Proposal notes

Proposal – Submit by Friday the 25th/10/19 by 17:00

|  |
| --- |
| no longer than four pages. This limit includes all figures/tables, but excludes any list of references and (if you want to use one) a cover page with project title, your name/supervisor's name and a short abstract. |
| The proposal should describe how the project will be executed and give details of intended timescales.  Real projects invariably start with a proposal to bid for resources/funding so you need to explain clearly the purpose of the project: why it is important to do it at all, where your work fits into the previous/current work, how you are going to do something in the time available. |
| Aims of the project.  Background to the project, including the project motivation and a summary of previous relevant work.  Methods/approach to be adopted.  Results expected (and any preliminary results already obtained).  A (simple) project plan - if you find it helps to explain this, you can include a table/Gantt chart |
| For a computing project, you will need to explain how your program will simulate the physical system that you are modelling.  you should consider how you will analyse and interpret your results. Your project plan may need to include alternative directions for the project to account for future events, e.g. data not being available when you need it, not being able to make some device work as well as anticipated |
| Use M67 published ages – hopefully analyse more clusters to improve accuracy (100 Myrs)  Get ages for lots more stars in potentially other clusters, to do that we need models of stellar evolution i.e. grid models  For that interpolation is done but that is slow, so instead we could calculate the tracks for each point (evolves a star to that point and then the sample moves to a new position and evolves again)  Alternatively, machine learning is built on linear algebra (almost exclusively) and is very fast, if we can build a machine learning algorithm that emulates stellar evolution, here is a grid of models, go learn it.  If we take the grid model that spans a region in space that is applicable to a cluster and we tell a neural net to learn (put in an age and mass and out comes observables e.g. delnu and numax)  Numax is currently an unknown sort of, so we will kind of ignore it  We will be getting numax values but they will likely but unreliable so in our final report we should have a disclaimer:  “We’re gonna use numax but we understand that there is no formal theoretical relationship between numax and the models we are calculating therefore we approximate it with this parameter this parameter that was published in this paper[citation needed], this is a major limitation of “the current things” but in future a better calibration and scaling relations… ”  We care about the method: machine learning, hierarchical models  We’re going to do a generative model: “put in fundamental and get observables”  Initially we will mostly ignore the systematic uncertainties that plague astrophysics  If we have enough time we can return to the systematics and try to solve some of those problems  Which essentially allows for endless scope beyond what is required. |
| Universal approximation function = it is mathematically guaranteed that you can approximate any function using a neural net of sufficiently large architecture.  (though it doesn’t mean you can train the neural net) |
| Proposal stuff: what do scientists do when they plan projects  For Abstract:  Context  Aims  Methods  Results  Conclusions  Implications  Stellar clusters are really exciting but we don’t know much about them  So we are going to model them to estimate the age spread  Context: other people have done this (looked at clusters and populations) before so mention the paper we found with the ages in it. Other people have used methods like…but not in the same context e.g. there are hierarchical models (see oli’s paper?)  Methods: neural net emulater and hierarchical modelling to find the spread  Results: Hopes and Aspirations,  Gant charts? E.g. on a timescale of this and this we are going to do x,y,z  Put thought into it despite knowing it may become completely obsolete  after proposal submission  Conclusions: if we get a results like this we can get that  Implications: doing this for more clusters would yield better results  Try to detail how we intend to do a better job than the researchers that came before us.  Which comes down to the square root of the number of data points  If are estimating the age of a star to that of 1Gyr but we have 100 stars, we might hope to resolve detail to 100Myrs.  At some point if we get some estimate for 1 cluster we could improve the no. data points by doing this on 10 clusters to get 100Myrs down to 10Myrs!  DON’T BE TOO OPTOMISTIC, there are of course other limiting factors so we couldn’t get down to 1Myr.  Prefis numbers with “in the best scenario”, “of course we won’t reach this because…”  Read up on neural nets and look up “sibenco’s universal function approximation theorem”  “look there is a universal approximation theorem that was proved 30 years ago so what we plan should work” – sounds good i.e. what we plan on doing will because of pre-existing knowledge  Add science and referencing |
| Age Distribution of Open Clusters  Generative model Heirarchical modelling  Machine learning |
| Things to research:  Background of machine learning+neural nets  Tensor flow, karas  Who has done machine learning and stellar evolution before  Mass loss function?  Open clusters  HMC, nuts  Heirarchical modelling |
| Machine learning in APOGEE Identification of stellar populations through chemical abundances  By:Garcia-Dias, R; Prieto, CA; Almeida, JS; Palicio, PA  <http://apps.webofknowledge.com.ezproxye.bham.ac.uk/full_record.do?product=WOS&search_mode=GeneralSearch&qid=1&SID=D58blpN5FsFfCOHatiQ&page=1&doc=5&cacheurlFromRightClick=no>  explore the possibility of using unsupervised clustering algorithms to separate stellar populations with distinct chemical patterns.  APOGEE is spectroscopically observing hundreds of thousands  of stars, primarily giants, which emphasizes regions that are  obscured by dust: the Galactic plane and the central parts of  the Galaxy  Only a few  of the clusters targeted by APOGEE have tens of observed stars  (e.g., NGC 2420, NGC 6791, the Pleiades, or M 67; see, e.g.  Cunha et al. 2015; Linden et al. 2017; Souto et al. 2016, 2018)  About 20 clusters include 5–30 stars each.  They used: FERRE is a data analysis code written in FORTRAN90. It matches models to data, taking a set of observations and identifying the model parameters that best reproduce the data, in a chi-squared sense.    N\* = no. cluster members  Membership was determined based on radial velocities and on the distribution of chemical abundances  Hogg et al. (2016)  machine-learning algorithms to stellar abundances was also employed by da Silva et al. (2012), Ting et al. (2012), Jofré et al. (2017), Anders et al. (2018), Boesso & Rocha-Pinto (2018), and Price-Jones & Bovy (2019).  Bayesian Gaussian mixture models  The algorithms are an affinity propagation (Frey & Dueck 2007), agglomerative clustering (Fränti et al. 2006), DBSCAN (Daszykowski & Walczak 2010), K-means (Macqueen 1967), mini-batch Kmeans (Bouveyron et al. 2007), spectral clustering (Ng et al. 2002), Gaussian mixing models (Bouveyron et al. 2007), and Bayesian Gaussian mixing models (Neal 1992; Lartillot &  Philippe 2004). These are all the nonhierarchical clustering algorithms available in the scikit-learn library (Thirion et al. 2016). |
| X-ray Study of Spatial Structures in Tycho's Supernova Remnant Using Unsupervised Deep Learning  <https://watermark.silverchair.com/stz1990.pdf?token=AQECAHi208BE49Ooan9kkhW_Ercy7Dm3ZL_9Cf3qfKAc485ysgAAAngwggJ0BgkqhkiG9w0BBwagggJlMIICYQIBADCCAloGCSqGSIb3DQEHATAeBglghkgBZQMEAS4wEQQMBjj73baGYgIxqEdpAgEQgIICKyDe8oc0FjHkTGrgJKZUCArtT_mnXbo4v8IxXGJ1Bmdw2jJXOUapMTn-i7fhWKsWw9hs0p26hPLdFiA5dKYPYCdDa-wymO2xzBseVDgpWxREdWAghkKT5ruGC-LkHN2MRO86NPjCELMDCuLZlNGvZNVbYVh8fXy5fnVJpj3cf6ygQPBjkeUKY9JZZ7UCzhlrVoE6wwH7THl4D0ToCQGeW4EhMnVfW-mEtfYwMSFQOyxTGMFqDvwlZH9sTBO7A_jP7IRC4DnyHqbeucHKCSeg-A5bsPk5UupE2Pc2CRAe5nsL4jLTVuR8vTYHm4gANZy3bWwEVCV4Og-ssJ8VdRYUtqjOnmLZCXJRwUjhYpwiLGng70lrmM_WupWL9rvcdKzy6W0s6SsIWffZmg4Ffni0lZq7KbH8REyzsadOXTUyf-lfDSRYgw1iPR-qv8Po9fdp5kWS-DgJ-Dwuym1whJ387C9i95ZtPWj6oMEHNkbpBK6lroEUNHemmhN3gEFvT0wLuwR1aOv9IexIPpJWgfvhpeNB-KJEBg-gXLnX8H39CahUS39dGSLBQdNzRmhmneypzVAuDvjg7HDEWX2MlG8I6IGx6eoylO7WnIu4L1lsrxnU89CdyRQpkmc1ANoVWffe8RZcNrKfCHu6Cl4_ztJKQmY9phVKvl7MXhivLZw1l-MReDXqwCGS02ZL0bWuRmt6_rIucm5rvGVvUlE-fSX_7rbpjqbvFNaK35Ml-A>  unsupervised learning method combining the VAE and a Gaussian mixture model (GMM) |
| Galaxies in X-ray selected clusters and groups in Dark Energy Survey data – II. Hierarchical Bayesian modelling of the red-sequence galaxy luminosity function  <https://watermark.silverchair.com/stz1612.pdf?token=AQECAHi208BE49Ooan9kkhW_Ercy7Dm3ZL_9Cf3qfKAc485ysgAAAnUwggJxBgkqhkiG9w0BBwagggJiMIICXgIBADCCAlcGCSqGSIb3DQEHATAeBglghkgBZQMEAS4wEQQMImPCCnUMT5iUhyEyAgEQgIICKN8RhBhqAc5iZcNbNlQoDcH3DYx7jbEZ3vzv8IfMr2gUzX-91U9hdkMrwKgdQDQLaVs3pDdwxLTqM3mCfmp-MrsDZYnj1aPKbW7bEA4lmsM-muE3qe8SJs3V0_UVno01m8doyKOC6qsBg-7rSNh1_hmFzR7dw2nL07hgt1Qio3xZy--POXJbhm0PSiRlOIrnB3p7V7eS49PqWRKD0yksYrDWuo6O7amPS72XbMbX9KNy34xN-xnfO0pXVLA54G6tckbJ31cI5cBbn6t4bmDEB8UShybWIaSKB4wfORoL_BJ-uy3R81J5DxllwX2sc9nDGpJR0-7Pd8LpKRMiYRz5VDoUTl5Ot1O9QXuyWDt9GQZsHO-OJA5f4aDxjSHu7Nwzsq7xfulphIxJ_Z-5MoCR4boAhcXwSsRibQFrE-7PBGKnNPctZYrgTFpHbIv_NqScEBkBUE7KawKKXij0ZeZ5Kul8yoIGw3MTSlKADKaUqZuYTn6vubFQxK4GQdXxI4ZKfEv8VIVySSPYIkGfvkp5534t9juMLGX8PV_woo__0S2QC3rHuctfnjXvNEU_OzY3askBOv8rvvfG8ga3iWMowKHul_eXdNQUBlqsVLUgn4bkE2NLrjVx3tCDyHmCkHP3J6yqAk2oV_anVjkcWrEvqliZSCAiPlkwforf7xhfmzTbYYsB-E7Gp526gwQMt88y5YgQ23UqqpCSzki2lgOi1xUKj97aDemRsg>  hierarchical Bayesian method (application examples to cosmology can be found in Loredo & Hendry 2010)  A hierarchical Bayesian method Given a model with a set of parameters Ω that describes the distribution of observables, Bayesian theory provides a framework for inferring Ω with a set of observed quantities {x} |
|  |
| Membership and fundamental parameters of the Praesepe cluster based  on Gaia-DR2  <https://watermark.silverchair.com/stz1213.pdf?token=AQECAHi208BE49Ooan9kkhW_Ercy7Dm3ZL_9Cf3qfKAc485ysgAAAngwggJ0BgkqhkiG9w0BBwagggJlMIICYQIBADCCAloGCSqGSIb3DQEHATAeBglghkgBZQMEAS4wEQQMTOcfchIzW2QMBjvKAgEQgIICK1YK7yuCmyV0EQmMb8iPwqxejFJQqyrwDDSsZcsCPYizp5aBpqVaYeiEXqLeZ2bRbmk-i3pw12PVjljhdFu-3kUXfZFX5EpTTYgfbOedC5jhHoaCwDDJA_vprrLk_9Q0fHYbs0jjKtSJmPil1mQlsgQOnhArBiMWLREuVNRX-F1BMA2hmnhb-vlcWcooo6RPOJ-ApEDtroDVsS96ILC7jxh-aF3klKpkjVwO-phs5XOIAcmIn2lziVUiMFFokpYkBefHKhJt0siJOl_spb-_qnE59P9Jd03OOrDP0B8N9R3nzlu7ZVR1xDIgrGUl1RNTEBnfIt5vZoA6KN_94IxqwwXOLm5HDtnNUl3vfDvmVsnS6-bU4czLnui72AdjDygtQoFvXD7_xYeYtGLy5Wg6ypVKplI-KKF_ALxTTma28Fm124XoMlQP88-mU3IE4M2ZZMgUdygnTz-l5VzLPkA_D2LgDD0EKXK_at3tqahJNkC7BleXS4nFoHLTvTVAgKPUlCLHlruV5xqcX4kmOw07coGF8QR2x4eF97fZCgVwzTi8TXSqpjn2VyTqeAU4CseECEKRBhjaK4Z542StI1qGppCM2Onl0fHnGsEjHJQw_x7RSm07qOoIw8_EFCzfwgJT6WA3xwo2CHqLFubVke0xc0jLci4Iy35RkPf6uwAbWTrN4q8CCkRkuwxyXn2MIE3Q2pZ_Y_ObIrYfYW0pt3zBygwUxhhxcDRFLIPdaw>  The combined ML method includes two key steps (Gao 2018a):  (1) constructing a reliable training set based on a two-component  (cluster and field stars) Gaussian mixture model (GMM); and (2)  predicting membership probabilities for the sample stars based on a  random forest (RF; Breiman 2001).  use a two-component GMM (Gaussian Mixture Model) to construct  a reliable training set in a 5D parameter space (proper motions,  positions, and parallax) |
| Mass and Age of Red Giant Branch Stars Observed with LAMOST  and Kepler  <https://arxiv.org/pdf/1712.09779.pdf>  a KPCA-based multivariate linear regression method  LAMOST spectra  Wu et al. (2018)  Xiang et al. (2017c)  Age and mass estimates from C, N abundance |
| “We further investigate the feasibility of deducing ages and  masses directly from LAMOST spectra with a **machine learning method based on kernel**  **based principal component analysis**, taking a sub-sample of these RGB stars as a training  data set” |
| **Deep Learning Applied to the Asteroseismic Modeling of Stars with Coherent**  **Oscillation Modes**  <https://iopscience-iop-org.ezproxye.bham.ac.uk/article/10.1088/1538-3873/aaeeec/pdf>  We develop a novel method based on machine-learning principles to achieve optimal initiation of CPU-intensive  computations for forward asteroseismic modeling in a multi-dimensional parameter space. A deep neural network is trained on a precomputed asteroseismology grid containing about 62 million coherent oscillation-mode frequencies derived from stellar evolution models. These models are representative of the core-hydrogen-burning stage of intermediate-mass and high-mass stars. The evolution models constitute a 6D parameter space and their predicted low-degree pressure- and gravity-mode oscillations are scanned using a genetic algorithm. A software  pipeline is created to find the best-fitting stellar parameters for a given set of observed oscillation frequencies. The proposed method finds the optimal regions in the 6D parameter space in less than a minute, hence providing the optimal starting point for further and more detailed forward asteroseismic modeling in a high-dimensional context.  Our method is open source and can be freely used by anyone  Comparison between asteroseismic data and stellar models is  done by computing the spectrum of oscillation modes predicted  by the theoretical models.  Statistical inference in asteroseismology can conveniently be done from forward modeling. Usually, one adopts a grid-based approach and considers millions of stellar structure models of different evolutionary stage. Applications to low-mass stars in the core-hydrogen-burning stage can be found in the pioneering papers by Miglio & Montalbán (2005) and Quirion et al. (2010), while Brassard et al. (2001) developed a methodology to treat coherent gravity modes in core-helium-burning subdwarf stars.  A prominent application of machine learning (ML) to lowmass star asteroseismology based on damped pressure modes was developed by Bellinger et al. (2016). It is based on the combined use of spectroscopic and asteroseismic input and relies on the principle of random forests.  Here, we focus on intermediate- and high-mass corehydrogen-burning stars in the mass range  2-20 Solar masses with detected coherent oscillation modes. Our aim is to develop and provide an optimal deep-learning tool as a critical start-up ingredient for efficient forward asteroseismic modeling in a +6D parameter space. In particular, we wish to explore the performance of a deep-learning network in order to gain computational efforts, because the “manual” asteroseismic modeling applied to a few stars with coherent modes has so far required months of analysis work, with a high risk of missing local minima in parameter space (see Ausseloos et al. 2004 for an example of such a systematic effort).  This is not practical for future applications to samples of tens to hundreds of stars anticipated by future space missions, such as TESS (Ricker et al. 2016) and PLATO (Rauer et al. 2014).  Future work will consider deep-learning methods for asteroseismology of fast rotators.  Their grid was constructed in the framework of the European FP 7 SpaceInn Project.  It was computed with the open-source code MESA  The grid is based on the same input physics as used and described by Schmid & Aerts (2016); the MESA inlist is provided in that paper and therefore omitted here to avoid repetition.5 In brief, the models are based on the following input physics: solar metal mixture by Asplund et al. (2009), OPAL opacity tables by Iglesias & Rogers (1996), mixinglength theory of convection with a fixed mixing-length parameter of 1.8 (expressed in local pressure scale heights), and the Schwarzschild criterion for convective stability  For chosen time steps along the evolutionary track, or equivalently Xc values, the equilibrium models are perturbed in a linear framework to compute oscillation modes and their accompanying frequencies.  Here, we limit ourselves to frequencies of coherent low-degree (l) zonal (m = 0) gravity and pressure modes, computed for each of the ∼70 values of Xc between the ZAMS and the TAMS. These oscillation modes were computed with the public code GYRE (Townsend & Teitler 2013; Townsend et al. 2018), adopting the Cowling and adiabatic approximations as a good approach, as argued in Paper I.  The deep neural network (DNN) provided here is a first step to improving upon manual analyses in terms of computational cost; it will need future upgrades to deal with correlations among the parameters and to deliver appropriate error estimation in a Bayesian framework, just as for the manual analyses.  In this work, a DNN is trained on an asteroseismological grid of intermediate-mass and highmass stars to predict the frequencies of their coherent oscillation modes. This is a typical regression problem.  The neural network architecture used in this work is a fully connected or dense network. A schematic view of a dense network is shown in Figure 1.    The input vector in this example is six-dimensional and the output is one-dimensional. In between are two layers that are n-dimensional. The layers have 6, n, n, and 1 neurons, respectively, in this case. The value of each neuron, except the input neurons, is determined by the following function:    In order to add non-linearities to the neural network, the activation function must be non-linear. A typical choice is the Rectified Linear Unit (ReLU; Nair & Hinton 2010). This function is defined as f (x) = x if x > 0 and f (x) = 0 if x < 0. By starting from the input layer and calculating all values of all neurons layer-by-layer, the output is finally computed. Initially, the weights and biases are set to random numbers and during training the weights are adjusted to correctly predict a predetermined training set.  To validate that the network has correctly learned to predict the occurring features, the network is evaluated on a separate validation set. Typically, the validation set is 10% to 20% of the total data set.  For regression problems, this typically is a mean-squared error between the prediction by the neural network and the actual output belonging to a particular training sample. By backpropagating the error through the network, the weights and biases are updated to improve the cost function in the next iteration (Chauvin & Rumelhart 1995).  A typical caveat of training neural networks is overfitting. This means the neural network learns to replicate the exact results of the training set, instead of generalizing over the particular samples and learning the underlying correlations. This can be solved by adding a penalty term for high-valued weights, which is called regularization (Ng 2004). One can verify if a network is overfitted by comparing the training and validation set accuracy. If the training loss is decreasing but the validation loss is increasing, the network is overfitting. A more thorough introduction into deep learning and the validation approach we followed here is given in the Appendix of Caron et al. (2018), to which we refer for further details.  The aim of this work is to train a neural network that is able to predict the frequencies of the coherent pulsation modes of a given star with detected oscillation frequencies belonging to identified modes (l, npg). We want to achieve the capacity to quickly search the parameter space spanned by the stellar parameters to find the optimal values for given measurements of the oscillations, without relying on any other information. The choice of using deep neural networks for this problem has several reasons. First, the size of the data set is quite large. From the grid, around 62 million oscillation modes are extracted and the objective is to predict the frequency belonging to specific modes given by a wide range of stellar parameters. The aim is not to train on parameters of a particular star, but for all possible stellar parameters that are available in the data set in one go. This forces the ML algorithm to generalize over all parameters and learn the underlying structure. It also means that the network only has to be trained once, and then can be applied to all stars that reside somewhere inside the training grid. Although training can take some time (around six hours for this work), evaluation is typically very fast (for this work on the order of microseconds per inference) and the network size is not very large (the compressed network file of this work is 370 MB). Furthermore, by adding a batch normalization layer as the first layer inside the network, the issue of incorrectly weighting parameters as stated in Bellinger et al. (2016) is no longer present. The normalizing layer normalizes all data before it is fed to the trainable parameters. In Bellinger et al. (2016) it is also stated that deep-learningbased regression is unconstrained. In this work that is indeed the case. However if the network accuracy is high enough this does not have to be a problem. If it were, it is trivial to design a constrained activation function that is bounded between two values. For example the hyperbolic tangent activation function is bounded between ±1 and can trivially be scaled and translated to get a bound between two values.  A downside of deep learning compared to random forests is that it is fairly easy with random forests to get insights into why particular predictions are being made, while this information is not accessible from a DNN. Training a DNN is also typically more difficult due to the high number of hyperparameters that need to be set correctly to optimally train on the data. In this work, a DNN is trained using TensorFlow (Abadi et al. 2015) and TFLearn (Damien et al. 2016) on an asteroseismic grid of intermediate- and high-mass stars to predict the frequencies of their coherent oscillation modes. The neural network architecture is shown in Table 1, which has a total number of 53,777,033 trainable parameters. Different neural network architectures have been evaluated and the architecture described in Table 1 provides the best accuracy.  In addition, different hyperparameter combinations have been evaluated. The network is trained using the Adam optimizer (Kingma & Ba 2014) with a learning rate of 10−3 and a batch size of 2048.    1. start with a random sample of N points;  2. take the n points with the least error (lowest cost);  3. sample m points around the n points taken from the previous step, using a Gaussian prior with as mean the parameter values and standard deviation σ1 times the mean;  4. repeat steps 2 and 3 I times;  5. repeat steps 2, 3, and 4 also with a standard deviation of σ2 and σ3 times the mean.  The cost function mentioned above is different from the cost function of the neural network. In this step, the neural network is already trained and no longer optimized. In addition, the loss of the neural network is not propagated to the cost function of the PF. This is because the error of the neural network is constant over all predictions. In a future update the neural network output will also contain uncertainty information. |
| **FUNDAMENTAL PARAMETERS OF MAIN-SEQUENCE STARS IN AN INSTANT WITH MACHINE LEARNING**  <https://iopscience.iop.org/article/10.3847/0004-637X/830/1/31/pdf>  In this work, we consider the constrained multiple-regression problem of inferring fundamental stellar parameters from observable quantities. We construct a random forest of decision tree regressors to learn the relationships connecting observable quantities of main-sequence (MS) stars to their zero-age mainsequence (ZAMS) histories and current-age structural and chemical attributes.  We validate our technique by inferring the parameters of simulated stars in a hare-and-hound exercise, the Sun, and the well-studied stars 16 Cyg A and B. Finally, we conclude by applying our method on a catalog of Kepler objects-of-interest.  We explore various model physics by considering stellar evolutionary tracks that are varied not only in their initial mass and chemical composition, but also in their efficiency of convection, extent of convective overshooting, and strength of gravitational settling. We compare our results to the recent findings from GBM (Silva Aguirre et al. 2015), ISO (Metcalfe et al. 2015), interferometry (White et al. 2013), and asteroseismic glitch analyses (Verma et al. 2014) and find that we obtain similar estimates but with orders-of-magnitude speed-ups.  We seek a multiple-regression model capable of characterizing observed stars. To obtain such a model, we build a matrix of evolutionary simulations and use machine learning to discover relationships in the stellar models that connect observable quantities of stars to the model quantities that we wish to predict. The matrix is structured such that each column contains a different stellar quantity and each row contains a different stellar model.  Uses MESA  We train a random forest regressor on our matrix of evolutionary models to discover the relations that facilitate inference of stellar parameters from observed quantities.  A schematic representation of the topology of our random forest regressor can be seen in Fig 3.    **A random forest** is an ensemble regressor, meaning that it is composed of many individual components that each perform statistical regression, and the forest subsequently averages over the results from each component (Breiman 2001). The components of the ensemble are decision trees, each of which learns a set of decision rules for relating observable quantities to stellar parameters. An ensemble approach is preferred because using only a single decision tree that is able to see all of the training data may result in a regressor that has memorized the training data and is therefore unable to generalize to as yet unseen values. This undesirable phenomenon is known in machine learning as over-fitting, and is analogous to fitting n data points using a degree n polynomial: the fit will work perfectly on the data that were used for fitting, but fail badly on any unseen data. To avoid this, each decision tree in the forest is given a random subset of the evolutionary models and a random subset of the observable quantities from which to build a set of rules relating observed quantities to stellar parameters. This process, known as statistical bagging (Hastie et al. 2005, Section 8.7), prevents the collection of trees from becoming over-fit to the training data, and thus results in a regression model that is capable of generalizing the information it has learned and predicting values for data on which it has not been trained.  We choose random forests over any of the many other nonlinear regression routines (e.g., Gaussian processes, symbolic regression, neural networks, support vector regression, etc.) for several reasons. First, random forests perform constrained regression; that is, they only make predictions within the boundaries of the supplied training data (see e.g., Hastie et al. 2005, Section 9.2.1). This is in contrast to other methods like neural networks, which ordinarily perform unconstrained regression and are therefore not prevented from predicting non-physical quantities such as negative masses or from violating conservation requirements. Second, due to the decision rule process that is explained below, random forests are insensitive to the scale of the data. Unless care is taken, other regression methods will artificially weight some observable quantities like temperature as being more important than, say, luminosity, solely because temperatures are written using larger numbers (e.g., 5777 versus 1, see for example Section 11.5.3 of Hastie et al. 2005 for a discussion). Consequently, solutions obtained by other methods will change if they are run using features that are expressed using different units of measure. For example, other methods will produce different regressors if trained on luminosity values expressed in solar units verses values expressed in ergs, whereas random forests will not. Commonly, this problem is mitigated in other methods by means of variable standardization and through the use of Mahalabonis distances (Mahalanobis 1936). However, these transformations are arbitrary, and handling variables naturally without rescaling is thus preferred.  Thirdly, random forests take only seconds to train, which can be a great benefit if different stars have different features available. For example, some stars have luminosity information available whereas others do not, so a different regressor must be trained for each. In the extreme case, if one wanted to make predictions for stars using all of their respectively observed frequencies, one would need to train a new regressor for each star using the subset of simulated frequencies that correspond to the ones observed for that star. Ignoring the difficulties of surface-term corrections and mode identifications, such an approach would be well-handled by a random forest, suffering only a small hit to performance from its relatively small training cost. On the other hand, it would be infeasible to do this on a star-by-star basis with most other routines such as deep neural networks, because those methods can take days or even weeks to train. Finally, as we saw in the previous section, random forests provide the opportunity to extract insight about the actual regression being performed by examining the importance of each feature in making predictions.  Uncertainty: There are three separate sources of uncertainty in predicting stellar parameters. The first is the systematic uncertainty in the physics used to model stars. These uncertainties are unknown, however, and hence cannot be propagated. The second is the uncertainty belonging to the observations of the star. We propagate measurement uncertainties σ into the predictions by perturbing all measured quantities n = 10,000 times with normal noise having zero mean and standard deviation σ. We account for the covariance between asteroseismic separations and ratios by recalculating them upon each perturbation.  The final source is regression uncertainty. Fundamentally, each parameter can only be constrained to the extent that observations are able to bear information pertaining to that parameter. Even if observations were error-free, there still may exist a limit to which information gleaned from the surface may tell us about the physical qualities and evolutionary history of a star. We quantify those limits via cross-validation: we train the random forest on only a subset of the simulated evolutionary tracks and make predictions on a held-out validation set. We randomly hold out a different subset of the tracks 25 times to serve as different validation sets and obtain averaged accuracy scores. |
| **AN IN-DEPTH STUDY OF GRID-BASED ASTEROSEISMIC ANALYSIS**  <https://iopscience.iop.org/article/10.1088/0004-637X/730/2/63/pdf>  GRID BASES MODELLING  As had been reported earlier, we find that it is relatively easy to get very precise values of stellar radii using grid-based techniques. However, we find that there are small, but significant, biases that can result because of the grid of models used. The biases can be minimized if metallicity is known. Masses cannot be determined as precisely as the radii and suffer from larger systematic effects. We also find that the errors in mass and radius are correlated. A positive consequence of this correlation is that log g can be determined both precisely and accurately with almost no systematic biases. Radii and log g can be determined with almost no model dependence to within 5% for realistic estimates of errors in asteroseismic and conventional observations. Errors in mass can be somewhat higher unless accurate metallicity estimates are available. Age estimates of individual stars are the most model dependent. The errors are larger, too. However, we find that for star clusters, it is possible to get a relatively precise age if one assumes that all stars in a given cluster have the same age.  We use three grids of models to test the model dependence of grid asteroseismology results. We use the Yale–Birmingham (YB) pipeline (Basu et al. 2010; described in Section 2.2) as the basis. We also estimate mass and radius directly from Δν, νmax, and Teff to determine whether or not grid methods are really needed to estimate these quantities. We test whether errors in the estimated mass and radius are correlated. And although determining ages of single stars is difficult (and model dependent), we examine whether seismic data allow us to do better than conventional fitting of evolutionary tracks. |
| Asteroseismic fundamental properties of solar-type stars observed by the NASA Kepler Mission  <https://arxiv.org/pdf/1310.4001.pdf>  Grid based modelling using 6 different models and 2 input parameters |
| **AUTOMATIC DETERMINATION OF STELLAR PARAMETERS VIA ASTEROSEISMOLOGY OF STOCHASTICALLY OSCILLATING STARS: COMPARISON WITH DIRECT MEASUREMENTS**  <https://iopscience.iop.org/article/10.1088/0004-637X/725/2/2176/pdf>  Grid based modelling  Here we present the SEEK package developed for the analysis of asteroseismic data from the Kepler mission. A central goal of the package is to obtain a fast and automatic determination of the stellar radius and other parameters in a form that is statistically well defined. The algorithms are tested by comparing the results of the analysis with independent measurements of stellar radius and mass for a sample of well-observed stars. We conclude that the SEEK package fixes stellar parameters with accuracy and precision |
| Metcalfe et al. (2009)  the authors combine a pipeline using a genetic algorithm to find the best model in the parameter space and a local analysis, based on linearization around this solution, to refine the fit and estimate the error in that solution.  Metcalfe et al. (2009) assumes that the solution is linear within the error bars and that no nearby local minima can be deep enough to contribute to the error analysis; this can lead to underestimation of the errors. In our view, only an error estimator, Bayesian or frequentist, looking at various models around the best solutions may provide fully reliable results. This is particularly true in a problem, as the present, where the valley of solution is extended. |
| Basu et al. (2010)  find the best model with a grid of stellar models, not unlike SEEK but using a smaller grid, while the errors are determined with a frequentist approach, using a Monte Carlo simulation of synthetic observations. |
| PROPERTIES OF 42 SOLAR-TYPE KEPLER TARGETS FROM THE ASTEROSEISMIC MODELING PORTAL  <https://arxiv.org/pdf/1402.3614.pdf> |
| Modelling the solar twin 18 Sco∗  <https://arxiv.org/pdf/1810.01387.pdf>  Methods. We set up a Bayesian model that relates the statistical properties of the data to the probability density of the stellar parameters. Special care is given to the modelling of the likelihood for the seismic data, using Gaussian mixture models. The probability densities of the stellar parameters are approximated numerically using an adaptive MCMC algorithm. From these approximate distributions we proceeded to a statistical analysis. We also performed the same exercise using local optimisation. |
| A review of Markov Chain Monte Carlo and information theory tools for inverse problems in subsurface flow  <https://link-springer-com.ezproxye.bham.ac.uk/content/pdf/10.1007%2Fs10596-011-9249-z.pdf>  The basic principle of these techniques is the application of the convergence property to a specific probability density function presented by the Markov Chains, provided that they comply with the conditions of reversibility and ergodicity [67]. This means that if it is possible to develop an algorithm that generates an ergodic reversible Markov Chain over a continuous space, then the states (points) of this chain will converge asymptotically to a sample of the function considered, also called stationary distribution.  Amar, J.G.: The Monte Carlo method in science and engineering. Comput. Sci. Eng. 8(2), 9–19 (2006) 66. Beichl, I., Sullivan, F.: Monte Carlo methods. Comput. Sci. Eng. 8(2), 7–8 (2006)  Gamerman, D., Freitas Lopes, H.: Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference. Chapman and Hall/CRC, Boca Raton (2006) |

With our project we intend to create a neural network capable of fast determinations of age distribution in multiparameter space.

Research how to optimize neural nets

|  |
| --- |
| **FUNDAMENTAL PARAMETERS OF MAIN-SEQUENCE STARS IN AN INSTANT WITH MACHINE LEARNING**  stellar ages cannot be measured directly; instead, they depend on indirect determinations via stellar modelling.  Traditionally, to determine the age of a star, procedures based on iterative optimization (hereinafter IO) seek the stellar model that best matches the available observations (Brown et al. 1994). IO is computationally intensive in that it demands the calculation of a large number of stellar models (see Metcalfe et al. 2009 for a discussion).  Several search strategies have been employed, including:  **grid-based modelling**, see Gai et al. 2011; Chaplin et al. 2014)  GBM by way of interpolation in a high-dimensional space, on the other hand, is sensitive to the resolution of each parameter and thus requires a very fine grid of models to search through (see e.g., Quirion et al. 2010, who use more than five million models that were varied in just four initial parameters). Additional dimensions such as efficiency parameters (e.g., overshooting or mixing length parameters) significantly impact on the number of models needed and hence the search times for these methods. As a consequence, these approaches typically use, for example, a solar-calibrated mixing length parameter or a fixed amount of convective overshooting. Since these values in other stars are unknown, keeping them fixed therefore results in underestimations of uncertainties. This is especially important in the case of atomic diffusion, which is essential when modelling the Sun (see e.g., Basu & Antia 1994), but is usually disabled for stars with M Mε > 1.4 because it leads to the unobserved consequence of a hydrogen-only surface (Morel & Thévenin 2002).  **in situ optimization (hereinafter ISO)** such as genetic algorithms (Metcalfe et al. 2014, Basu et al. (2010)), ISO requires that new stellar tracks are calculated for each target, as they do not know a priori all of the combinations of stellar parameter values that the optimizer will need for its search. They furthermore converge to local minima and therefore need to be run multiple times from different starting points to attain global coverage.  **Markov-chain Monte Carlo** (Bazot et al. 2012)  **downhill simplex algorithm** (Paxton et al. 2013;  see Silva Aguirre et al. 2015 for an extended discussion on the various methods of dating stars) |
| Deep Learning Applied to the Asteroseismic Modeling of Stars with Coherent Oscillation Modes  = genetic algorithm, forward machine learning? |
| FUNDAMENTAL PARAMETERS OF MAIN-SEQUENCE STARS IN AN INSTANT WITH MACHINE LEARNING = created grid of evolutionary tracks and used MESA, random trees, tested by blind Hare and Hound  “We seek a multiple-regression model capable of characterizing observed stars. To obtain such a model, we build a matrix of evolutionary simulations and use machine learning to discover relationships in the stellar models that connect observable quantities of stars to the model quantities that we wish to predict.” |
| Pont & Eyer (2004) and Jørgensen & Lindegren (2005). These investigations compared the Bayesian approach to other means used for the determination of stellar age. We refer the reader to these papers and to the textbook by Gregory (2005) for technical questions involved in the use of Bayesian statistics. |
| **Hamiltonian Monte Carlo** algorithm (also known as hybrid Monte Carlo), is a Markov chain Monte Carlo method for obtaining a sequence of random samples which converge to being distributed according to a target probability distribution for which direct sampling is difficult.  Hamiltonian Monte Carlo reduces the correlation between successive sampled states by proposing moves to distant states which maintain a high probability of acceptance due to the approximate energy conserving properties of the simulated Hamiltonian dynamic when using a symplectic integrator. |
| **The No-U-Turn Sampler: Adaptively Setting Path Lengths**  **in Hamiltonian Monte Carlo**  <https://arxiv.org/pdf/1111.4246.pdf>  Hamiltonian Monte Carlo (HMC) is a Markov chain Monte Carlo (MCMC) algorithm that avoids the random walk behavior and sensitivity to correlated parameters that plague many MCMC methods by taking a series of steps informed by first-order gradient information.  These features allow it to converge to high-dimensional target distributions much more quickly than simpler methods such as random walk Metropolis or Gibbs sampling. However, HMC’s performance is highly sensitive to two user-specified parameters: a step size and a desired number of steps L. In particular, if L is too small then the algorithm exhibits undesirable random walk behavior, while if L is too large the algorithm wastes computation  the No-U-Turn Sampler (NUTS), an extension to HMC that eliminates the need to set a number of steps L. NUTS uses a recursive algorithm to build a set of likely candidate points that spans a wide swath of the target distribution, stopping automatically when it starts to double back and retrace its steps.  We also derive a method for adapting the step size parameter on the fly based on primal-dual averaging. NUTS can thus be used with no hand-tuning at all.  simple methods such as random-walk Metropolis (Metropolis et al., 1953) and Gibbs sampling (Geman and Geman, 1984) may require an unacceptably long time to converge to the target distribution. This is in large part due to the tendency of these methods to explore parameter space via inefficient random walks (Neal, 1993).  The cost of HMC per independent sample from a target distribution of dimension D is roughly O(D5/4 ), which stands in sharp contrast with the O(D2 ) cost of random-walk Metropolis (Creutz, 1988).  HMC’s increased efficiency comes at a price. First, HMC requires the gradient of the log-posterior. Computing the gradient for a complex model is at best tedious and at worst impossible, but this requirement can be made less onerous by using automatic differentiation (Griewank and Walther, 2008).  Secondly is that poor choices in the input parameters can cause bad results  Appendix A. Estimating Effective Sample Size… |
| Isochrone ages for field dwarfs: method and application to  the age-metallicity relation<https://arxiv.org/pdf/astro-ph/0401418.pdf>  For a general introduction to Bayesian data analysis see Sivia (1996), for a very clear and detailed presentation see Jaynes (2003), and for a rapid overview centered on astrophysical applications see Loredo (1990) |
| Jørgensen & Lindegren (2005)  <https://www-aanda-org.ezproxye.bham.ac.uk/articles/aa/pdf/2005/22/aa2185-04.pdf> |
| Bayesian statistics use in astrophysics  FROM LAPLACE TO SUPERNOVA SN 1987A: BAYESIAN INFERENCE IN ASTROPHYSICS  <https://bayes.wustl.edu/gregory/articles.pdf> |
| Bayesian Hierarchical Modelling of Initial-Final Mass Relations Across Star Clusters  <https://arxiv.org/pdf/1806.06733.pdf>  Bayesian hierarchical modelling (Gelman 2006; Gelman et al. 2013) is a statistical method that simultaneously fits object-specific parameters for multiple objects by pooling their data under one overall model. The resulting estimates from hierarchical models are shrinkage estimates (Si et al. 2017a) that generally have better statistical properties than do their unpooled counterparts. Bayesian hierarchical models have been used in numerous projects in astrophysics  (e.g., Jiao et al. 2016; Shariff et al. 2016; Mandel et al. 2017; Si et al. 2017a,b; Si & van Dyk 2018). In the context of constraining the IFMR, Andrews et al. (2015) pooled 142 wide double WDs in a hierarchical model.  In summary, the Bayesian hierarchical approach with a two-stage sampler produces shrinkage estimates of the IFMR parameters that have smaller RMSEs than the case-by-case analysis. Also, if the case-by-case samples are available, it only takes the two-stage algorithm ten minutes to obtain the estimates under the hierarchical model. So we recommend readers to use our two-stage algorithm to fit hierarchical models. |
| In statistics, **a shrinkage estimator** is an estimator that, either explicitly or implicitly, incorporates the effects of shrinkage. In loose terms this means that a naive or raw estimate is improved by combining it with other information. The term relates to the notion that the improved estimate is made closer to the value supplied by the 'other information' than the raw estimate. In this sense, shrinkage is used to regularize ill-posed inference problems. |
| Prior distributions for variance parameters in hierarchical models(Comment on Article by  Browne and Draper)<https://projecteuclid.org/download/pdf_1/euclid.ba/1340371048>  Hierarchical (multilevel) models are central to modern Bayesian statistics for both  conceptual and practical reasons. On the theoretical side, hierarchical models allow a  more “objective” approach to inference by estimating the parameters of prior distributions from data rather than requiring them to be specified using subjective information  (see James and Stein, 1960, Efron and Morris, 1975, and Morris, 1983). At a practical level, hierarchical models are flexible tools for combining information and partial  pooling of inferences (see, for example, Kreft and De Leeuw, 1998, Snijders and Bosker,  1999, Carlin and Louis, 2001, Raudenbush and Bryk, 2002, Gelman et al., 2003).  A hierarchical model requires hyperparameters, however, and these must be given their own prior distribution. |
| AIMS - A new tool for stellar parameter determinations  using asteroseismic constraints  <https://arxiv.org/pdf/1901.02663.pdf>  Asteroseismic Inference on a Massive Scale (AIMS) is an open source code that uses Bayesian statistics and a Markov Chain Monte Carlo approach to find a representative set of models that reproduce a given set of classical and asteroseismic constraints. These models are obtained by interpolation on a pre-calculated grid, thereby increasing computational efficiency.  We test the accuracy of the different operational modes within AIMS for grids of stellar models computed with the Li`ege stellar evolution code (main sequence and red giants) and compare the results to those from another asteroseismic analysis pipeline, PARAM.  AIMS uses Bayesian statistics and a Markov-Chain-MonteCarlo (MCMC) algorithm (emcee, Foreman-Mackey et al. 2013) to select models representative of the input data by interpolating in a pre-defined grid. The combination of these techniques allows for an efficient, comprehensive search of the parameter space defined by the grid parameters. Userdefined priors and the likelihood function resulting from the input constraints shape the exploration of the parameter space. AIMS initialises the grid search in the region of a set of models with the highest posterior probability. This increases the efficiency of the parameter space exploration, which in turn helps the MCMC algorithm converge faster. |
| Abstract  Markov chain Monte Carlo (MCMC) is a technique for estimating by simulation the expectation of a statistic in a complex model. Successive random selections form a Markov chain, the stationary distribution of which is the target distribution. It is particularly useful for the evaluation of posterior distributions in complex Bayesian models. In the Metropolis–Hastings algorithm, items are selected from an arbitrary “proposal” distribution and are retained or not according to an acceptance rule. The Gibbs sampler is a special case in which the proposal distributions are conditional distributions of single components of a vector parameter. Various special cases and applications are considered.  Encyclopedia of biostatistics  <https://onlinelibrary.wiley.com/doi/abs/10.1002/0470011815.b2a14021?deniedAccessCustomisedMessage=&userIsAuthenticated=false> |
| A Very Brief Introduction to Machine Learning With Applications to Communication Systems  <https://arxiv.org/pdf/1808.02342.pdf>  What is machine learning  The problem of interest is studied in detail, producing a mathematical model that capture the physics of the set-up under study. Based on the model, an optimized algorithm is produced that offers performance guarantees under the assumption that the given physics-based model is an accurate representation of reality.  In contrast, in its most basic form, the machine learning approach substitutes the step of acquiring domain knowledge with the potentially easier task of collecting a sufficiently large number of examples of desired behaviour for the algorithm of interest. These examples constitute the training set. |
| Introduction to Neural Network based Approaches for Question Answering over Knowledge Graphs  <https://arxiv.org/pdf/astro-ph/0102224.pdf>  “Artificial neural networks” is a relatively loose term referring to mathematical models which have some kind of distributed architecture, that is, consist of processing nodes (analogous to neurons) with multiple connections (analogous to dendrites and axons). These connections generally have adaptable parameters which modify the signals which pass along them. There are numerous types of artificial neural networks for addressing many different types of problems, such as modelling memory, performing pattern recognition, and predicting the evolution of dynamical systems. Most networks therefore perform some kind of data modelling, and they may be split into two broad classes: supervised and unsupervised. The former refers to networks which attempt to learn the relationship between a data and a parameter domain, while the latter refers to networks used to find “natural” groupings with a data set independently of external constraints. What they have in common is the idea of learning about a problem through relationships intrinsically present in data, rather that through a set of predetermined rules.  feedforward multilayer perceptron. The term perceptron is historical, and refers to the function performed by the nodes. Feedforward means that there is a definite input and output, and a flow of data in one direction. This is in contrast to recurrent neural networks in which data flows in a loop  The weights, w, appearing in equations 1–3 are the free parameters of the network. Clearly, in order for the network to yield appropriate outputs for given inputs, the weights must be set to suitable values. This is done by training the network on a set of input vectors for which the ideal outputs (or targets) are already known. This is supervised learning. |
| Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks  <https://www.amazon.com/Neural-Smithing-Supervised-Feedforward-Artificial/dp/0262527014/ref=as_li_ss_tl?s=books&ie=UTF8&qid=1523836940&sr=1-1&keywords=%22neural+smithing%22&linkCode=sl1&tag=inspiredalgor-20&linkId=70977b12ae60ed55d5e64a2480c5697c>  Input Layer: Input variables, sometimes called the visible layer.  Hidden Layers: Layers of nodes between the input and output layers. There may be one or more of these layers.  Output Layer: A layer of nodes that produce the output variables.  Specifically, the universal approximation theorem states that a feedforward network with a linear output layer and at least one hidden layer with any “squashing” activation function (such as the logistic sigmoid activation function) can approximate any Borel measurable function from one finite-dimensional space to another with any desired non-zero amount of error, provided that the network is given enough hidden units.  A node, also called a neuron or Perceptron, is a computational unit that has one or more weighted input connections, a transfer function that combines the inputs in some way, and an output connection. |
| Each input has an associated weight (w), which is assigned on the basis of its relative importance to other inputs. The node applies a function f (defined below) to the weighted sum of its inputs    The output Y from the neuron is computed as shown in the Figure 1. The function f is non-linear and is called the Activation Function. The purpose of the activation function is to introduce non-linearity into the output of a neuron. This is important because most real world data is non linear and we want neurons to learn these non linear representations.  <https://ujjwalkarn.me/2016/08/09/quick-intro-neural-networks/> |
| Kepler, Borucki et al. 2010; K2, Howell et al. 2014) |
|  |
|  |
| HMC is a mathematical version of hierarchical modelling  Hyperprior: (mean, uncertainty on the mean, standard deviation, uncertainty on the standard deviation)  You can have multiple hyperpriors and each one contains these 4 parameters pertaining to some property e.g. age, mass, radius  A prior is generated given the set of hyperpriors which contains a set of generated mean and standard deviation for each of the given parameters.  The HMC is given real data and the output of the prior (after in our case put through a neural net to output observable properties) is compared against the real data. The outcome of the comparison is the posterior  Step 1: generate a prior and look at the posterior. Change the prior a small amount and check how it affects the posterior. Repeat many times to sample the distribution of how priors effect the posterior. The HMC figures out how changing the prior effects the posterior and uses that to determine how to optimally change the prior.  Step 2: NUT sampler, moves through the distribution created by step 1 to figure out the best prior that gives the best posterior (I think).  MCMCs In Bayesian statistics, the recent development of Markov chain Monte Carlo methods has been a key step in making it possible to compute large hierarchical models that require integrations over hundreds or even thousands of unknown parameters. |