Community feedback on the initial ESSP 4 plan

Collected through

- Google form (4 responses)
- Zoom meeting (about 20 participants)

Proposed Timeline

Most people on the Google form noted that the timeline seemed a bit tight. One suggested that the target is to have results that can be discussed at EPRV 6 rather than pushing to have final results to present at EPRV 6.

Note that the ESO 2024 proposal deadline was March 21. The deadline for 2025 hasn't been released yet (call for proposals link), but we should maybe keep a similar time clear for this year.

Number of Data Sets

The google form had general support for "fewer, more targeted data sets for each data challenge but with more data challenges." People cited this as the way to go to prioritize comparing different methods, producing a focused paper, and giving participants adequate time to process all data sets. This can also highlight issues addressed by subsequent papers. Better for comparing different methods.

One concrete suggestions was: "we make a list of specific questions that could be usefully addressed by a data challenge, sort them in a reasonable order, and then try to make progress at a rate of ~0.5-2 per year (depending on how involved each is)"

It should be noted, however, that having a more focused question for each data challenge does not necessarily map to fewer data sets. The support was for "**more focused data challenges**," not necessarily for less data sets. For instance, to truly test detectability with a reasonable measure of false alarm probability is asking for some 10e3 data sets to be tested. Fewer data sets are really just to assess "which methods are ready to go and which need more refining."

The number of data sets may define what people attempt to apply. For instance, too many different data sets and people may shy away from methods that are too computationally expensive or require specific tuning. We definitely don't want to limit our comparison in this way.

It was asked that this data challenge be mindful of the efforts of the PLATO data challenge and any RCN work. To me this suggests that we should all make the effort to establish that this is exactly the goal of the new executive committee.

Kinds of Data Sets

The simpler the system architecture the better (i.e. zero or single planet, circular orbits, or even just random Doppler shifts). Getting too complicated may shift the focus to how to best explore the parameter space of possible system architectures rather than separating out stellar

variability. Jointly-fitting multi-Keplerian models with stellar variability may be premature at this stage.

Random Doppler shifts needlessly disadvantages methods that jointly fit a Keplerian fit and is perhaps too much of an abstraction from the goal. The goal is not purely to measure Doppler shifts but rather to enable the recovery of a planet signal. Similarly, injecting no signals is a useless test on its own.

Note that systems with more than two or three planets are much more difficult to constrain. The parameter space increases dramatically with each new planet. Injecting too many planets may make the challenge more about recovering planets than separating out stellar signals.

Consider instituting a minimum period (i.e. 2+ days) to keep computational costs down.

Consider tuning the period to harmonics of the stellar rotation period.

We should consider whether different orbital inclinations of the planet could produce a detectable change in the observed spectra. This is perhaps more appropriate for a future data challenge or perhaps even an initially independent study.

How Much is Known About the System

We could consider doing some semi-blind tests that are specifically tuned to better target understanding when planet recovery methods fail.

Do we want to provide a system for which the planet parameters are known? This can help people tune their methods. This could, for example, emulate a system with a transiting planet for which the period is known. However, it would make overfitting the problem much too easy.

Indeed, perhaps all data sets should have a small injected planetary signal (below the detectability threshold) to emulate how in real life we can never be 100% sure that all true Doppler shifts are removed from a star.

Transit Follow-Up

A lot was discussed about considering whether injected planets transited or not. From my impression, this felt like a response to Michaël and Nathan's excellent and thorough work with the PLATO collaboration and is likely not something we need to focus on with the ESSP. It was mentioned that the ESSP line of work, as it is focused on the EPRV end of things, is unlikely to have transiting planets as a high priority as the distance of Earth-like planets greatly decreases transit likelihood.

Some relevant comments made on this topic include the idea of including non-transiting planets as a form of periodic, correlated "noise." There was also the suggestion of using a real system with both transiting and non-transiting planets, of which CoRoT-7 is a classic case.

Data Products, Standardization, and Sampling

The main suggestion was to try to expand the data set beyond thirty days. Thirty days is much too close to the rotation rate of the Sun and likely not enough data to really test mitigation methods. It also sounds like HARPS(-N) data has recently been reprocessed, so we should try to get the most current data products.

We can use an initial data drop to figure out what meta data/data standardization is needed.

Merged 1D spectrum should be on a log wavelength scale to preserve velocity spread (i.e. as opposed to a uniform wavelength grid)

Responses to the google form suggested that a telluric model would be useful but not essential. Someone suggested YARARA pre-processing of the spectra and CCF data. We will not provide this data product for now to keep things simpler. Especially as there is not a unique recognized method for removing tellurics.

Most people agreed that we don't need photometry at this stage. There were some suggestions of providing transit ephemerides, but others mentioned that the question of transit follow-up was not the primary focus of this data challenge.

We could consider obscuring what instrument observations are from. If we're going to do this, it should be accompanied by a formal question that we are trying to answer. For instance, our goal is not to test a method's ability to run on all methods equally well, and so in that regard it makes sense to help participants tune their methods to different instruments. On the other hand, we don't want participants to abuse the knowledge.

If there is a risk that information is misused, we should err on the side of hiding that information. Otherwise, participants may use the information to varying degrees and cloud the resulting analysis.

There was the suggestion that we provide observations paced 2-3 hours to combat (super-)granulation. HARPS-N operates in this way, though the observations usually just get averaged together. I guess we are kind of already doing this by providing observations from the four different instruments at high airmass.

We should consider how we can use different data sets (for example with different cadences) to test how good methods are at accounting for different stellar signals, such as (super)granulation. This can form the basis of future data challenges.

We can consider incorporating the equivalent width of the CCF as an activity indicator (see eq 1)

Proposed Analysis

There is a difference between (1) testing the ability to properly constrain mass/orbital parameters and (2) the ability to fully explore the parameter space of possible planet systems here. I may be extrapolating too far from other statements that were said, but I think the idea here was that the ability to (1) mitigate stellar signals so that a Keplerian signal becomes clear is different from (2) the ability to separate between different possible Keplerian signals. For instance, it was suggested that an initial test could be to measure the masses and orbits of known planets.

It was noted that the results are not going to be a binary (i.e. was the planet recovered?) but more a question of confidence. To this end, participants of the PLATO challenge were also asked to supply the posterior of their Keplerian fit. This helped that challenge better assess its goal, which was focused on characterizing the mass accuracy of the recovered planet signal.

Planet Recovery Pipeline

We should consider whether a standard recovery pipeline is even necessary. We want to ensure that the goal remains testing mitigation methods, not a test of the planet recovery pipeline. We should stay conscious of the fact that separating out the signals and properly fitting the resulting planet signal are separate skills (and indeed researchers typically specialize in one over the other). Of course, the two processes are not always fully independent, and correlations between the two must be taken into account.

If we are going to have a set recovery pipeline, it is likely we will have to change the proposed procedure. The proposed pipeline for planet recovery is dangerous in the case where there are correlations between the different sources of periodic signals. We will consult João for a better way of processing data. Appendix 3 of Nelson+ 2020 outlines an alternative method that was also suggested.

Requested Resources

We should provide a snippet of code that demonstrates how to work with the standardized data format. This code should show how to open and plot the data. We will provide a python example and accrue other coding languages (e.g. Julia, R, IDL) as necessary. We will make sure the data will be shared in a data format that is widely accessible (i.e. FITS, csv).

It would be helpful for participants to understand what they are and aren't "allowed" to do. This could include specific examples; for instance requesting that if they use additional data to train their models that they only make use of publicly available data. Such examples could be added to as we get more questions. This should be accompanied with some general test describing the purpose of the challenge to help participants determine what is in the "spirit" of the guidelines.

One avenue to consider is how much the sun should be treated as a star. We will never understand another star in the universe to the degree that we understand the Sun, but we want

the results from this project to be widely applicable to exoplanet hosting stars. It is likely that participants will make use of solar information even unconsciously.
(We might even want to obscure the data somewhat, like multiplying all times by some dilation factor!)

Some of the things we want to enforce should be emphasized in the questions we ask to describe each method. For example, the method document could ask them to specify any external data used.

We should mention that similar methods may be asked to join forces or specify how they are different from other methods.