Submission Guidelines

This document details what should be submitted for each group and method as well as the conventions for each file. This will help decrease confusion and streamline results by standardizing file locations, file names, and table structure. Questions and suggestions are always welcome through **lily.zhao@yale.edu**.

The same Yale Box folder that contains the data files will also feature a folder for each group (where the name of the folder will be the provided name of the group) with a sub-folder for each of the provided targets for method results. All information from a group about members, methods, and results should be uploaded to this folder. Files should be named as specified. For tables, column names are specified with one example row given for each table.

1   GROUP INFORMATION

Each group should provide a `.csv` file with the name `<group name>_members.csv` with the following information for each group member:

Table 1: File Name: `<group name>_members.csv`

Name	E-mail	ORCID iD	Affiliations
Lily Zhao	lily.zhao@yale.edu	0000-0002-3852-3590	Yale University, . . .
⋮	⋮	⋮	⋮

Additional affiliations can be listed in following columns. The provided names and affiliations should reflect how members wish the information to appear in the final published report.

2   METHOD DESCRIPTION

We ask that each group submit a summary table describing some basic properties of all methods tried. Each method should also have a separate document containing answers to more detailed questions. Of course, it is reasonable to submit just one document for methods that are similar (e.g. indicators using the same principal but on different lines, GPs trained on different indicators, etc.).

2.1   TABLE OF METHOD BASICS

Basic information about each of the methods tried by a group will be collected into one table. For each method, please provide the following basic information in a `.csv` table named `<group name>_methods.csv`.

Table 2: File Name: `<group name>_methods.csv`

Name	Nick Name	Type	Input	Re-Calculated	Run On	Run Time	Results
CCF FWHM	ccf_fwhm	IND	spectra	None	Laptop	Seconds	ex_ccf_fwhm_results.csv
⋮	⋮	⋮	⋮	⋮	⋮	⋮	

The **Name** column should have a full, descriptive name for each method that can be used to refer to it in text. The **Nick Name** column is for a shorter, file-name friendly version (e.g. no spaces, special characters, etc.) that will be incorporated into file names. Variations of a similar method can, for example, have the same base name with different numbers, but each distinct method yielding results that deserve to be analyzed separately should be given a separate row.

The **Type** column will be used to group methods in the final report. Please categorize each proposed method into one of the following four categories:

1. **IND** - a type of activity indicator

2. **GP** - an application of Gaussian Processes to model the activity signal

3. **RV** - a new way of deriving RVs

4. **OTH** - (i.e. other) the method is too creative to fit into these oppressive categories

Methods categorized as "other" can be further explained within the detailed questions.

The **Input** column should list the data products required by the method, e.g. 2D spectra, photometry, specific indicators, etc. If, in the analysis that produced the results submitted, your group re-generated any of the required inputs (i.e. did not use the provided values), please specify what was re-calculated in the **Re-Calculated** column.

The **Run On** and **Run Time** columns will be used to get a qualitative sense of how much processing power is required to implement the proposed method. For **Run On**, we are not looking for model numbers, just general device type, e.g. laptop, desktop, cluster, etc. For **Run Time**, we are looking just for the appropriate unit really, e.g. seconds, hours, days, etc.

In the **Results** column, please specify the name of the file containing results for that method without the star name. Result files should follow the convention `<star number>_<group name>_<method name>_results.csv`; for this column, the star number does not need to be specified. It is reasonable to combine results from similar methods, for example if your group plans to submit many types of indicators. In that case, the file combining the results from many methods should be specified, e.g. `<group name>_<method name>_indicators.csv`.

2.2 DETAILED METHOD QUESTIONS

For each method, we also ask a series of question to garner more detailed information. The questions can be found in the "Method Questions[.tex/.docx]" documents, which can be used as a template to be filled out for each method. Each document should specify the group name and relevant method name in the header.

These documents should be uploaded to your group's folder on Yale Box in `PDF` or `TEX` format with the file name: `<group name>_<method name>.[pdf/tex]`. Methods can be grouped into one document if their guiding principals are similar enough that the answers to most of the questions are the same. If your team is submitting many methods, feel free to sort method descriptions and results into sub-folders based on category.

3   RESULTS

Result should be organized into different folders by star (i.e. all results for HD 101501 should be in a `101501` folder). We plan to compare the results of different methods using the RMS of cleaned, activity-less RVs and the presence (or lack thereof) of activity-driven periodicities in the results. Therefore, for each method wherever possible, please provide:

Table 3: File Name: `<star no.>_<group name>_<method name>_results.csv`

Time [MJD]	RV_C [m/s]	e_RV_C [m/s]	RV_A [m/s]	e_RV_A [m/s]	Ind. [U]	e_Ind. [U]
58639.459	55.027	0.345	NaN	NaN	7741.16	27.68
⋮	⋮	⋮	⋮			

The **Time [MJD]** column should mirror the time stamps given. We ask that results are returned for all observations provided, even if there exists outliers. Undiagnosed outliers should be brought to the attention of the `EXPRES` team, who will do a closer investigation of the data provided.

Column **RV_C [m/s]** should reflect the cleaned RVs for each exposure with corresponding errors given in column **e_RV_C [m/s]**. Columns **RV_A [m/s]** and **e_RV_A [m/s]** should contain RVs and errors that are attributed to stellar activity/variability or noise(what was once called "dirty" RVs in V1) though we understand that only some methods will be able to deliver this. All RVs should be given in m/s.

Missing values should be filled in with the string `'NaN'`. There will be a place to discuss reasons why data may be missing in the detailed questions document for each method.

For indicator based methods, we also ask that the value of the indicators be provided along with errors for each value. Please be sure to specify the unit of the indicator in the column header whenever possible.

If the results of multiple methods are being combined into one results table, please repeat the relevant columns for each method. Additionally, please prepend the name of the method to the front of the column name for each method.

4   CHALLENGES

4.1   INJECTED PLANET TESTS

Note: the data for these tests were not uploaded until February. These results will not be due until **March 14, 2021**.

Data sets with injected simulated planet orbits are provided to test whether or not a method preserves planetary signals while mitigating activity signals. A method for mitigating activity can not be considered successful if it also washes out planetary signals. Simulated systems could have any number of planets with a variety of periods and eccentricities.

The data for this test will be housed in the "Challenge" folder, but the results should otherwise be submitted in the same way. The provided resultant RV measurements will all be fit in post-processing (on the Yale side of things) by the same planet-fitting code to allow for a consistent comparison of returned planet parameters.

4.2 TRAINING/VALIDATION TESTS

Guided training/validation frameworks are provided to establish a better statistical understanding of how methods are performing in a way that regularizes these test across teams. For each target, a set of masks is available. There are 60 masks for each target each with as many entries as there are observations of that target. These masks should be used to exclude observations from the analysis wherever the mask is False.

Results of these tests should be submitted similar to the general results with each training/validation set appending another set of columns to the results table. Other than "Time [MJD]", column names should be pre-pended with the number of the mask using zero indexing. For example, the results from implementing the first mask should read:

Table 4: File Name: `<star no.>_<group name>_<method name>_results_tv.csv`

Time [MJD]	0RV_C [m/s]	0e_RV_C [m/s]	0RV_A [m/s]	0e_RV_A [m/s]	0Ind. [U]	0e_Ind. [U]
58639.459	55.027	0.345	NaN	NaN	7741.16	27.68
⋮	⋮	⋮	⋮			

An example of the first several columns of a method that only returns RVs is given below.

Table 5: File Name: `<star no.>_<group name>_<method name>_results_tv.csv`

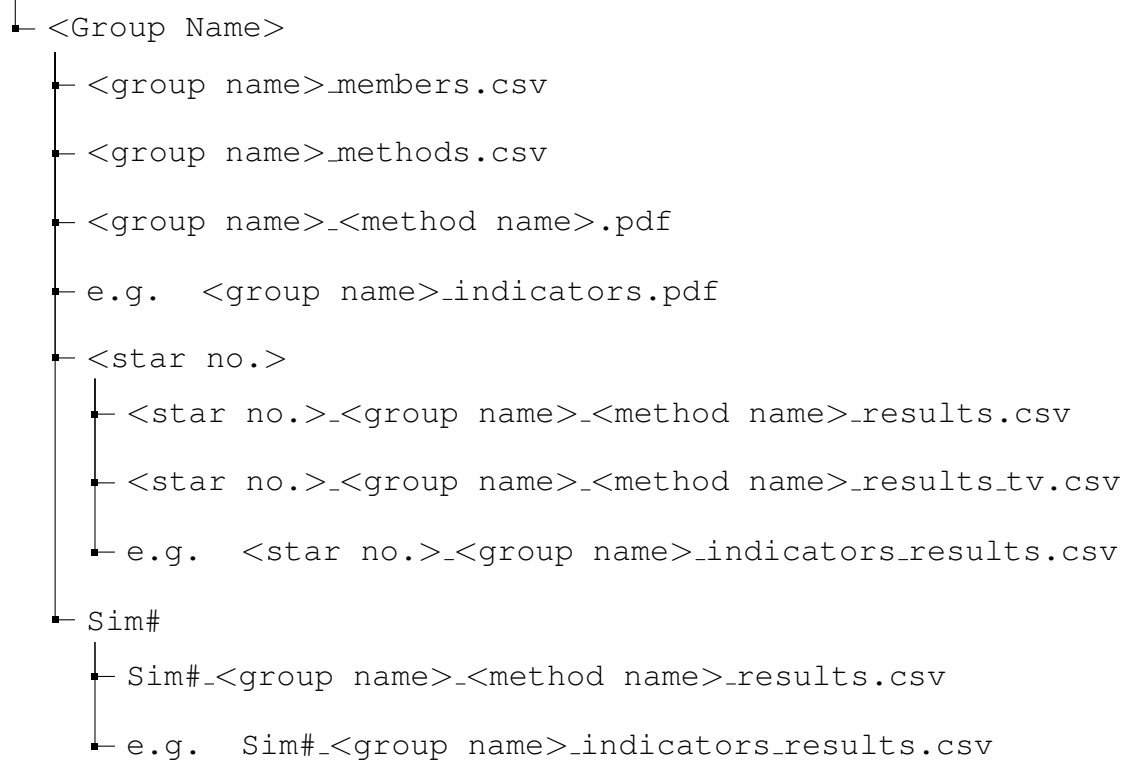| Time [MJD]	0RV_C [m/s]	0e_RV_C [m/s]	1RV_C [m/s]	1e_RV_C [m/s]
58639.459	55.027	0.345	55.027	0.345
⋮	⋮	⋮	⋮	

5 SUMMARY FILE STRUCTURE

In summary, the directory structure and file names for the Yale Box folder is recreated below. An example file name for a table that collects the information of all indicators tried by a group is included.

5.1 DATA FOLDER

```
Data
├─ <star no.>
│  ├─ <star no.>_spectra.zip
│  ├─ <star no.>_activity.csv
│  ├─ <star no.>_stellar_parameters.pdf
│  ├─ CCFs<star no.>
│  │  └─ <star no.>_YYMMDD.####_ccf.fits
│  ├─ Photometry <star no.>
│  │  ├─ <star no.>_photometry.csv
│  │  └─ photometry notes and plots
├─ Challenges
│  ├─ Training_Validation
│  │  ├─ <star no.>_masks.csv
│  │  ├─ <star no.>_validation.png
│  │  └─ validation_masks.png
│  ├─ Injected_Planet (to come)
│  │  └─ Sim#
│  │     ├─ Sim#_spectra.zip
│  │     ├─ Sim#_activity.csv
│  │     ├─ CCFs#
│  │     │  └─ Sim#__YYMMDD.####_ccf.fits
├─ Read Noise
│  └─ 190506.read_noise.fits
```

5.2   TEAMS FOLDER

```
Teams
│
└─ <Group Name>
   │
   ├─ <group name>_members.csv
   │
   ├─ <group name>_methods.csv
   │
   ├─ <group name>_<method name>.pdf
   │
   ├─ e.g.  <group name>_indicators.pdf
   │
   ├─ <star no.>
   │  │
   │  ├─ <star no.>_<group name>_<method name>_results.csv
   │  │
   │  ├─ <star no.>_<group name>_<method name>_results_tv.csv
   │  │
   │  └─ e.g.  <star no.>_<group name>_indicators_results.csv
   │
   └─ Sim#
      │
      ├─ Sim#_<group name>_<method name>_results.csv
      │
      └─ e.g.  Sim#_<group name>_indicators_results.csv
```

5.3   NOTES FOLDER

```
Notes
│
├─ Submission_Guidlines.pdf
│
├─ Method_Questions.tex
│
├─ Method_Questions.docx
│
├─ 201009_CheckIn1_Notes.pdf
│
├─ 201113_CheckIn2_Notes.pdf
│
├─ 201113_CheckIn2_Slides.pdf
│
├─ 201113_CheckIn2_ZoomChat.txt
```