MSAP3_01- Power excess detection test
PLATO-UBI-PSM-DN-0017

September 29, 2023

Contents

Prepared by	Date
Martin B. Nielsen	September 29, 2023
Emily Hatt	September 29, 2023
William J. Chaplin	September 29, 2023
Guy R. Davies	September 29, 2023
Checked by	
Approved by	
Authorized by	

Issue	Date	N° change description	Page(s)	Paragraph(s)
1.0		Initial release	All	All
1.1		Text and figure update	7, 14, 16, 18-24	–
1.2		Updated statistics in Section 3 and figures in 8.2	13-18, 20-24	–
1.3	25-02-2022	Rearranged the internal structure of the MSAP3 sub-modules.	All	All
1.4	29-09-2023	Update to the I/O descriptions	7-10	

1 Introduction

1.1 Scope of the document

This document aims to provide a description of the baseline algorithm for MSAP3_01 sub-module, which has the purpose of detecting a power excess in the AARPS due to solar-like oscillations. This document provides technical details (inputs, outputs, data types) as well as the functional description of the sub-module. The justification for the choice of this specific algorithm and a description of its scientific performance is provided in RD2. Moreover, the exact position of this algorithm within the data processing pipeline is described in RD1.

1.2 Nomenclature

Term	Description
Baseline algorithms	Initial, minimum set of algorithms needed to perform the essential scientific analysis tasks of the EAS or SAS pipelines. As they are the 'initial' (or 'baseline') set of algorithms, they could be updated later on. By 'essential' scientific analysis we mean the analysis / data processing that is necessary for fulfilling the scientific goals of Plato.
Legacy code	It is defined as a code that has been previously developed, used, and validated in contexts other than Plato. To be used in the Plato framework, it must be well-documented, extensively tested, and too heavy to be implemented from scratch by the PDC.
Prototype	Preliminary software that can help testing the algorithm, specially input and output data
Pseudo-code	a Schematic description of the algorithm or code written in a simplified language
Training set	A set of data compiled currently from short-cadence *Kepler* observations. This data set is used to determine default parameters at points in the MSAP3_01 algorithm. The data set and the derivation of these parameters is described in RD2.
Nyquist frequency	The highest sampled frequency in the AARPS
Short-cadence *Kepler* observations	Observations form the *Kepler* mission with a cadence of approximately 58 seconds.
p-mode envelope	The distribution of power in the AARPS due to solar-like oscillations. The distribution is approximately Gaussian.
Notional envelope	A predicted p-mode envelope as it would appear in the AARPS at a given central test frequency.
Power excess	Power in the AARPS that deviates significantly from the background noise level.
Unit test	A test of individual functions in a code or algorithm with the purpose of ensuring reproducibility of the code, and testing for failures in normal and limit cases.

Standard data types are defined in the following table:

Type	Size	Values
bool	1 bit	0 (false) 1 (true)
enumeration		Label 1 = value 1, Label 2 = value 2, ...
signed byte	8 bits	[-128, 127]
signed short	16 bits	[–32768, 32767]
signed int	32 bits	[-2147483648, 2147483647]
signed long	64 bits	[–9223372036854775808, 9223372036854775807]
unsigned byte	8 bits	[0,255]
unsigned short	16 bits	[0,65535]
unsigned int	32 bits	[0,4294967295]
unsigned long	64 bits	[0,18446744073709551615]
float (= float simple precision)	32 bits	3.4E +/- 38 (7 digits)
double (=float double precision)	64 bits	1.7E +/-308 (15 digits)
string	128 char (1 char = 1 byte)	

1.3 Reference documents

RD1 PLATO-LESIA-PSPM-DD-0021, Work and data flows of the stellar L1/L2 processing pipeline

RD2 Justification document for the MSAP3_01 algorithm [Nielsen et al., 2022].

RD3 PLATO-UBI-PSM-DN-0018, Description of the MSAP3_02 method.

1.4 Abbreviations

LC	light curve
PLATO	PLAnetary Transits and Oscillations of Stars
SNR	Signal-to-Noise Ratio
TBC	To Be Confirmed
TBD	To Be Defined
TBW	To Be Written
PSD	power spectral density
PSLS	Plato Solar-like lightcurve simulator
AARPS	Asteroseismic analysis ready power spectrum
AARLC	Asteroseismic analysis ready light curve

2 General overview

2.1 Name of the algorithm and status

The description below and deliverables are baseline versions of the MSAP3_01 algorithm. The output is in the form of a posterior probability as a function of frequency in the AARPS, which denotes the probability of a power excess being consistent with an p-mode envelope. In addition, quality control metrics are also provided as output.

2.2 Synopsis

The objective of MSAP3_01 is to estimate the probability that a p-mode envelope is present in the AARPS which is consistent with main-sequence and sub-giant solar-like oscillators (see definitions in RD2). This posterior density is then used by MSAP3_02 and MSAP3_03. The performance metrics and justification for MSAP3_01 are presented in RD2.

2.3 Model

MSAP3_01 evaluates the probability that a power excess in the SNR spectrum (the AARPS divided by an estimate of the background noise level) is inconsistent with noise, while at the same time being consistent with the power expected from solar-like oscillations.

This posterior probability and the associated result from MSAP3_02 are passed forward to MSAP3_03 where a decision is made for whether or not the global asteroseismic parameters can be measured. The flow-diagram of MSAP3_01 is presented in Fig. 3, see RD3 for details on MSAP3_02.

Both MSAP3_01 and MSAP3_02 produce detection probabilities for each test frequency in the AARPS. Thresholds determining a detection are set in MSAP3_03. The results are summarized as follows:

- MSAP3_01 is True, MSAP3_02 is True : Firm detection. The AARPS shows a power excess and a repeating pattern consistent with solar-like oscillations.
- MSAP3_01 is True, MSAP3_02 is False : Weak detection. The AARPS shows a power excess, potentially from an envelope, but a repeating pattern is not evident.
- MSAP3_01 is False, MSAP3_02 is True : Weak detection. The AARPS shows a repeating pattern, but the power excess is inconsistent with that expected from the scaling relations.
- MSAP3_01 is False, MSAP3_02 is False: No detection. The AARPS shows no discernible power excess or repeating pattern.

2.4 Complete list of inputs

Parameters

Name	IDP_SAS_AARPSD
Description	AARPSD[WP12PDP_I32]: Asteroseismic Analysis-ready power spectrum. Frequency-power spectra of the AARLC (AARPS), including information on the spectral window function
Reference	
Related to	Source: MSAP1

Columns

Name	Description	Type, dim	Unit
frequency	Frequency points of the AARLC power spectral density (PSD)	float64,1	μHz
power	Power points of the AARLC PSD	float64,1	ppm^2/μHz

Parameters

Name	IDP_SAS_AARPSD_METADATA
Description	Metadata for AARPSD[WP12PDP_I32]: Metadata for the frequency-power spectra of the AARLC (AARPS), including information on the spectral window function, and the duty cycle.
Reference	
Related to	Source: MSAP1

Columns

Name	Description	Type, dim	Unit
NT	Number of time stamps in the time series with real-valued flux.	int64,0	
Nyquist	The Nyquist frequency of the time series.	float64,0	Hz
amplitude	The amplitude spectrum.	float64,1	ppm
dT	The length of the time series in time.	float64,0	days
df	The frequency resolution of the amplitude, power and power density spectra.	float64,0	Hz
dt	The median sampling rate of the time series.	float64,0	days
dutyCycle	The duty cycle of the observations compared to the nominal time series length and cadence	float64,0	
fit_mean	Optional argument for the spectrum calculation to include the mean as a fit variable.	bool,0	
freqHz	Frequency points in Hz	float64,1	Hz
indx	Mask array for removing nan and/or -inf values from the time series.	bool,1	
ls	Class instance of astropy.timeseries.LombScargle	—S87,0	
normfactor	Normalization factor to use in computing the PSD	float64,0	[TBC]
power	The power spectrum.	float64,1	ppm^2
timeConversion	Conversion factor to use so that the time stamps are in seconds. Depends on the units of the time series being analyzed.	int64,0	s/day

Parameters

Name	PDP_D_SAS_TEFF_SAPP
Description	Teff SAPP: Value of Teff selected after statistical analysis: combination of values yielded from the SAPP pipeline, after validation by WP125200.
Reference	
Related to	Source: MSteSci1

Columns

Name	Description	Type, dim	Unit
PDP_D_SAS_TEFF_SAPP	Temperature established from classical observables.	float64,1	K

Parameters

Name	PDP_D_SAS_RADIUS_CLASSICAL
Description	Radius determined from classical methods (non seismic).
Reference	
Related to	Source: MSteSci2

Columns

Name	Description	Type, dim	Unit
PDP_D_SAS_RADIUS_CLASSICAL	Radius established from classical observables.	float64,0	$R_\odot$

2.5 Complete list of outputs

Parameters

Name	IDP_SAS_POWER_EXCESS_PROBABILITY
Description	Power excess probability: The merit and probability density of Nu max using MSAP3-03 to assess detection flag
Reference	
Related to	

Columns

Name	Description	Type, dim	Unit
MSAP3_01 merit	Power excess test merit values for different frequency bins across the spectrum.	float64,1	
MSAP3_01 numaxpdf	probability density function (PDF) of numax given the posterior probability of the envelope power density matching the expected scaling relations.	float64,1	
Test frequency	Frequencies in the AARPS test for the presence of p-mode oscillations.	float64,1	μHz

Parameters

Name	IDP_SAS_POWER_EXCESS_METRICS
Description	Power excess probability: The metrics and criteria used to compute the probability density of Nu max
Reference	
Related to	

Columns

Name	Description	Type, dim	Unit
Amax	Values of predicted maximum envelope amplitude as a function of test frequency with varying degrees of uncertainty.	float64,2	$ppm^2/\mu Hz$
Binned background	Binned values of the PSD background level.	float64,1	$ppm^2/\mu Hz$
Binned power	Binned values of the PSD.	float64,1	$ppm^2/\mu Hz$
H0 likelihood	The likelihoood that power in a frequency bin is consistent with the background.	float64,2	
H1 likelihood	The likelihood that the power in a frequency bin is consistent with the expectation from scaling relations.	float64,2	
H1 posterior	pseudo-posterior calculation of the probability that an envelope is observed at a test frequency.	float64,2	
Observed SNR	The binned PSD divided by the binned background level.	float64,2	
Predicted SNR	The predicted signal-to-noise ratio based on scaling relations.	float64,2	
Test frequency	Frequencies in the AARPS test for the presence of p-mode oscillations.	float64,1	μHz
Threshold prior	The probability that the envelope power will exceed a given false-alarm-probability.	float64,2	
numax prior	The calculated prior on numax based on classical observables.	float64,2	

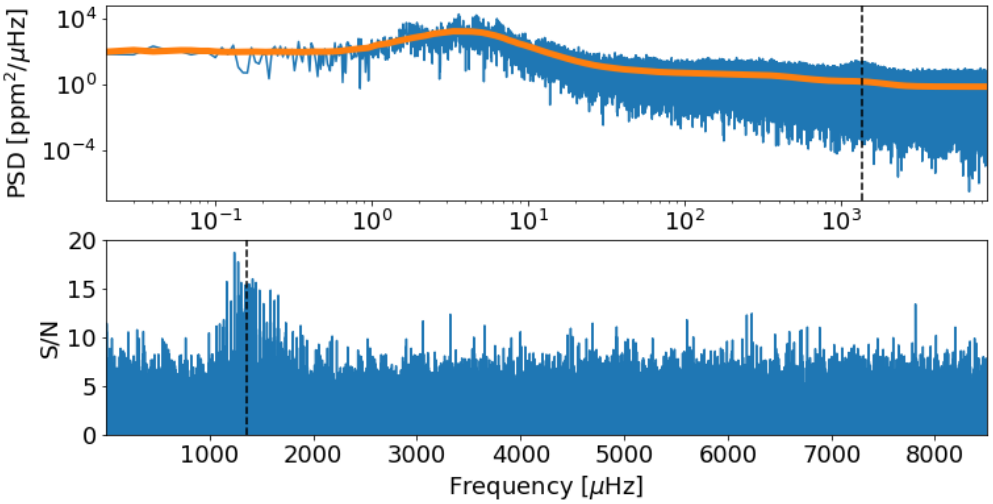

Figure 1: Top: The background estimate (orange) of the power spectral density (blue) of the star KIC1435467. Bottom: the residual or S/N of the oscillations. In both frames the dashed black line denotes the literature value of $\nu_{\rm max}$.

3 Processing description

3.1 Preparing the AARPS for analysis

MSAP3_01 uses the S/N spectrum to estimate the detection probability of a frequency bin containing power consistent with a p-mode envelope. To compute the S/N spectrum we first estimate the background noise level of the spectrum. The background consists of a frequency independent noise level (called white noise or shot noise) and a number of frequency dependent red noise terms caused by granulation and other long period brightness fluctuations on the stellar surface. We opt for a non-parametric form where the background level is approximated by computing the median power around a series of frequencies in the spectrum, evenly spaced in log-frequency from the lowest observed frequency in the spectrum to the Nyquist frequency. We let the range around each frequency, in which the median power is estimated, follow the width of a notional p-mode envelope at that frequency (see Eq. 1). A linear interpolation is then used to represent the background variation between these points. Note that this is exactly the same procedure as in MSAP3_02.

In MSAP3_01 we use the binned power spectrum, which is an average of the spectrum in discrete bins in frequency. The width of the bins is somewhat arbitrary, and in the following we chose a width of 1 μHz. High frequency resolution is not strictly required to determine the presence of solar-like oscillations with MSAP3_01, and binning the spectrum simply reduces the computational time required for the analysis, and so the amount of binning may be reduced if facilities allow. Excessive binning on the other hand will impact the precision with which $\nu_{\rm max}$ can be determined.

3.2 The Power excess probability

At each frequency in the binned AARPS, the probabilities of two hypotheses are evaluated: first, the H0 hypothesis which asks what the probability is that the power in a range around that frequency bin is consistent with the background noise level in the spectrum; second, the H1 hypothesis which asks what the probability is that the power in that same range is consistent with the power expected from scaling relations derived from previous asteroseismic observations.

Using the probability of the H0 hypothesis we can identify a power excess in the spectrum, i.e. frequency bins where the probability of H0 is low. With the H1 hypothesis we evaluate the probability that this excess stems from what we expect an actual p-mode envelope to look like, and not another source of variability in the time series. These could be other types of pulsators, exoplanet transits, rotational variability, or uncorrected systematic effects. The combination of these two probabilities is used to evaluate the posterior probability of the power in a frequency bin being due to the presence of a p-mode envelope.

3.2.1 Evaluating the H0 probability

The H0 hypothesis asks whether the power at a particular frequency is consistent with the background noise level. We evaluate this probability at all the frequencies in the binned spectrum. However, we sum the observed power in a range around each frequency, where the range is equivalent to the expected width, $\Gamma_{\rm E}$, of a p-mode envelope at that frequency. A notional p-mode envelope is expected to appear as a roughly Gaussian distribution of power, with a envelope full width at half maximum of [Mosser et al., 2012, Schofield, 2019]

$$\Gamma_{\rm E} = \begin{cases} 0.66\,\nu^{0.88} & T_{\rm eff} <= 5600K \\ 0.66\,\nu^{0.88}\,(1 + 6\times\,10^{-4}(\,T_{\rm eff} - \,{\rm T}_{\rm eff,\odot})) & T_{\rm eff} > 5600K \end{cases} \tag{1}$$

The power in the spectrum is χ^2 distributed with $N_{\rm dof} = 2N_{\rm env}$ degrees of freedom, where $N_{\rm env}$ is the number of frequency bins in a notional p-mode envelope in the unbinned AARPS. For a frequency bin ν_i with power p_i, we can then compute the likelihood of H0 given the background level b_i

$$\mathcal{L}(S_i|H0, N_{{\rm env},i}) = \frac{S_i^{2N_{{\rm env},i}-1}}{\gamma(2N_{{\rm env},i})}\exp(-S_i), \tag{2}$$

where γ is the γ-function, $S_i = \sum_{i-N_{{\rm env},i}/2}^{i+N_{{\rm env},i}/2} p_i/b_i$ is the S/N at $i = 0, 1, 2 \ldots M$, and M is the total number of frequency bins to be tested for the presence of an envelope. Since $\Gamma_{\rm E}$ increases with frequency, so does $N_{\rm env}$. We require that $N_{\rm env} \geq 1$ since for low frequencies and low resolution the number of bins may be < 1. Similarly at $\nu \approx \nu_{\rm Nyquist}$, $N_{\rm env}$ decreases as the frequency range exceeds the Nyquist frequency limit of the spectrum.

3.2.2 Evaluating the H1 probability

The H0 hypothesis is useful for detecting a power excess in the spectrum. However, we are also interested in how well this excess compares to what we expect a p-mode envelope to look like. Similar to our approach to computing the likelihood of H0, we test a range around each frequency in the binned AARPS given by $\Gamma_{\rm E}$, but now we compare the observed power with that predicted by asteroseismic scaling relations. Note that in this section, since we test each frequency as if it were $\nu_{\rm max}$, we will use $\nu_i = \nu_{\rm max}$ for clarity in relation to literature sources.

The distribution of power in the oscillation spectrum of a Sun-like star is approximated by a Gaussian, centered on $\nu_{\rm max}$, with a full width at half maximum $\Gamma_{\rm E}$ and a height $H_{\rm E}$. As shown in Eq. 1 the envelope width scales with $\nu_{\rm max}$ and $T_{\rm eff}$. Similarly, a scaling relation can be derived for $H_{\rm E}$ as is shown in detail in [Chaplin et al., 2011]. Here we will briefly cover the parts of this derivation as they relate to the MSAP3_01 algorithm.

The envelope height in the PSD spectrum is defined as

$$\frac{H_{\rm E}}{H_{{\rm env},\odot}} = \frac{\Delta\nu_\odot}{\Delta\nu}\left(\frac{A_{\rm max}}{A_{{\rm max},\odot}}\right)^2, \tag{3}$$

where $H_{{\rm env},\odot} = 0.1\,{\rm ppm}^2/\mu{\rm Hz}$ is the envelope height of the Sun, and

$$\frac{A_{\rm max}}{A_{{\rm max},\odot}} = \beta\left(\frac{L/\,{\rm L}_\odot}{M/\,{\rm M}_\odot}\right)\left(\frac{T_{\rm eff}}{{\rm T}_{{\rm eff},\odot}}\right)^{-2}, \tag{4}$$

which is the amplitude of a notional radial mode at $\nu_{\rm max}$. $L/\,{\rm L}_\odot$, $M/\,{\rm M}_\odot$, and $T_{\rm eff}/\,{\rm T}_{{\rm eff},\odot}$ are the stellar luminosity, mass, and effective temperature respectively in terms of the equivalent solar quantities[1]. The empirically determined correction factor β is given by

$$\beta = 1 - \exp\left(-\frac{T_{\rm red} - T_{\rm eff}}{\Delta\mathcal{T}}\right), \tag{5}$$

where $T_{\rm red}$ is the red edge temperature of the δ-scuti instability strip for a luminosity, L, and is given by

$$T_{\rm red} = T_{{\rm red},\odot}\left(\frac{L}{{\rm L}_\odot}\right)^{-0.093}. \tag{6}$$

[1] Where applicable we use the same solar values at in Schofield et al. [2019]

We use $T_{\mathrm{red},\odot} = 8907$K, and $\Delta\mathcal{T} = 1550$K is an empirical parameter determined by Chaplin et al. [2011]. Equation 5 corrects the predicted mode amplitude for the observed reduction of the mode amplitudes for stars where ν_{max} and T_{eff} approach the red edge of the δ-scuti instability strip [Chaplin et al., 2011]. We note that when using Eq. 5 and 6, β may become ≈ 0 for a notional envelope placed at a low test frequency ν_{red}. The predicted S/N will therefore be approximately equivalent to 1 and any resulting detection probability will appear high. However, typically $\nu_{\mathrm{red}} \sim 10\,\mu$Hz, which is much lower than the frequency range that we are concerned with for main-sequence and sub-giant stars.

Lastly in order to compute the predicted A_{max} and subsequently H_{E}, we rewrite Eq. 3 and 4 using the scaling relation for the large separation

$$\frac{\Delta\nu}{\Delta\nu_{\odot}} = \left(\frac{M}{\mathrm{M}_{\odot}}\right)^{0.5}\left(\frac{T_{\mathrm{eff}}}{\mathrm{T}_{\mathrm{eff},\odot}}\right)^{-1.5}, \tag{7}$$

follow by

$$\frac{\Delta\nu}{\Delta\nu_{\odot}} = \frac{\nu_{\mathrm{max}}}{\nu_{\mathrm{max},\odot}}^{a}. \tag{8}$$

The exponent $a = 0.791$ is determined based on the *Kepler* training set defined in RD2.

We then have

$$A_{\mathrm{max}} = A_{\mathrm{max},\odot}\beta\left(\frac{\nu_{\mathrm{max}}}{\nu_{\mathrm{max},\odot}}\right)^{-1}\left(\frac{T_{\mathrm{eff}}}{\mathrm{T}_{\mathrm{eff},\odot}}\right)^{1.5}. \tag{9}$$

Observations of A_{max} show a degree of variance around Eq. 4 [see, e.g, Huber et al., 2019]. This variance is approximately Gaussian around A_{max} at a given value of ν_{max}, and so to account for this variance in the following, we use

$$\tilde{A}_{max} = \frac{1}{\sqrt{2\pi\sigma_A^2}}\exp\frac{-(\log A' - A_{max})^2}{2\sigma_A^2}, \tag{10}$$

where A' is a parameter that allows us to marginalize over the amplitude uncertainty. We set $\sigma_A = 0.1$, but this will need to be re-evaluated as this value may not be appropriate to account for the uncertainty at all test frequencies.

We can then write the envelope height as

$$H_{\mathrm{E}} = \eta^2 D^2 H_{\mathrm{env},\odot}\left(\frac{\nu_{\mathrm{max}}}{\nu_{\mathrm{max},\odot}}\right)^{-\alpha}\tilde{A}_{max}^2. \tag{11}$$

Here we have applied two additional correction terms: η and D. The first correction, η, reduces the predicted envelope height due to the attenuation of an oscillation observed at discrete intervals, and is given by

$$\eta = \mathrm{sinc}\left(\frac{\pi\nu_{\mathrm{max}}}{2\nu_{Nyquist}}\right). \tag{12}$$

The second correction, D, is due to the dilution of the flux in the aperture of the target star caused by nearby sources such as background stars or binary companions. For an isolated star with only faint background stars $D \approx 1$, but may be less in crowded fields for example. In the following sections where *Kepler* targets are shown, we have assumed that D can be represented approximately by the crowding factor [see Thompson et al., 2016, for details].

We can now write the predicted power as

$$p_{\mathrm{pred}} = H_{\mathrm{E}}\exp\left(-\frac{(\nu - \nu_{\mathrm{max}})^2}{2c^2}\right), \tag{13}$$

where,

$$c = \frac{\Gamma_{\mathrm{E}}}{2\sqrt{2\ln 2}}, \tag{14}$$

and subsequently the likelihood of the H1 hypothesis given the ratio of the observed and predicted power

$$\mathcal{L}(R_i|H1, N_{\mathrm{env},i}) = \frac{R_i^{2N_{\mathrm{env},i}-1}}{\gamma(2N_{\mathrm{env},i})}exp(-R_i), \tag{15}$$

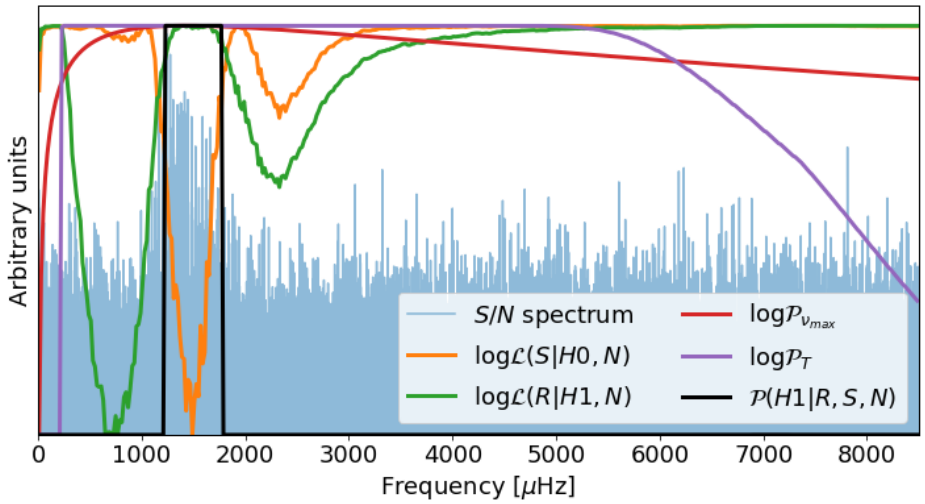

Figure 2: Illustration of the likelihoods and priors involved in computing the posterior posterior probability, $\mathcal{P}(H1|R, S, N)$, of H1 (black), for the star KIC1435467. Note that the likelihoods and priors are shown on a log-scale for clarity, while the S/N and posterior are shown on a linear scale.

where $R_i = \sum_{i-N_{\mathrm{env},i}/2}^{i+N_{\mathrm{env},i}/2} p_i/p_{i,\mathrm{pred}}$. Using equation 2 and 15 allows us to compute the posterior probability of the H1 hypothesis

$$\mathcal{P}(H1|R_i, S_i, N_{\mathrm{env},i}) = \int_A \frac{\mathcal{P}(H1)\mathcal{L}(R_i|H1, N_{\mathrm{env},i})}{\mathcal{P}(H1)\mathcal{L}(R_i|H1, N_{\mathrm{env},i}) + \mathcal{P}(H0)\mathcal{L}(S_i|H0, N_{\mathrm{env},i})} dA', \tag{16}$$

where each term has an associated prior probability, $\mathcal{P}(H0)$ and $\mathcal{P}(H1)$. Here the integral over A' marginalizes over the uncertainty in the envelope amplitude. Marginalizing over the uncertainties of the envelope width scaling relation, for example, should in principle also be carried out, but this is not yet done for this algorithm.

3.2.3 Applying priors

We propose using two priors to compute the posterior probability of H1. The first estimates the probability that the predicted power exceeds a false alarm probability threshold. This requires that we compute a S/N threshold at each frequency, which the predicted power must exceed if the false alarm probability is, say, $\mathcal{P}_{\mathrm{thr}} = 0.01$. At each test frequency we therefore solve

$$\mathcal{P}_{\mathrm{thr}} = \frac{S_{\mathrm{thr},i}^{2N_{\mathrm{env},i}-1}}{\gamma(2N_{\mathrm{env},i})} \exp(-S_{\mathrm{thr},i}), \tag{17}$$

for $S_{\mathrm{thr},i}$, where $S_{\mathrm{thr},i} = \sum_{i-N_{\mathrm{env},i}/2}^{i+N_{\mathrm{env},i}/2} p_{\mathrm{thr},i}/b_i$ is the S/N threshold that the predicted power must exceed to yield a false alarm probability less than $\mathcal{P}_{\mathrm{thr}}$. Given the predicted power by Eq. 13 we can then compute the probability that this threshold is exceeded by

$$\mathcal{P}_{\mathrm{T},i} = \int_{R_{\mathrm{thr},i}}^{\infty} \frac{R_{\mathrm{thr},i}^{2N_{\mathrm{env},i}-1}}{\gamma(2N_{\mathrm{env},i})} \exp(-R_{\mathrm{thr},i}), \tag{18}$$

where $R_{\mathrm{thr},i} = \sum_{i-N_{\mathrm{env},i}/2}^{i+N_{\mathrm{env},i}/2} S_{\mathrm{thr},i}/S_{\mathrm{pred},i}$. The effect of $\mathcal{P}_{\mathrm{T}}$ is to penalize frequency bins where the predicted power is so small that, given the background we do not expect to see power that with certainty can be attributed to a p-mode envelope.

The second prior that we apply estimates ν_{max} of the target from measurements that are independent of the power spectrum. We use Eq. 7 and Eq. 8 along with

$$\frac{\nu_{\mathrm{max}}}{\nu_{\mathrm{max},\odot}} = \frac{M}{\mathrm{M}_\odot} \left(\frac{R}{\mathrm{R}_\odot}\right)^{-2} \left(\frac{T_{\mathrm{eff}}}{\mathrm{T}_{\mathrm{eff},\odot}}\right)^{-0.5} \tag{19}$$

to obtain an approximate guess $\tilde{\nu}_{max}$

$$\frac{\tilde{\nu}_{max}}{\nu_{\rm max,\odot}} = \left(\frac{R}{\rm R_\odot}\right)^{\frac{0.5}{0.5-\alpha}} \left(\frac{T_{\rm eff}}{\rm T_{eff,\odot}}\right)^{\frac{-0.25}{0.5-\alpha}}. \tag{20}$$

We can then assign a weight to each frequency bin as

$$\mathcal{P}_{\nu_{\rm max},i} = \exp\left(\frac{-\log(\nu_i/\tilde{\nu}_{max})^2}{2\sigma^2_{\nu_{\rm max}}}\right). \tag{21}$$

The width of this distribution is conservatively set to $\sigma_{\nu_{\rm max}} = 0.5$, based on the residuals of using Eq. 20 and the training set of *Kepler* stars, which is shown in RD2.

We can then finally compute the prior probability on H1 by

$$\mathcal{P}(H1) = 0.5\mathcal{P}_{\nu_{\rm max}}\,\mathcal{P}_{\rm T}, \tag{22}$$

and thereby

$$\mathcal{P}(H0) = 1 - \mathcal{P}(H1). \tag{23}$$

This prior penalizes frequency bins where we either do not expect $\nu_{\rm max}$ to be based on independent observables, or where the predicted power of the oscillations would be too low to produce a signal of any significance. Equation 22 is scaled such that it tends to 0.5 when the joint probability $\mathcal{P}_T\,\mathcal{P}_{\nu_{\rm max}} \approx 1$. This ensures that the prior simply excludes frequencies where we do not expect the envelope to be located.

4 Type of delivery

The algorithm exists as a prototype `Python` code (see Section 6). The individual functions each have associated doc strings describing their respective inputs and outputs. An `iPython` notebook is provided with working examples of the algorithms.

5 Algorithm maturity

The algorithm concept is defined and interfaces (inputs/outputs) are stable, but not all processing steps are stable in the sense that the individual functions in the code are complete but not rigorously tested. A homogeneous testing set that spans the entire range of stars that will be observed (see RD2) needs to be developed. Furthermore, a complete set of unit tests need to be developed, and exceptions need to be caught with appropriate error handling.

6 Algorithm source

We provide the source code for the prototype algorithm presented above, along with a set of inputs and outputs for the example targets shown in section 8.

Table 1: List of delivered items

Filename	Description
MSAP3_01_Description.pdf	This document
MSAP3_01.zip	Archive containing the files below
.../MSAP3_01.py	File containing the MSAP3_01 algorithm
.../MSAP3_01_examples.ipynb	iPython notebook with examples for a few targets
.../MSAP3_01_examples.pdf	A pdf print of the examples.ipynb
.../extras.py	Additional helper functions. Not part of MSAP3.
.../inputs	Subdirectory containing example input data
.../.../KIC3656476.pow	AARPS of KIC3656476
.../.../KIC8006161.pow	AARPS of KIC8006161
.../.../KIC4465324.pow	AARPS of KIC4465324
.../.../KIC8174981.pow	AARPS of KIC8174981
.../.../sim0_04429_0000704.pow	Simulated AARPS from PSLS
.../.../sim1_04429_0006704.pow	Simulated AARPS from PSLS
.../.../sim0_02154_0005115.pow	Simulated AARPS from PSLS
.../.../sim1_02154_0011115.pow	Simulated AARPS from PSLS
.../.../example_targets.csv	Table with target input parameters
.../.../KeplerArtefactList	File with artifact frequencies in *Kepler* data. Not part of MSAP3.
.../outputs	Subdirectory with output from the example targets
.../.../KIC3656476.MSAP3_01_output.pydict	Pickled Python dictionary output for KIC3656476
.../.../KIC8006161.MSAP3_01_output.pydict	Pickled Python dictionary output for KIC8006161
.../.../KIC4465324.MSAP3_01_output.pydict	Pickled Python dictionary outputs for KIC4465324
.../.../KIC8174981.MSAP3_01_output.pydict	Pickled Python dictionary output for KIC8174981
.../.../sim0_04429_0000704.MSAP3_01_output.pydict	Pickled Python dictionary output for sim0_04429_0000704
.../.../sim1_04429_0006704.MSAP3_01_output.pydict	Pickled Python dictionary output for sim1_04429_0006704
.../.../sim0_02154_0005115.MSAP3_01_output.pydict	Pickled Python dictionary output for sim0_02154_0005115
.../.../sim1_02154_0011115.MSAP3_01_output.pydict	Pickled Python dictionary output for sim1_02154_0011115
.../pbjam	Published Python package for peakbagging.

7 Flow-diagram

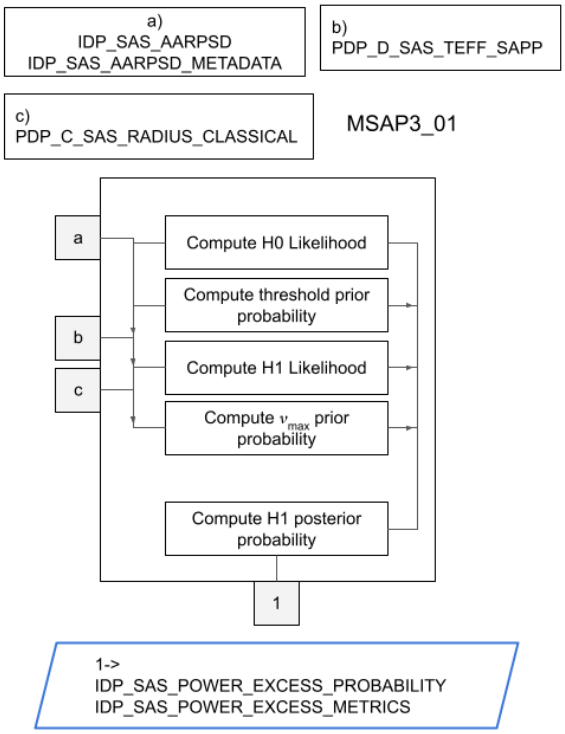

Figure 3: Summary diagram of the MSAP3_01 sub-module. Input *a* is the AARPS and related metadata from the MSAP1 module. Input *b* is the aperture flux dilution. The source module of the flux dilution is currently TBC. Input *c* is the estimated $T_{\rm eff}$, and input *d* is the radius estimate from classical parameters. Output 1 is the resulting posterior probability for $\nu_{\rm max}$ and the corresponding metrics. See RD1 for the precise location of MSAP3_01 in the MSAP3 pipeline.

8 Test Cases

Here we provide a number of test cases based on targets observed by *Kepler* and a set of simulated Plato targets. The *Kepler* targets were selected from the training set described in RD2. The simulated Plato AARPS were constructed using PSLS [Samadi et al., 2019], and a description of the *Kepler* training set is provided in RD2. We use *Kepler* and simulated Plato targets as scientific and technical test cases.

8.1 Technical test cases

The results for a number of test cases are provided in the supporting material, and are shown in section 8.2. The results are the posterior probabilities at each test frequency as computed by MSAP3_01. Unit tests for the algorithm functions are not currently implemented.

8.2 Scientific test cases

8.2.1 Kepler targets

Figures 4 – 7 show some examples of the results from MSAP3_01 for stars observed by *Kepler*.

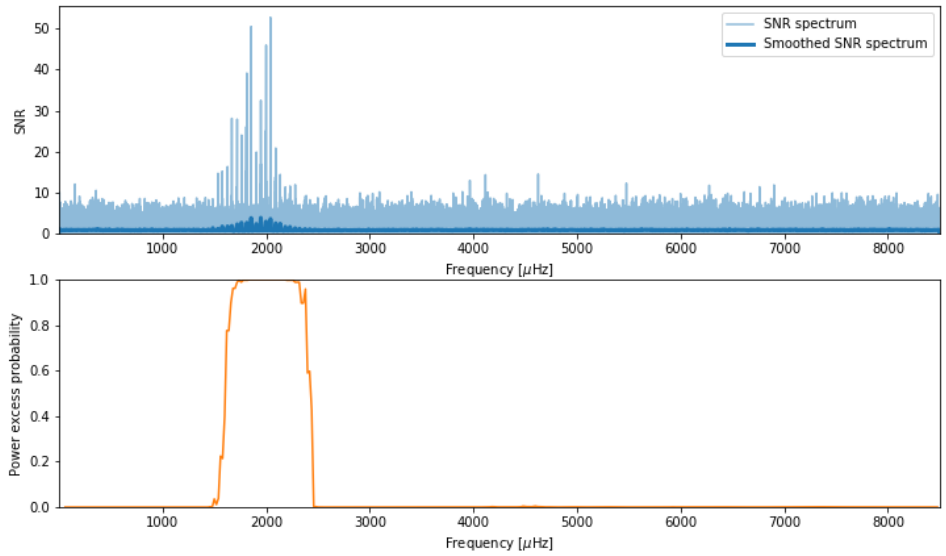

Figure 4: Top: S/N of the power spectrum for KIC3656476 in blue and the smoothed S/N in black. Bottom: The posterior probability that a power excess is inconsistent with the noise background and also consistent with the p-mode scaling relations. KIC3656476 is a young sub-giant with $T_{\rm eff} \approx 5690$K and $m_V = 9.55$, observed by *Kepler* for ~ 3 years. The power excess probability shows a clear detection.

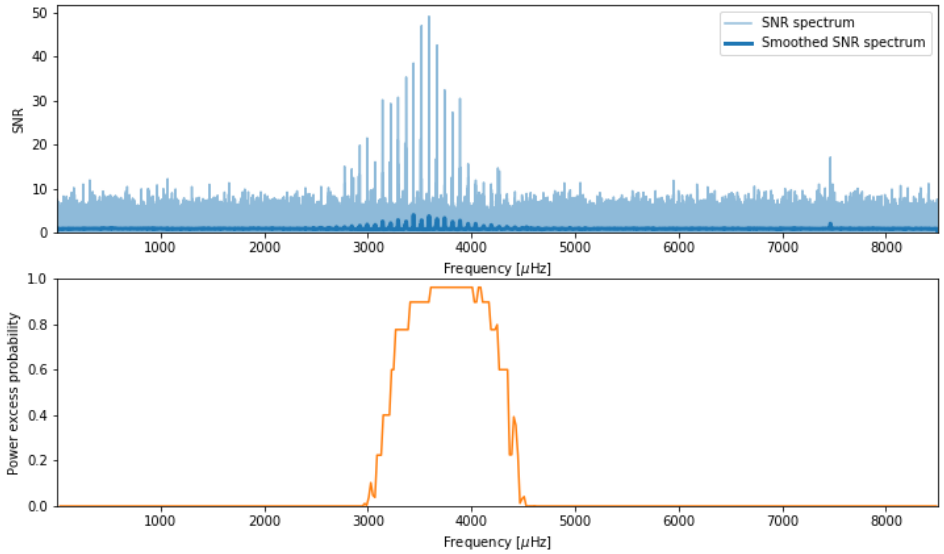

Figure 5: Same as Fig. 4 but for KIC8006161. KIC8006161 is a cool main-sequence star with an effective temperature of $T_{\rm eff} \approx 5380$K and $m_V = 7.52$, and was observed by *Kepler* for ~ 3 years. The power excess probability shows a clear detection.

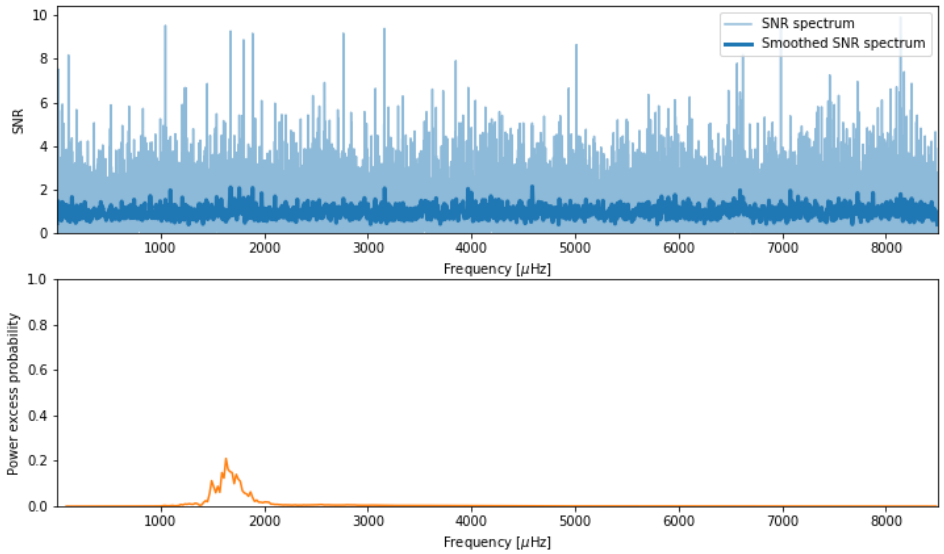

Figure 6: Same as Fig. 4 but for KIC4465324, which is slightly more evolved than the Sun, and has an effective temperature of $T_{\rm eff} \approx 5850$K and is relatively faint with an $m_V = 11.49$. It was observed by *Kepler* for 1 month. MSAP3_01 assigns a low probability of detection in this case.

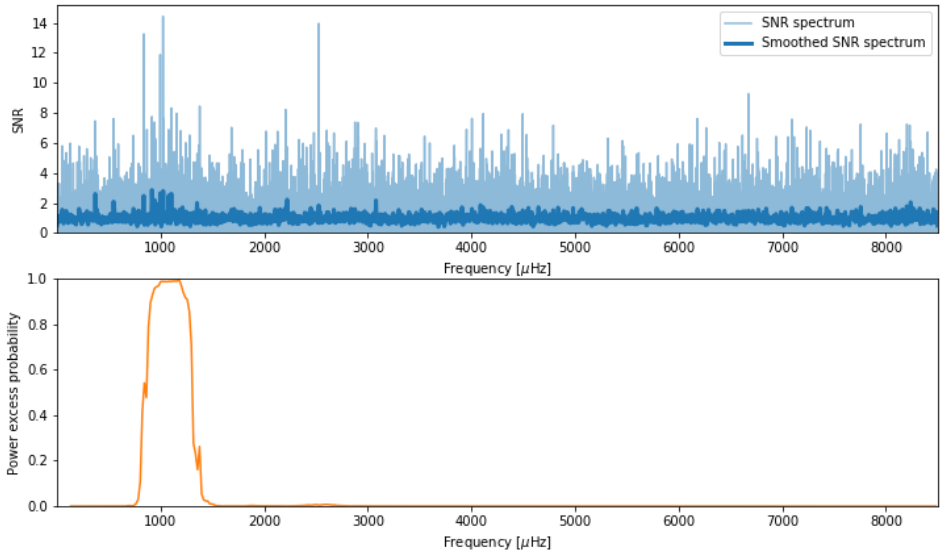

Figure 7: Same as Fig. 4 but for KIC8174981, which is an F-type main-sequence star, with an effective temperature of $T_{\rm eff} \approx 6065$K and $m_V = 10.78$. It was observed by *Kepler* for 1 month. The power excess probability shows a clear detection. A slightly elevated probability of a power excess is visible at $\sim 3000\,\mu$Hz but no large separation is detected in this range (see RD3).

8.2.2 Simulated Plato targets

Figures 8 – 11 show examples of results from MSAP3_01 for two simulated targets as they would be observed by Plato, both with and without oscillations injected. The simulations were performed using PSLS [Samadi et al., 2019].

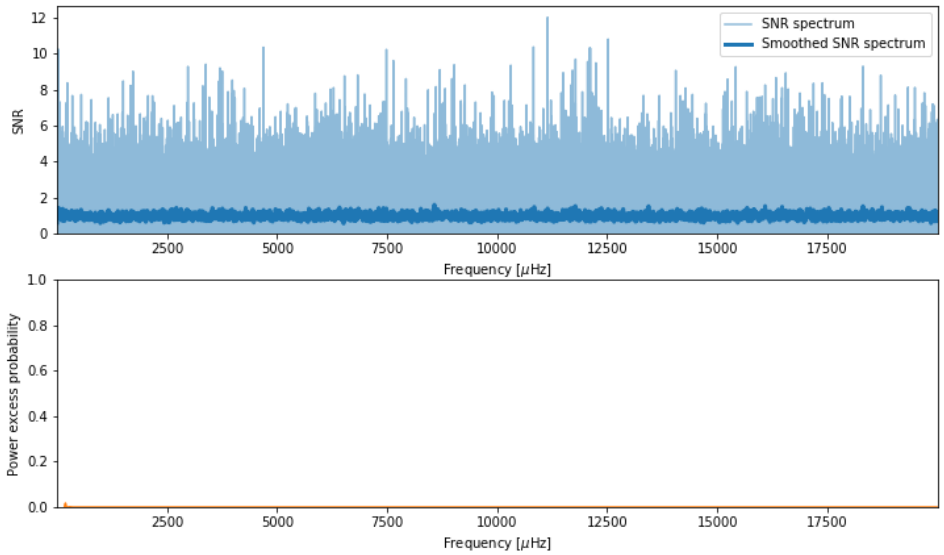

Figure 8: Same as Fig. 4 but for a simulated Plato target (sim0_04429_0000704 in Table 1) without injected oscillations, so that the spectrum only contains the red and white noise components. The simulated star has a temperature of $T_{\rm eff} = 6205$K, and a $m_V = 8.85$. The spectrum was simulated using PSLS to be that of a time series observed by Plato, with a length of 90 days.

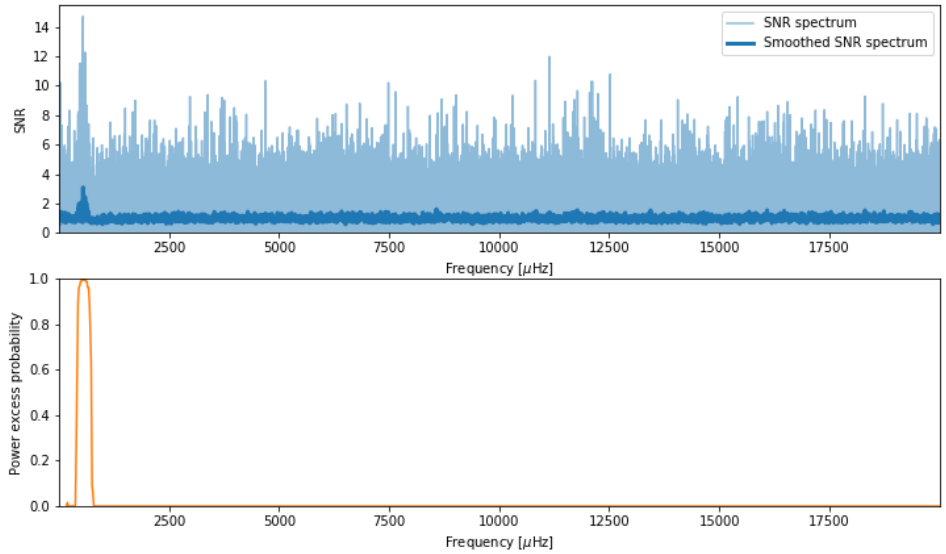

Figure 9: Same as Fig. 8 but with oscillations injected (sim1_04429_0006704 in Table 1).

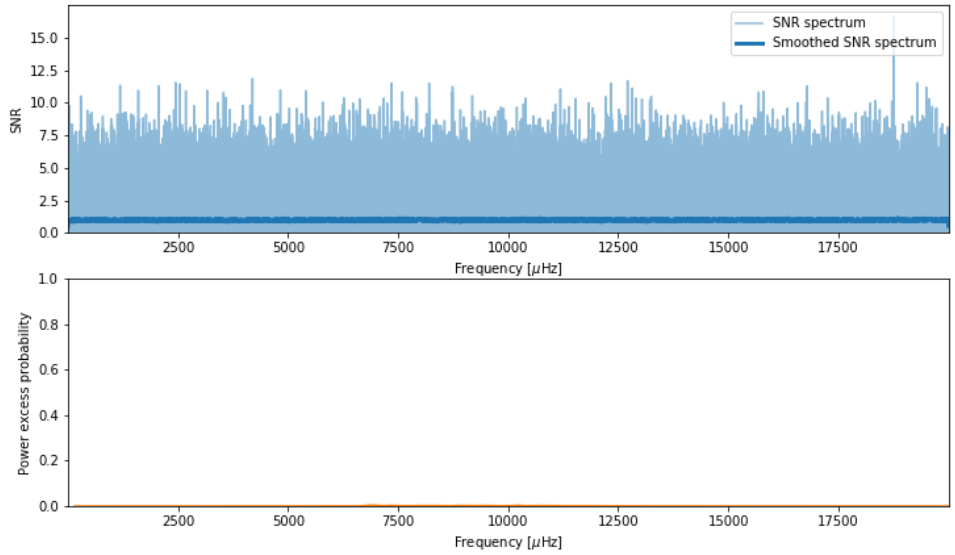

Figure 10: Same as Fig. 4 but for a simulated Plato target (sim0_02154_0005115 in Table 1) without injected oscillations, so that the spectrum only contains the red and white noise components. The simulated star has a temperature of $T_{\rm eff}$ = 6156K, and a m_V = 3.47. The spectrum was simulated using PSLS to be that of a time series observed by Plato, with a length of 2 years.

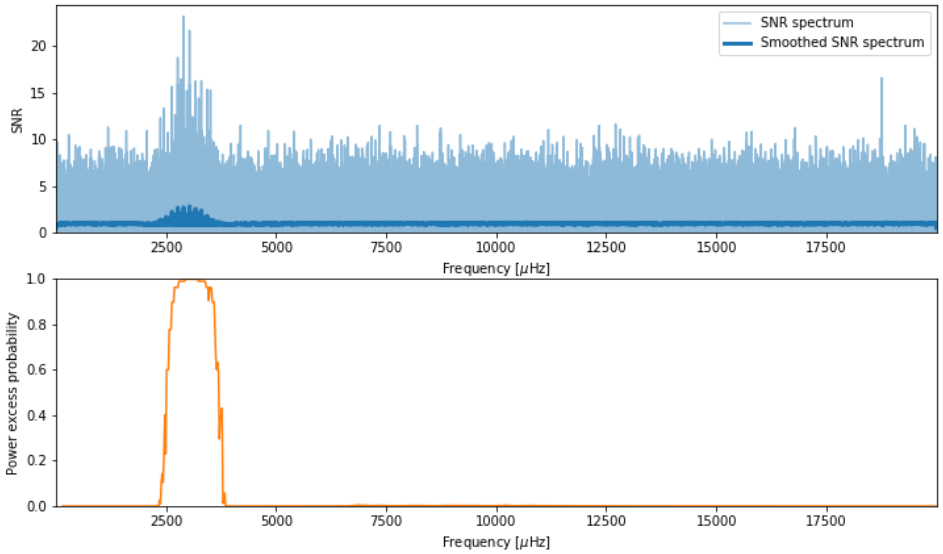

Figure 11: Same as Fig. 4 but with oscillations injected (sim0_02154_0005115 in Table 1).

References

W. J. Chaplin, H. Kjeldsen, T. R. Bedding, J. Christensen-Dalsgaard, R. L. Gilliland, S. D. Kawaler, T. Appourchaux, Y. Elsworth, R. A. García, G. Houdek, C. Karoff, T. S. Metcalfe, J. Molenda-Żakowicz, M. J. P. F. G. Monteiro, M. J. Thompson, G. A. Verner, N. Batalha, W. J. Borucki, T. M. Brown, S. T. Bryson, J. L. Christiansen, B. D. Clarke, J. M. Jenkins, T. C. Klaus, D. Koch, D. An, J. Ballot, S. Basu, O. Benomar, A. Bonanno, A. M. Broomhall, T. L. Campante, E. Corsaro, O. L. Creevey, L. Esch, N. Gai, P. Gaulme, S. J. Hale, R. Handberg, S. Hekker, D. Huber, S. Mathur, B. Mosser, R. New, M. H. Pinsonneault, D. Pricopi, P. O. Quirion, C. Régulo, I. W. Roxburgh, D. Salabert, D. Stello, and M. D. Suran. Predicting the Detectability of Oscillations in Solar-type Stars Observed by Kepler. *ApJ*, 732(1):54, May 2011. doi: 10.1088/0004-637X/732/1/54.

Daniel Huber, William J. Chaplin, Ashley Chontos, Hans Kjeldsen, Jørgen Christensen-Dalsgaard, Timothy R. Bedding, Warrick Ball, Rafael Brahm, Nestor Espinoza, Thomas Henning, Andrés Jordán, Paula Sarkis, Emil Knudstrup, Simon Albrecht, Frank Grundahl, Mads Fredslund Andersen, Pere L. Pallé, Ian Crossfield, Benjamin Fulton, Andrew W. Howard, Howard T. Isaacson, Lauren M. Weiss, Rasmus Handberg, Mikkel N. Lund, Aldo M. Serenelli, Jakob Rørsted Mosumgaard, Amalie Stokholm, Allyson Bieryla, Lars A. Buchhave, David W. Latham, Samuel N. Quinn, Eric Gaidos, Teruyuki Hirano, George R. Ricker, Roland K. Vanderspek, Sara Seager, Jon M. Jenkins, Joshua N. Winn, H. M. Antia, Thierry Appourchaux, Sarbani Basu, Keaton J. Bell, Othman Benomar, Alfio Bonanno, Derek L. Buzasi, Tiago L. Campante, Z. Çelik Orhan, Enrico Corsaro, Margarida S. Cunha, Guy R. Davies, Sebastien Deheuvels, Samuel K. Grunblatt, Amir Hasanzadeh, Maria Pia Di Mauro, Rafael A. García, Patrick Gaulme, Léo Girardi, Joyce A. Guzik, Marc Hon, Chen Jiang, Thomas Kallinger, Steven D. Kawaler, James S. Kuszlewicz, Yveline Lebreton, Tanda Li, Miles Lucas, Mia S. Lundkvist, Andrew W. Mann, Stéphane Mathis, Savita Mathur, Anwesh Mazumdar, Travis S. Metcalfe, Andrea Miglio, Mário J. P. F. G. Monteiro, Benoit Mosser, Anthony Noll, Benard Nsamba, Jia Mian Joel Ong, S. Örtel, Filipe Pereira, Pritesh Ranadive, Clara Régulo, Thaíse S. Rodrigues, Ian W. Roxburgh, Victor Silva Aguirre, Barry Smalley, Mathew Schofield, Sérgio G. Sousa, Keivan G. Stassun, Dennis Stello, Jamie Tayar, Timothy R. White, Kuldeep Verma, Mathieu Vrard, M. Yıldız, David Baker, Michaël Bazot, Charles Beichmann, Christoph Bergmann, Lisa Bugnet, Bryson Cale, Roberto Carlino, Scott M. Cartwright, Jessie L. Christiansen, David R. Ciardi, Orlagh Creevey, Jason A. Dittmann, Jr. Do Nascimento, Jose-Dias, Vincent Van Eylen, Gabor Fürész, Jonathan Gagné, Peter Gao, Kosmas Gazeas, Frank Giddens, Oliver J. Hall, Saskia Hekker, Michael J. Ireland , Natasha Latouf, Danny LeBrun, Alan M. Levine, William Matzko, Eva Natinsky, Emma Page, Peter Plavchan, Masoud Mansouri-Samani, Sean McCauliff, Susan E. Mullally, Brendan Orenstein, Aylin Garcia Soto, Martin Paegert, Jennifer L. van Saders, Chloe

Schnaible, David R. Soderblom, Róbert Szabó, Angelle Tanner, C. G. Tinney, Johanna Teske, Alexandra Thomas, Regner Trampedach, Duncan Wright, Thomas T. Yuan, and Farzaneh Zohrabi. A Hot Saturn Orbiting an Oscillating Late Subgiant Discovered by TESS. *AJ*, 157(6):245, Jun 2019. doi: 10.3847/1538-3881/ab1488.

B. Mosser, M. J. Goupil, K. Belkacem, J. P. Marques, P. G. Beck, S. Bloemen, J. De Ridder, C. Barban, S. Deheuvels, Y. Elsworth, S. Hekker, T. Kallinger, R. M. Ouazzani, M. Pinsonneault, R. Samadi, D. Stello, R. A. García, T. C. Klaus, J. Li, S. Mathur, and R. L. Morris. Spin down of the core rotation in red giants. *Astronomy & Astrophysics*, 548:A10, December 2012. doi: 10.1051/0004-6361/201220106.

M. B. Nielsen, E. Hatt, W. J. Chaplin, W. H. Ball, and G. R. Davies. A probabilistic method for detecting solar-like oscillations using meaningful prior information. Application to TESS 2-minute photometry. *Astronomy & Astrophysics*, 663:A51, July 2022. doi: 10.1051/0004-6361/202243064.

R. Samadi, A. Deru, D. Reese, V. Marchiori, E. Grolleau, J. J. Green, M. Pertenais, Y. Lebreton, S. Deheuvels, B. Mosser, K. Belkacem, A. Börner, and A. M. S. Smith. The PLATO Solar-like Light-curve Simulator. A tool to generate realistic stellar light-curves with instrumental effects representative of the PLATO mission. *Astronomy & Astrophysics*, 624:A117, April 2019. doi: 10.1051/0004-6361/201834822.

M. Schofield. *The asteroseismic potential of the NASA TESS satellite*. PhD thesis, University of Birmingham, 2019.

Mathew Schofield, William J. Chaplin, Daniel Huber, Tiago L. Campante, Guy R. Davies, Andrea Miglio, Warrick H. Ball, Thierry Appourchaux, Sarbani Basu, Timothy R. Bedding, Jørgen Christensen-Dalsgaard, Orlagh Creevey, Rafael A. García, Rasmus Handberg, Steven D. Kawaler, Hans Kjeldsen, David W. Latham, Mikkel N. Lund, Travis S. Metcalfe, George R. Ricker, Aldo Serenelli, Victor Silva Aguirre, Dennis Stello, and Roland Vanderspek. The Asteroseismic Target List for Solar-like Oscillators Observed in 2 minute Cadence with the Transiting Exoplanet Survey Satellite. *ApJS*, 241(1):12, Mar 2019. doi: 10.3847/1538-4365/ab04f5.

Susan E. Thompson, Dorothy Fraquelli, Jeffrey E. Van Cleve, and Douglas A. Caldwell. Kepler Archive Manual. Kepler Science Document KDMC-10008-006, May 2016.