Asteroseismology of δ Scuti stars: emulating model grids using a neural network

Owen J. Scutt [1]★ Simon J. Murphy [2]★ Martin B. Nielsen [1] Guy R. Davies,[1] Timothy R. Bedding [3] and Alexander J. Lyttle [1]

[1]*School of Physics & Astronomy, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK*
[2]*Centre for Astrophysics, University of Southern Queensland, Toowoomba, QLD 4350, Australia*
[3]*Sydney Institute for Astronomy, School of Physics, University of Sydney, Sydney, NSW 2006, Australia*

ABSTRACT

Young δ Scuti (Sct) stars have proven to be valuable asteroseismic targets, but obtaining robust uncertainties on their inferred properties is challenging. We aim to quantify the random uncertainties in grid-based modelling of δ Sct stars. We apply Bayesian inference using nested sampling and a neural network emulator of stellar models, testing our method on both simulated and real stars. Based on results from simulated stars, we demonstrate that our method can recover plausible posterior probability density estimates while accounting for both the random uncertainty from the observations and neural network emulation. We find that the posterior distributions of the fundamental parameters can be significantly non-Gaussian and multimodal, and have strong covariance. We conclude that our method reliably estimates the random uncertainty in the modelling of δ Sct stars and paves the way for the investigation and quantification of the systematic uncertainty.

Key words: asteroseismology – methods: data analysis – methods: statistical – stars: fundamental parameters – stars: variables: δ Scuti.

1 INTRODUCTION

Stellar ages for individual stars are notoriously difficult to measure (Soderblom 2010). One method is to model a cluster with isochrones, which is particularly sensitive to high-mass stars at the main-sequence (MS) turn-off (Lipatov, Brandt & Gossage 2022). Other techniques, such as the lithium depletion boundary (e.g. Galindo-Guil et al. 2022) or kinematics (Miret-Roig et al. 2022; Žerjal et al. 2023), are able to use low-mass stars, which are much more abundant. However, methods that utilize intermediate-mass stars for measuring stellar ages have been lacking.

Asteroseismology – the study of stellar oscillations – is highly sensitive to age and has long held promise as an independent method for age determination (e.g. Aerts 2015). Like other techniques, asteroseismology is model dependent, but the physics of those models is generally different from the high- and low-mass stars (Soderblom 2010); hence, the techniques are highly complementary (Kerr et al. 2022a, b). Until recently, however, asteroseismology of intermediate-mass stars [the so-called δ Scuti (Sct) variables] has been hampered by the difficulties in identifying which modes are excited. The discovery of regular patterns in the pulsation mode frequencies of some δ Sct stars (Bedding et al. 2020) has opened up a pathway to determine their masses, ages, and metallicities,

without the requirement that the star resides in a cluster or association.

In recent years, oscillations in large numbers of δ Sct stars have been measured using white-light photometry from space telescopes such as *CoRoT* (e.g. Paparó et al. 2013; Michel et al. 2017; Barceló Forteza, Roca Cortés & García 2018), *Kepler* (e.g. Uytterhoeven et al. 2011; Balona, Daszyńska-Daszkiewicz & Pamyatnykh 2015; García Hernández et al. 2017; Bowman & Kurtz 2018; Guzik 2021), and *TESS* (e.g. Antoci et al. 2019; Hasanzadeh, Safari & Ghasemi 2021; Barac et al. 2022; Chen et al. 2022). Observed oscillation frequencies can be compared against grids of model frequencies to find a best-fitting set of parameters (Murphy et al. 2021, 2022). It is somewhat more challenging to understand the resulting uncertainties, which are not uniquely determined by the spacing of the model grid (Pedersen 2020), and instead depend more strongly on the underlying physics (Steindl et al. 2021). Part of the challenge is that models can be computationally expensive and calculating new evolutionary tracks on the fly for Monte Carlo sampling is prohibitive.

In order to treat the uncertainties more robustly, we aim to convert a discrete grid of stellar models into a continuous function. We use a neural network to emulate a grid of stellar models that has been pre-computed over the range of expected stellar parameters. We combine the trained neural network with a Bayesian sampler to formally treat random uncertainties in the observables. This yields estimates for the posterior probability density of the fundamental properties, which quantifies their uncertainties. It also allows us to infer viable frequencies for modes that were not detected, but which might exist in the data at low signal-to-noise ratio.

★ E-mail: oxs235@student.bham.ac.uk (OJS); simon.murphy@usq.edu.au (SJM)
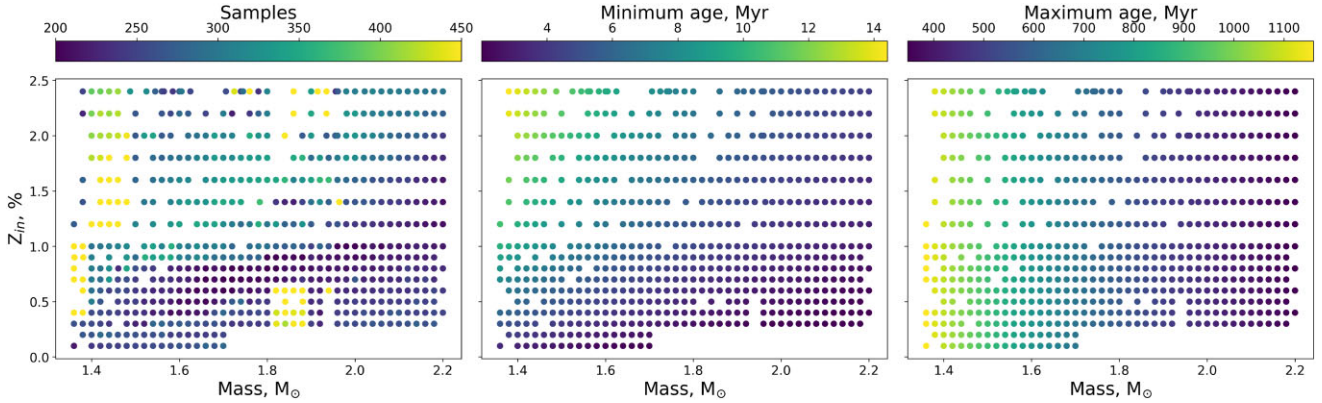
Figure 1. The model grid used in this work, where each symbol represents an evolutionary track with a particular metallicity (Z_{in}) and mass. Colour coding indicates the number of models along each track (left), and the minimum (middle) and maximum (right) ages for which pulsation frequencies were calculated (see Section 2).

In the following section, we describe the grid of stellar models on which the neural network is trained, and in Section 3 we discuss the details of the network architecture and training method. In Section 4, we present the method used to perform the Bayesian inference, and show results for a selection of simulated and real sets of observations (Section 5).

2 THE STELLAR MODEL GRID

We used the model grid described in Murphy et al. (2023), consisting of evolutionary tracks computed with MESA (r15140; Paxton et al. 2011, 2013, 2015, 2018, 2019) and pulsation models calculated with GYRE (v6.0.1; Townsend & Teitler 2013). Provisional versions of this grid have already been used to model the pulsations of *δ* Sct stars (Kerr et al. 2022a, b; Murphy et al. 2022; Currie et al. 2023), and the physics of the models are described in Murphy et al. (2022).

A well-sampled grid was needed to train the neural network emulator. Here, evolutionary tracks were spaced by 0.02 $M_\odot$ in mass M and 0.001 in initial metallicity Z_{in}. For $Z_{in} > 0.010$, the spacing was increased to 0.002. The grid is shown in mass–metallicity space in Fig. 1. A common problem in MESA is that pre-MS models sometimes fail to converge and the evolution is terminated (see e.g. Steindl et al. 2021). In such cases, we attempted to recalculate the track with a slightly increased mass ($M+ = 0.001$) up to five times before abandoning that track. Abandoned tracks appear as gaps in the grid in Fig. 1.

It is also important to ensure that the tracks are sampled well in age. Computational errors are minimized by keeping the time interval small throughout the evolution, even if not all time-steps are saved as outputs. The internal sampling is described in Murphy et al. (2023). For outputs, we saved evolutionary and pulsation models every 0.05 Myr from 2 until 10.5 Myr, in order to adequately sample the rapid evolutionary changes that occur on the pre-MS. After this, the evolution is somewhat slower, and sampling of 3 Myr was deemed adequate up to ages of 40 Myr. Beyond this, the tracks were instead sampled according to changes in position on the HR diagram (limits of $\Delta \log T_{eff} = 0.0006$ and $\Delta \log L = 0.002$), with an upper limit of 100 Myr between samples. The minimum and maximum ages sampled for each track are shown in Fig. 1. Where large gaps occurred in the grid, or when the specific M–Z_{in} combination demanded it, we manually recalculated tracks with finer sampling. This explains the variations in the number of samples per track in Fig. 1. The resulting

grid spans classical observable ranges of $\log L = 0.3$–1.8 dex and $T_{eff} = 6000$–14 000 K.

For each pulsation model, we computed the frequencies of radial modes (spherical degree $\ell = 0$) having radial orders n from 1 to 11, and dipole ($\ell = 1$) modes having $n \sim 1$–10. This encompasses the range of radial orders observed for real stars (e.g. Bedding et al. 2023), with oscillation frequencies spanning $\nu = 4$–150 d^{-1}. We calculated the mean frequency separation between radial orders ($\Delta \nu$) using the radial modes having $n = 5$–9, by fitting a straight line to the mode frequencies as a function of n (see White et al. 2011), using

$$\nu = \Delta \nu (n + \ell/2 + \epsilon). \tag{1}$$

The variable ϵ is the intercept of that line with the y-axis, and describes the distance of the radial mode ridge from the y-axis in an échelle diagram. In addition to the individual mode frequencies, we stored the values of $\Delta \nu$ and ϵ for each model in the grid, since these asteroseismic quantities relate to astrophysical quantities (Murphy et al. 2023).

To reduce the effect of the strong covariance between stellar age τ and mass M, and ease the training of the neural network, we used the assumption that the MS lifetime is approximately proportional to $M^{-3.2}$ and defined the scaled age (e.g. Davies & Miglio 2016)

$$\mathcal{K} = 10^{-4} \tau \, (M/M_\odot)^{3.2}. \tag{2}$$

This scaled age serves as an estimate of the fractional MS age of our models.

3 CONSTRUCTING THE NEURAL NETWORK

To overcome the discretely sampled nature of the model grid, we used a neural network consisting of a series of fully connected dense layers in place of standard interpolation for continuous stellar model emulation. The network was trained on the model grid, learning to predict observable parameters given stellar model input parameters. This way, the network learned the map from model parameters to observables and could be used for likelihood estimation during inference. To this end, we used the fundamental parameters M, Z_{in}, and scaled age ($\mathcal{K}$) as inputs for parameter augmentation and network training. Outputs consist of the classical observables (L and T_{eff}), asteroseismic quantities ($\Delta \nu$ and ϵ), and 11 radial and 10 dipole mode frequencies. We refer to these 25 outputs collectively as the 'observable parameters'.
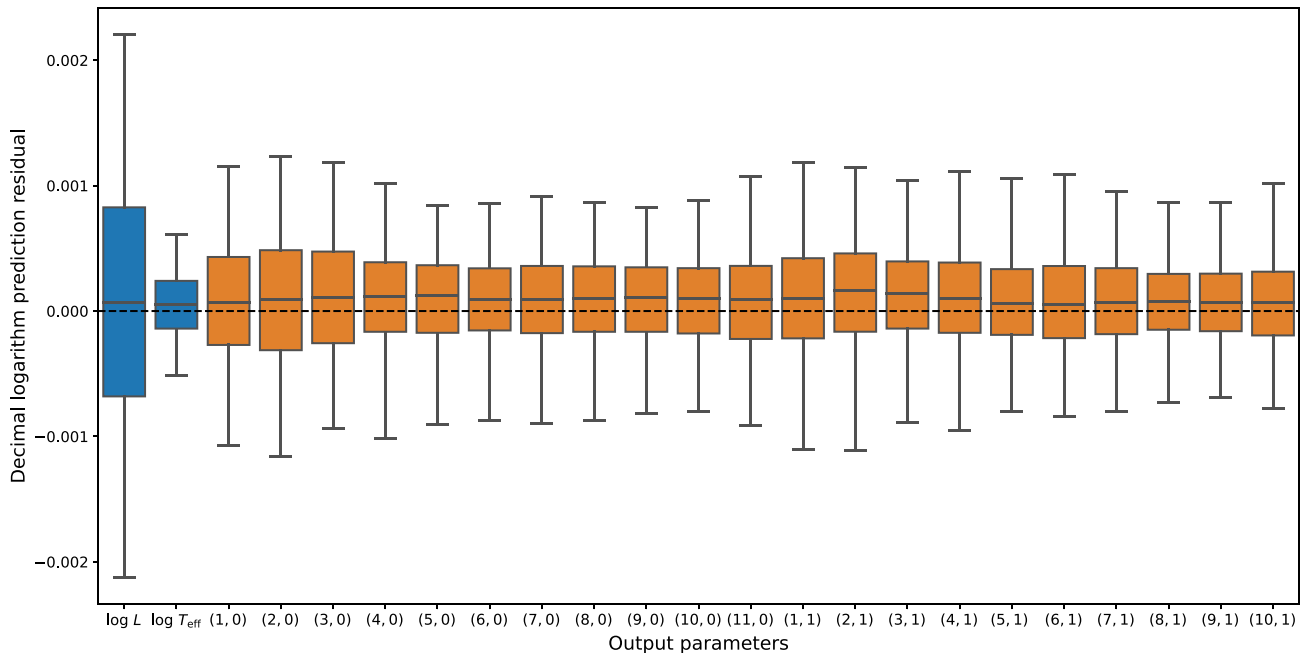
Figure 2. Residual distributions of predictions from the neural network. The residuals are that of the decimal logarithms of $T_{\rm eff}$, L (blue), and mode frequencies (n, ℓ) (orange). The boxes show a central line at the median value of the distribution, with edges at the lower and upper quartiles. Whiskers extend to the 5th and 95th percentile range. The dashed line indicates complete agreement between the network predictions and model grid values.

Once the input and output parameters were defined, we carried out data set-wide parameter augmentation to improve the training of the network. We converted all parameters (excluding ϵ) to the decimal logarithm and applied a Z-score standardization to all parameters (including ϵ). Both of these operations restricted all parameters to similar ranges, to avoid the neural network assigning erroneously high importance to parameters spanning several orders of magnitude during training. We found the combination of the two operations to be optimal for this investigation.

To further simplify the training process, we adopted a process of pre-training dimensionality reduction and in-network dimensionality reprojection similar to that described in Mancini et al. (2022). We performed principal component analysis on the observable parameters in the model grid, as follows. For all models, we calculated the covariance matrix of all 25 observable parameters. The eigenvectors of the resulting covariance matrix, or 'principal components', were ranked in order of descending eigenvalue, returning a list of principal components explaining the most to the least variance in the observable parameters. We determined how many principal components to include using the explained variance ratio, which describes the percentage of the variance of the observable space present in just the chosen principal components. We found that nine principal components explained all but 10^{-4} per cent of the total variance. This sufficiently explained the covariance of the 25 parameters in the full observable space.

The use of principal components presents the neural network with a simpler map to learn – replacing the fundamental parameters by the reduced dimensions of the 'latent parameters' – and also removes covariance information from the observables, which is redundant for the neural network. We then added a custom non-trainable layer to the neural network, which projects the latent parameters back into the full observable parameter space before the network outputs predictions.

Finally, we split the model grid into a training and a testing set for the neural network. The training set was randomly selected to comprise 80 per cent of the model grid, to be seen by the network during training. The test set, composed of the remaining 20 per cent of the grid, was unseen by the network during the training process and was used solely for evaluation of network prediction performance. This served as a check that the network is capable of model grid interpolation – the training set became a sparser representation of the original model grid, with the test set providing models guaranteed to hold combinations of parameters previously unknown to the network.

In addition to the data augmentation prior to training, the hyperparameters of a neural network can be tuned to promote faster and more stable learning. To quantify network performance for comparison between different hyperparameter permutations, we compared their validation loss profiles over multiple network training sessions. We adopted a 'grid search' method for testing potential combinations of network hyperparameters. This involved creating a grid of potential values for the number of fully connected dense layers (ranging from three to eight in steps of 1), activation functions, optimizers, learning rates, loss functions, and batch sizes. We populated a grid with these hyperparameters, and then tested the resulting network at each position in the hyperparameter grid for successful validation loss minimization.

We found the optimal network consisted of six fully connected dense layers of 64 neurons, each using an exponential linear unit activation function (Clevert, Unterthiner & Hochreiter 2015), followed by the custom layer for projection from latent to observable parameters, and a final dense output layer with linear activation function. We used the ADAM optimizer (Kingma & Ba 2014) with a learning rate of 10^{-4}, and the mean squared error loss function. For our training set of 1.4×10^5 models, a batch size of 6×10^4 models provided a good compromise between speed and training stability. We used a validation split of 25 per cent of the training
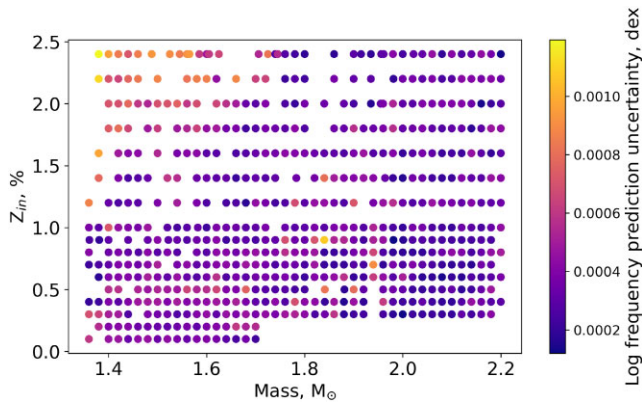
Figure 3. Stellar model grid in model parameter space, coloured according to mean decimal logarithm of network uncertainty on mode frequency prediction averaged over each track.

set – where the test set is used to evaluate neural network success after training, a 'validation set' is used for evaluation of neural network success during training. Once primary training was complete with the learning rate above, we saved and recompiled the network with a slower but less volatile learning rate of 7×10^{-5}, and restarted training until no validation loss reduction was observed for 10^4 training epochs. The network and custom latent-to-observable projection layer were constructed using the TENSORFLOW sequential API (Abadi et al. 2015).

Once the optimal network from the grid search was trained, we evaluated the network performance across the full observable parameter space. Using the test set previously removed from the model grid, we plotted distributions of the decimal logarithm prediction residuals for each parameter. This allowed us to visualize any bias and uncertainty inherent in the neural network predictions. We used the median absolute deviation of these prediction residual distributions, shown in Fig. 2, to quantify network prediction uncertainty for observable parameters. We found an uncertainty in network predictions of 8×10^{-4} dex and 2×10^{-4} dex for $\log L$ and $\log T_{\rm eff}$, respectively. We found a mean prediction uncertainty of $\sim 3 \times 10^{-4}$ dex for the decimal logarithm of the mode frequencies averaged over the entire test set. As shown in Fig. 3, neural network prediction uncertainty increased into the low mass, high metallicity domain. While the vast majority of the emulation uncertainty over tracks are well explained by our adopted uniform value, we note that neural network prediction uncertainty could be improved by better populating this region of the grid with more models for training.

4 INFERRING THE FUNDAMENTAL STELLAR PARAMETERS

To perform the Bayesian inference on the input model parameters, $\theta = \{M, \log Z_{\rm in}, \log \mathcal{K}\}$, for a given observed set of mode frequencies, we sampled the posterior distribution

$$P(\theta|D) = \frac{\mathcal{P}(\theta)\mathcal{L}(D|\theta)}{\mathcal{E}(D)}. \tag{3}$$

Here, $\mathcal{P}(\theta)$ is the prior on the input model parameters, and $\mathcal{L}(D|\theta)$ is the likelihood of observing a set of parameters (D), given the model parameters. The evidence, $\mathcal{E}(D)$, is calculated at each step during the sampling.

In addition to the input model parameters, we also included a variable offset term, Δn, as input to account for the possible ambiguity in assigning the radial orders of the observed modes. This ambiguity arises because the radial orders of a set of modes cannot always be determined from the observed mode frequencies alone, and are typically decided by comparison to stellar models. Including Δn in the sampling allows us to marginalize over this uncertainty when estimating the posterior distribution of the input model parameters. We expect this uncertainty to only lead to an error of ± 1 radial order, and so we chose the prior on Δn to be a set of δ-functions at $\Delta n = -1$, 0, and 1.

Table 1 lists a summary of the prior functions. For the priors on M, $\log Z_{\rm in}$, and $\log \mathcal{K}$, we chose to use β distributions, since they can be bounded to match the limits of the stellar model grid. In addition, the shape parameters of the β distributions may be chosen such that the priors reflect our expectation of the distribution of real observations of M, $Z_{\rm in}$, and $\mathcal{K}$.

Our choice of range and shape of the prior on $\mathcal{K}$ is motivated by the age range and distribution we expect to target, and also be able to observe. Mode identification for δ Sct stars is currently possible up to approximately one third of the MS age, after which the coupling between the buoyancy dominated and acoustic modes spoils the regular mode frequency patterns. Furthermore, at older ages the physics of mixing and overshooting become more important, and these were not treated as variables in the model grid. Hence, the prior on age extends to approximately one third of the expected MS lifetime. The lower limit on the prior on the scaled age was chosen because stars in our mass range of interest (see below) do not evolve to cross the δ Sct instability strip until ages ≥ 2 Myr. Due to the motion of stars through the δ Sct instability strip, the age prior is biased towards lower ages, with a fall-off in the age prior distribution towards older stars. We note that a minority of points defined by the prior on $\log \mathcal{K}$ do fall outside of the range of the grid. We define a β distribution as our prior on $\log \mathcal{K}$ with upper and lower limits of -3 and -0.3, respectively, whereas the model grid has a range of -3 to -0.36. This is not of concern as this high region of $\log \mathcal{K}$ is extremely underpopulated – for example, the range $-0.4 < \log \mathcal{K} < -0.3$ contains just 0.4 per cent of points in the entire grid. Furthermore, the neural network is capable of emulating points that lie outside the ranges of the training grid. In this regime, interpolation becomes extrapolation, and predictions are made with increased uncertainty. The prior region that lies outside of the grid is rapidly diminished during inference, and thus should not have bearing on the resulting model parameter posteriors.

The prior on $Z_{\rm in}$ ranges from approximately 0.07 to 1.5 times the solar metal mass fraction of 1.42 per cent used in the models (Asplund et al. 2009), which covers the metallicity distribution of stars forming within approximately 1 kpc of the Sun at the current age of the Galaxy (Hayden et al. 2020). The existence of metal-poor δ Sct stars in modern star-forming regions (e.g. HD 139614 in Upper Centaurus–Lupus; Murphy et al. 2021) suggests that slightly subsolar metallicities are more common than slightly supersolar metallicities in young δ Sct stars. We therefore skewed the prior probability density towards subsolar values.

Finally, the mass range was chosen to ensure that models exist both within and on either side of the instability strip (Dupret et al. 2004; Murphy et al. 2019). Our slight skew towards lower masses accounts for the similar skew present in the stellar initial mass function (Krumholz 2014).

Fig. 4 shows samples drawn from these prior density distributions, both in terms of the sampled variables and those transformed to M, $Z_{\rm in}$, and age. These priors are applied to the inference performed for

Table 1. Prior density functions used in equation (3). The priors on M, $\log Z_{\text{in}}$, and $\mathcal{K}$ are given by β_b^a, where a and b are the shape parameters of the β-distributions, and the prior on Δn is a series of δ-functions at integer values. In all cases, the arguments to the distribution functions denote lower and upper limits.

Parameter	Prior function
$M[\text{M}_\odot]$	$\beta_3^2(1.3, 2.3)$
$\log Z_{\text{in}}$	$\beta_2^6(-3.1, -1.6)$
$\log \mathcal{K}$	$\beta_{1.2}^2(-3, -0.3)$
Δn	$\in_R\{-1, 0, 1\}$
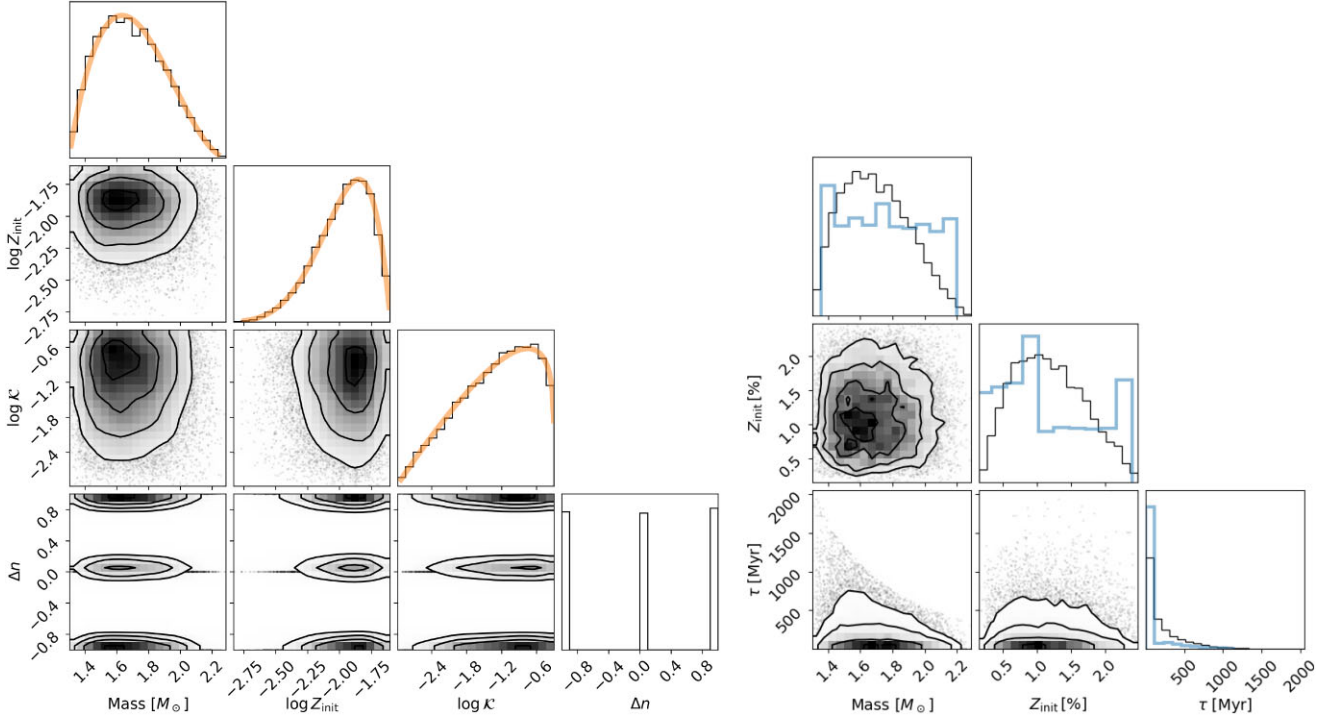
Figure 4. Left: Samples drawn from the one-dimensional prior distributions of the sampled parameters. The diagonal frames show the marginalized distributions (black) and the functions used to draw the samples (orange). The priors used for Δn are δ-functions at $\Delta n = -1, 0$, and 1. The off-diagonal frames show the two-dimensional distributions of the input parameters. Right: The samples from the left frames (black) transformed to the same units as the output from the stellar model grid (blue).

all targets (see below). Additional priors may be applied on a target-by-target basis if, for example, the mass can be constrained by other sources such as orbiting companions, or limits can be placed on the metallicity by spectroscopy.

For each of the samples drawn from the prior distributions, the neural network produces the following outputs: a set of mode frequencies, the effective temperature, and the luminosity. Given a set of outputs, we then evaluated the likelihood of the observations by

$$\log \mathcal{L}(D|\theta) = \log \mathcal{L}(D_{\text{S}}|\theta) + \log \mathcal{L}(D_{\text{C}}|\theta). \quad (4)$$

We separate the log-likelihood into the seismic variables, D_{S}, and the classical (non-seismic) variables, D_{C}. The contribution to the likelihood of the mode frequencies is given by

$$\log \mathcal{L}(D_{\text{S}}|\theta) = \sum_i \log \mathcal{N}\left(\nu_i^{\text{obs}}, \sqrt{\sigma_{\nu_i^{\text{obs}}}^2 + \sigma_{\nu_i^{\text{NN}}}^2}\right), \quad (5)$$

and that of the classical observables is given by

$$\log \mathcal{L}(D_{\text{C}}|\theta) = \log \mathcal{N}\left(\log L^{\text{obs}}, \sqrt{\sigma_{L^{\text{obs}}}^2 + \sigma_{L^{\text{NN}}}^2}\right)$$
$$+ \log \mathcal{N}\left(T_{\text{eff}}^{\text{obs}}, \sqrt{\sigma_{T^{\text{obs}}}^2 + \sigma_{T^{\text{NN}}}^2}\right). \quad (6)$$

The width of the probability densities used in the inference is given by two terms that specify the observational uncertainty (superscript 'obs'), and the noise due to the precision of the neural network's ability to emulate the model grid (superscript 'NN'). Based on the spread of the residuals presented in Figs 2 and 3, this emulation uncertainty is approximately 4×10^{-4} dex, which equates to a relative uncertainty of ≈ 0.1 per cent on the output parameters. This additional uncertainty was added in quadrature to the uncertainty of the observed mode frequencies.

In the following, we will use simulated frequencies corresponding to those obtained from two sectors of data from the *TESS* mission (Ricker et al. 2015). We therefore adopted an uncertainty on the
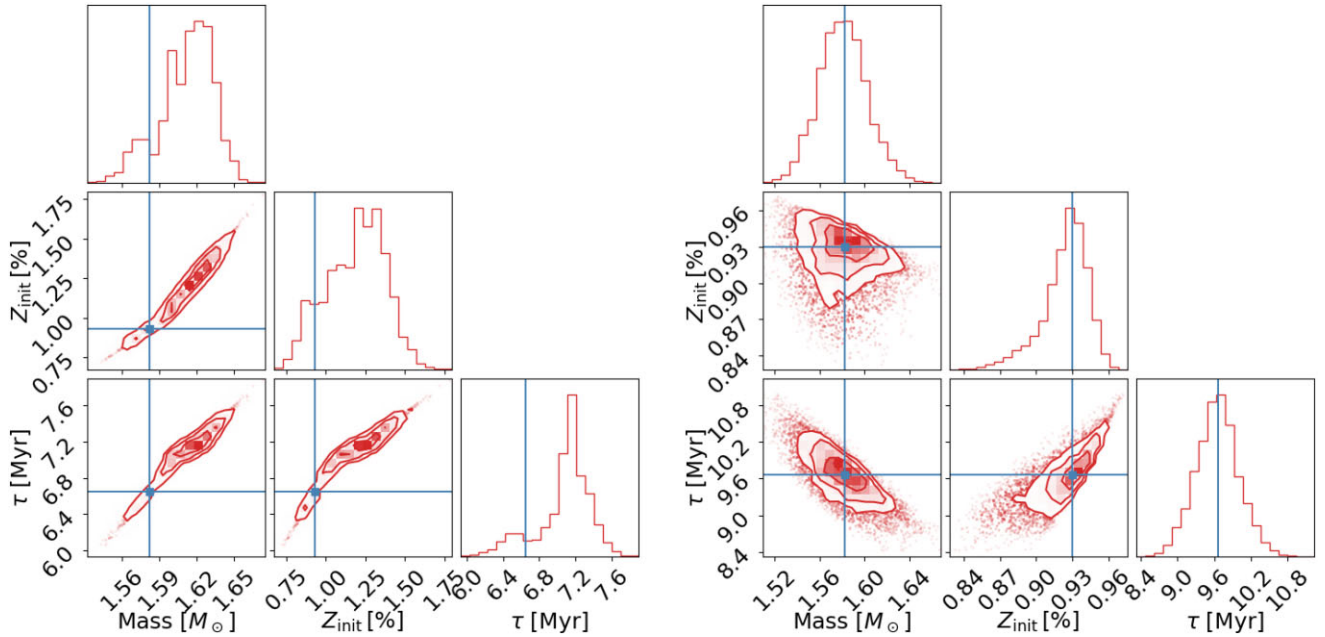
Figure 5. Samples drawn from the posterior distributions of exemplar simulated stars 5 (left) and 6 (right). For clarity, we transform the initial metallicity to per cent and $\mathcal{K}$ to stellar age τ. The model input values used to generate the simulated star data are shown in blue.

mode frequencies of $0.02\,\mathrm{d}^{-1}$, which is the frequency resolution of the resulting power spectra. The uncertainties on $\log L$ and $\log T_{\mathrm{eff}}$ depend on the target in question, but for the simulations shown below these were fixed to $\sigma_L = 0.05$ dex and $\sigma_T = 200$ K.

The neural network residuals also showed a bias of $\sim 10^{-5}$ dex, which equates to an offset of 0.01 per cent on each of the output parameters. This offset is small compared to the combination of the observed and neural network uncertainties, and so we did not consider it in the analysis. However, if either of these sources of uncertainty were decreased by, for example, improving the estimates of the observed mode frequencies, the importance of this bias would need to be re-evaluated.

We performed the sampling using the nested sampling method from the DYNESTY PYTHON package (Skilling 2004; Speagle 2020). Nested sampling determines iso-likelihood contours in the input parameter space, which were iteratively redefined until samples were consistently drawn around the global likelihood maximum. In the DYNESTY package, this process is terminated when the change in model log-evidence $\Delta \log \mathcal{E}$ is less than a pre-defined value chosen according to the DYNESTY documentation.[1] The method presented above is not restricted to using DYNESTY, and other sampling methods may be used, such as MULTINEST (Buchner et al. 2014) or EMCEE (Foreman-Mackey et al. 2013).

5 RESULTS

5.1 Simulated stars

In order to test our methodology and validate the accuracy of the neural network emulator, we have performed tests based on 25 simulated stars in a 'hare-and-hounds' exercise. To produce these simulated stars, we proceeded as follows. Values of stellar mass and initial metallicity were selected to lie in between values in the

grid, but still within the defined parameter range of the grid. We calculated stellar models and pulsation frequencies using MESA and GYRE, using the same settings as for the grid. We selected ages from the newly calculated tracks, which then defined the truth values for mass, metallicity, and age for our simulated stars and their associated 'true' observables. We only selected a subset of the calculated modes, to better reflect typical observations of δ Sct stars. We selected modes at four consecutive radial orders for each degree, within the bounds of $n = 1$–8.

To simulate noisy observations, we added noise to the observable parameters of the simulated stars. These random offsets were drawn from a normal distribution, with mean of zero and a standard deviation of: $0.02\,\mathrm{d}^{-1}$ for the mode frequencies, 200 K for the effective temperatures, and 0.05 dex for the log luminosity.

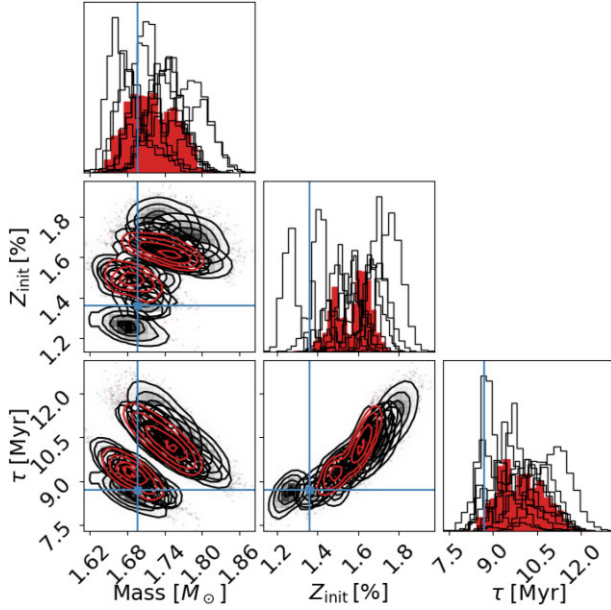#### 5.1.1 Exemplar simulated stars

Fig. 5 shows the posterior probability estimates for simulated stars with some of the most promising results. The posterior samples for simulated star index 5 demonstrate our ability to quantify random uncertainty on our inferred properties from the posterior and shows that the true properties of this simulated star lie comfortably with the posterior distribution. In this case our method is performing as required but it is worth noting that, even for this exemplar, the posterior still contains significant correlation between the parameters.

We see that the posterior distribution is not well described by a series of separable one-dimensional normal distributions. Instead, there are strong covariances between inferred parameters, which are to be expected from stellar evolution theory. However, the one-dimensional marginal posterior distributions show evidence of not being normally distributed and, in the case of the stellar age, even somewhat multimodal. To examine the degree of accuracy and precision of the results, we will study the summary statistics of the posterior distribution. This does not capture all the detail that is of value, but is none the less useful as a test of our methods.
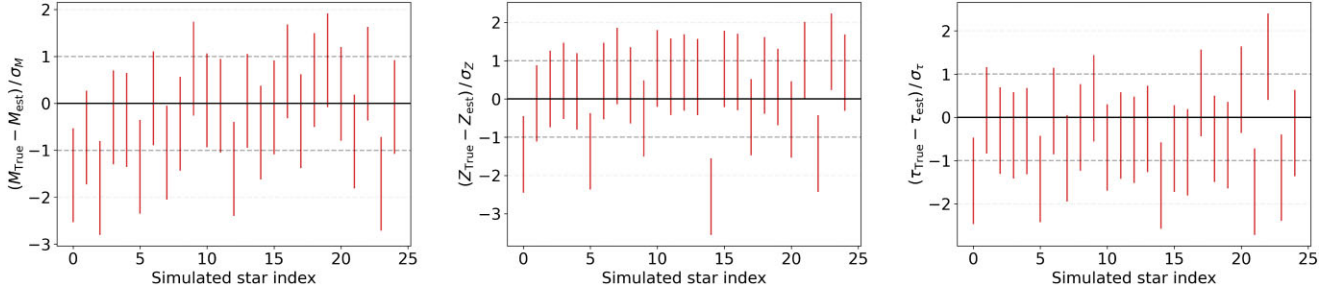
Figure 6. Difference between the inferred and model values of M, Z_{in}, and age relative to the uncertainty of the inference, for a set of 25 simulated stars. The uncertainty is taken as the standard deviation of the marginalized posterior distributions of each of the parameters (see Section 5.1.2 for details).

For each parameter of each star, we computed the difference between the truth value and the inferred value (the mean of the posterior samples for that parameter) divided by the uncertainty (the standard deviation of the posterior samples for that parameter). If our inference is perfect, and our posterior distributions are well behaved, then this metric should be drawn from a normal distribution with zero mean and unit variance. However, multimodality, non-Gaussianity, and other pathological behaviour in the posterior can bias our metrics away from our assumed normal distribution.

Fig. 6 shows the metric for each simulated star and each input property of the star. It is clear that the majority of our simulated star results are consistent with the truth value given the uncertainty. And broadly, the numbers of metrics at the 1 and 2 sigma levels are consistent with expectations. There are however some outliers or results which we will discuss. The index for the most significant outlier in terms of metallicity is 14 and an interesting behaviour in the age distribution is observed in simulated star index 21.
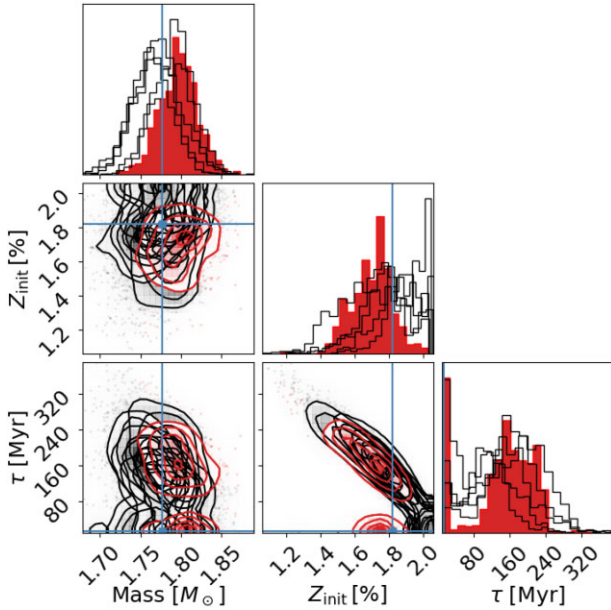

Figure 7. Samples drawn from the posterior distribution of simulated star 14 are shown in red. Subsequent runs using the same truth values ($M = 1.696 \, M_\odot$, $Z_{in} = 0.0136$, $\tau = 8.76$ Myr) but different noise realizations are shown in black. The truth values are shown in blue (see Section 5.1.3 for details).

The posterior samples for simulated star index 6 show an ideal example of posterior probability estimation for a simulated star. Each model parameter posterior distribution is unimodal, with peaks ($M = 1.58 \, M_\odot$, $Z_{init} = 0.93$ per cent, $\tau = 9.65$ Myr) comfortably centred on the truth values used to generate simulated star 6 ($M = 1.58 \, M_\odot$, $Z_{init} = 0.93$ per cent, $\tau = 9.67$ Myr).

5.1.2 Results for 25 simulated stars

While examining simulated stars individually is useful, it is hard to draw conclusions on the validity of our approach because we are looking at a single realization of noise on the observables. We now consider all 25 simulated stars, including the exemplars above (simulated star indices 5 and 6), to look at the statistics of our posterior probability distributions when compared to the truth values of the input properties. As part of our method, we fitted a parameter to account for our uncertainty in our assumption of the radial order label n. In our tests on simulated stars, we recovered the correct radial order in all cases with no meaningful uncertainty on the posterior of the radial order labels.

5.1.3 Further tests of simulated star 14

Simulated star 14 appears as an outlier in metallicity by $\sim 2.5\sigma$. To examine this behaviour, we have produced 10 more realizations of this simulated star. That is, we have taken the same truth values as inputs, but redrawn the simulated noise on each observable parameter using the same noise distributions.

Fig. 7 shows the posterior samples for the original and subsequent runs for simulated star 14. Firstly, it is clear that the posterior for Z_{in} is multimodal and contains significant covariance. Secondly, there is a bias of the posterior distributions away from the truth value in both metallicity and age that cannot be explained simply as a result of the realization noise on the observables.

We have checked for the possible origins of this bias. We examined the prior probability distribution, but found it to be smooth and nearly flat over the region of the posterior. We have performed multiple realizations of the noise and still observed this bias and therefore also exclude the noise or likelihood as the source of the bias.

A possible source of error is in the neural network emulation producing differences in the predicted mode frequencies. While these errors are typically small, of the order of 3×10^{-4} dex, the noise from the neural network is not random noise that would be expected to reduce with more realizations of the observables. Instead, the error is systematic and will produce a bias. The systematic error will always be present in emulation and this will lead to a bias, but it is the magnitude of this bias that is interesting. For this simulated star, the bias is similar to the reported uncertainty, which is about 1.5 Myr. However, this error can be reduced by extending the training time of

Figure 8. Samples drawn from the posterior distribution of simulated star 21 are shown in red. Subsequent runs using the same truth values ($M = 1.777\,M_\odot$, $Z_{in} = 0.0182$, $\tau = 12.97\,$Myr) but different noise realizations are shown in black. The truth values are shown in blue (see Section 5.1.4 for details).

the neural network, or increasing the grid search density around the optimal neural network architecture.

5.1.4 *Further tests of simulated star 21*

Simulated star 21 shows an interesting behaviour in the age posterior. Fig. 8 shows the posteriors for the original simulated star 21 and for 10 more realizations, as we did for simulated star 14 above. No meaningful bias is observed in the posteriors for mass or metallicity, given the priors we apply.

The true age of the simulated star was 12.96 Myr, which corresponds to the pre-MS evolution stage. The age posterior is clearly bimodal, with solutions around 10 and 160 Myr. This behaviour is observed in all the realizations, lending confidence that this is not a result of the noise being added. In fact, this bimodality is consistent with our understanding of the evolution of these stars and illustrates the difficulty of distinguishing the phase of the MS evolution where the track crosses its pre-MS evolution in the HR diagram.

5.2 Application to HD 99506

We applied our methods to HD 99506, which is one of the high-frequency δ Sct stars discussed by Bedding et al. (2020). We used the following inputs: $T_{\rm eff} = 7970 \pm 250\,$K and $L/L_\odot = 7.58 \pm 0.37$ (taken from table 1 of Bedding et al. 2020), and the mode frequencies that we have measured and listed in Table 2. We chose only the mode frequencies that were obvious, leaving out tentatively identified modes such as the $n = 4$ and $n = 9$ radial modes, and the $n = 1$ and $n = 8$ dipole modes. The identified modes span two radial orders more than any of the simulations and so, despite the gaps at some orders, they provide tighter constraints. The resulting posteriors on M, $Z_{\rm in}$, and age are unimodal and indicate a per cent-level random uncertainty (Fig. 9). The inferred age ($9.71 \pm 0.31\,$Myr) corresponds to the pre-MS phase, before the onset of pp-chain H-burning but

after the temporary pre-MS carbon-nitrogen-oxygen (CNO) burning phase.

The calculation of well-sampled posteriors for uncertainty estimates is a marked improvement on what is possible using discrete grid points and χ^2 minimization (e.g. Kerr et al. 2022a), where an arbitrary threshold in χ^2 needs to be adopted. It is especially useful that the posteriors are marginalized, given the aforementioned correlation in astrophysical parameters demonstrated with the simulated stars.

The neural network is also able to generate posterior predictions for each mode frequency, using the posterior samples as inputs. This can be useful for estimating the validity of uncertain mode identifications. In Fig. 10, we see that the leftmost (lower frequency) of two close peaks at the $n = 4$ radial mode is a good match and could perhaps be identified. The weak peak at $n = 9$ would also have fitted well. Inclusion of these would have resulted in tighter posteriors. On the other hand, none of the missing dipole mode frequencies, nor the $n = 1$ or 2 radial modes, would have been good additions. If we had supplied those modes as input, the posteriors would have broadened markedly.

6 CONCLUSIONS

We have presented a method for performing Bayesian inference on fundamental stellar properties of δ Sct stars using a neural network. This method emulates the stellar model and oscillation codes, MESA and GYRE, by learning from a grid of models that encompasses the physical properties of stars in or near the δ Sct instability strip. We used a nested sampling method to estimate the posterior distribution of the fundamental stellar properties, given a set of mode frequencies and classical observables. The resulting posterior distribution reflects the random observational uncertainty as well as the uncertainty of the neural network. By providing samples from the posterior probability density, which might be multimodal, non-Gaussian, and strongly covariate, we formally quantified the statistical uncertainty in the fundamental stellar properties. This improves our ability to investigate the systematic uncertainty in the stellar models.

We used a test set that was initially unseen by the training algorithm to evaluate the performance of the trained neural network. We found that the neural network is capable of reaching an average frequency precision of $\approx 3 \times 10^{-4}\,$dex, with an offset of $\approx 5 \times 10^{-5}\,$dex. These performance metrics may improve if the network were retrained with, for example, additional grid points or with the aim to reach a lower target loss. However, the flexibility of neural networks allows for the extension of the model grid to include additional variables, such as initial helium abundance or convective overshoot, by increasing the number of neurons in the network architecture (see e.g. Hendriks & Aerts 2019; Lyttle et al. 2021).

We applied our method to 25 simulated stars to quantify our accuracy and precision in the recovery of our input stellar properties. We showed the method to be capable of faithfully reproducing the true input parameters with the exception of simulated star 14. On investigation, we found a bias in the reported metallicity and age of this simulated star to be a result of the error in prediction by the neural network. The bias is of the order of 1.5 Myr in age for this star, but was confirmed to be systematic in nature. Further improvements in the accuracy of the neural network emulation would reduce the size of this effect.

Finally, we have applied our method to observations of a real δ Sct star, HD 99506. We found this star to be in the pre-MS stage of evolution and report a random uncertainty of only 3 per cent in age. Systematic uncertainty, such as that arising from missing or imperfect physics, has not been accounted for in this number, but our methods pave the way for the quantification of the systematic

Table 2. Identified modes for HD 99506 used as inputs in the modelling.

Frequency (d^{-1})	n	ℓ
33.48997	3	0
46.46549	5	0
53.51968	6	0
60.60165	7	0
67.65639	8	0
35.99870	3	1
50.04504	5	1
57.18150	6	1
64.19957	7	1
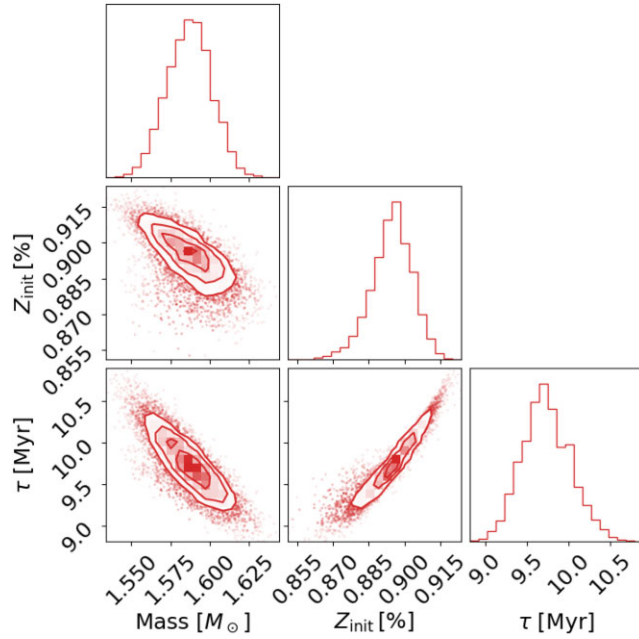
Figure 9. The posterior distributions in M, Z_{in}, and age for HD 99506, based on *TESS* observations.

uncertainty in future work. Similarly, the inclusion of additional physics such as rotation, and additional modes such as those of higher degree or different azimuthal order, would be trivial extensions to this framework in future.

SOFTWARE

Additional software used in this work, which have not explicitly been mentioned above, are listed below:

(i) PYTHON (Van Rossum & Drake Jr 1995)
(ii) MATPLOTLIB (Hunter 2007)
(iii) NUMPY (Harris et al. 2020)
(iv) SCIPY (Virtanen et al. 2020)
(v) PANDAS (Reback et al. 2020)
(vi) CORNER (Foreman-Mackey 2016)
(vii) LIGHTKURVE (Lightkurve Collaboration 2018)
(viii) ECHELLE (Hey & Ball 2020)

ACKNOWLEDGEMENTS
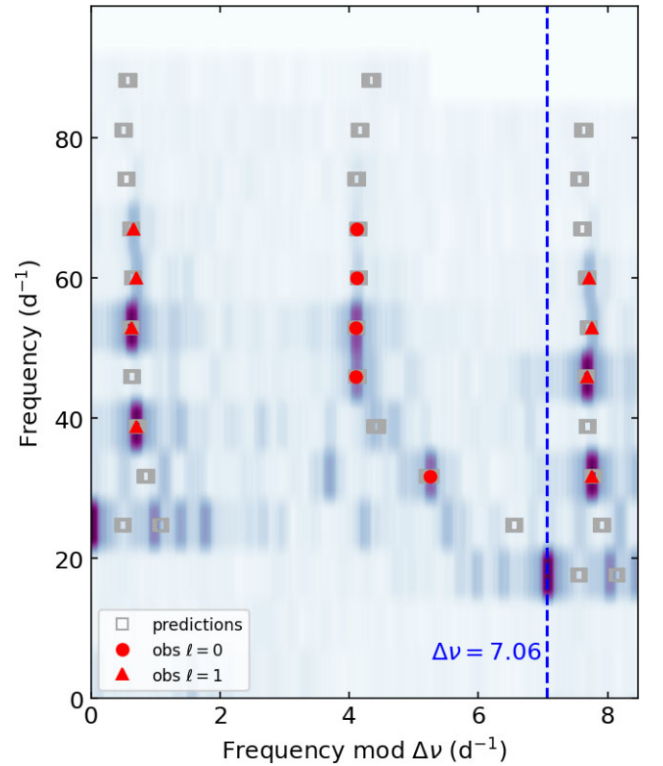
Figure 10. The posterior predicted frequencies overlaid on observed mode frequencies for HD 99506. The greyscale is the observed amplitude spectrum smoothed by a Gaussian of width four times the frequency resolution. For each radial mode, multiple samples are taken from the retrieved model parameter posteriors and predictions for each sample are plotted.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Abadi M. et al., 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, https://www.tensorflow.org/ (accessed February 5, 2023)

Aerts C., 2015, Astron. Nachr., 336, 477

Antoci V. et al., 2019, MNRAS, 490, 4040

Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, ARA&A, 47, 481

Balona L. A., Daszyńska-Daszkiewicz J., Pamyatnykh A. A., 2015, MNRAS, 452, 3073

Barac N., Bedding T. R., Murphy S. J., Hey D. R., 2022, MNRAS, 516, 2080

Barceló Forteza S., Roca Cortés T., García R. A., 2018, A&A, 614, A46

Bedding T. R. et al., 2020, Nature, 581, 147

Bedding T. R. et al., 2023, ApJ, 946, L10

Bowman D. M., Kurtz D. W., 2018, MNRAS, 476, 3169

Buchner J. et al., 2014, A&A, 564, A125

Chen X. et al., 2022, ApJS, 263, 34

Clevert D.-A., Unterthiner T., Hochreiter S., 2015, preprint (arXiv:1511.07289)

Currie T. et al., 2023, Science, 380, 198

Davies G. R., Miglio A., 2016, Astron. Nachr., 337, 774

Dupret M.-A., Grigahcène A., Garrido R., Gabriel M., Scuflaire R., 2004, A&A, 414, L17

Foreman-Mackey D., 2016, J. Open Source Softw., 1, 24

Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306

Galindo-Guil F. J. et al., 2022, A&A, 664, A70

García Hernández A. et al., 2017, MNRAS, 471, L140

Guzik J. A., 2021, Frontiers Astron. Space Sci. , 8, 55

Harris C. R. et al., 2020, Nature, 585, 357

Hasanzadeh A., Safari H., Ghasemi H., 2021, MNRAS, 505, 1476

Hayden M. R. et al., 2020, MNRAS, 493, 2952

Hendriks L., Aerts C., 2019, PASP, 131, 108001

Hey D., Ball W., 2020, Echelle: Dynamic Echelle Diagrams for Asteroseismology, https://doi.org/10.5281/zenodo.3629933 (accessed February 5, 2023)

Hunter J. D., 2007, Comput. Sci. Eng., 9, 90

Kerr R., Kraus A. L., Murphy S. J., Krolikowski D. M., Offner S. S. R., Tofflemire B. M., Rizzuto A. C., 2022a, ApJ, 941, 49

Kerr R., Kraus A. L., Murphy S. J., Krolikowski D. M., Bedding T. R., Rizzuto A. C., 2022b, ApJ, 941, 143

Kingma D. P., Ba J., 2014, preprint (arXiv:1412.6980)

Krumholz M. R., 2014, Phys. Rep., 539, 49

Lightkurve Collaboration, 2018, Astrophysics Source Code Library, record ascl:1812.013

Lipatov M., Brandt T. D., Gossage S., 2022, ApJ, 934, 105

Lyttle A. J. et al., 2021, MNRAS, 505, 2427

Mancini A. S., Piras D., Alsing J., Joachimi B., Hobson M. P., 2022, MNRAS, 511, 1771

{Michel} E. et al., 2017, in EPJ Web Conf., 160, 3001

Miret-Roig N., Galli P. A. B., Olivares J., Bouy H., Alves J., Barrado D., 2022, A&A, 667, A163

Murphy S. J., Hey D., Van Reeth T., Bedding T. R., 2019, MNRAS, 485, 2380

Murphy S. J., Joyce M., Bedding T. R., White T. R., Kama M., 2021, MNRAS, 502, 1633

Murphy S. J., Bedding T. R., White T. R., Li Y., Hey D., Reese D., Joyce M., 2022, MNRAS, 511, 5718

Murphy S. J., Bedding T. R., Gautam A., Joyce M., 2023, preprint (arXiv:2306.13142)

Paparó M. et al., 2013, A&A, 557, A27

Paxton B., Bildsten L., Dotter A., Herwig F., Lesaffre P., Timmes F., 2011, ApJS, 192, 3

Paxton B. et al., 2013, ApJS, 208, 4

Paxton B. et al., 2015, ApJS, 220, 15

Paxton B. et al., 2018, ApJS, 234, 34

Paxton B. et al., 2019, ApJS, 243, 10

Pedersen M. G., 2020, PhD thesis, KU Leuven

Reback J. et al., 2020, pandas-dev/pandas: Pandas 1.0.0

Ricker G. R. et al., 2015, J. Astron. Telesc. Syst., 1, 014003

Skilling J., 2004, in Fischer R., Preuss R., Toussaint U. V.,eds, AIP Conf. Proc. Vol. 735, Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering. Am. Inst. Phys., New York, p. 395

Soderblom D. R., 2010, ARA&A, 48, 581

Speagle J. S., 2020, MNRAS, 493, 3132

Steindl T., Zwintz K., Barnes T. G., Müllner M., Vorobyov E. I., 2021, A&A, 654, A36

Townsend R. H. D., Teitler S. A., 2013, MNRAS, 435, 3406

Uytterhoeven K. et al., 2011, A&A, 534, A125

Van Rossum G., Drake Jr F. L., 1995, Python Tutorial. Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands

Virtanen P. et al., 2020, Nat. Methods, 17, 261

White T. R., Bedding T. R., Stello D., Christensen-Dalsgaard J., Huber D., Kjeldsen H., 2011, ApJ, 743, 161

Žerjal M., Ireland M. J., Crundall T. D., Krumholz M. R., Rains A. D., 2023, MNRAS, 519, 3992

This paper has been typeset from a TEX/LATEX file prepared by the author.