**­­Response to reviewer report: FIESTA II. Disentangling stellar and instrumental variability from exoplanetary Doppler shifts in Fourier domain**

We thank sincerely all the science/data/statistics reviewers for their helpful and insightful comments. The reviewer comments are repeated below in *shadow\_and\_italics*, followed by the response (black) and the changes to the manuscript (purple texts or figures/s­napshots wrapped in purple frame).

1. *Introduction: the CCF is introduced by saying that it is a cross-correlation “of the stellar spectrum with a template spectrum or synthetic mask,” but the type of CCF used throughout this work appears to be specifically the EPRV standard cross-correlation with a weighted binary mask, which produces a CCF with the shape of an average spectral line. It’s not clear how the results & interpretation of the FIESTA method would change with a different kind of CCF. If the authors have thoughts on this, it would be an interesting discussion point! Regardless of that, the introduction should be more specific about the definition of the CCF that’s adopted in the paper, with a reference to Pepe et al. (2002) or a similar paper that goes into more detail.*

We agree that a different choice of CCF mask would inevitably change the CCF shape and thus result in different FIESTA outputs. While CCFs based on a template spectrum can improve the robustness of RVs for some stars (e.g., low SNR, broad absorption features in cool stars), we expect a narrow, weighted binary mask to preserve more line shape information. However, this paper rather focuses on the FIESTA method in its general form, and the method does not depend on the choice of CCF mask. While it could be interesting to study the effects of different types of masks used to construct the CCFs on the FIESTA outputs, that is beyond the scope of this paper.

For clarity, we have updated the sentence in question:

“In the context of radial velocity detection of exoplanets, the signal is usually the cross-correlation function (CCF) of the stellar spectrum with a template spectrum or synthetic mask. By combining information from many spectral lines, the CCF has significantly higher signal-to-noise ratio (SNR) than individual spectral lines. In this study, we validate and test FIESTA using CCFs based on a weighted binary mask (Baranne et al. 1996; Pepe et al. 2002).”

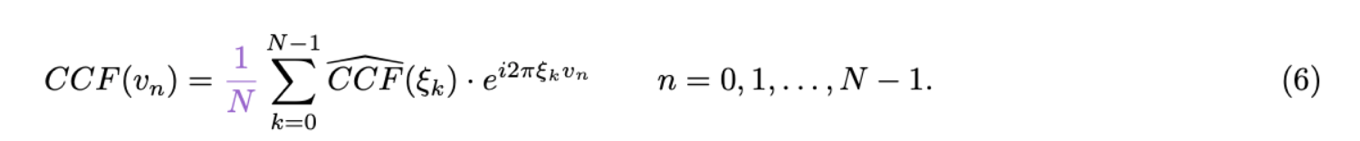
We also add the new Section 7.3.3 for such a discussion.

Text

Description automatically generated

1. *Section 2.1: Equation 6 may be missing a factor of 1/N.*

Indeed. 1/*N* was missing and is now added.

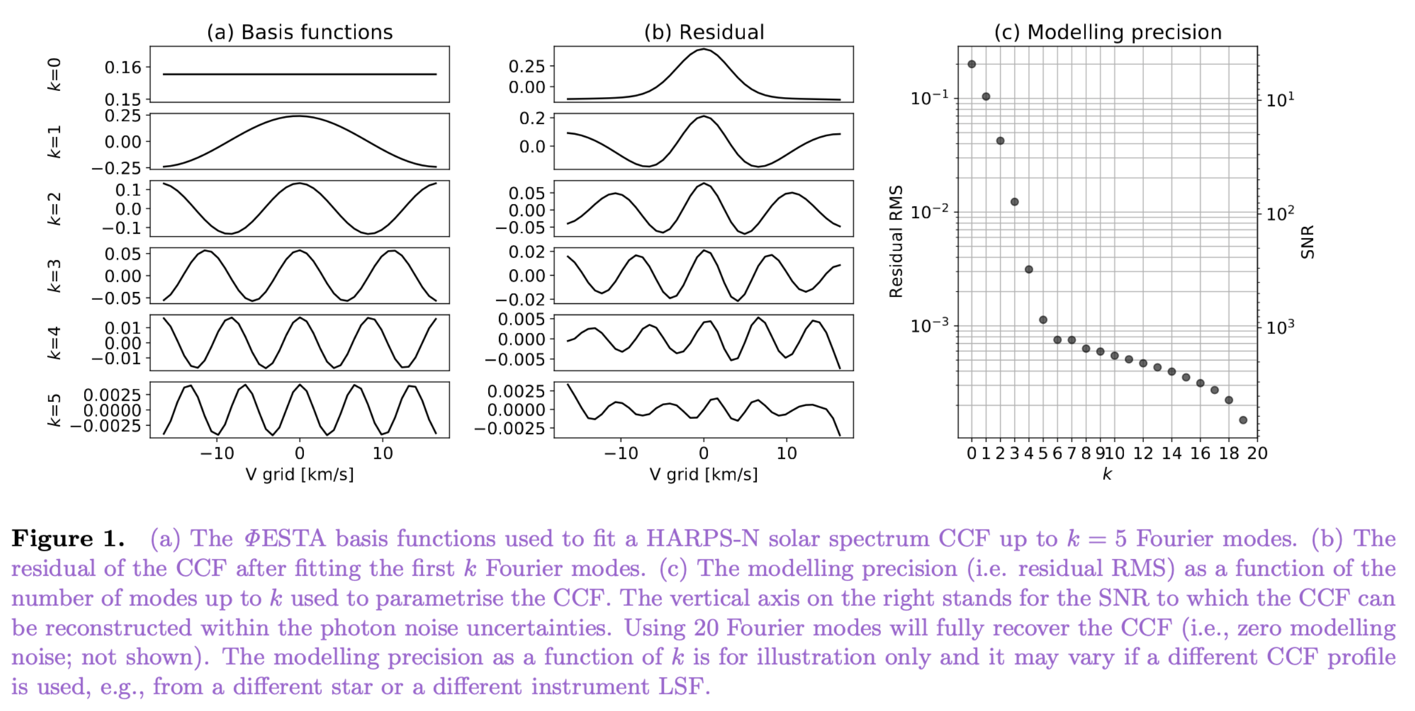


Meanwhile, we notice that the discrete Fourier transform in Equation 2 has an extra term 1/*N*, which is now removed.

1. *Figure 1 is not referenced in the text. It’s a very useful figure for building understanding of the method and should definitely be discussed!*

Figure 1 was referenced in Section 3.1, but we realise that it was inadequately explained in the text, especially the left half of the original figure, which is the current (a) and (b). We expanded the texts and figure captions.

“For example, Fig. 1 shows the first 5 *Φ*ESTA basis functions (or 6 including the zeroth mode, which is the mean function of the CCF) used to fit an arbitrary but typical HARPS-N solar spectrum CCF. For details of the solar spectrum, refer to Dumusque et al. (2021) and Section 5.”



1. *Section 3 needs structural editing: currently it states that there are “five aspects to consider…” before enumerating six questions and elaborating in five subsections (plus summary). I think that item 5 in the list of questions could be merged with item 4 to reconcile this. Also, the summary suggests that users adopt the minimum of 4 different k limits, is there one missing?*

*“…item 5 in the list of questions could be merged with item 4 to reconcile this.*” – Indeed. The original item 4 and item 5 basically express the same idea. They are merged and rewritten as the current item 4.

“From which *k* is the variability in the *Ak* and *ϕk* time-series dominated by the measurement uncertainties of *Ak*’s and *ϕk*’s?”

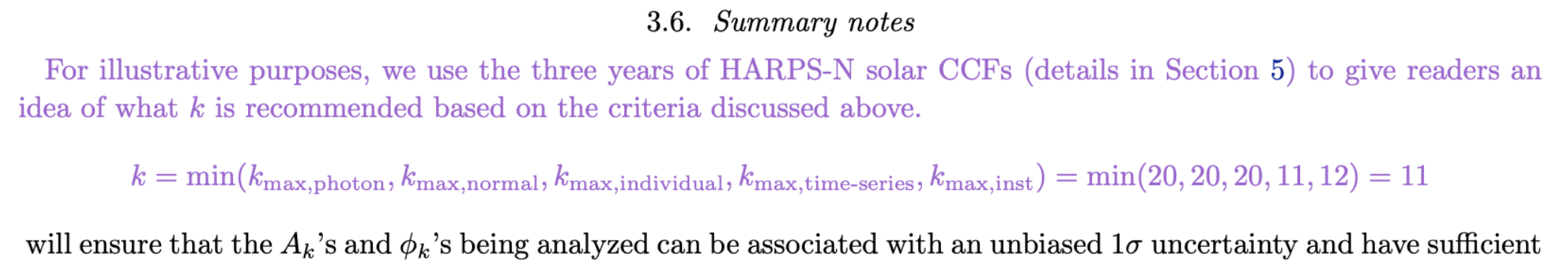
“*Also, the summary suggests that users adopt the minimum of 4 different k limits, is there one missing?*” – Nothing is missing as item 2 “What is the distribution of errors in *Ak* and *ϕk* due to photon noise?” is not a constraint for *k*.

1. *(also) Section 3: It would be useful for illustrative purposes to state what k limits are obtained for the analyses done in this study. Maybe this could be a table?*

We have updated in Section 3.1



And in Section 3.6



1. *Section 3.2: The first two sentences refer to “linear projection of the spectra” and “Gaussian noise in the spectra” where it should read “CCF” for “spectra.” The CCFs themselves are already transformations of the original spectra, so the assumption of Gaussian noise in the CCF needs to be explicitly justified.*

The original text was correct. While the reviewer is correct that we were taking a linear projection of CCFs, the CCFs themselves are a linear projection of the spectra. The linear projection of a linear projection is still a linear projection. Therefore, Gaussian noise in the original spectrum directly implies Gaussian noise in both the CCF and the estimated coefficients and .

Nevertheless, we took the referee’s advice and make the expression more explicit.



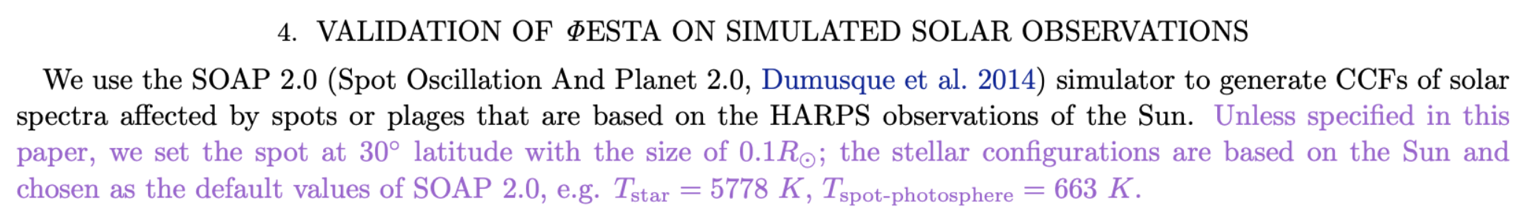
1. *Section 3.5: This limit doesn’t seem quite right… shouldn’t the number of spectral lines used to construct the CCF also factor into the velocity resolution? A HARPS CCF clearly has meaningful & coherent structure on length scales smaller than 2.5 km/s, or else it would be impossible to measure RVs at the precision HARPS achieves.*

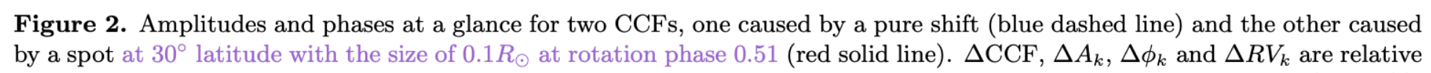
This subsection talks about the effect of instrumental resolution on parametrising CCFs. One can measure the location of the CCF minimum (“the RV”) with a precision much better than the width of CCF. The improvement due to combining multiple lines is reflected in the signal-to-noise of the CCF. We now use “instrumental resolution” to clarify our intent in the text.



1. *Table 1 is not necessary as a table; it could be stated in the text or as a footnote.*

Changed as suggested in Section 4. The other places (Figure 2, 3) that referenced Table 1 have also been updated accordingly.



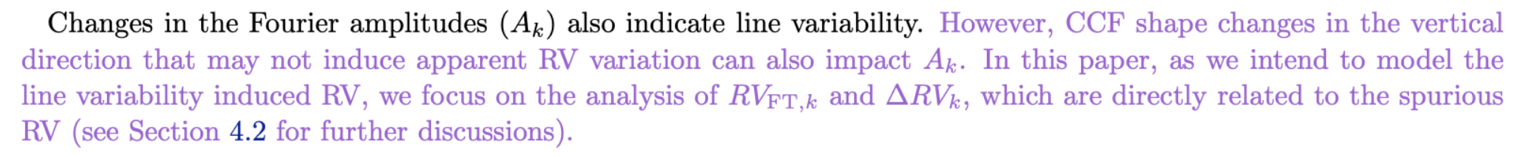


1. *Figure 2: I would interpret the top-right panel as saying that at small k, the Fourier amplitude is more informative than the phase in disentangling the spot-driven signal from the true RV shift, however the amplitude information is not employed in most of the following analysis. Can this choice be discussed/justified more in the text?*

Both A*k*’s and ∆*RVk*’s are sensitive to line deformations, as shown in Figure 2. The advantage of ∆*RVk* is that they are zeros for true Doppler shifts, even if the line depth or continuum level changes. In contrast, A*k* can also be sensitive to line depth and continuum changes not directly linked to RV variation. Therefore, the ∆*RVk*’s are more suitable for identifying line shape changes that contribute to apparent RV shifts.

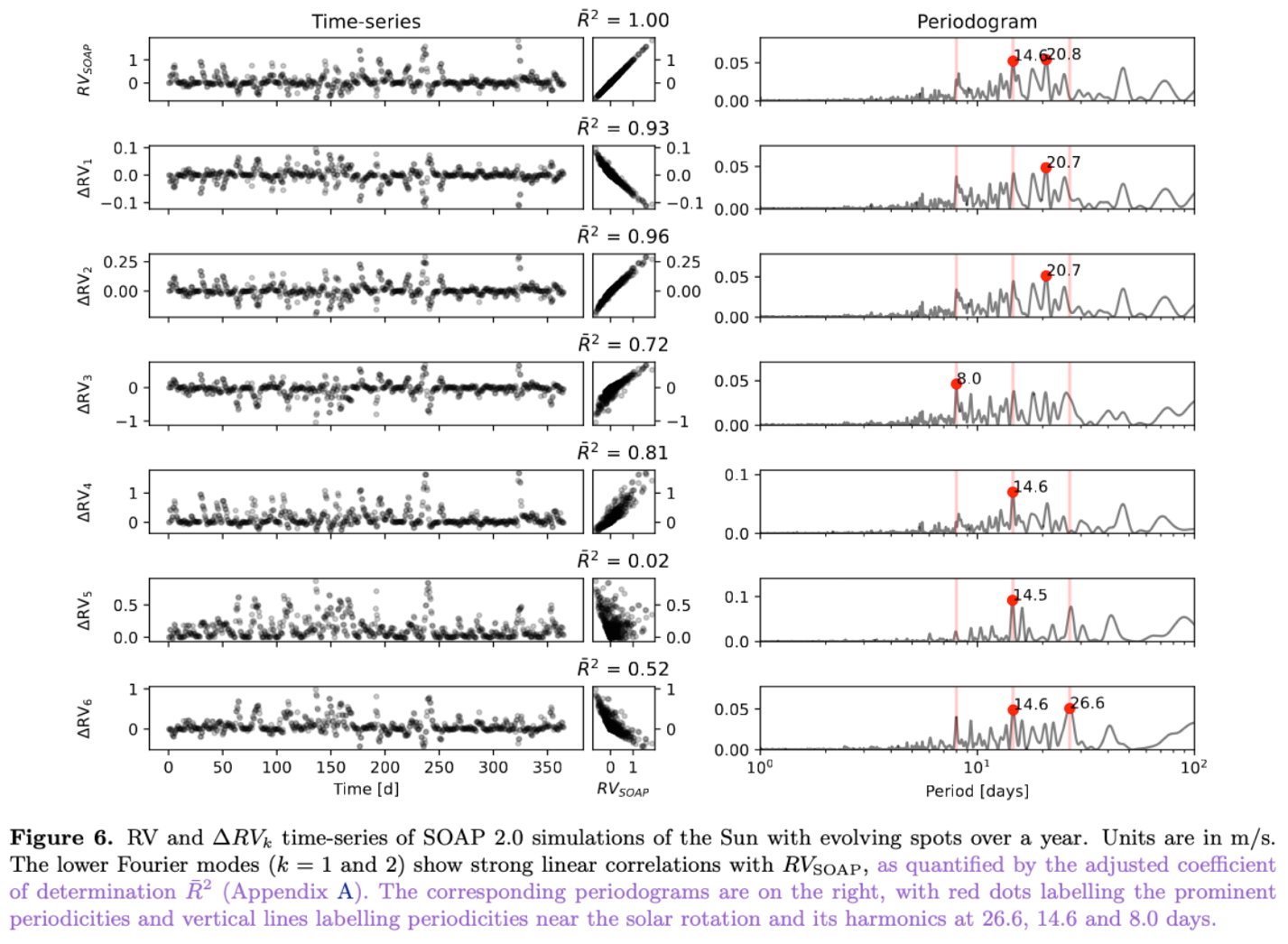
In this paper, our intention is to model the line variability induced RVs. For the reasons above, we opt to use ∆*RVk* in our analysis. However, if we were to study the line deformation, e.g., due to the LSF changes that may not introduce apparent RV shifts, we may as well need the additional amplitude information.

We have expanded on our explanation of why ∆*RVk* is preferred over *Ak* in Section 2.5.



1. *Section 4.4: This section and Figure 6 give really nice context for the following analysis. It may be worth showing or at least commenting on the periodograms of the various k components, for maximum comparability to Figures 7 & 8.*

While we had inspected the periodograms of the SOAP RVs and the ∆*RVk,*, we opted to exclude them from the original manuscript. However, considering the readers might ask the same question, we have added the following figure.



1. *I think it would actually be worth going through the entire analysis on this simulated data, including the PCA and subsequent filtering steps; if the intuition is correct that reducing dimensionality with PCA is motivated by the limited number of mechanisms changing the CCF, then the simulated data with only spots should have fewer high-scoring PCA components and it should be entirely contained in the S-PC short-term variable components, right? Is that the case?*

If one applies PCA to ∆*RVk*’s from SOAP, the first (i.e., most prominent) PC score would explain over 96% of the variance when applied to the simulated data (see for and 2) and the rest PC scores account for less than 4% of the variance and won’t be needed.

In addition, the simulated data do not exhibit long-term variability, such as the solar magnetic cycle or long-term instrumental variations that we observe in the HARPS-N observations, and so S-PC1 alone would be enough (i.e., >96% variance explained) to represent the RV variability from the SOAP simulation.

We do not want to leave the reader the impression that there is only way we make use of the FIESTA outputs. While we describe a standard FIESTA algorithm, the approach for analyzing the outputs does not necessarily need to be a standard procedure. Indeed, in the case of simulated data in Section 4.4, there is no need to proceed with PCA, separating out the long-term and short-term variabilities, and multiple linear regression modelling as there is for HARPS-N solar observations in Section 5. By including both approaches, the paper illustrates different ways FIESTA can be put to practice.

1. *Related to the above point: can you test for overfitting by injecting & recovering a true Doppler shift? I assume that doing the linear regression modeling from Section 5.5 ought to remove the spot effects while preserving true Doppler signals, so can you show that is actually the case?*

We think the referee is talking about injecting a true planetary signal and recovering it in the time series analysis. Otherwise the tests from Figure 2 are sufficient to demonstrate the method successfully preserves and recovers true Doppler shifts, as identified with ∆*Ak*and *RVFT,k*.

While injecting and recovering a true Doppler signal could be a useful test, we focus on using cross validation to ensure that we are not overfitting the stellar variability indicators when applying multiple linear regression.

1. *Figures 7 & 8: These figures might be easier to interpret if you add semi-transparent shaded regions in different colors around the significant periodicities enumerated in Section 5.2, so the viewer can see at a glance whether different Fourier components correspond in part to different mechanisms.*

Figures 7 and 8 (as well as 6) are updated as suggested.

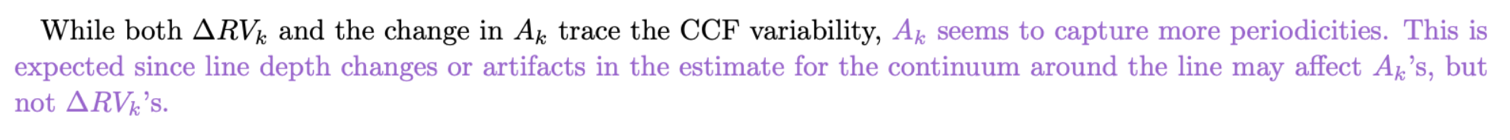
Table

Description automatically generatedTable

Description automatically generated

1. *Section 5.2: I am not sure I understand the point about the A\_k periodograms being noisier because the A\_k values are affected by continuum normalization, can you elaborate? Or potentially address this with similar plots of A\_k for the simulated data?*

We regret having used “noisier” and have rephrased the sentence and added the requested elaboration:



Below we attach the *Ak* periodogram of the simulated data for the reviewer’s reference (not presented in the manuscript). It does capture more periodicities, but these extra periodicities do not have a straightforward interpretation, which led us to originally referred them as “noise”.

A picture containing table

Description automatically generated

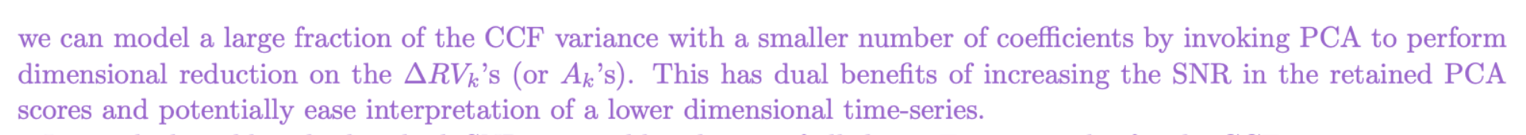
1. *Section 5.3: The PCA step of analysis needs to be motivated a bit more. The reasoning is that certain aspects of the line deformation might map into multiple Fourier components, so PCA could separate them out, but then the PCA components do not in fact show an obvious interpretation (e.g. “PCA component #1 looks like rotational modulation”). So why continue on with the PCA parameterization in Sections 5.4 and 5.5, instead of returning to the chief Fourier components and analyzing those? If neither option is particularly interpretable, it seems generally better to stay closer to the data rather than adding in an extra transformation.*

First, as an unsupervised learning algorithm, one should not expect that PCA would separate physically distinct signals. It is designed to explain the most variance in the signal for a given number of components. In contrast, PCA is extremely effective at dimensional reduction. While it is challenging to visualize the behavior of a 3-vector as a function of time, it is extremely difficult to visualize the behavior of a 20-dimensional vector as a function of time.

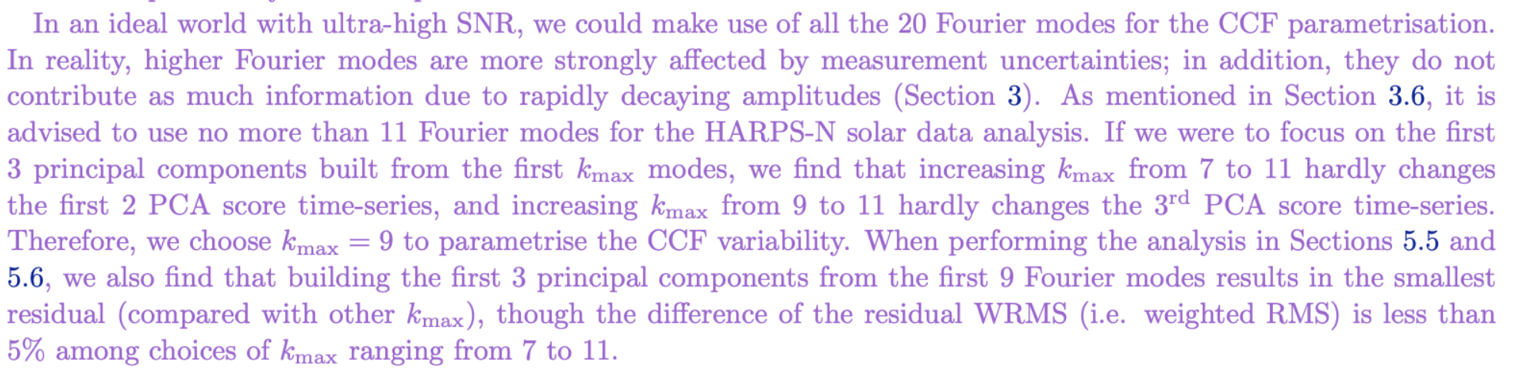
Second, as shown in Figure 2, an active region does not affect just a single Fourier component. By performing PCA, we concentrate the SNR in the leading PCA components by aggregating information from many ∆*RVk*’s into a smaller number of components. If we were to compare two analyses either a given number of ∆*RVk*’s or the same number of PCs, then the PCA approach is guaranteed to explain the same or more fraction of the variance. Indeed, we find that the PCA approach helps with the interpretation of the Fourier modes by preserving most of the periodicities with only 3 principal components instead of 20 Fourier modes.

In Section 7.2.1 we suggest that future work combining PCA with future regularisation of the PCA scores has the potential to separate out the physical processes of different timescales.

We have rephrased the text and added following in Section 5.3:



In addition, we added the next paragraph to justify the choice of *k*max for PCA.



1. *(also) Section 5.3: Why do PCA on Delta RV\_k, as opposed to something like Delta RV\_k multiplied by A\_k (which might naturally resolve the need to downweight the higher frequency modes), or both Delta RV\_k and Delta A\_k?*

As shown in Section 3.2, we can accurately propagate the uncertainties for ∆*RVk* from the CCF. Therefore, the inverse squared uncertainties provide the appropriate. Since *Ak*’s are affected by line depth variations, either multiplying by them or including them as inputs to the PCA would introduce a dependence of the PC scores on line depths and continuum normalization.

1. *Section 5.4: The filtering step here could benefit from a little bit more explanation/discussion, since at this point in the paper there are a LOT of different “frequencies” in play!*

We hope the following changes can better explain what we do with the kernel filtering.

Text

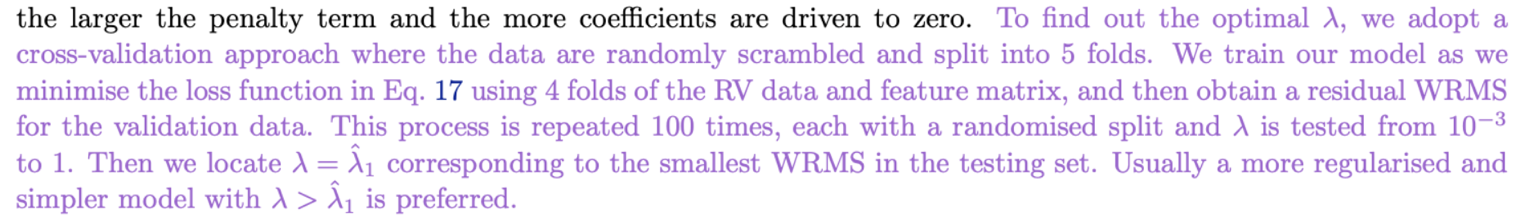
Description automatically generated

For clarity, all “Fourier frequency modes” or “frequency modes” have been rewritten as “Fourier modes” in the manuscript.

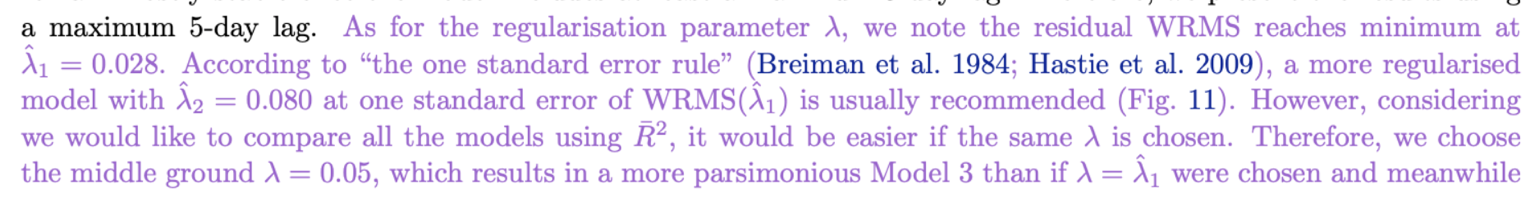
1. *Section 5.5: The justification of the regularization amplitude choice is a bit too vague. Could you use a cross-validation approach?*

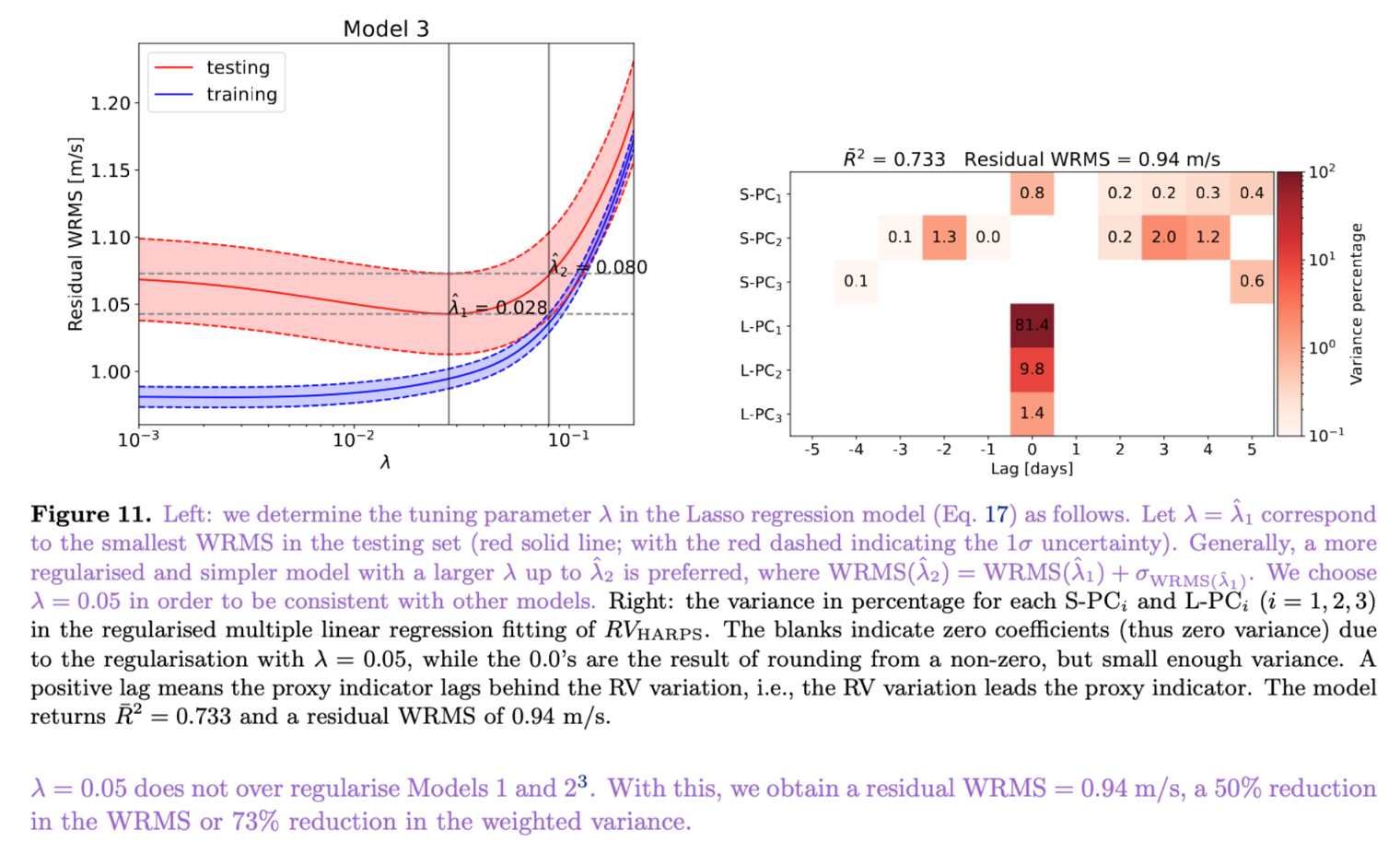
Thanks for the suggestion. We have adopted a cross-validation approach in determining the tuning parameter .

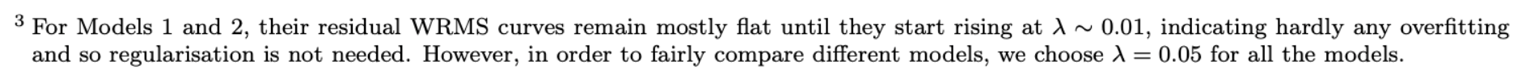
(Section 5.5)



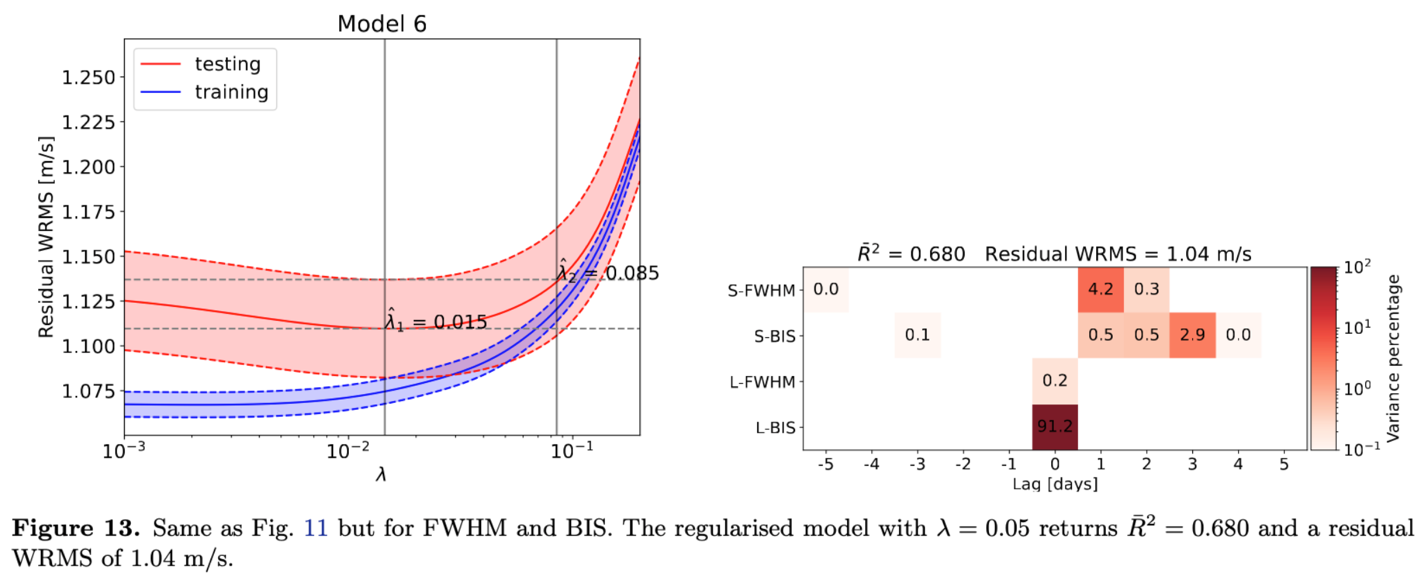
(Section 5.6)







(Section 5.7)



Here are the WRMS - plots for the other Models 1, 2, 4 and 5 for the reviewer’s reference. They are mentioned in the texts (e.g. no overfitting and so regularisation is not needed but is chosen to be 0.05 for model comparisons) but not shown in the paper.



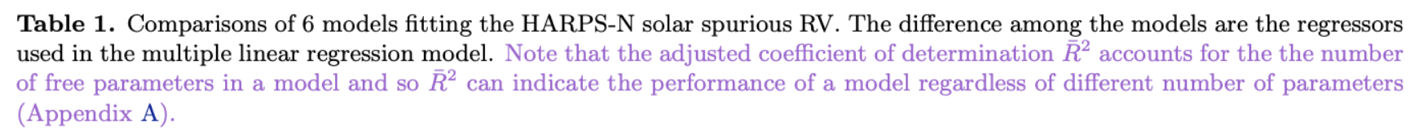


1. *Figures 11 & 13: The colorbar needs a label, and would be better remapped to log scaling to show detail.*

They are both updated as suggested (see the corresponding figures in the previous reply).

1. *Section 5.6: Can you demonstrate that the lagged multiple linear regression improves performance beyond the level expected naturally from adding extra free parameters to the model?*

We compared the performance of different models using the adjusted coefficient of determination which accounts for the number of free parameters in a model.



1. *Section 7: The summary of the core FIESTA method is very clearly stated here. I would add to it a concise description of how the FIESTA results may be used to analyze/denoise RV time series, since this aspect is where the manuscript is more difficult to follow.*

We combine the discussion and the summary session into one, which is now Session 7. The detailed approach for analysing the RV time-series is now followed by the summary of the core FIESTA method.

Some minor changes have been made.

“HARPS-N solar data — In Section 5, we applied ΦESTA to 3 years of HARPS-N solar observations (Fig. 7 and 8). As neighbouring Fourier modes may show high correlations with each other, we are motivated to use PCA for dimension reduction.”

1. *Section 7.2.2: Is the point about the potential use of temporally correlated noise models supported by the results of the lagged regression in Section 5.6? (I think the answer is yes, so it may be worth mentioning here!)*

Yes, it is. We more explicitly address the idea as follows in the new Section 7.3.2:

Text

Description automatically generated

1. *General comment: The manuscript needs language editing for British/American spelling inconsistencies (for example, normalize and normalise are both used).*

normalize 🡪 normalise

modeling 🡪 modelling

parameterize 🡪 parametrise

summarize 🡪 summarise

analyze 🡪 analyse

standardize 🡪 standardise

characterize 🡪 characterise

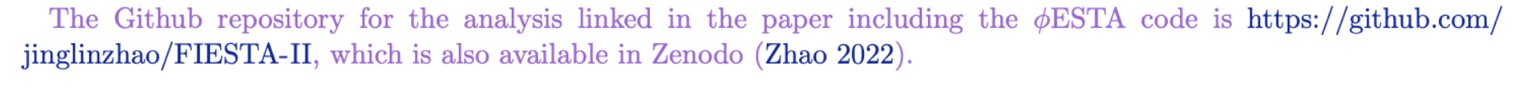
minimize 🡪 minimise

regularize 🡪 regularise

recognize 🡪 recognise

1. *Another general comment: it is great to see the Github repository for the analysis linked in the paper!*

We add the following at the end of the paper (before the appendix).



1. *We recommend that living code on github repositories place a "frozen" version on Zenodo (or other 3rd party repositories that issue DOIs) and then cite them in the article.*

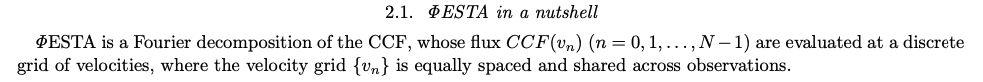
Addressed above.

1. *Can equations 1-5 be related to the 'cross-correlation theorem', a generalization of the Wiener-Khinchin theorem? See*[*https://mathworld.wolfram.com/Cross-CorrelationTheorem.html*](https://mathworld.wolfram.com/Cross-CorrelationTheorem.html)*and*[*http://www.ee.ic.ac.uk/hp/staff/dmb/courses/E1Fourier/00800\_Correlation.pdf*](http://www.ee.ic.ac.uk/hp/staff/dmb/courses/E1Fourier/00800_Correlation.pdf)*.*

Equations 1-5 are definitions of discrete Fourier transform, its inverse form and the related quantities. The form can be replaced by any one-dimensional array input and is not specific to how the spectrum cross-correlation function is derived. The input CCF has finite support and is not periodic, so thinking of Fourier transforms may help provide intuition, but is not directly applicable.

Other changes not mentioned above.

* Added co-author Chris Tinney.
* The description of the CCF velocity grid is made clear and concise at the beginning of Section 2.1.



* ­Abstract has been updated.

