Using Bayesian Hierarchical Modelling to Determine Age Distributions in Open Clusters

H. Leung, H. Westwood

Date: 25/10/19

Abstract

Context. Stellar formation and evolution are some of the most highly studied areas of astronomy but are still not very well understood. Given the ever increasing amount of asteroseismic data available for stars, it's only logical to attempt harness this information to study such fields, due to asteroseismologies usefulness in determining the ages of stellar objects. This naturally requires development of new tools, that give accurate results and are not computationally intensive.
Aims. To develop a framework consisting a Bayesian hierarchical model, with a feedforward neural network and Hamiltonian Monte Carlo, in order to accurately calculate ages of stars in open clusters.
Methods. Utilizing a Bayesian hierarchical model, using hyperpriors determined from literature, fundamental stellar parameter data from the priors will be passed through a supervised neural network trained on a stellar evolution grid model to give modelled observable data. With a Hamiltonain Monte Carlo, we can then acquire posterior density functions, constrained by both the hyperpriors and posterior to get accurate fundamental stellar parameter data for a population of stars in an open cluster.
Preliminary results. From a Bayesian hierarchical model using age estimates from 31 red giants [Stell et al. 2016] we calculated an estimate for the mean (3.89 ± 0.32 Gyr) and spread of (1.010.42 Gyr) of the age of open cluster M67.
Implications. With this framework we intend to develop, it will be possible to get fast and accurate age determinations of open clusters and the stars therein (with good enough asteroseismic data).

1 Introduction

Open clusters (OCs) have been an invaluable resource for probing the mechanics of Galactic formation and evolution for many decades. The nature of their usefulness comes from the fact that the stars within the OC form from the same giant molecular cloud and do so at almost the same time (C. J. Lada and E. A. Lada, 2003). The result is that the stars in the OC have very similar ages and chemical compositions, allowing for the study of stellar formation mechanics of a sample on the order of 100 stars, that formed at a particular epoch. The advantage of using OCs for this is that having a fairly large sample which all have similar ages and compositions, helps reduce the uncertainty on the mean age of the cluster.

Asteroseismology has become a very popular way of dating stars, in large part due to the vast amount of data collected from missions such as Kepler (Borucki et al., 2010), K2 (Howell et al., 2014) and CoRoT (Baglin et al., 2008) missions and the development of tools used for analysing such data, like MESA and GYRE. Asteroseismology allows us to investigate the interior of stars by measuring the star's surface pulsation modes (William J. Chaplin and Miglio, 2013), which can then be converted into stellar properties (mass, radius, age etc.) with scaling relations or with a combination of MESA and GYRE.

Numerical models of stellar evolution have also been very impactful in recent years, given the widespread increase in computational power, allowing models to become progressively more accurate year after year. With regards to determining the ages of stars their purpose is to enable mapping from observable properties (ν_{max}, $\Delta\nu$, T_{eff}) to fundamental stellar properties (age, mass, radius) and vice versa. Two applications of such models are grid modelling and recursive stellar evolution simulations. The grid models, use a pre-calculated grid, containing many evolutionary tracks, calculated from some given initial conditions. Each track contains the properties of a single star, for a large number of snapshots along the life of a star (Paxton et al., 2010). Grid models are powerful because very little computation is required to access the information contained within, which makes them very fast. This data is very useful when used to map observational data onto stellar properties, like the backward grid modelling package SEEK, designed to quickly calculate stellar radii using Kepler data.(Quirion, Christensen-Dalsgaard, and Arentoft, 2010) The weakness of grid modelling appears because the grid itself has a finite number of points and as such the observational data is unlikely to map directly onto one of these points, thus requiring the use of interpolation between grid points, producing uncertainties on the resulting stellar properties (Pont and Eyer, 2004)(Jørgensen and Lindegren, 2005) (for more examples see W. J. Chaplin et al. (2013), Kallinger et al. (2010), Lagarde et al. (2012)). The second method is by recursively computing stellar evolution tracks and slightly changing the initial conditions of the track, to create one which has properties somewhere along it, that match some observed properties to a desired accuracy. The drawback of this method is that it is time consuming and computationally intensive to calculate all the tracks (for more examples see Appourchaux et al. (2015), W. H. Ball and Gizon (2017), W H Ball, Themeßl, and Hekker (2018)).

A superior alternative to using the aforementioned numerical modelling methods, is Bayesian hierarchical modelling (BHM), which uses Bayesian statistics to simultaneously constrain, parameters on a set of objects, by considering all the data as part of a single model.(Gelman, 2006) In general BHMs, have better estimates on their parameters than methods that separate the data into multiple models.(Si et al., 2018) This is due to how BHM constrain using hyperpriors to assist the convergence, as well as observed data. Previously BHM has been implemented by Si et al. (2018) and Andrews et al. (2015) to derive accurate initial-final mass relations of white dwarfs in open clusters (Hyades, M67, NGC 188, NGC 2168, and NGC 2477), as well as studies by Shariff et al. (2016) and Hinton et al. (2019), who used BHM to study type 1a Supernovae (among others). However, BHM hasn't yet been used for determining the age distribution of stars in open clusters.

Using BHM alone is not enough to discard the numerical modelling methods mentioned above, as there still needs to be some algorithm to convert between fundamental properties and observables. Machine learning allows for the desired conversion without suffering the same drawbacks as the numerical modelling methods. In the most basic sense, machine learning enables for the development of an algorithm that describes the system under study without having to spend time to understand the system, referred to as collecting domain knowledge. Although the time is instead spent collecting examples that display the desired behaviour (e.g. values of fundamental stellar parameters and their equivalent observable properties), known as the training set (Simeone, 2018).

Neural network, are a broad category of mathematical models, under the umbrella of machine learning, that have some architecture where signals are sent between nodes. Supervised neural networks (SNN are used to determine relationship between a data, and parameter domain. Feedforward neural networks (FNN) are a type of SNN, where feedforward describes how the data flows in one direction from input to output.(Coryn A.L. Bailer-Jones, 2001) FNNs have an input layer where data flows in, some number of hidden layers and an output layer where some new altered data created by the neural network flows out. Each of the non-input layer nodes have 4 features: inputs, input weights, bias and activation function. A node takes inputs from some number of nodes in a layer preceding theirs, the weighting of an input is based on its importance relative to the other weight inputs to this node. The bias aids in optimizing the activation function of that node. The role of the activation function is to combine the other 3 features of the node to create an output. Through training, the neural network alters the weights of inputs and biases in order to sort itself into an algorithm that successfully accomplishes the given task. An example FNN is shown in figure 1.

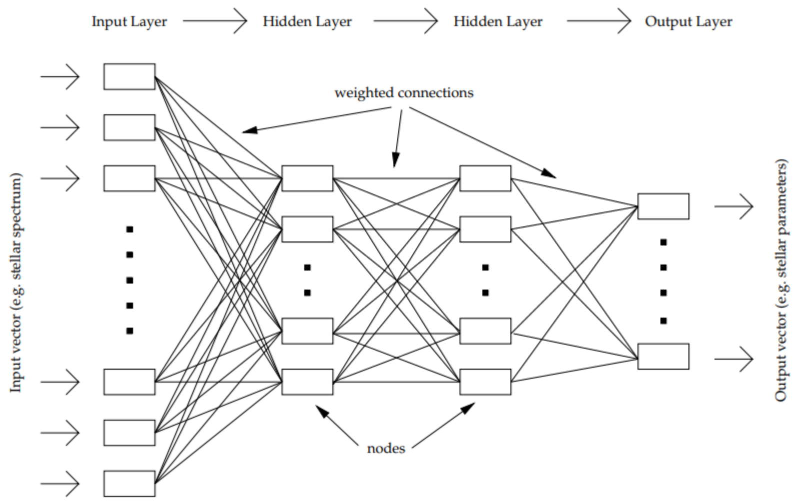

Figure 1: Feedforward neural network architecture. "...each layer has an extra node that holds a constant value (and has no inputs) and provides an offset to the subsequent layers. These are necessary for the network model to function properly." Figure by Coryn A.L. Bailer-Jones (2001)

Previously neural networks have been implemented by Hendriks and Aerts (2019) to do forward modelling on coherent oscillations of intermediate mass stars, as well as Verma et al. (2016) who did backward modelling on solar like oscillations.

A popular addition to BHM comes in the form of the Markov chain Monte Carlo (MCMC): a statistical method of evaluating posterior distributions through simulation, particularly useful in complex Bayesian models (Gilks, 2005). An upgraded version of an MCMC is a Hamiltonian Monte Carlo (HMC), which does what MCMCs do but in a more refined way, which makes it less computationally intensive (for more details see Hoffman and Gelman (2011) and for an application of MCMCs see Rendle et al. (2019)).

2 Method

2.1 Inputs and Hyperpriors

The inputs for the BHM will be a set of stellar parameters as hyperpriors (age, mass, radius, log g, T_{eff}). Each set of hyperpriors will be derived from the analysis of an open cluster, like that of M67 done by Stello et al. (2016). We will also use observed asteroseismic data from the same OC used to generate one of the sets of N, hyperpriors. An individual hyperprior will contain 4 values describing how a stellar parameter X, is distributed in a OC: μ_X the mean of X, $\delta_{\mu,X}$ the error on the mean of X, and similarly for the standard deviation of X, σ_X and $\delta_{\sigma,X}$.

2.2 Priors

A list of normally distributed priors, from a set of hyperpriors. Each prior contains a mean and standard deviation for each stellar parameter.
Additionally each parameter in the prior $P_i(\mu_1, \sigma_1, \mu_2, \sigma_2, ...\mu_N, \sigma_N)$, is generated from 2 log-normal distribution $logN(\mu_X, \delta_{\mu,X})$ and $logN(\sigma_X, \delta_{\sigma,X})$ (to prevent the prior values from being negative). From each prior a simulated population of stars is generated and for each star in that simulated population its properties are taken from a normal distribution $N(\mu_X, \delta_X)$.

2.3 Neural Network

We will create a feedforward neural network, trained on a pre-calculated grid of models that maps fundamental stellar parameters onto observable properties (like those previously mentioned). The FNN will be validated by evaluating the loss function using the mean absolute error (MAE), because it is the preferred method when the data does not have noise, which is true of the grid models. Once properly trained the FNN should be able to do the same job as the grid model at a similar speed but without the systematic error sustained from interpolation between grid points and with better accuracy.

2.4 Posterior

From the trained FNN given the inputted generated stellar population from the i-th prior, we get some modelled true stellar parameters, $Y_{true,i,j}$. "Y" refers to some observable parameter, "true" because they are exact values as they don't have uncertainties, it's generated from the i-th prior and it's for the j-th star. Using the observational uncertainty on the observed asteroseismic data obs, we can add artificial noise to modelled true observables, to give modelled observables $Y_{obs,i,j}$. We can then compare the modelled observational data against the actual observed data to determine a posterior for the i-th prior.

2.5 Hamiltonian Monte Carlo (HMC)

The HMC's purpose is to determine how the prior influences the posterior in order to generate a set of posterior density functions, one for the mean and standard deviation for each of the parameters that made up the set of hyperpriors. The peak of each of the posterior density functions gives the most likely value of the variable corresponding to the posterior density function.

3 Results and Implications

M67 is one of the well studied open clusters, with a number of people finding age estimates for the OC and the stars therein. Previous age estimates for M67 have been found to be 3.6-4.6 Gyr by VandenBerg and Stetson (2004) and 3.8-4.3 Gyr Barnes et al. (2016). By building a basic BHM and using individual age estimates for 31 red giants, found by Stello et al. (2016) we acquired an age estimate of 3.89 ± 0.32 Gyr with a spread of 1.010.42 Gyr. This is consistent with the literature values. By refining the BHM, expect to be able to be able to resolve the mean ages of individual stars in multiple open clusters, in the best case, down to an uncertainty on the order of 10s of Myrs. Using this framework should allow for greater study of galactic archaeology and stellar evolution.

4 Timeline

Table 1: Project timeline overview

Autumn week 5-7	Creating a BHM for M67. Additional literature reading for data on other OCs
Autumn week 8-11	Research most optimal stellar evolution grid model to train our FNN.
Term Break	Research on optimal FNN architecture. FNN development plus refinement.
Spring week 1-5	Training, validating and tuning FNN. Getting results from the BHM framework.
Spring week 6-9	Additional data analysis using BHM (time permitting)
Spring week 10-11	Project write-up

References

Andrews, Jeff J. et al. (2015). "CONSTRAINTS ON THE INITIAL-FINAL MASS RELATION FROM WIDE DOUBLE WHITE DWARFS". In: *The Astrophysical Journal* 815(1), p. 63. DOI: 10.1088/0004-637x/815/1/63.

Appourchaux, T. et al. (2015). "A seismic and gravitationally bound double star observed byKepler". In: *Astronomy Astrophysics* 582, A25. DOI: 10.1051/0004-6361/201526610.

Baglin, Annie et al. (2008). "CoRoT: Description of the Mission and Early Results". In: *Proceedings of the International Astronomical Union* 4(S253), pp. 71–81. DOI: 10.1017/s1743921308026252.

Ball, W H, N Themeßl, and S Hekker (2018). "Surface effects on the red giant branch". In: *Monthly Notices of the Royal Astronomical Society* 478(4), pp. 4697–4709. DOI: 10.1093/mnras/sty1141.

Ball, W. H. and L. Gizon (2017). "Surface-effect corrections for oscillation frequencies of evolved stars". In: *Astronomy Astrophysics* 600, A128. DOI: 10.1051/0004-6361/201630260.

Barnes, Sydney A. et al. (2016). "ROTATION PERIODS FOR COOL STARS IN THE 4 Gyr OLD OPEN CLUSTER M67, THE SOLAR–STELLAR CONNECTION, AND THE APPLICABILITY OF GYROCHRONOLOGY TO AT LEAST SOLAR AGE". In: *The Astrophysical Journal* 823(1), p. 16. DOI: 10.3847/0004-637x/823/1/16.

Borucki, W. J. et al. (2010). "Kepler Planet-Detection Mission: Introduction and First Results". In: *Science* 327(5968), pp. 977–980. DOI: 10.1126/science.1185402.

Chaplin, W. J. et al. (2013). "ASTEROSEISMIC FUNDAMENTAL PROPERTIES OF SOLAR-TYPE STARS OBSERVED BY THE NASAKEPLERMISSION". In: *The Astrophysical Journal Supplement Series* 210(1), p. 1. DOI: 10.1088/0067-0049/210/1/1.

Chaplin, William J. and Andrea Miglio (2013). "Asteroseismology of Solar-Type and Red-Giant Stars". In: *Annual Review of Astronomy and Astrophysics* 51(1), pp. 353–392. DOI: 10.1146/annurev-astro-082812-140938.

Coryn A.L. Bailer-Jones Ranjan Gupta, Harinder P. Singh (2001). "An introduction to artificial neural networks". In: *n Automated Data Analysis in Astronomy.*

Gelman, Andrew (2006). "Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper)". In: *Bayesian Analysis* 1(3), pp. 515–534. DOI: 10.1214/06-ba117a.

Gilks, W. R. (2005). "Markov Chain Monte Carlo". In: *Encyclopedia of Biostatistics.* DOI: 10.1002/0470011815.b2a14021.

Hendriks, L. and C. Aerts (2019). "Deep Learning Applied to the Asteroseismic Modeling of Stars with Coherent Oscillation Modes". In: *Publications of the Astronomical Society of the Pacific* 131(1004), p. 108001. DOI: 10.1088/1538-3873/aaeeec.

Hinton, S. R. et al. (2019). "Steve: A Hierarchical Bayesian Model for Supernova Cosmology". In: *The Astrophysical Journal* 876(1), p. 15. DOI: 10.3847/1538-4357/ab13a3.

Hoffman, Matthew D. and Andrew Gelman (2011). "The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo". In: *Journal of Machine Learning Research*, pp. 1593–1623.

Howell, Steve B. et al. (2014). "The K2 Mission: Characterization and Early Results". In: *Publications of the Astronomical Society of the Pacific* 126(938), pp. 398–408. DOI: 10.1086/676406.

Jørgensen, B. R. and L. Lindegren (2005). "Determination of stellar ages from isochrones: Bayesian estimation versus isochrone fitting". In: *Astronomy Astrophysics* 436(1), pp. 127–143. DOI: 10.1051/0004-6361:20042185.

Kallinger, T. et al. (2010). "Asteroseismology of red giants from the first four months ofKeplerdata: Fundamental stellar parameters". In: *Astronomy Astrophysics* 522, A1. DOI: 10.1051/0004-6361/201015263.

Lada, Charles J. and Elizabeth A. Lada (2003). "Embedded Clusters in Molecular Clouds". In: *Annual Review of Astronomy and Astrophysics* 41(1), pp. 57–115. DOI: 10.1146/annurev.astro.41.011802.094844.

Lagarde, N. et al. (2012). "Thermohaline instability and rotation-induced mixing". In: *Astronomy Astrophysics* 543, A108. DOI: 10.1051/0004-6361/201118331.

Paxton, Bill et al. (2010). "MODULES FOR EXPERIMENTS IN STELLAR ASTROPHYSICS (MESA)". In: *The Astrophysical Journal Supplement Series* 192(1), p. 3. DOI: 10.1088/0067-0049/192/1/3.

Pont, Frédéric and Laurent Eyer (2004). "Isochrone ages for field dwarfs: method and application to the age–metallicity relation". In: *Monthly Notices of the Royal Astronomical Society* 351(2), pp. 487–504. DOI: 10.1111/j.1365-2966.2004.07780.x.

Quirion, Pierre-Olivier, Jørgen Christensen-Dalsgaard, and Torben Arentoft (2010). "AUTOMATIC DETERMINATION OF STELLAR PARAMETERS VIA ASTEROSEISMOLOGY OF STOCHASTICALLY OSCILLATING STARS: COMPARISON WITH DIRECT MEASUREMENTS". In: *The Astrophysical Journal* 725(2), pp. 2176–2189. DOI: 10.1088/0004-637x/725/2/2176.

Rendle, Ben M et al. (2019). "aims – a new tool for stellar parameter determinations using asteroseismic constraints". In: *Monthly Notices of the Royal Astronomical Society* 484(1), pp. 771–786. DOI: 10.1093/mnras/stz031.

Shariff, Hikmatali et al. (2016). "BAHAMAS: NEW ANALYSIS OF TYPE Ia SUPERNOVAE REVEALS INCONSISTENCIES WITH STANDARD COSMOLOGY". In: *The Astrophysical Journal* 827(1), p. 1. DOI: 10.3847/0004-637x/827/1/1.

Si, Shijing et al. (2018). "Bayesian hierarchical modelling of initial–final mass relations acrossstar clusters". In: *Monthly Notices of the Royal Astronomical Society* 480(1), pp. 1300–1321. DOI: 10.1093/mnras/sty1913.

Simeone, Osvaldo (2018). "A Very Brief Introduction to Machine Learning With Applications to Communication Systems". In: *IEEE Transactions on Cognitive Communications and Networking* 4(4), pp. 648–664. DOI: 10.1109/tccn.2018.2881442.

Stello, Dennis et al. (2016). "THEK2M67 STUDY: REVISITING OLD FRIENDS WITHK2REVEALS OSCILLATING RED GIANTS IN THE OPEN CLUSTER M67". In: *The Astrophysical Journal* 832(2), p. 133. DOI: 10.3847/0004-637x/832/2/133.

VandenBerg, Don A. and P. B. Stetson (2004). "On the Old Open Clusters M67 and NGC 188: Convective Core Overshooting, Color-Temperature Relations, Distances, and Ages". In: *Publications of the Astronomical Society of the Pacific* 116(825), pp. 997–1011. DOI: 10.1086/426340.

Verma, Kuldeep et al. (2016). "Asteroseismic determination of fundamental parameters of Sun-like stars using multilayered neural networks". In: *Monthly Notices of the Royal Astronomical Society* 461(4), pp. 4206–4214. DOI: 10.1093/mnras/stw1621.