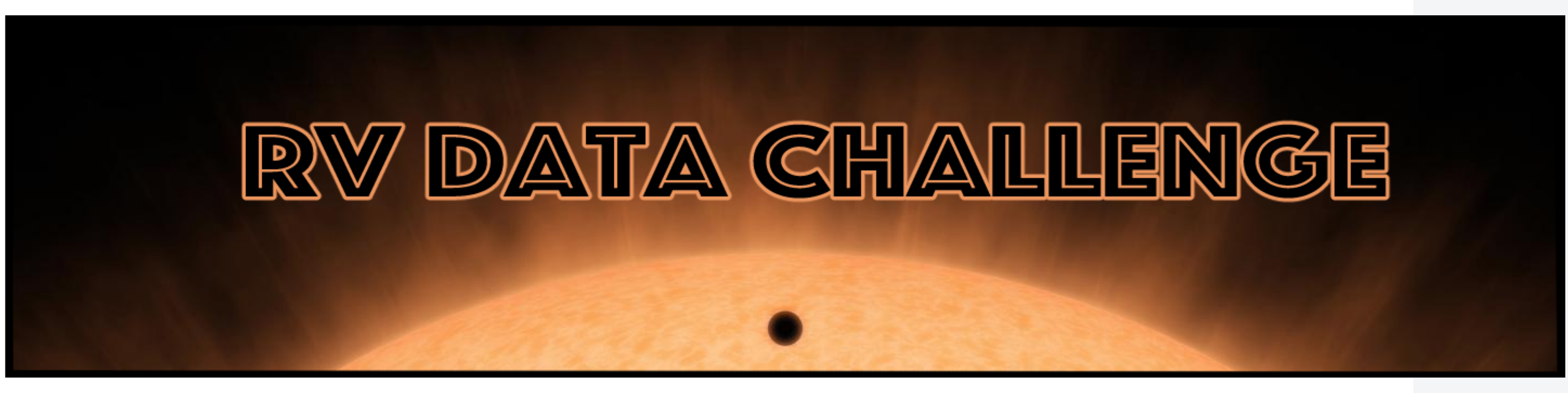
SVWG -- RV data challenge

Created by Sophia Sulis, last modified by Michael Cretignier on 21 Mar, 2025



Contact

Michael Cretignier (michael.cretignier@physics.ox.ac.uk),

Nathan Hara (nathan.hara@lam.fr)

Goal

RV data challenge aims to give insights on the best data analysis practices to achieve precise mass measurements of Earth analogs with RV follow-up of PLATO candidates.

The challenge will consist of analysing **2 G-type stars** (early and late), one mid-active (Sun) and one very inactive (Tau Ceti).
For each star, **3 different planetary systems** (2 x 3 = 6 datasets) will be injected at the spectrum level. Planetary systems are not communicated to participants for the blind phase.

The goal is to retrieve the orbital elements of the planets, in particular the mass.

In the summer 2024, a single dataset will be provided to the participants, so that they can test their methods (see summer data challenge dataset below). The proper data challenge will start in September 2024.

If you intend to participate please email Michael Cretignier and Nathan Hara (see above), and we recommend you star this confluence page.

Output from participants

For the meta analysis, the participants are asked to provide the **posterior distribution** of the orbital elements of the planets:

- RV Semi-amplitude
- Period
- Eccentricity
- Mean anomaly at t=0
- Argument of periastron

The posterior distribution can be given as an MCMC chain or nested sampling output.

Participants are also asked to provide their **"cleaned" RV time series** in a .csv table format N x 3 [time, RV_cleaned, RV_err].

Data products

The data products available for the participants are:

1. S1D spectra time series
2. CCFs time series
3. RV + proxies time series

All instruments, tend to contain heavy instrumental systematics that can, or not, significantly affect the RV extraction afterwards depending on if their signature is diluted once a CCF is computed.

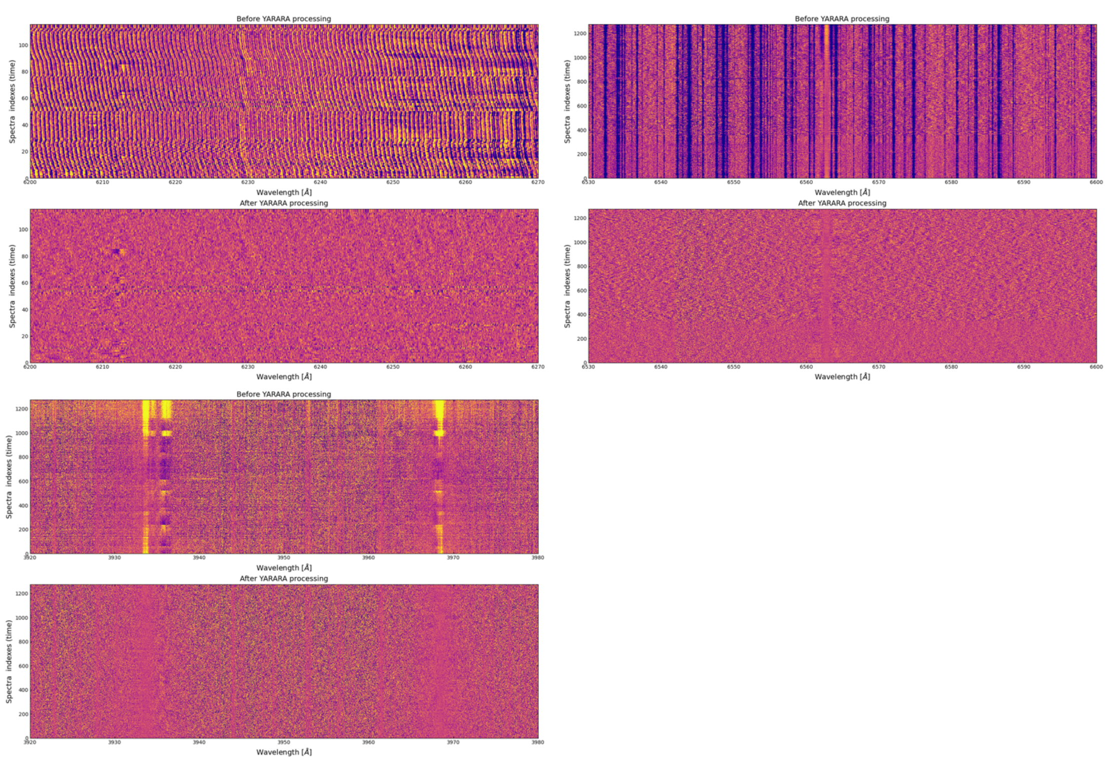


Figure 1: Example of instrumental systematics corrected by YARARA. ESPRESSO systematics (left), HARPS-N telluric lines (middle), HARPS-N ghosts (right).

Data are provided in **two flavours** of processing by YARARA (Cretignier et al. 21):

- Uncorrected of instrumental systematics (**YV0**)
- Corrected of instrumental systematics but not of activity (**YVA**)

S1D Spectra:

The spectra time series are all interpolated on a common wavelength grid in the air sampled with an equidistant bin of 0.01 .
Solar spectra are in the heliocentric rest frame, Tau Ceti spectra are in the barycentric rest frame.

CCFs:

For both stars, CCFs are obtained using the generic line list of the G2 CCF HARPS DRS (v.3.5).

Products description:

Transit_information.csv: Table containing the P and Tc of transiting planets.

**_DACE_YV0.rdb: A DACE-compatible table that can be imported in DACE.unige.ch of the YV0 time-series (RVs + proxies)*

**_DACE_YVA.rdb: A DACE-compatible table that can be imported in DACE.unige.ch of the YVA time-series (RVs + proxies)*

**_CCF_YV0.p:* CCFs from the YV0 products

**_CCF_YVA.p:* CCFs from the YVA products

**_flux_YV0.npy:* The YV0 spectra time-series flux normalised.

**_flux_YVA.npy:* The YVA spectra time-series flux normalised.

**_flux_err.npy:* The flux normalised uncertainties time-series.

**_continuum_YV0.npy:* The continua time-series. It can be used on the flux normalised spectra time-series to return to the absolute flux units.

**_Analyse_material.p:* Storage of wavelength grid for the spectra ('wave' keyword)

**_Analyse_summary.csv:* Table containing the time entry of the observations + activity proxies

**_Kitcat_mask.p:* Table containing all the stellar lines identified on the spectrum with a VALD atomic database crossmatch

**_YVA_PSF_vectors.csv:* Table containing the vectors used by YARARA to correct the change of the instrumental PSF

Reading data (useful Python functions):

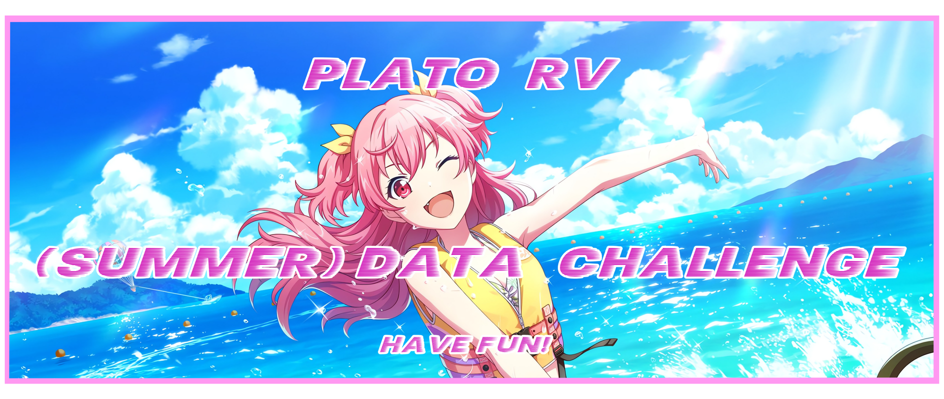
pickle (.p) → data = pd.read_pickle('file.p')

csv (.csv) → data = pd.read_csv('file.csv',index_col=0)

numpy (.npy) → data = np.load('file.npy')

Summer data challenge dataset (Mystery Star)

24th June - ~~10th September~~ 22th September



↑ Data can be obtained by clicking on the banner here above↑

Observed by HARPS, 362 spectra, SNR ~ 200

Stellar parameters:

- Sp. type = K1V
- Teff = 5154 +/- 80 K
- Logg = 4.55 +/- 0.08 dex
- FeH = -0.30 +/- 0.08 dex
- RV_sys = 2.78 km/s
- Mass = 0.79 +/- 0.04 Msol
- Radius = 0.77 +/- 0.04 Rsol

RV Challenge Datasets

Dataset 1 (Sun):

14th January - ~~28th March~~ - 25th April

Observed by the **HARPS-N solar telescope** (Dumusque et al. 15, Collier Cameron et al. 19, Dumusque et al. 21), **1275 spectra** spread over **8 years**.

Daily cadence is created using a 15min (3 x 5 minutes) integration time by stacking the spectra (**SNR ~450** [375 - 525]).

Stellar parameters:

- Sp. type = G2V
- Teff = 5778 +/- 1 K
- Logg = 4.44 +/- 0.01 dex
- FeH = 0.00 +/- 0.01 dex
- RV_sys = 0.00 km/s
- Mass = 1.00 +/- 0.01 Msol
- Radius = 1.00 +/- 0.01 Rsol

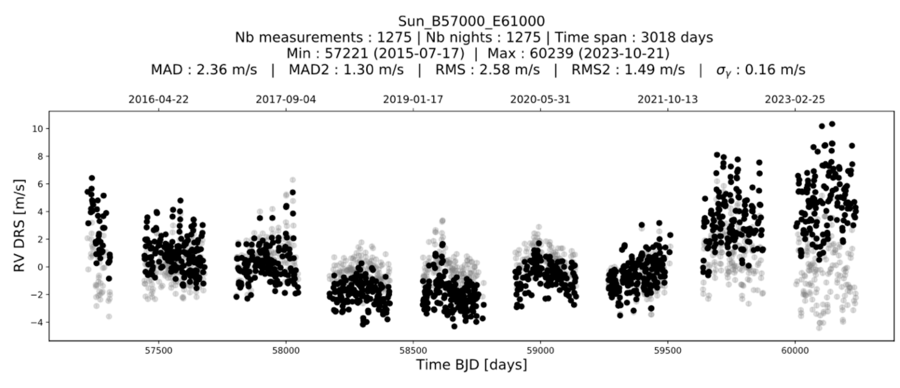


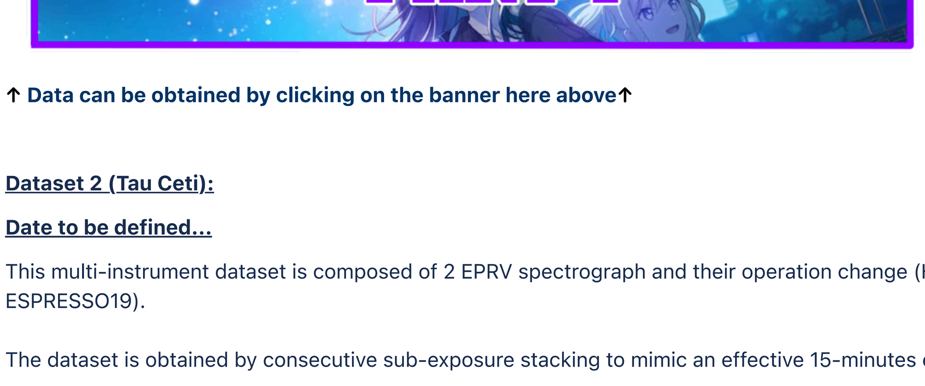
Figure 2: Summary of one of the three HARPS-N solar dataset.

Further information to participants:

All the planetary systems are known to be perfectly circular (e=0).
The **ephemerides of transiting exoplanets** (P, Tc) are provided in the X_Transit_information.csv files.

The dataset SUN_B57001_E61001-FallChallenge1 contains a mono-transit case with only the Tc provided.

The period of the mono-transiting planets is only known to be longer than 100 days.



↑ Data can be obtained by clicking on the banner here above↑

Dataset 2 (Tau Ceti):

Date to be defined...

This multi-instrument dataset is composed of 2 EPRV spectrograph and their operation change (HARPS03, HARPS15, ESPRESSO18 and ESPRESSO19).

The dataset is obtained by consecutive sub-exposure stacking to mimic an effective 15-minutes exposure time.

ESPRESSO18: 30 nightly epochs

ESPRESSO19: 109 nightly epochs

HARPS03: 432 nightly epochs

HARPS15: 220 nightly epochs

CORALIE14: