Stellar parameter inference in Galactic O-type stars using evolutionary model computations and Neural Networks

Claudio Cuervo Rodríguez

MÁSTER DEGREE IN ASTROPHYSICS. FACULTY OF PHYSICS
UNIVERSITY OF LA LAGUNA

Master Thesis in astrophysics

07/09/2021

Supervisors:

Sergio Simón Díaz
Ricardo Dorda Laforet

Resumen en castellano

Contexto: Debido a los próximos grandes surveys espectroscópicos de estrellas masivas, se necesitan de nuevas técnicas capaces de manejar estas grandes cantidades de datos. A esto debe sumarse los avances en el campo de la evolución estelar, los cuales establecen que parámetros estelares tales como la masa o el radio no pueden ser calculados conociendo únicamente los valores de temperatura efectiva y luminosidad, siendo necesario un análisis multidimensional. Una de las herramientas usadas para determinar los parámetros estelares es BONNSAI. Esta herramienta consiste en un análisis estadístico bayesiano que hace coincidir los datos empíricos con los modelos de evolución estelar de Bonn. Como alternativa a esta herramienta, este trabajo realiza un estudio exploratorio en el uso de redes neuronales para la inferencia de masas, radios y luminosidades. Esta metodología se presenta como una forma nueva de hallar estos parámetros estelares a partir de modelos evolutivos con el fin de complementar las medidas obtenidas a través de los análisis espectroscópicos cuantitativos.

Objetivos: Este trabajo consiste en estudiar la viabilidad de las redes neuronales en la inferencia de masas, radios y luminosidades de una muestra empírica de estrellas tipo O de la base de datos de IACOB. Se buscan además cuales son las condiciones óptimas del entrenamiento para un análisis futuro más profundo.

Metodología: Realizamos inferencias a varias muestras de estrellas. La red neuronal se entrena para todas las inferencias con estrellas sintéticas generadas con el código de SYCLIST. Los parámetros de entrada se distribuyen en tres grupos: el $\mathcal{L}$-training que incluye la temperatura efectiva, luminosidad espectroscópica y velocidad de rotación proyectada, el L-training que incluye la temperatura efectiva, luminosidad y velocidad de rotación proyectada y el $L\&\mathcal{L}$-training que incluye la temperatura efectiva, luminosidad, luminosidad espectroscópica y velocidad de rotación proyectada.

En primer lugar se definen dos métodos de entrenamiento de la red neuronal que se denominan entrenamiento global y entrenamiento local. Se estudia cual de dos métodos da menores errores intrínsecos en la inferencia. El entrenamiento global se basa en muestrear y entrenar toda la región de interés del diagrama espectroscópico de Hertzsprung-Russell (definida por la muestra empírica de estrellas tipo O), mientras que el entrenamiento local busca un mejor resultado disminuyendo la región de entrenamiento para no sobrecargar a la red neuronal. Para el entrenamiento global se representa la distribución del error relativo de cada parámetro objetivo y se obtiene su valor medio y desviación estándar. Además, también se estudia si la inferencia presenta tendencias en las zonas de alta luminosidad y baja luminosidad. Para el entrenamiento local se toman 8 estrellas de prueba y se les realiza la inferencia con muestra de entrenamiento formada por todas las estrellas encerradas en una caja centrada en los valores de temperatura efectiva y luminosidad espectroscópica de la estrella problema.

El siguiente paso consiste en estudiar el efecto de los errores observacionales. Para ello se añaden errores aleatorios a los parámetros de entrada asumiendo una distribución gaussiana con desviación estándar igual al promedio de los errores de los parámetros de entrada de

la muestra empírica. Para esta inferencia se vuelven a obtener las distribuciones de errores relativos y con los resultados del análisis de errores intrínsecos se obtiene la distribución de los errores observacionales añadidos. Finalmente se realiza la inferencia en una muestra empírica de estrellas masivas de tipo O, que posteriormente se compara con los resultados obtenidos espectroscópicamente por Holgado et al. (2020).

Resultados y conclusiones: El entrenamiento local presenta errores relativos muy dependientes de la posición de la estrella en el diagrama espectroscópico de Hertzsprung-Russell. El entrenamiento global resulta en incertidumbres más bajas y consistentes que las obtenidas por el entrenamiento local. Por ello, concluimos en que el entrenamiento global es mejor que el entrenamiento local. Como las incertidumbres de los errores intrínsecos del entrenamiento $L\&\mathcal{L}$-training son las menores, se obtiene que el mejor grupo de entrenamiento $L\&\mathcal{L}$-training es el mejor entrenamiento del primer análisis. Dividiendo la muestra de prueba en dos regiones (una de alta luminosidad y otra de baja luminosidad) se obtiene que la inferencia tiende a infraestimar el valor de las variables objetivo en la región de luminosidad alta frente a los valores de la región de luminosidad baja para el entrenamiento global. Además, las incertidumbres de la región de baja luminosidad son menores que aquellas de la región de alta luminosidad.

El análisis de los errores observacionales añadidos muestran que los mejores grupos de variables de entrenamiento son el $\mathcal{L}$-training y el L-training por tener las menores incertidumbres totales (intrínseca y observacional). Estas incertidumbres se encuentran en valores inferiores al 13% en todas las variables objetivo. Este análisis muestra también que las distribuciones de errores relativos de los errores observacionales añadidos son casi idénticas a las distribuciones que consideran errores observacionales e intrínsecos. Por ello, los errores dominantes en la inferencia son los observacionales

Al realizar la inferencia con la muestra empírica de estrellas O, los resultados obtenidos discrepan con los resultados del análisis espectroscópico cuantitativo. Este resultado está relacionado con la cuestión abierta de la discrepancia de masas, que establece que las masas obtenidas a partir de los modelos de evolución estelar y aquellas obtenidas a partir del análisis espectroscópico cuantitativo no son iguales. Se piensa que este problema esta causado por una determinación errónea de las gravedades en el análisis espectroscópico cuantitativo o por deficiencias de los modelos de evolución estelar. Como mejoras a realizara para un futuro análisis se comenta que es necesario poseer más valores de velocidad de rotación respecto a la velocidad de rotación crítica para cubrir mejor este parámetro. A la hora de obtener los cúmulos, una IMF constante ayudaría a la homogenización del espacio de parámetros y es necesario considerar un rango de metalicidades en lugar de un solo valor (metalicidad solar). Se concluye finalmente que el método de inferencia de parámetros estelares empleando redes neuronales da resultados aceptables cuando se evalúan estrellas sintéticas. Sin embargo, las imperfecciones en los modelos de evolución estelar y la discrepancia de masas provocan que al inferir los parámetros con estrellas reales no se obtengan resultados similares a los del análisis espectroscópico.

Contents

Acknowledgments

Chapter 1

Introduction

This work aims to serve as a pilot study to investigate the viability of the use of neural networks to infer certain stellar parameters in high-mass (O-type) main sequence stars. In particular, the ultimate objective is to be able to provide a full characterization of the physical properties of these stars. This is done by complementing the empirical information which can be extracted from the quantitative analysis of their optical spectra (such as the effective temperature and the surface gravity) with other important stellar parameters which determination requires the use of the outcome of evolutionary model computations (e.g. the stellar mass and radius).

High-mass stars (such as, e.g., those with O spectral types) are those with mass above 8 $M_\odot$ and they can produce elements heavier than carbon in their core. The CNO cycle is the dominant reaction in their their cores while they are on the main sequence, meaning that their life expectancy is much smaller compared to those stars of less mass, which burn hydrogen into helium via the pp-chain. high-mass stars, at the end of their lives, produce a neutron star or a black hole as a remnant of the supernova that occurs in the final collapse of their core, when the radiation pressure generated by the nuclear reactions is surpassed by the gravitational force.

Generally speaking, the physical parameters of any star (including, for example, the effective temperature, radius, luminosity, or surface gravity) mainly depend on their initial mass and age. In fact, the initial mass is the most critical parameter determining the

evolution of stars, since it determines its inner structure, the nuclear reactions occurring in their cores along their lifetimes, and and even the type of stellar remnant once the star dies. Hence, in order to properly characterize a given star, it is needed to obtain empirical information about all the parameters mentioned above, and, in the case of high-mass stars, also about other important quantities such as their spin rates and the amount of mass that the star is loosing due to stellar winds.

The traditional method of obtaining the mass of a star consist in comparing the position of the star in the Hertzsprung-Russell Diagram (HRD) with the nearest evolutionary stellar track, or by interpolating the closest tracks. However, recent studies showed that the evolution of high-mass star is critically affected by its initial rotational velocity (see Figure 1.1 and explanation below), meaning that having access to the effective temperature and luminosity of the star may not necessarily provide enough information to obtain the present mass of a star by locating it in the HRD (Maeder & Meynet, 2000). As a consequence, a multidimensional analysis of the stellar parameters is needed to characterize the stars (see, e.g., Schneider et al., 2014). This implies that more complex techniques need to be developed, being a machine learning technique, neural networks, the chosen one for this work.

Depending on the specific type of star (mainly depending on its effective temperature or spectral type) different techniques are required to derive information about its mass. In the case of high-mass OB-type stars, parameters such as the projected rotational velocity ($V sini$), the effective temperature (T_eff), or the surface gravity ($\log(g)$) are commonly obtained by means of quantitative spectroscopy (see notes in Simón-Díaz, 2020). Having access to other parameters such as the luminosity, radius and mass requires extra information: the apparent magnitude, interstellar extinction and the distance to the star. Alternatively, if there is no reliable information about distance, these three stellar parameters can be inferred by using evolutionary tracks provided by a stellar evolution code.

While this approach is relatively straightforward in the case of intermediate and low

mass stars, it becomes more delicate as we move to the high-mass star domain. This is due to the combination of two important ingredients affecting the evolution of high-mass stars. On the one side, these stars can reach much larger rotational speeds than in the case of lower mass stars (e.g. the distribution of projected rotational velocities in main sequence stars with masses above 8 $M_\odot$ have been found to reach values up to 600 km/s Ramírez-Agudelo et al., 2013). On the other side, in those cases with significant initial equatorial velocities, the rotation of the star produces non-negligible effects on the evolutionary tracks. This effect can be seen in Figure 1.1 the stellar evolutionary tracks differ depending on the considered initial equatorial velocity[1]. These differences are more important at higher mass.

As a consequence, the so-called evolutionary mass[2] cannot be univocally determined by considering the effective temperature (T_{eff}) and the luminosity of the star (or, alternatively, T_{eff} and the spectroscopic luminosity $\log(\mathcal{L}/\mathcal{L}_\odot)$). The considered initial equatorial velocity in the evolutionary tracks plays an important role whenever the star is rotating above $\sim$20% of its critical velocity. The multivaluation in mass can be seen in Figure 1.1 by looking at the dotted track of 40 $M_\odot$ and the solid track of 85 $M_\odot$, which coincide at $T_{\mathrm{eff}} = 32\,500$ K and $\log(\mathcal{L}/\mathcal{L}_\odot) = 4.2$ dex, giving a difference in the estimated mass of 45 $M_\odot$.

1.1 High-mass stars observational current state

The knowledge on high-mass stars has its applications in other fields beyond the stellar physics. As these stars are a source of ionizing photons, the development in the high-mass stars field serves as an improvement in the study of HII regions where they are formed (Tenorio-Tagle et al., 2006). Also, as these stars have stellar winds, increasing the knowledge of the surface enrichment and their wind properties (Trundle & Lennon, 2005) improve the

---

[1]The representation used in this figure is similar to the Hertzsprung-Russell diagram (HRD) which is the so-called spectroscopic Hertzsprung-Russell diagram (sHRD), being the difference between both representations that the sHRD uses the quantity $\mathcal{L}$ (defined as $\mathcal{L} = T_{\mathrm{eff}}^4/g$). This quantity allows to construct a diagram with a similar structure as the HRD with the advantage that a measure of the distance of the star is not needed, since only T_{eff} and $\log(g)$ (two quantities that are obtained from the analysis of the spectrum) are considered.

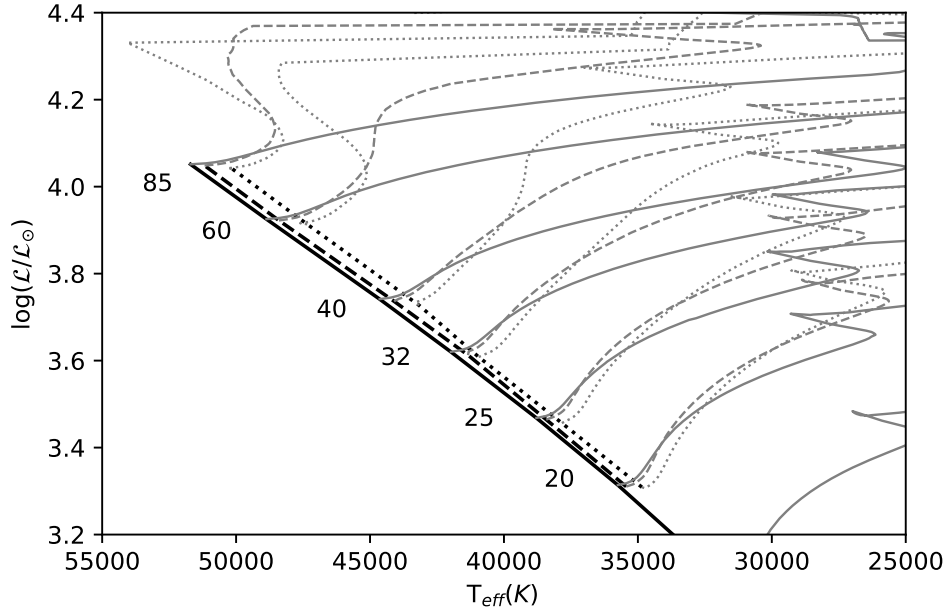[2]The value of the mass derived using the evolutionary tracks.

Figure 1.1: *Spectroscopic Hertzsprung-Rusell diagram (sHRD) of the evolutionary tracks given by the Geneva models for a solar metallicity. Solid, dashed and dotted lines represent the tracks for a star with the indicated initial mass with a ratio of equatorial velocity over the critical velocity (v_{eq} /v_{crit}) of 0, 0.2 and 0.4, respectively.*

understanding of the chemodynamical evolution of galaxies.

As new multi-object spectrographs are commissioned, it is expected to have a considerable increase in the number of known high-mass stars in the foresight future. This is the case of WEAVE (William Herschel Telescope Enhanced Area Velocity Explorer) which is a multi-object spectrograph that is under development and is expected to have its first light by the begining of 2022. This instrument will allow to gather hundreds spectra of high-mass stars in a given pointing thanks to the feature of being a multi-object spectrograph allowing to measure multiple stars that are in a given field and obtain the spectra of all these stars by moving their light to the spectrograph though different fibers. These instruments are very important for the development of the long term surveys of Galactic O-stars that will be carried out in the next years. But, before then, we already have accesible data from spectroscopic surveys which have provided high-quiality spectra for thousand of high-mass

OB stars. One of these surveys is IACOB (P.I., Simón-Díaz, Spain), whose objective is to build a large database of high-resolution multi-epoch spectra of Galactic OB stars to investigate the physical properties and evolution of high-mass stars. Another survey studying high-mass stars is VLT-Flames Tarantula Survey (VFTS, P.I., C. Evans, UK). It is focused in obtaining multi-epoch optical spectroscopy of OB stars in the 30 Doradus region, in the Large Magellanic Cloud. Its objective is to detect binary high-mass stars by measuring variations in their radial velocity between different epochs.

This work has been developed in the framework of IACOB. This project has currently compiled high quality multi-epoch spectroscopic observations for a sample of more than 1000 Galactic high-mass O and B-type stars, part of which are represented in Figure 1.2. While the quantitative spectroscopic analysis of the stars surveyed by the IACOB project is allowing us to obtain estimates for their projected rotational velocities, effective temperatures and surface gravities, the question is now: how other parameters (such as the mass), for which we need to make use of the outcome of stellar evolutionary model computations, are going to be obtained (specially taking into account the caveats mentioned above). As an answer to this, this work presents an exploratory study on this by using machine learning techniques, specifically Artificial Neural Networks (ANN).

At the moment there are two main methods for measuring the mass and radius of a star. These methods use the parameters derived through quantitative spectroscopic analysis together with distance and photometric information (see e.g. Simón-Díaz, 2020) while others use the stellar evolution models, comparing the observed parameters with the evolutionary tracks.

Ideally speaking, the masses determined by means of both methodologies should agree. However, a long-standing issue in the field of high-mass stars is the so-called mass discrepancy problem (Herrero et al.,1992). This mass has proven to be model dependent and also depends on the mass range. This discrepancy is thought to be caused either by the large uncertainties obtained in the luminosity and gravity or by a flaw in the implementation of

5
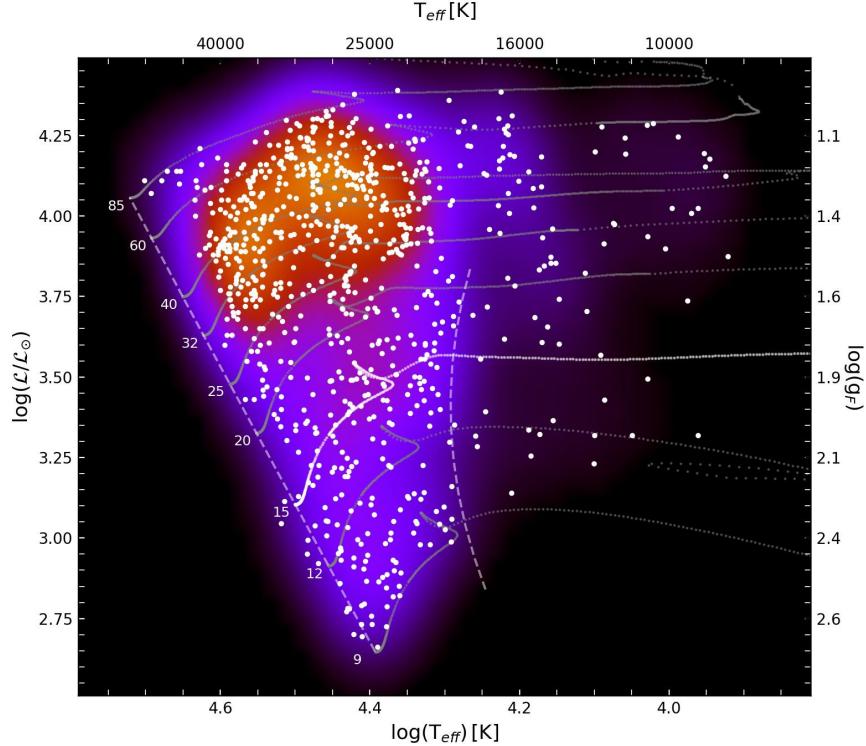
Figure 1.2: *IACOB sample of OB stars in the spectroscopic Hertzsprung-Rusell diagram. The color scale shows the density of stars in the sample. Courtesy of Abel de Burgos (IAC/ULL).*

the physical processes when computing the evolutionary models (Markova et al. 2018).

1.2 Objectives and structure

Stellar parameter inference using machine learning had not been done previously; however, a big data analysis tool focused on inferring stellar parameters was developed. This tool is BONNSAI (Schneider et al. 2014), which is a bayesian statistical method that matches observational data with a given stellar evolution model (Bonn Models in this case). The aims of this inference tool are to obtain a probability distribution of the stellar parameters able to predict the values of stellar parameters that had not been yet measured, and to test the stellar evolution models.

This work presents a starting point to create an alternative to the BONNSAI tool, since

one of its main drawbacks is that it requires high computation times. Supervised machine learning includes many tools that can be used to infer the stellar parameters (e.g different types of Neural Networks, Random Forest algorithms). The chosen method in this work is a Multilayer Perceptron which is one of the simplest forms of Artificial Neural Networks (ANN). This is the type of Neural Network available with the Orange Data Mining Software (Demšar et al. 2013), which is the software used to train the Neural Network. These Neural Networks are characterized for having at least three layers of neurons (input, output and hidden layers) whose training is based on the backpropagation technique. The advantage of using an ANN is that it only needs to be calculated once per training sample, meaning that once obtained an optimal sample it is only need to be applied to a problem set of stars, resulting in a method faster than the bayesian method.

The main objective of this work is to explore the use of ANNs in the inference of the radius (R), mass (M) and luminosity ($\log(L/L_\odot)$) of a sample of Galactic O-type stars, also identifying the optimal conditions for a future deeper analysis. This inference is made to complement the spectroscopically obtained measurements of gravity and effective temperature. The training sample is made of *synthetic* stars obtained from the stellar evolution models. This is also a preparatory work for the use of real stars in the training, once their number is high enough to cover the whole parameter space.

This work is structured in 5 chapters. Chapter 2 consist in a description of the tools and samples (*empirical* and *synthetic*) that are used ; an optimization of the *synthetic* sample for the training of the ANN; the training variables that are considered in the different training variable sets considered and the definition of two initial training methods. Chapter 3 shows the analysis made with the sample of synthetic stars and the study of how observational errors would affect in these synthetic stars. In Chapter 4, this study is done with a sample of Galactic O-type stars of which there is available data on T_{eff}, $\log(g)$ and $\log(L/L_\odot)$, and comparing it to the results obtained by a previously done quantitative spectroscopic analysis in Holgado et al. (2020). Chapter 5 presents the main conclusions of this work.

7

Chapter 2

Methodology

In this chapter we describe the tools we have used to create and analyze the various working samples used in training and testing the ANN. We also describe the various training methods and inference tests that we have explored in this work.

2.1 Tools

The tools used in this work are two: SYCLIST and the Orange Data Mining Software. SYCLIST (Georgy et al., 2014) is a cluster generator code which was developed by the University of Geneva, based in the Geneva evolutionary models (Ekström et al., 2012). The aim of the development of these evolutionary models is to provide more precise results by considering the stellar parameters variations given recent observations and the insight given by techniques such as asteroseismology, interferometry or spectropolarimetry about the stellar interior structure, rotation and stellar winds. These models give as an output the following parameters: effective temperature (T_{eff}), luminosity ($\log(L)$), initial mass (M_{ini}), actual mass (M), surface abundances of helium, carbon, nitrogen, oxygen, equatorial velocity (v_{eq}) and velocity over the critical velocity[1] (v_{eq}/v_c) . The online version of SYCLIST[2] can give three outputs: interpolated evolutionary tracks, isochrones and synthetic coeval stellar populations (Georgy et al., 2014). The output chosen is the synthetic coeval stellar

---

[1]Defined as the velocity at which the centrifugal force surpasses the gravitational force, causing that the star can not be gravitationally bounded.

[2]https://www.unige.ch/sciences/astro/evolution/en/database/syclist/

populations (where the number of synthetic stars requested for each cluster are 10^5) because it presents a more dense grid of values than evolutionary tracks and isochrones. This allows to have the HRD space more sampled than using the other two outputs.

The Orange Data Mining Software is a Python based platform aimed to help with the management and analysis of large datasets. It includes tools of supervised machine learning (e.g. random forest algorithm or artificial neural networks) as well as unsupervised machine learning tools (e.g. principal component analysis or manifold learning). These tools are displayed in a graphic environment, making its usage intuitive for novel users. However the number of hyperparameters are limited compared to other Machine Learning tools (e.g. SciKit-learn) but provides a good interface for entering in the Machine Learning discipline and for developing an exploratory work.

2.2   The *empirical* sample

The *empirical* sample analyzed in this work is part of the database of O-type stars (Holgado et al. (2020) surveyed by the IACOB and OWN projects (Figure 2.1). These are two complementary surveys aimed at obtaining empirical information about Galactic OB stars. The IACOB project is the responsible of obtaining and performing the quantitative spectroscopic analysis of a large database of high-resolution multi-epoch of OB-type stars in the Northern Hemisphere, while OWN is a long-term survey whose objective is to monitor O-Stars and WN-type stars in the Southern Hemisphere, with the aim of detecting and studying spectroscopic binaries. The spectrographs used in these surveys are FIES at NOT2.56m and HERMES at MERCATOR1.2m in the Northern Hemisphere; and FEROS at MPG/ESO-2.2m.

This empirical sample comprises 123 O-type stars including both, likely single (LS) and single line spectroscopic binaries (SB1). A quantitative spectroscopic analysis made by Holgado et al. (2020) yielded the following parameter range: projected equatorial velocity ($V sini$) between $10-250$ km/s, $\log(\mathcal{L}/\mathcal{L}_\odot)$ between $3.4-4.8$ dex and T_{eff} between $27\,400-$

50 300 K, being these stars in the Main Sequence evolutionary phase. Other spectroscopically obtained parameters of these stars are the spectral type, luminosity class, macroturbulence, microturbulence, effective temperature, gravity, spectroscopic luminosity, luminosity, helium abundance, spectroscopic radius and wind strength parameter. The parameters used in this work are the effective temperature (T_{eff}), the spectroscopic luminosity ($\log(\mathcal{L}/\mathcal{L}_\odot)$), the luminosity ($\log(L/L_\odot)$) and projected equatorial velocity ($V\sin i$) (summarized in Table A.1). The mean uncertainties of these parameters are 700 K, 0.08 dex and 0.02 dex[3] respectively.
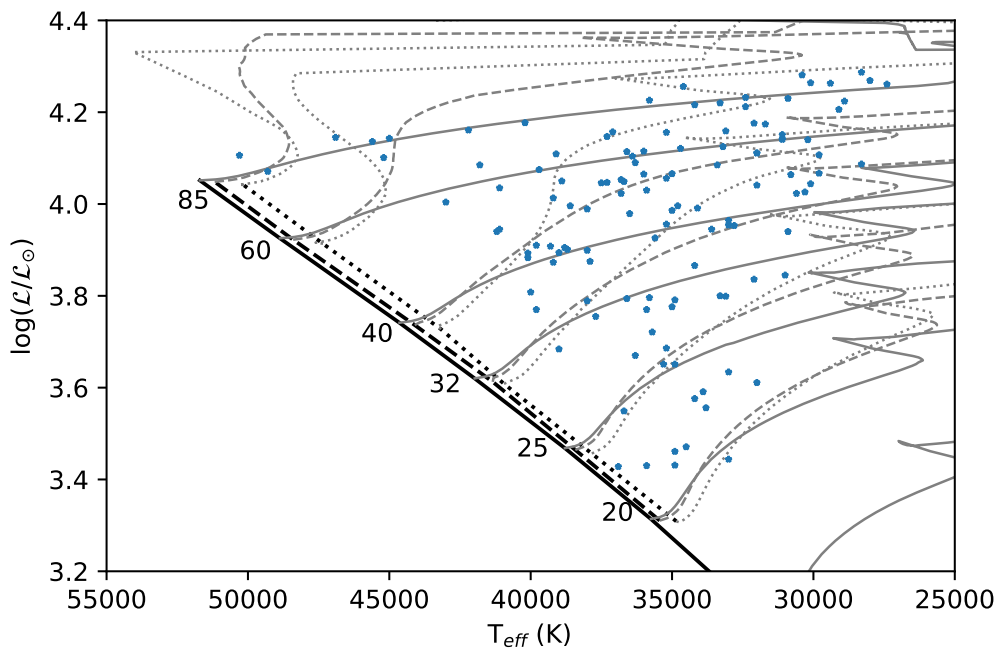
Figure 2.1: *Spectroscopic Hertzsprung-Rusell diagram of the* empirical *sample used in this work. Solid, dashed and dotted line represents the tracks for a star with a ratio of equatorial velocity over the critical velocity (see Chapter 2) of 0,0.2 and 0.4, respectively.*

2.3   The *synthetic* sample

The initial *synthetic* sample is created by using SYCLIST, and is represented in Figure 2.2. As stars evolve faster as the initial mass increase, it is needed to have a smaller difference in

---

[3]These uncertainties for $\log(L/L_\odot)$ do not include the distance uncertainties.

age in the younger clusters in order to have a good sampling of the high luminosity region of the HRD. For this reason the age span is separated into two groups: the first one goes from 1 Myr to 3 Myr in steps of 0.25 Myr while the second goes from 3 Myr to 8.5 Myr in steps of 0.5 Myr. Stars younger than 1 Myr are not computed because this age is the lower limit considered in SYCLIST computations and 8.5 Myr is taken as an upper limit because this age correspond to the turnoff point of stars with mass close to 15 $M_\odot$. Cluster having an age older that 8.5 Myr are hence not expected to have any star above 15 Msol in the main sequence, our main region of study (see Figure 2.1). These clusters follow a Salpeter IMF (Initial Mass Function), meaning that the density of stars in the HRD/sHRD shows a gradient with respect to mass.
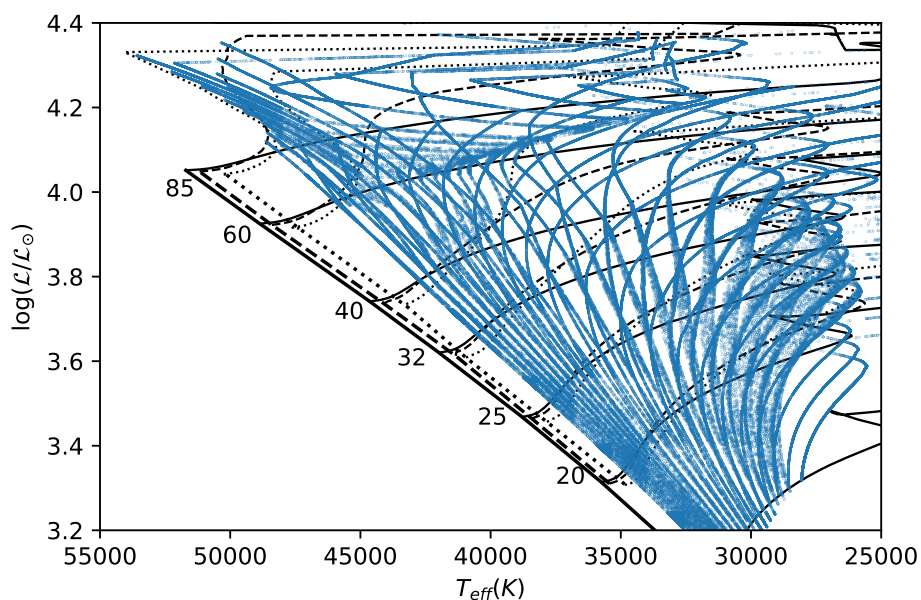
Figure 2.2: *Spectroscopic Hertzsprung-Rusell diagram of the initial* synthetic *sample built joining together stars from all considered clusters (with varying age and initial rotational velocities, see text for explanation). Solid, dashed and dotted line represents the tracks for a star with a ratio of equatorial velocity over the critical velocity of 0,0.2 and 0.4, respectively.*

As the mass is multivaluated in the HRD/sHRD since it depends strongly with v_{eq}/v_c for the higher mass stars, rotation becomes an important parameter to be considered. Because

the evolutionary tracks at the highest masses follow different paths in the HRD/sHRD depending on the value of v_{eq}/v_c, clusters are obtained for the three available v_{eq}/v_c (velocity groups), each one having a total of 20 ages for a total of 60 clusters. This large amount of clusters ensure that all the main region of study (i.e. that covered by the sample of O-type stars presented in Figure 2.1) is well sampled with stars of the three velocity groups.

The *synthetic* sample includes stars from 15 to 85 $M_\odot$ at every evolutionary stage. However, it is needed first to limit the region since the interest of this work are the Main Sequence O-stars. Evolutionary tracks for the highest mass stars return to the region delimited by the *empirical* sample in later evolutionary stages (e.g. Wolf-Rayet stars). A rejection of these evolved stars is made by excluding stars which actual mass is lower than 70%[4] of the initial mass. It is considered a range in T_{eff} of $25\,000 - 55\,000$ K and $3.2 - 4.4$ dex for $\log(\mathcal{L}/\mathcal{L}_\odot)$, being this domain slightly larger than the range of values of the *empirical* sample to minimize border effects in the inference of the outermost stars parameters.

The high density of stars in the low-luminosity regions of the HRD/sHRD of Figure 2.2 (due to the effect of the considered IMF) may lead to an overload of redundant information in the inference and and to an increase in the computation time. To avoid this situation, we decided to make a pre-processing of the initial sample to ensure homogenization of data density across the whole parameter space. This is needed because the star density of the low-luminosity region bins is $\sim 5 \times 10^5$ stars/bin without the homogenization. This homogenization consist in a binning of the HRD/sHRD where a given density of stars inside each bin is forced. This density corresponds to the density of the lowest populated bins in order to maximize the sampling while maintaining density homogeneity between the bins (Figure 2.3). The bin size considered corresponds to a 2σ, being σ the mean value of the uncertainties obtained from the measurements of both T_{eff} and $\log(\mathcal{L}/\mathcal{L}_\odot)$ from the *empirical* sample. This bin sizes are $1\,400$ K and 0.16 dex for T_{eff} and $\log(\mathcal{L}/\mathcal{L}_\odot)$ of, respectively. As the measured rotational velocity of the stars is not the actual v_{eq} but the projected equatorial

---

[4]This remaining mass threshold ensures that only main sequence stars are considered.

velocity ($V sini$), a projection of v_{eq} is being made to match the available empirical variables. This is made by creating a homogeneous distribution in $sin(i)$, being i the angle between the angular velocity of the star and the line-of-sight (LOS), and projecting v_{eq} by multiplying it by $sin(i)$.
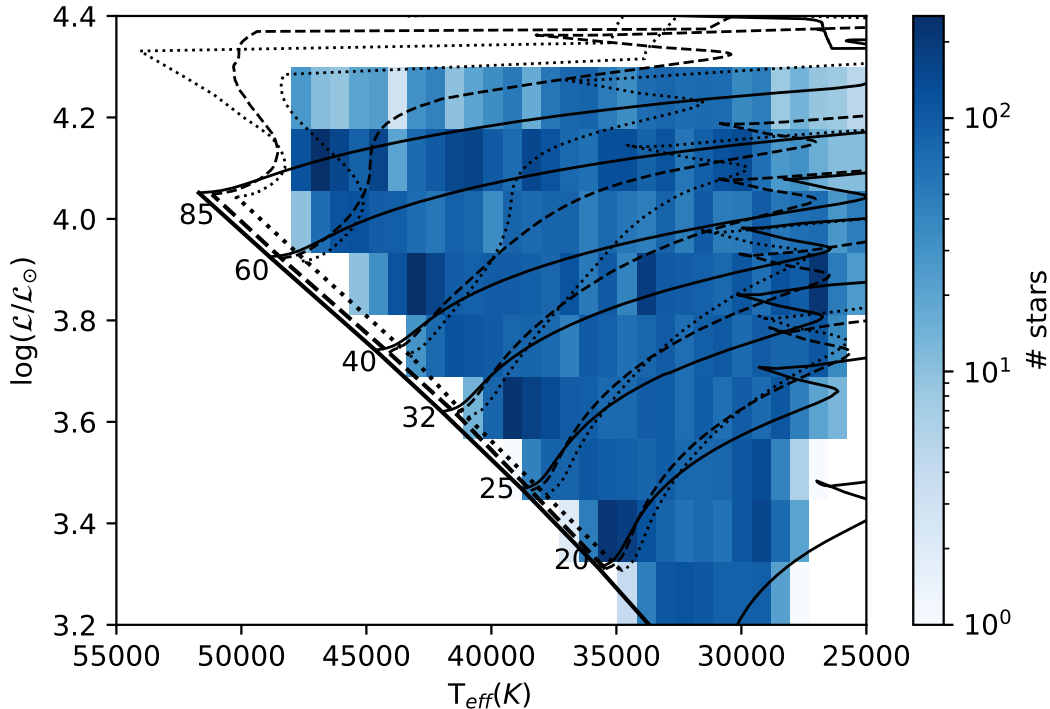
Figure 2.3: *Spectroscopic Hertzsprung-Rusell diagram of the* synthetic *sample showing the binning. The density of stars in each bin is around 200 stars.*

2.4 Training Parameters and methods

It is necessary to state that the training is made using synthetic stars because the number of observed O-type stars is still too sparse to cover the whole parameter space. The input parameters used in the inference are $\log(T_{\text{eff}})$, $\log(\mathcal{L}/\mathcal{L}_{\odot})$ and the $V sini$ from the spectroscopic analysis and $\log(L)$ which is obtained by using the distance to the star. The input parameters $\log(T_{\text{eff}})$, $\log(\mathcal{L}/\mathcal{L}_{\odot})$ and $\log(L)$ are used because they are the variables that form

the HRD and sHRD, and because they are used in the quantitative spectroscopic analysis and track interpolation to derive the values of mass and therefore radius. The $V sini$ is taken to include a component associated with the rotation of the stars.

Three combinations of these input parameters are used to train the ANN. The first training set consist in $\log(T_{\text{eff}})$, $V sini$ and $\log(\mathcal{L}/\mathcal{L}_\odot)$ ($\mathcal{L}$-training); the second uses $\log(T_{\text{eff}})$, $V sini$ and $\log(L/L_\odot)$ (L-training), and the third uses $\log(T_{\text{eff}})$, $V sini$ and both $\log(L/L_\odot)$ and $\log(\mathcal{L}/\mathcal{L}_\odot)$ ($\mathcal{L}\&L$-training). These different training sets are considered to check how having access to empirical information about the stellar luminosities (which is only available when we have reliable information about the distance to the star)[5] modifies the inference results by comparing the inference with $\log(L)$ with the inference with $\log(\mathcal{L}/\mathcal{L}_\odot)$. The last configuration is taken in an attempt of getting better results by increasing the number of input variables since $\log(\mathcal{L}/\mathcal{L}_\odot)$ and $\log(L)$ are obtained through different methods and carry different information. The target variables of the inference are the actual mass (M_{inf}) and radius (R_{inf}). In addition, the stellar luminosity $\log(L/L_\odot)^{inf}$ will be also considered as target variable in those cases in which it is not used as input variable.

The ANN viability study is performed by doing the following inference test, using the three training sets for all the tests. Firstly it is studied which of the following methods provide the lowest systematic errors when applied to a synthetic test sample:

- a training considering the whole region of interest (see Sect. 2.3) where the O-type stars are located (in terms of T_{eff} and $\log(\mathcal{L})$),

- a local training considering only considering a small region in the parameter space around the star whose parameters are going to be inferred.

After choosing the method with the lowest systematic errors, the next analysis consist in a comparison between an inference of synthetic stars and an inference of the same synthetic

---

[5]And therefore a measure of $\log(L)$

14

stars with an added error to their T_{eff}, L and $\log(g)$[6] based on the averaged uncertainties from the *empirical* sample. Lastly, the trained ANN is applied to the *empirical* sample and the inferred target variables are compared with the values obtained by the quantitative spectroscopic analysis.

---

[6]Remember that $\mathcal{L} = T_{\text{eff}}^4/g$. Apart from T_{eff}, variations in $\log(g)$ affects directly into the value of $\log(\mathcal{L}/\mathcal{L}_\odot)$.

Chapter 3

Mass, radius and luminosity inference: intrinsic errors and observational uncertainties.

In this chapter, we investigate the intrinsic errors as well as the effect of considering observational uncertainties associated with the inference of stellar masses, radii and luminosities. These uncertainties sources set the minimum uncertainty obtainable of the ANN used in this work, assuming that stellar evolution models are correct. This chapter presents then three analyses about training the ANN and inferring stellar parameters using the *synthetic* sample described in Sect. 2.3 and the three training sets ($\mathcal{L}$-training, L-training and $L\&\mathcal{L}$-training). The first two analyses are made to quantify the intrinsic errors of two training methods while the last analysis consist in quantifying the effect of the observational uncertainties in the inference. This quantification is made by inducing uncertainties based on the average uncertainties of the *empirical* sample.

3.1 Quantification of intrinsic errors

In order to know which are the intrinsic errors associated to the inference of stellar parameters using ANN, a study is done using the *synthetic* sample (Sect. 2.3). This study consist in using two methods for the training: a global training where its domain in the sHRD is defined by the *empirical* sample (see Sect 2.3) and a local training that uses only neigh-

bouring stars of each problem star. The motivation behind the local training is to check if a training using small grids of stars around each problem star is more efficient than the global training and yield better results. For this reason, a comparison between both methods is made, choosing the best method for the rest of the work.

3.1.1 Global training analysis

For the global training method, the *synthetic* sample is splitted in two groups: the *synthetic* training sample and the *synthetic* test sample (see each of the two samples represented in the sHRD in Figure 3.1). The *synthetic* training sample comprises 90% of the *synthetic* sample and the *synthetic* test sample includes the remaining 10%. The ANN is trained with the *synthetic* training sample and it is applied to the *synthetic* test sample to infer the target variables.
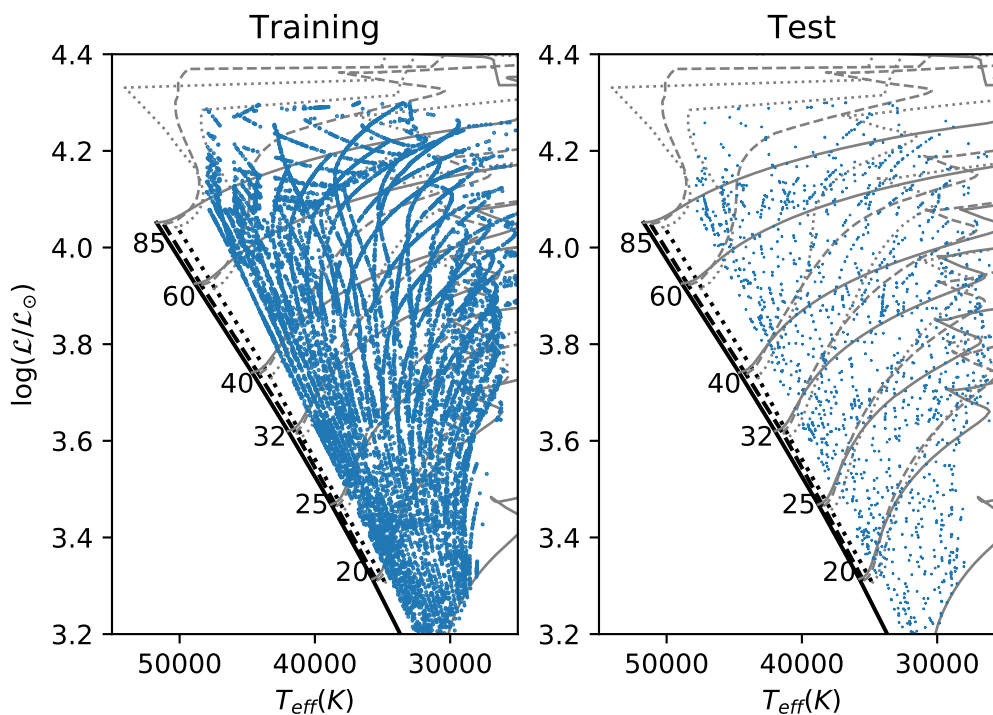

Figure 3.1: *sHRD illustrating the distribution of the the training (left) and test sample (right).*
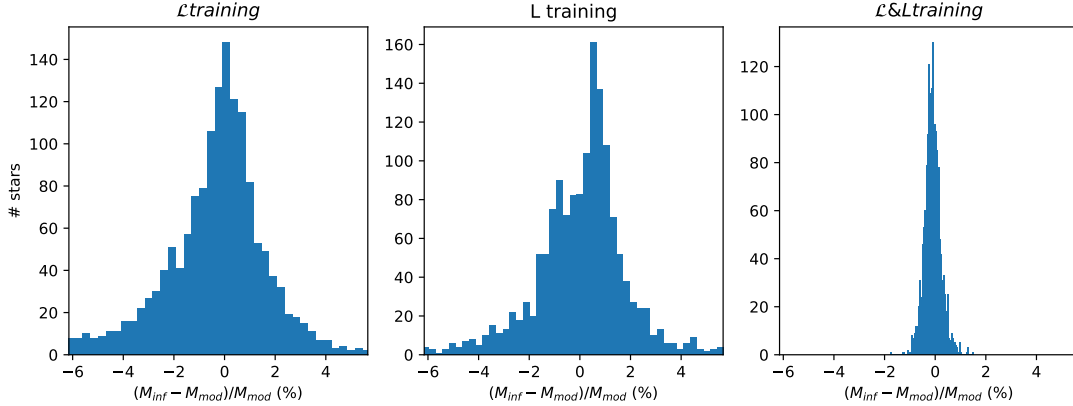
Figure 3.2: *Results of the global training analysis. These histograms show the relative error for the inferred mass (M_{inf}) with respect to the input value given by the models (M_{mod}) for the three types of training variable sets mentioned in Sect. 2.4. The bin size follows the Freedman-Diaconis rule (Freedman & Diaconis 1981).*

The results of the global training method are presented in Figure 3.2, Figure A.1 and Figure A.2 for the mass (M), radius (R) and luminosity ($\log(L)$) respectively. These figures show the error distribution for each target variable. These errors are presented as relative errors for M and R, and the absolute error for the logarithm of the luminosity [1]. Mean values and standard deviations of the distributions are summarized in Table 3.1.

It can be seen in Figure 3.2 that the training considering both types of luminosities gives the distribution with the lowest intrinsic errors in the mass inference. This is seen quantitatively in the values given by the Table 3.1. The standard deviation of the mass inference is similar when using only one type of luminosity and is one order of magnitude lower in the $\mathcal{L}\&L$-training. The improvement in the intrinsic errors associated with the $\mathcal{L}\&L$-training inference is caused by the increase of information of the stars (since both $\log(L)$ and $\log(\mathcal{L})$ are considered). The mean values of the distributions are ~ 0 meaning that there is not any significant global trend in under/overestimating the value of the mass.

The inclusion of $\log(L)$ in the inference results in a remarkable improvement in the inference of the radius (see L-training and $\mathcal{L}\&L$-training results in Table 3.1 and figure A.1. This

---

[1]The absolute error of $\log(L)$ shows the relative error of L since: $\Delta log(L) \approx \Delta L/L$

Target variable	$\mathcal{L}$-training		L-training		L & $\mathcal{L}$-training	
	$<x>$	σ	$<x>$	σ	$<x>$	σ
$\Delta M/M(\%)$	0	3	0	3	-0.1	0.3
$\Delta R/R(\%)$	-1	3	0.2	0.9	0.0	1.1
$\Delta \log(\mathrm{L}/L_\odot)$ (dex)	0.00	0.03				

Table 3.1: *Mean ($<x>$) and standard deviation (σ) of the relative error distribution of each inferred stellar parameters for all the training sets. Note that the absolute error of $log(L)$ shows the relative error of L since $\Delta log(L) \approx \Delta L/L$.*

is because the luminosity, radius and effective temperature are correlated by the equation of the luminosity (eq. 3.1):

$$\frac{L}{L_\odot} = \left(\frac{R}{R_\odot}\right)^2 \left(\frac{T_{\mathrm{eff}}}{T_{\mathrm{eff}_\odot}}\right)^4 . \qquad (3.1)$$

As the value of the intrinsic error in the radius is $\sim 2\%$ lower for both L-training and $\mathcal{L}\&L$-training than the value given by $\mathcal{L}$-training, the two first mentioned training sets are the more convenient to infer the radius.

It is also worth to study how the intrinsic errors varies with respect to the position of the test stars in the sHRD. In order to quantify these variations, the *synthetic* test sample is divided in two groups, the upper sHRD stars (up-sHRD) and the lower sHRD (low-sHRD) stars. These groups contain those star that have $\log(\mathcal{L}/\mathcal{L}_\odot)$ greater than 3.8[2] dex and $\log(\mathcal{L}/\mathcal{L}_\odot)$ lower than 3.8 dex respectively. Histograms of the relative error of the target variables are obtained for these two groups and the complete test sample. They are represented in Figure 3.3, Figure A.3 and Figure A.4 for inference of the mass, radius and luminosity respectively. Mean values and standard deviations of these distributions are summarized in Table 3.2.

The mass inference is underestimated in up-sHRD stars (red) respect to low-sHRD stars (blue) in all the training sets. The standard deviations are higher in the up-sHRD than in the low-sHRD because evolutionary tracks are less affected by v_{eq} in the low-sHRD region,

---

[2]This value is taken because its approximately the mean value of the spectroscopic luminosity domain.
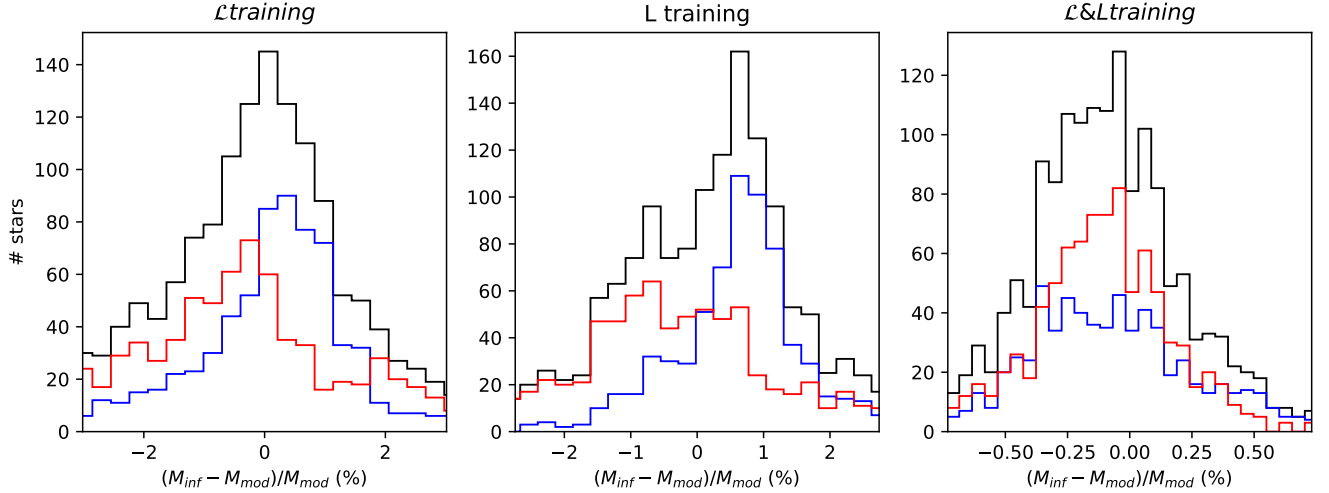
Figure 3.3: *Histograms representing the distribution of intrinsic uncertainties associated with the inference of the mass using the three considered training sets. Red/blue distributions represent the case of the stars located in the upper/lower part of the sHRD (see Figure 3.1 right panel). In black results for the global sample.*

Target variable	$\mathcal{L}$-training		L-training		L & $\mathcal{L}$-training	
Up-sHRD stars	$<$x$>$	σ	$<$x$>$	σ	$<$x$>$	σ
$\Delta M/M(\%)$	-1	4	0	4	-0.3	0.3
$\Delta R/R(\%)$	-1	4	0.2	0.7	-0.0	1.1
$\Delta \log(\mathrm{L}/L_\odot)$ (dex)	0.00	0.04				
Low-sHRD stars	$<$x$>$	σ	$<$x$>$	σ	$<$x$>$	σ
$\Delta M/M(\%)$	0.2	1.4	0.6	1.0	-0.1	0.4
$\Delta R/R(\%)$	-0.5	1.2	0.3	1.1	-0.1	1.0
$\Delta \log(\mathrm{L}/L_\odot)$ (dex)	-0.003	-0.017				

Table 3.2: *Mean ($<$x$>$) and standard deviation (σ) of the relative error distribution of each inferred stellar parameters for all the training sets. This is calculated for the up-sHRD stars and the low-sHRD stars.*

as it corresponds to the least massive stars. The radius inference has similar results for the *L*-training and the *L&$\mathcal{L}$*-training between both groups. In the case of the *L*-training the inference of the radius is worse for the up-sHRD as the mean value and deviations of the relative error are higher in this case. The luminosity inference behaves similarly to the other two target variables: the intrinsic errors are much lower in the low-sHRD. The explanation

found for this is that low-luminosity stars are less affected by rotational effects; thus the values of the effective temperature and both luminosities for a given mass do not differ as much as for the high-luminosity stars.

From these results we conclude that there are trends in the inference. The target variables of the stars are underestimated in the up-sHRD region when compared to the low-sHRD region. Also the low-sHRD stars standard deviations are smaller than those from the up-sHRD stars.

3.1.2   Local training method

The other method proposed for the inference of the target variables is the local training. For this method an specific training sample is made for each problem star. Each training sample consist in all the training stars inside a box in the sHRD centered in the value of T_{eff} and $\log(\mathcal{L})$ of the star and with a width of $1\,400$K for T_{eff} and 0.16 dex for the $\log(\mathcal{L})$. This width corresponds to a 2σ value of the uncertainties of the *empirical* sample (Sect. 2.2).

This training method needs to create an specific sample for each star and the sample creation process is not automatized. Thus, a group of eight mock *synthetic* stars that are distributed across the sHRD are used to analyze this method (Figure 3.4). These mock stars positions in the sHRD were chosen to cover appropriately all the domain. This analysis is also made using the three training sets ($\mathcal{L}$-training, L-training and $L\&\mathcal{L}$-training). The results of the inference of these *synthetic* mock stars are summarized in Table 3.3.

Results presented in Table 3.3 show different behaviours depending on the target variable and the used training set. The $\mathcal{L}$-training shows that for the mass the relative errors are lower for stars with lower T_{eff} (excluding S6). For the radius, the trend observed is that the relative error gets smaller as $\log(\mathcal{L})$ and T_{eff} are higher, i.e. as the stellar mass increases. The values of $\log(L)$ do not show any clear behaviour.

In the L-training, the mass and radius inferences behave similarly to those from the $\mathcal{L}$-training. However, the $L\&\mathcal{L}$-training shows the opposite behaviour in the mass inference
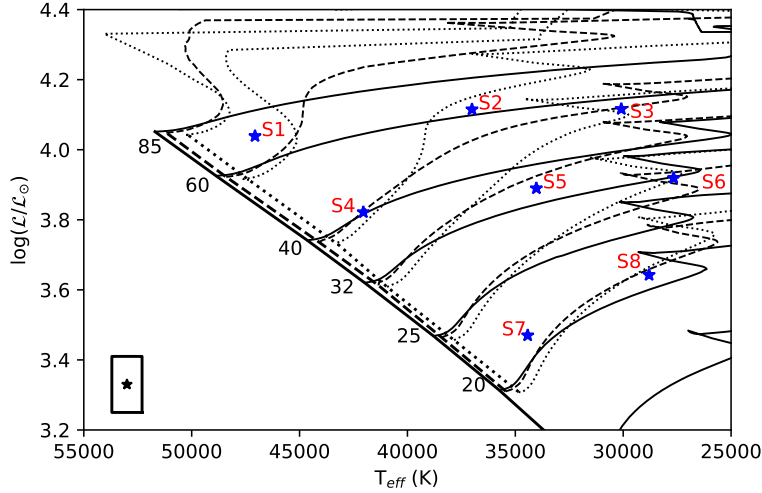
Figure 3.4: *Positions in the sHRD of the eight stars used in the local training method analysis. The box in the bottom-left corner show the training box size which is centered at the star position.*

$\mathcal{L}$-training									
	S1	S2	S3	S4	S5	S6	S7	S8	Average
$\Delta M/M(\%)$	7	5	-3	4	0.4	11	27	0.8	7±3
$\Delta R/R(\%)$	4	-14	-7	16	18	-3	17	19	6±5
$\Delta log(L/L_\odot)$	0.12	0.10	-0.024	0.3	-0.06	0.1	1.4	0.04	0.25±0.17
L-training									
	S1	S2	S3	S4	S5	S6	S7	S8	Average
$\Delta M/M(\%)$	13	3	-4	6	-0.4	0.6	30	-0.7	6±4
$\Delta R/R(\%)$	-5	-1.9	-4	8	14	-0.8	15	15	6±3
$L\&\mathcal{L}$-training									
	S1	S2	S3	S4	S5	S6	S7	S8	Average
$\Delta M/M(\%)$	13	-1.0	-4	13	0.6	0.5	-1.8	-1.5	2.4±2.4
$\Delta R/R(\%)$	1.1	-2.9	-3	-18	13	2.5	30	11	4±5

Table 3.3: *Relative errors of the target variables using the local training method for the three training sets described in Sect. 2.4 of the stars represented in Figure 3.4.*

for the stars with lower masses (S7 and S8) while for the remaining stars the behaviour is the same as the other two training sets.

One interesting case to be mentioned is the inference results for S7. The relative errors of the mass inference in the *L*-training and $\mathcal{L}$-training are much higher that for the rest of

the stars. It can be thought that the local training sample used with this star is wrong. In order to check if the initial local training sample of S7 is wrong, other training samples were obtained giving similar results. The errors of the target variables may vary a lot depending of the position of the star (e.g. S5 and S6 radius inference for all training sets), so **we conclude that the local training highly depends on the position in the sHRD of the star.**

It can be concluded then that this method of training may give intrinsic errors that are much higher than the results obtained in the global training. They do not follow an specific trend, so these types of results can not be avoided.

$\mathcal{L}$-training								
	S1	S2	S3	S4	S5	S6	S7	S8
$\Delta M/M(\%)$	3	-0.15	-4	0.04	0.9	1.2	-0.9	-0.3
$\Delta R/R(\%)$	2.4	5	-6	0.10	3	2.2	-0.16	0.05
$\Delta log(L/L_\odot)$	-0.02	0.016	-0.05	-0.02	0.015	-0.004	0.010	0.023
L-training								
	S1	S2	S3	S4	S5	S6	S7	S8
$\Delta M/M(\%)$	7	3	-5	-0.03	1.1	1.0	-2.4	-0.78
$\Delta R/R(\%)$	-1.2	-0.5	-0.9	0.4	-0.7	0.03	0.8	-0.7
$L\&\mathcal{L}$-training								
	S1	S2	S3	S4	S5	S6	S7	S8
$\Delta M/M(\%)$	0.23	0.3	-0.09	0.09	-0.3	0.021	-0.11	0.4
$\Delta R/R(\%)$	1.1	0.04	-1.2	0.011	0.7	0.4	3	1.2

Table 3.4: *Relative errors of the target variables using the global training method for the three training sets described in Sect. 2.4 of the stars represented in Figure 3.4.*

A comparison of these results to those obtained by applying the global training method to these stars (Table 3.4) is made. It is seen that the intrinsic errors in the case of the global training do not show values of the relative errors one order of magnitude higher than the rest of stars. They are therefore more consistent. Other critical difference between both methods is the computation time: the training time for the global training method is ∼22 seconds while for each specific training sample of the local training method it is around ∼8 seconds. This means that for a problem sample of more than 2 stars the global training become the

most time-efficient method. **Because of the more consistent intrinsic errors and the more efficient time usage, it is concluded that the global training is better than the local training.** For this reason, the global training is considered in the forthcoming analyses.

3.2 Inference of the *synthetic* sample with added observational errors.

While the objective of the analysis presented in Sect. 3.1 was to find the lowest intrinsic errors among two training methods and three training sets, this section presents the results of the analysis of how observational errors propagates through the inference.

For this exercise we build a new test sample in which each data-point in the original test sample (see Sect. 3.1.1) is modified by adding a value randomly obtained from a Gaussian distribution for each dimension. The standard deviation of these distributions are the corresponding averaged uncertainties of the *empirical* sample (see Sect. 2.2). These mean uncertainties are 700K for T_{eff}, 0.02 for $\log(L)$ and 0.08 for $\log(\mathcal{L}/\mathcal{L}_\odot)$. The new test sample is generated by adding these randomly generated errors to the values of the test sample defined in Section 3.1.1.

The training is done with the same training sample than in Sect 3.1.1 and it is applied to the new test sample (see Figure 3.5). The results are showed in Figure 3.6, Figure A.5 and Figure A.6 for the inference of the mass, radius, and luminosity, respectively. Mean values and standard deviations of the relative error distribution are summarized in Table 3.5. Although both luminosities are related by the mass[3], they are obtained through different methods (quantitative spectroscopy and by using distances for $\log(\mathcal{L})$ and $\log(L)$ respectively). For this reason, the errors in $\log(L)$ and $\log(\mathcal{L})$ are added independently. This implies that the new values of the luminosities may contradict the relationship between $\log(\mathcal{L})$ and $\log(L)$ that the ANN has made during the training, and hence the $L\&\mathcal{L}$-training

---
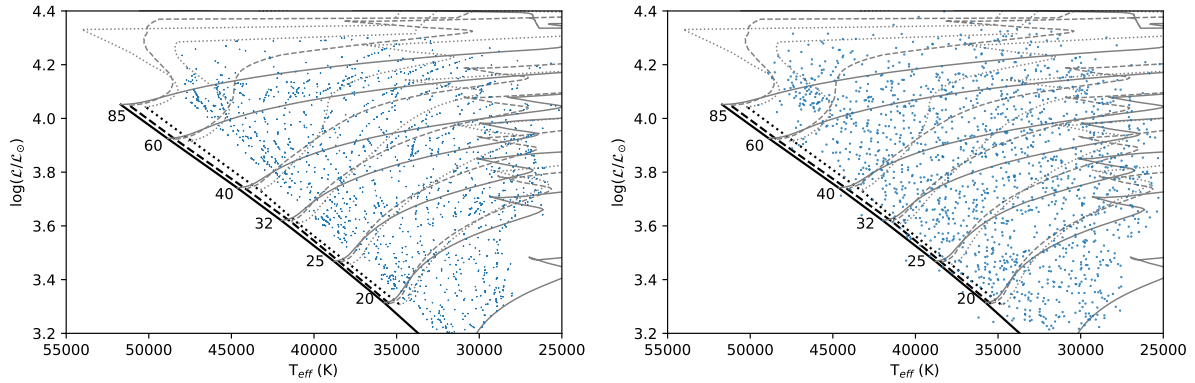
[3]$\mathcal{L}=L/M$ in solar units.

Figure 3.5: *(Left) Sample used in the analyses of intrinsic errors and (Right) same sample with added errors for the analysis of inference with uncertainties. Note that the left figure shows the structure of the clusters that forms the initial* synthetic *sample while the right figure stars presents the dispersion introduced by adding errors.*

worsen when compared to the $\mathcal{L}$-training and the L-training. In this test the training with the lowest uncertainties is the one that uses L. This is a expected result because the uncertainty in $\log(L)$ is four times lower than the uncertainty in $\log(\mathcal{L})$.
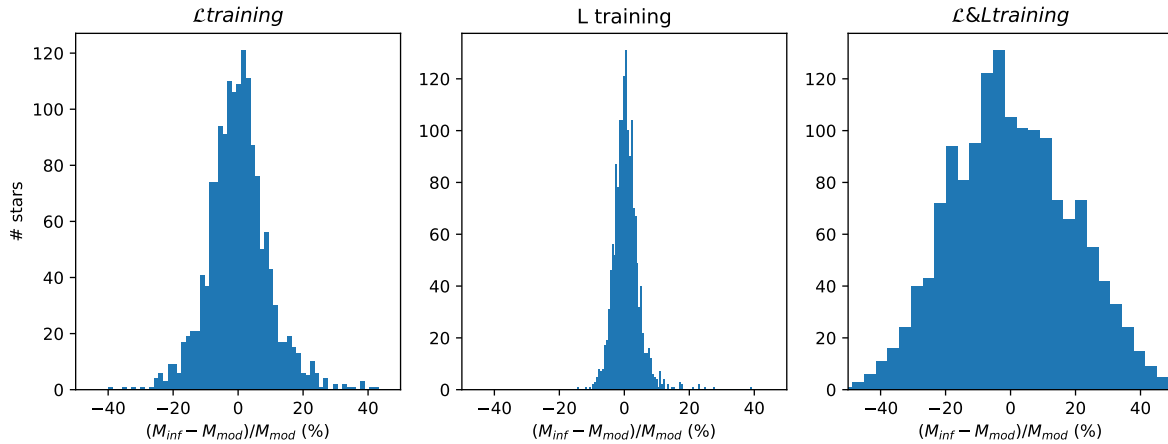
Figure 3.6: *Results of the global training with added observational errors. These histograms show the relative error of the inferred mass (M_{inf}) with respect to the input value given by the models (M_{mod}) for the three types of training variable sets mentioned in Chapter 2. The bin size follows the Freedman-Diaconis rule (Freedman & Diaconis 1981).*

After comparing these results with those obtained in Section 3.1.1, **it can be concluded**

Target variable	$\mathcal{L}$-training		L-training		$L\&\mathcal{L}$-training	
	$<$x$>$	σ	$<$x$>$	σ	$<$x$>$	σ
$\Delta M/M(\%)$	0	9	-1	4	-1	19
$\Delta R/R(\%)$	0	13	0	4	0	6
$\Delta \log(\mathrm{L}/L_\odot)$ (dex)	0.00	0.11				

Table 3.5: *Mean and standard deviation of the relative error distribution of each target variables for all the training variable sets defined in Chapter 2 with added observational errors.*

that as the L-training is the least affected by observational uncertainties, it results in the best training. This is because the observational uncertainty of $\log(L)$ is smaller than that of $\log(\mathcal{L})$. Since the uncertainty of $\log(L)$ considered in the previous test do not take into account distance uncertainties it is interesting to repeat the analysis after calculating the typical value for the uncertainties of $\log(L)$ when the typical distance uncertainty is considered.

In order to obtain the typical uncertainty of $\log(L)$, it is needed to obtain a relationship between the distance and the luminosity. This relationship can be derived from the following equations:

$$M_V = V - 5\ log(d) + 5, \tag{3.2}$$

$$M_V - M_{V\odot} = -2.5\ log(L/L_\odot), \tag{3.3}$$

where eq. 3.2 is the distance modulus and eq. 3.3 is the subtraction of the absolute magnitude in V (M_V) of a star and the solar absolute magnitude in V-band. The distance is noted as d and the apparent magnitude in V-band as V.

The errors propagation of eq. 3.2 and eq. 3.3 are:

$$\Delta M_V = \Delta V - \frac{5}{ln10}\frac{\Delta d}{d}, \tag{3.4}$$

$$\Delta log(L/L_\odot) = \frac{\Delta M_V}{2.5}. \tag{3.5}$$

Taking into account that $\Delta V \sim 0.01$ mag (which is negligible), and substituting eq. 3.4 in eq. 3.5 we obtain:

$$\Delta log(L/L_\odot) \approx 0.9\frac{\Delta d}{d} \tag{3.6}$$

which gives a value of $\Delta log(L/L_\odot) \approx 0.09$ dex by asuming that the values of the relative error of the distance given by GAIA for O-type stars are $\frac{\Delta d}{d} \sim 10\%$[4] (Holgado et al. in prep.).

The inference results when considering an error for $\Delta log(L/L_\odot) \approx 0.09$ dex summarized in Table 3.6 and the corresponding histograms for the mass and the radius are presented in Figure 3.7 and Figure A.7, respectively. The $\mathcal{L}$-training is not considered in this part of the analysis because it is not affected by any modification of the considered uncertainties associated with the luminosity when building the test sample.
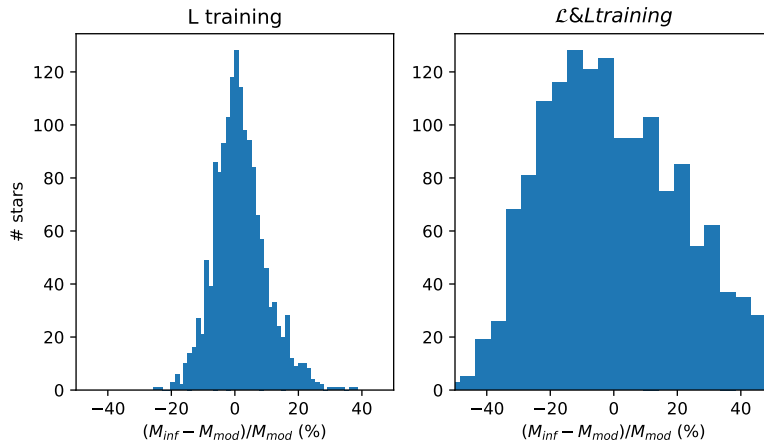

Figure 3.7: *Representation of the relative error distribution of the mass for the L-training and the L&$\mathcal{L}$-training considering $\Delta log(L/L_\odot) = 0.09$ dex. The bin size follows the Freedman-Diaconis rule (Freedman & Diaconis 1981).*

---

[4]The relative error actually goes from 5% to 15% so it is taken the mean value as a typical uncertainty.

Target Variable	L-training		$L\&\mathcal{L}$-training	
	$<$x$>$	σ	$<$x$>$	σ
$\Delta M/M(\%)$	-1	7	0	30
$\Delta R/R(\%)$	-1	11	-1	9

Table 3.6: *Mean and standard deviation of the relative error distribution of each target variable for all the training variable sets defined in Chapter 2 when observational errors are added, and including $\Delta\log(L/L_\odot) = 0.09$ dex.*

The comparison between Table 3.5 and Table 3.6 shows, on the one side that a change of 0.07 in $\Delta\log(L/L_\odot)$ worsen the target variables uncertainties in the L-training (Table 3.6) up to values slightly lower (2% of difference) to those of the $\mathcal{L}$-training (Table 3.5). On the other side, the $L\&\mathcal{L}$-training results in a much worse mass inference, having an uncertainty of 30% (compared to 19% obtained for $\Delta\log(L/L_\odot) = 0.02$ dex). The radius inference does not change significantly meaning than the best training set to infer the radius is the $L\&\mathcal{L}$-training.

The $L\&\mathcal{L}$-training gives the best inference in radius but since its mass inference errors are more than three times larger than for the other training sets, it is not practical to use it. **We conclude that in the case of *synthetic* stars with added observational errors the best training sets are the L-training and $\mathcal{L}$-training.**

From the results obtained in this section and those presented in Table 3.1, it can be quantified the contribution of the added observational errors in the global training. On the one side the values provided in Table 3.1 are the systematic uncertainties. On the other side the values quoted in Table 3.5 (for $\mathcal{L}$-training) and Table 3.6 (for L-training and $L\&\mathcal{L}$-training) represent the combined effect of intrinsic errors and observational uncertainties. Hence, subtracting the relative errors obtained in Sect. 3.1.1 to those of this section yields the distribution of added observational errors (Figure 3.8, Figure A.8 and Figure A.9 for the mass, radius and luminosity, respectively). The distributions mean value and standard deviations are summarized in Table 3.7.

After comparing this results with those obtained in Table 3.5 and Table 3.6, we see
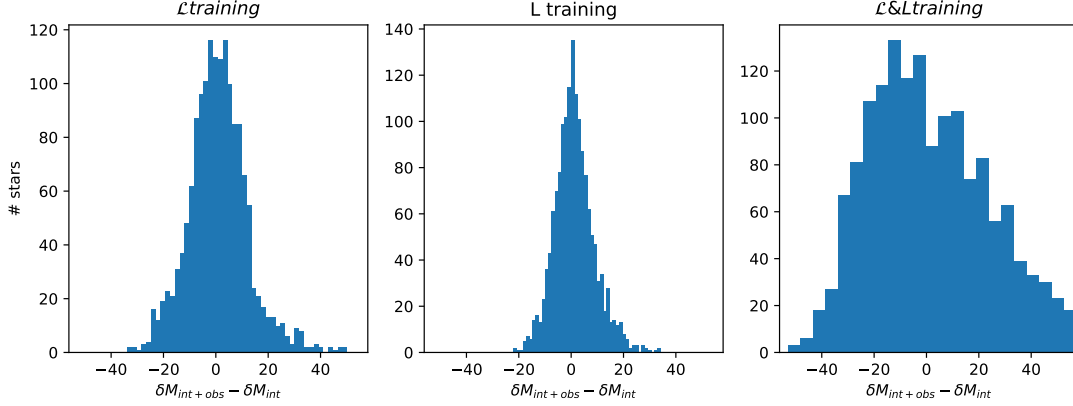
Figure 3.8: *Relative error (e.g. δM_{int} for the mass intrinsic errors) distributions of the added observational errors for the mass inference, for the three types of training variable sets mentioned in Chapter 2. The bin size follows the Freedman-Diaconis rule (Freedman & Diaconis 1981).*

Target variable	$\mathcal{L}$-training		L-training		$L\&\mathcal{L}$-training	
	$<x>$	σ	$<x>$	σ	$<x>$	σ
$\Delta M/M(\%)$	1	9	1	7	3	25
$\Delta R/R(\%)$	3	13	0	11	0	9
$\Delta \log(L/L_\odot)$ (dex)	0.00	0.11				

Table 3.7: *Standard deviation associated to the added observational errors distribution.*

that all the standard deviations but $\mathcal{L}$-training and $L\&\mathcal{L}$-training mass inference are equal. These two last mentioned inferences have standard deviations that are very close to the value obtained by considering intrinsic and added observational errors. These values for the $\mathcal{L}$-training mass inference are are 7% and 9% (for intrinsic and intrinsic+observational, respectively).

We conclude then than when considering intrinsic errors and observational errors, the training with the lower uncertainties is the _L_-training. Nevertheless, since the uncertainties of the $\mathcal{L}$-training are slightly worse ($\sim$ 2%), it proves to be a good alternative to the _L_-training when there is no a reliable measurement of the distance. We also conclude that the added observational errors are dominant in the inference.

Chapter 4

Application of ANN to the *empirical* sample.

In this chapter, we move a step further in our investigation of the applicability of ANN to infer stellar parameters by considering what we have called the *empirical* sample (see Chapter 2). We compare the results using the $\mathcal{L}$-training and L-training for the *empirical* sample with those obtained by Holgado et al. (2020) using an spectroscopic quantitative analysis. In order to understand how the spectroscopic results were obtained, a brief description is included before the analysis.

The spectroscopic analysis uses stellar atmosphere models to create a grid of synthetic spectra that are used to fit the problem spectra (e.g. Simón-Díaz et al. 2011). The main parameters obtained are effective temperature and gravity. These two parameters can be then complemented with other such as the luminosity, radius and mass only in the case that we have access to the distance to the star. In particular, the stellar luminosity can be obtained if we have empirical information about the bolometric magnitude, which can be obtained from the apparent visual magnitude, the extinction, the bolometric correction (which is a function of the effective temperature) and the distance. The stellar radius can be obtained by using the expression of the luminosity in terms of the radius and effective temperature (eq. 4.1) and hence the mass (eq. 4.2) (Herrero et al., 1992):

$$\log\left(\frac{L}{L_\odot}\right) = 2\log\left(\frac{R}{R_\odot}\right) + 4\log\left(\frac{T_{\text{eff}}}{T_{\text{eff}\odot}}\right) \tag{4.1}$$

$$\log\left(\frac{M}{M_\odot}\right) = 2\log\left(\frac{R}{R_\odot}\right) + \log\left(\frac{g}{g_\odot}\right) \tag{4.2}$$

which are defined as spectroscopic radius and spectroscopic mass respectively to be distinguished from those obtained through other methods.

4.1 Global training applied to the *empirical* sample.

Since the training sets resulting in the lowest uncertainties in the inference are $\mathcal{L}$-training and L-training (Sect. 3.2), we used them in this test. The results of the inference are shown in Figure 4.1, where the inferred values of the target variables (left column), their values obtained by Holgado et al. (2020) using a spectroscopic quantitative analysis (middle column) and the relative comparison between both results (right column) are represented.

The $\mathcal{L}$-training result in inferred evolutionary masses underestimated when compared to the evolutionary tracks (see top left panel in Figure 4.1). It is likely due to the uncertainties in $\log(\mathcal{L})$ since as the stars are more massive, their tracks are more horizontal meaning that this uncertainties in $\log(\mathcal{L})$ enclose a wide range of mass. This was initially thought to be a border effect in the inference domain. In order to confirm this, the mass range was extended up to 120 $M_\odot$; however, as the results did not change when extending the domain, this is not the cause of the underestimation.

The top-middle panel of Figure 4.1 presents the spectroscopic mass estimated by Holgado et al. (2020), and top right panel of Figure 4.1 the comparison between the inferred mass and the spectroscopic mass. The comparison between both spectroscopic and inferred mass shows that most stellar masses are highly overestimated (from 30 000 to 36 000 K) or highly underestimated (from 36 000 to 50 000 K). This discrepancy in the mass obtained from the quantitative spectroscopy and from the inference using the evolutionary models is an open question present in the stellar physics. The so-called mass discrepancy was addressed by
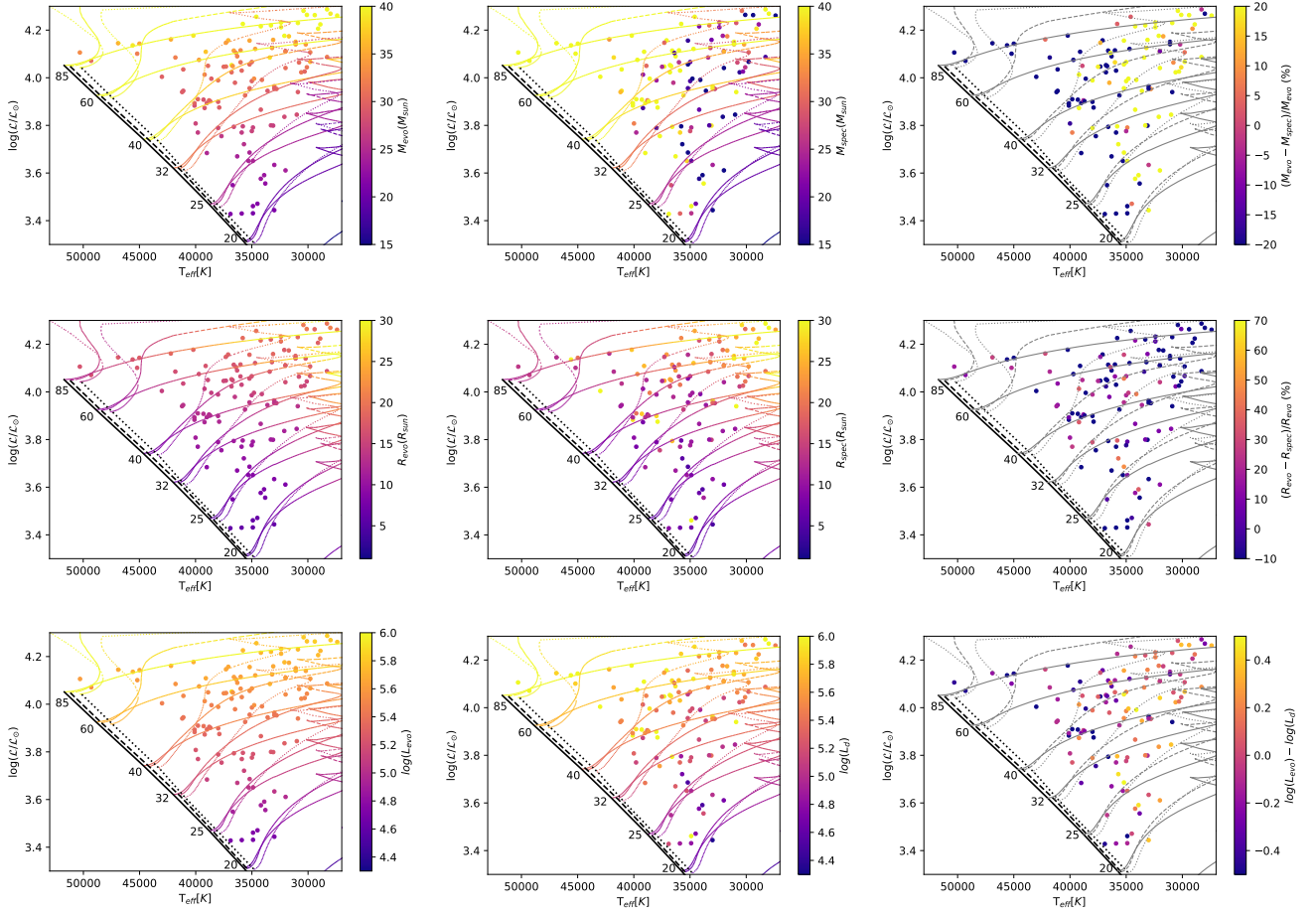
Figure 4.1: *Comparison of the target variables inferred using the $\mathcal{L}$-training set (left column) and the corresponding values using quantitative spectoscopic analysis (middle column). The right column shows the difference between both methods. Solid lines represent ZAMS and tracks of $v/v_v = 0$, dashed lines to $v/v_v = 0.2$, and dotted lines to $v/v_c = 0.4$. Since this training does not use L, it is also inferred.*

Herrero et al. (1992). Although it is not clear which of quantitative spectroscopy and evolutionary models is the cause of this, it is though that the mass discrepancy depends on the assumptions made to describe the physical processes of the stellar evolution (Markova et al., 2018).

Radius inference results are in general underestimated across the sHRD when compared to the spectroscopic results, which is more clear for those of higher masses evolved stars

(the top-right corner of the middle-right panel in Figure 4.1). In the spectroscopic results (central panel of Figure 4.1), there are few stars that are being shown with radius much larger than the evolutionary tracks close to them. Far from meaning that these are very high-mass stars, this is more related to a wrong measure in the distance of the stars. Since the distance is needed to compute the luminosity and by knowing the luminosity the mass and radius are computed, a wrong measure of the distance leads to values of these variables inconsistent with the prediction from adjacent evolutionary tracks.

The inferred luminosity (bottom pannels of Figure 4.1)) has a more stable behavior at higher $\log(\mathcal{L})$ than the other target variables and is more consistent with the evolutionary tracks.
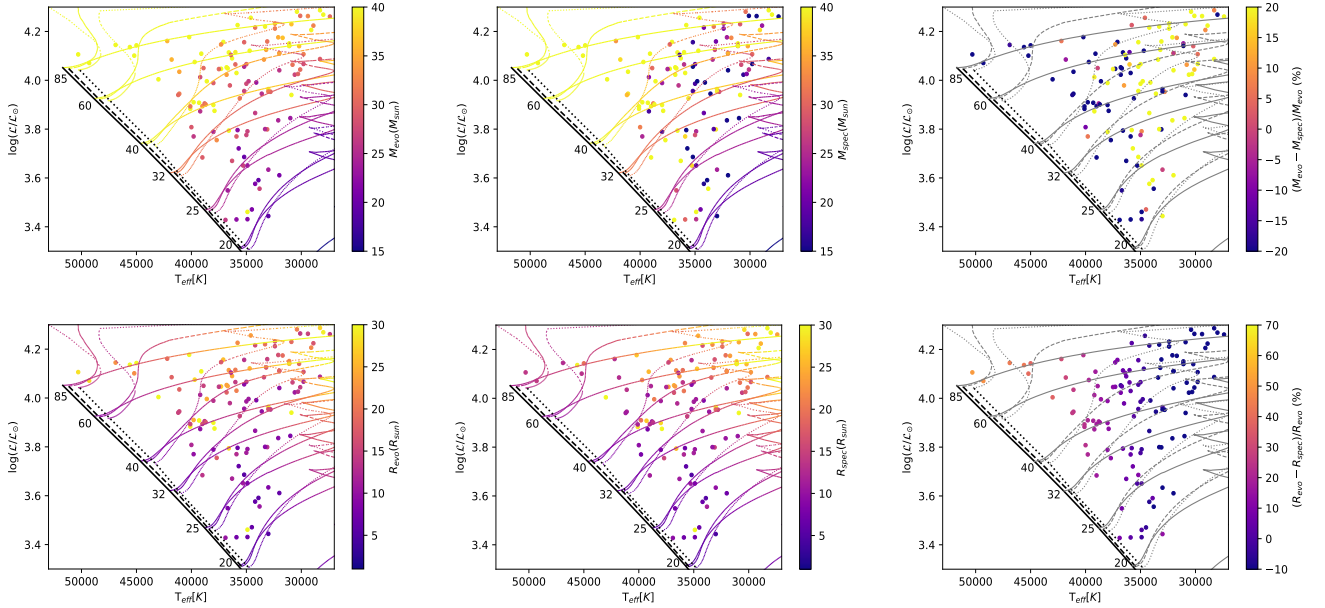
Figure 4.2: *Comparison of the target variables inferred using the* L*-training set (left column) and the corresponding values using quantitative spectroscopic analysis (middle column). The right column shows the difference between both methods. Solid lines represent ZAMS and tracks of $v/v_v = 0$, dashed lines to $v/v_v = 0.2$, and dotted lines to $v/v_c = 0.4$.*

When using L-training instead of $\mathcal{L}$-training there are some stars without a reliable measurement of their distances, meaning that their luminosities are not accurate. Figure 4.2

33

top-right panel shows that the inferred masses have two different trends depending on the region: for evolved stars their inferred masses are overestimated while for the younger stars is the contrary.

Radius on its part shows a clear trend when comparing both methods (middle-right panel in Figure 4.2). The stellar radii are mostly overestimated. This overestimations is lower for more evolved stars (those with cooler temperatures). The effective temperature at which the difference is close to zero is around 35 000 K, meaning that the trend in radius is anticorrelated with the inferred mass. The overestimation of the mass is associated with an underestimation of the radius in the evolved stars region and conversely when approaching the ZAMS.

The results of the inference can be compared with the evolutionary tracks and the color scale of Figure 4.1 and Figure 4.2. The $\mathcal{L}$-training inferred masses are clearly underestimated (top-left pannel in Figure 4.1 as the high mass stars around the 85 $M_\odot$ track have inferred masses around 30 $M_\odot$. The reason found to explain this is that the *synthetic* training sample is not appropriate to study real stellar populations yet. This is because:

- There are only three values of v_{eq}/v_c considered in the training. As v_{eq}/v_c is a continuous variable, it is needed to have a better sampling in this variable, which projection is an input in the training.

- The training sample was built using solar metallicity. As the value of the metallicity increases, the stellar effective temperatures become smaller.

- There are discrepancies between the models and the observational data due to missing ingredients in the stellar evolution theory.

The first and the third reason are the most critical for the inference. The first is important because the effect of the equatorial velocity critically affects to the evolutionary tracks, while the cooling effect caused by the variations of the metallicity is not that decisive. These limitations lie in the improvement of the stellar-evolution models, since SYCLIST

only provide three value of metallicity and v_{eq}/v_c for each (Georgy et al., 2014). Once more values of v_{eq}/v_c can be provided as more metallicities can be considered, as well as the models improve by fixing their flaws, it is expected to get then better results for the inference applied to observed stars.

4.2 Future improvements

In general, it was seen that for a sample of real stars the inference work correctly, although there are discrepancies between the inference results and the spectroscopic results which are caused by the assumptions in the models. This is caused by few factors: the equatorial velocity in the Geneva models drops rapidly into slow rotators, which means that the mass loss due to stellar winds occur in a short period of time. This is a problem in the models because measured projected equatorial velocities are far from the low values expected from these models, and is one of the parameters that do not correspond with observations. The number of critical velocities used in this work are the three available, but since the rotational velocity is a continuous variable, more rotations are needed.

It is also important to state that these stellar models are for single stars. In the case of multiple systems new evolutionary tracks are needed since the multiplicity may affect the stellar parameters of their individual components. One of this effects are an increase in the mass obtained through quantitative spectroscopy for a given star due to having a higher value of the luminosity as part of the companion light may be measured. For unresolved multiple systems this can lead to very high mass measurements. Other effects that may be present is the modification of the stellar rotation rates caused by tidal effect, mergers or mass transference, affecting to the evolution of the individual stars (see e.g. Sana & Evans, 2011)

To improve this method from what has been presented in this work, few ideas are proposed. The first one is to create a highly homogeneous sample in the sHRD by using a constant IMF so all regions can be equally sampled, considering also that for the considered

time span more ages are needed to fill the sHRD. The number of *synthetic* stars used in this work are 5.89 millions of stars, which were needed to fill the upper part of the sHRD, where the frequency of stars is much lower than at lower initial masses. The constant IMF allow to consider much less stars while considering more values of v/v_c, more ages and even more metallicities. Since these clusters were all computed for a solar metallicity, it is needed to consider at least a range of metallicities similar to the ones of the problem sample, as different metallicities move stars to cooler temperatures in the sHRD.

Chapter 5

Conclusions

This work shows that ANN are a tool to be considered when determining those stellar parameters based on the outcome of stellar evolutionary models (such as, e.g., the mass). The main advantage presented by this method is the adaptability of the model, meaning that as long as effective temperature, equatorial projected velocity and both types of luminosities are given, it is possible to train it and do the inference.

The methodology of the ANN has proven to give acceptable results considering the intrinsic and added observational errors. However, the lack of tools to create a better sample and the flaws in the stellar evolution models limits this ANN trained with *synthetic* stars to be applied to observational data.

The inference made by the global training presents trends, seen as underestimations in the target variables when comparing the high-luminosity stars with the low-luminosity stars. However, this training proves to have lower intrinsic uncertainties than the local training, because the local training depends highly in the position in the sHRD of the problem star. For the case where there are no added observational errors, the best training set is the $L\&\mathcal{L}$-training, presenting intrinsic uncertainties of 0.3% and 1.1% for the mass and radius, respectively.

From the comparison between intrinsic and observational uncertainties, we conclude that the second one is the dominant error source. As the uncertainties of both $\mathcal{L}$-training and L-training are the lowest and similar, these are the best training sets for added observa-

tional uncertainties, yielding maximum uncertainties of 9% and 13% in the mass and radius inference, respectively.

In the inference applied to the *empirical* sample there are underestimation trends in the mass and radius. The mass discrepancy between the evolutionary-models results and the spectroscopic results was addressed by Herrero et al. (1992). It is still not clear if the discrepancy is caused by wrong measurements of the gravity in the quantitative spectroscopic analysis or by the flaws of the evolutionary models. We think that this mass discrepancy is enhanced by the lack of variety of values of (v_{eq}/v_c) as only three values are accounted. However, this method can be improved by using a more appropriate sample for the training (see Sect. 4.2), proving wether more values of (v_{eq}/v_c) improve the inference results.

Bibliography

[1] Demšar, J., Curk, T., Erjavec, A., et al. 2013, Journal of Machine Learning Research, 14, 2349

[2] Ekström, S., Georgy, C., Eggenberger, P., et al. 2012, , 537, A146

[3] Freedman, D. & Diaconis, P. 1981, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 57, 453

[4] Georgy, C., Granada, A., Ekström, S., et al. 2014, , 566, A21

[5] Herrero, A., Kudritzki, R. P., Vilchez, J. M., et al. 1992, , 261, 209

[6] Holgado, G., Simón-Díaz, S., Haemmerlé, L., et al. 2020, , 638, A157

[7] Maeder, A. & Meynet, G. 2000, , 38, 143

[8] Markova, N., Puls, J., & Langer, N. 2018, , 613, A12

[9] Ramírez-Agudelo, O. H., Simón-Díaz, S., Sana, H., et al. 2013, , 560, A29

[10] Sana, H. & Evans, C. J. 2011, in Active OB Stars: Structure, Evolution, Mass Loss, and Critical Limits, ed. C. Neiner, G. Wade, G. Meynet, & G. Peters, Vol. 272, 474–485

[11] Schneider, F. R. N., Langer, N., de Koter, A., et al. 2014, , 570, A66

[12] Simón-Díaz, S. 2020, A Modern Guide to Quantitative Spectroscopy of Massive OB Stars, 155–187

[13] Simón-Díaz, S., Castro, N., Herrero, A., et al. 2011, in Journal of Physics Conference Series, Vol. 328, Journal of Physics Conference Series, 012021

[14] Tenorio-Tagle, G., Muñoz-Tuñón, C., Pérez, E., Silich, S., & Telles, E. 2006, , 643, 186

[15] Trundle, C. & Lennon, D. J. 2005, , 434, 677
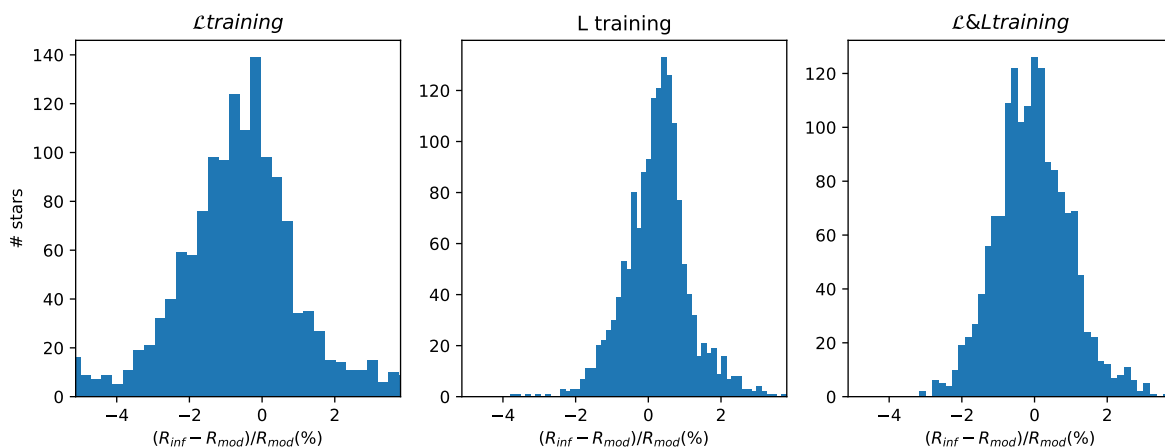
Appendix A

Complementary Figures and Tables

Figure A.1: *Results of the global training analysis. These histograms show the relative error for the inferred radius (R_{inf}) and the value given by the models (R_{mod}) for the three types of training variable sets mentioned in Chapter 2. These results are obtained without adding any uncertainty source and the bin size follows the Freedman-Diaconis rule (Freedman & Diaconis 1981).*

Name HD	$T_{\rm eff}$ (K)	$\log(\mathcal{L}/\mathcal{L}_\odot)$	$\log(L/L_\odot)$ (dex)	M_{spec} (M$_\odot$)	R_{spec} (R$_\odot$)
HD93129AaAb	45.6±1.1	4.136±0.108	6.55±0.03	737.0±74.3	29.5±0.3
CYGOB2-7	50.3±1.8	4.106±0.144	6.05±0.05	94.9±29.9	13.9±0.3
HD190429A	36.6±0.8	4.114±0.098	5.89±0.03	59.0±6.2	22.0±0.3
HD15570	40.2±0.8	4.177±0.069	6.21±0.02	106.1±14.8	26.4±0.4
HD14947	39.1±1.1	4.109±0.111	6.13±0.03	112.8±20.1	25.5±0.5
CYGOB2-9	40.1±1.0	3.883±0.109	6.27±0.03	243.8±57.6	28.4±0.4
CPD-472963AB	37.1±0.5	4.157±0.046	5.97±0.02	65.9±6.4	23.4±0.2
CYGOB2-11	37.3±1.5	4.047±0.156	5.99±0.05	91.4±25.1	23.8±0.6

HD169582	38.9±1.3	4.050±0.218	5.48±0.04	30.4±13.5	12.2±0.2
HD229196	36.8±1.0	4.053±0.093	6.00±0.03	74.2±16.6	24.9±0.4
HD163758	34.6±0.6	4.256±0.076	5.89±0.01	42.9±6.5	24.6±0.4
HD157857	36.7±0.7	4.049±0.087	5.91±0.01	63.9±5.2	23.1±0.3
HD69464	35.8±1.0	4.226±0.094	5.83±0.03	41.0±6.6	21.6±0.4
HD193514	35.9±0.5	4.030±0.047	5.78±0.01	56.7±7.6	20.1±0.2
HD94963	36.3±0.6	4.090±0.067	5.92±0.02	67.9±8.0	23.1±0.3
HD151515	36.0±1.0	4.065±0.139	5.45±0.02	25.7±6.9	13.7±0.3
HD188001	32.4±0.5	4.232±0.094	5.60±0.02	25.1±3.0	20.4±0.3
HD192639	34.7±0.7	4.121±0.115	5.96±0.02	70.0±17.3	26.5±0.3
HD156154	34.2±0.5	4.216±0.047	5.53±0.01	21.1±2.1	16.7±0.2
HD17603	33.3±0.8	4.220±0.090	5.73±0.02	31.3±5.1	21.9±0.4
HD171589	36.5±0.8	3.979±0.071	5.40±0.02	25.9±3.8	12.5±0.2
HD34656	36.0±0.5	4.115±0.047	5.76±0.01	44.3±3.6	19.4±0.2
HD225160	33.2±1.2	4.125±0.153	5.63±0.01	27.8±3.9	21.0±0.3
BD-114586	32.4±0.6	4.212±0.105	5.54±0.02	21.4±4.3	18.6±0.3
HD162978	35.0±0.5	4.066±0.047	5.35±0.02	19.3±1.9	12.9±0.1
HD303492	28.3±1.0	4.287±0.086	6.02±0.01	48.5±6.4	42.5±1.9
HD125241	32.1±0.5	4.176±0.048	5.58±0.01	23.7±2.5	19.8±0.2
HD75211	33.4±0.6	4.085±0.068	5.53±0.02	25.2±3.5	17.5±0.2
HD207198	33.1±0.5	4.159±0.048	5.47±0.01	20.6±2.1	16.6±0.2
HD30614	29.4±0.6	4.263±0.078	5.50±0.01	13.8±2.5	23.5±0.6
HD152249	31.1±0.5	4.151±0.049	5.81±0.01	44.1±4.1	27.8±0.4
HD202124	31.1±0.5	4.141±0.066	5.62±0.02	25.3±5.8	23.1±0.4
HD210809	30.9±0.5	4.230±0.057	5.69±0.01	28.8±3.7	24.4±0.4
HD209975	32.0±0.5	4.111±0.048	5.42±0.02	20.6±1.7	16.8±0.2
HD71304	32.0±0.5	4.111±0.048	5.67±0.02	35.8±2.9	22.3±0.2
HD57061AaAb	33.0±0.5	3.964±0.048	7.11±0.02	1472.0±70.9	110.3±1.2
HD152424	30.2±0.5	4.140±0.067	5.69±0.01	35.0±4.7	25.5±0.4
HD218915	31.1±0.5	4.141±0.049	6.02±0.02	73.6±7.0	35.2±0.2
HD154368	30.4±0.5	4.281±0.049	5.78±0.01	31.0±2.7	28.0±0.5
HD123008	31.7±0.5	4.174±0.057	5.70±0.02	32.8±4.3	23.4±0.3
HD76968	30.8±0.5	4.064±0.049	5.51±0.02	27.8±2.8	19.9±0.2
HD188209	30.1±0.5	4.264±0.095	5.43±0.01	14.8±3.2	19.2±0.3
HD105056	27.4±0.6	4.261±0.089	5.44±0.02	14.5±2.2	23.7±0.6
HD195592	28.0±0.5	4.269±0.051	6.07±0.01	65.9±7.3	46.8±0.4
HD149038	29.8±0.5	4.107±0.050	5.58±0.02	30.3±4.0	23.2±0.3
HD225146	28.3±0.6	4.087±0.097	5.34±0.02	17.6±3.3	19.6±0.3
HD104565	28.9±0.5	4.224±0.050	5.59±0.01	23.2±2.8	25.2±0.5
HD154811	29.8±0.5	4.067±0.050	5.28±0.02	15.1±1.6	16.6±0.2
HD152147	30.1±0.5	4.044±0.095	5.47±0.02	25.9±6.0	20.1±0.3
HD47432	29.1±0.5	4.206±0.050	5.82±0.02	38.5±3.5	32.2±0.6

HD68450	30.6±0.9	4.023±0.140	5.22±0.03	16.4±4.4	14.6±0.3
HD10125	30.9±0.5	3.940±0.057	6.26±0.02	193.5±27.8	47.0±0.5
HD152405	30.3±0.5	4.026±0.094	5.36±0.02	21.5±4.2	17.4±0.2
HD93843	37.3±0.6	4.147±0.057	5.62±0.02	29.7±3.3	15.6±0.2
HD152723AaAb	38.0±0.5	3.899±0.046	6.16±0.02	185.8±18.2	27.8±0.2
HD96946	39.0±0.5	3.894±0.073	5.59±0.01	50.5±9.2	13.8±0.1
HD156738AB	37.9±1.1	3.875±0.158	6.24±0.03	249.4±83.8	30.8±0.5
HD190864	37.5±0.9	4.046±0.090	5.43±0.03	24.5±4.4	12.3±0.2
HD93160AB	36.6±0.7	3.794±0.096	6.02±0.02	169.3±22.8	25.6±0.4
CYGOB2-4A	36.4±1.7	4.104±0.171	5.34±0.05	18.2±5.0	11.7±0.4
HD163800	35.2±0.5	4.156±0.047	5.37±0.02	16.7±1.6	13.0±0.2
HD319702	36.3±0.7	3.670±0.144	5.27±0.02	42.1±12.5	10.9±0.1
HD36861A	35.2±0.5	4.056±0.047	5.06±0.02	9.8±2.5	9.1±0.1
HD114737AB	35.7±0.5	3.721±0.047	4.69±0.02	9.6±1.7	5.8±0.1
HD218195A	34.1±0.5	3.991±0.094	4.94±0.02	9.1±1.5	8.4±0.1
HD93249A	32.8±0.5	3.953±0.084	5.37±0.02	27.5±5.0	15.1±0.2
HD24431	34.9±0.5	3.791±0.056	5.18±0.02	24.9±4.1	10.6±0.1
HD193443AB	33.0±0.5	3.954±0.048	5.56±0.02	39.3±4.0	18.5±0.2
HD16832	32.0±0.5	4.041±0.057	5.12±0.02	12.4±1.7	11.8±0.1
HD96264	33.1±1.0	3.799±0.159	5.36±0.04	39.0±12.0	14.7±0.3
HD154643	31.0±0.5	3.845±0.049	4.95±0.02	12.5±1.6	10.4±0.1
HD189957	32.1±0.5	3.836±0.066	5.03±0.02	15.3±2.4	10.5±0.1
HD64568	46.9±1.0	4.145±0.071	5.85±0.03	50.2±6.5	12.7±0.2
HD93128	49.3±2.2	4.071±0.178	6.03±0.06	100.5±34.0	14.3±0.4
HD46223	42.2±0.5	4.161±0.045	5.69±0.02	33.6±3.6	13.0±0.2
HD96715	45.2±1.2	4.101±0.129	5.83±0.03	55.1±13.8	13.3±0.3
HD93250AB	45.0±0.6	4.143±0.055	6.14±0.02	99.3±12.0	19.5±0.2
HD168076AB	43.0±1.8	4.004±0.167	6.79±0.05	574.0±125.2	44.2±1.0
HD15629	41.8±0.5	4.085±0.054	5.66±0.02	38.3±4.9	12.8±0.2
HD303308AB	41.1±0.9	3.945±0.089	5.52±0.03	38.6±6.5	11.5±0.1
HD168112AB	39.7±1.0	4.075±0.137	6.04±0.03	96.4±26.7	22.3±0.3
HD319699	41.2±0.8	3.940±0.078	5.49±0.02	36.3±6.1	11.0±0.1
HD46150	41.1±0.5	4.035±0.045	5.66±0.02	42.7±3.9	13.4±0.1
HD93204	39.2±0.7	4.013±0.068	5.27±0.02	22.5±4.3	9.5±0.2
HD101190AaAb	39.8±0.5	3.910±0.046	5.92±0.02	104.3±11.5	19.4±0.1
HD303311	40.1±0.7	3.893±0.067	5.51±0.02	41.4±5.3	11.8±0.2
CPD-592600	39.3±0.8	3.908±0.087	6.18±0.02	177.6±28.2	26.7±0.3
HD42088	40.0±0.5	3.808±0.046	5.31±0.02	32.1±3.3	9.4±0.1
HD322417	38.6±0.8	3.996±0.106	5.92±0.03	88.2±20.8	20.5±0.2
HD91572	38.8±0.5	3.905±0.083	5.40±0.02	31.8±5.7	11.1±0.1
HD167633	38.0±0.8	3.989±0.079	5.38±0.02	23.4±3.3	11.4±0.1
HD12993	39.2±0.6	3.873±0.075	5.11±0.02	17.7±2.9	7.9±0.1

HD93222AB	36.8±0.7	4.023±0.115	5.65±0.02	44.0±10.4	16.5±0.2
HD91824	39.8±0.9	3.770±0.136	5.14±0.03	24.6±7.1	7.8±0.1
HD93146A	38.7±0.6	3.901±0.075	5.58±0.02	48.9±7.0	13.8±0.1
HD242926	39.0±0.7	3.684±0.114	5.42±0.02	56.9±13.9	11.3±0.1
HD152590	38.0±0.5	3.789±0.074	5.01±0.02	17.5±2.3	7.4±0.1
HD35619	37.7±0.6	3.755±0.123	5.62±0.02	77.7±20.1	15.3±0.2
HD135591	35.0±0.5	3.986±0.047	5.24±0.02	18.0±1.9	11.4±0.1
HD94024	34.8±0.5	3.996±0.056	5.20±0.02	14.5±2.4	11.1±0.1
HD97848	35.6±0.6	3.926±0.067	5.51±0.02	39.2±5.9	15.1±0.1
HD101223	35.2±0.5	3.956±0.056	5.20±0.02	17.7±1.7	10.7±0.1
HD191978	35.8±0.6	3.796±0.085	4.87±0.02	12.6±1.7	7.1±0.1
HD46966	35.9±0.5	3.770±0.074	5.34±0.01	38.1±6.4	12.2±0.1
HD298429	33.6±1.1	3.945±0.151	5.11±0.04	14.1±4.2	10.6±0.2
HD46149	36.9±0.5	3.428±0.083	5.17±0.02	54.9±11.3	9.3±0.1
TRUMPLER14-9	36.7±0.7	3.549±0.115	4.98±0.02	28.2±6.5	7.6±0.2
HD14633AaAb	35.0±0.5	3.776±0.084	5.24±0.02	28.3±4.8	11.4±0.1
HD93028	35.3±0.7	3.651±0.087	5.22±0.02	38.1±7.5	10.9±0.1
CPD-417733	34.2±0.5	3.866±0.074	5.16±0.02	19.8±3.0	10.9±0.1
CPD-592551	34.9±0.5	3.651±0.056	4.92±0.02	17.7±2.3	7.9±0.1
HD214680	35.2±0.5	3.686±0.047	4.51±0.02	7.5±1.8	4.9±0.1
HD216898	35.9±0.6	3.430±0.114	4.89±0.02	29.9±7.9	7.2±0.1
HD96622	33.3±0.6	3.800±0.114	5.06±0.02	18.5±4.4	10.2±0.1
HD46202	34.9±0.5	3.431±0.074	4.82±0.02	25.6±4.5	7.1±0.1
HD12323	34.2±0.9	3.576±0.186	4.74±0.03	15.7±6.2	6.7±0.1
HD93027	33.8±0.5	3.556±0.075	5.19±0.02	44.1±7.8	11.5±0.1
HD192001	33.0±0.5	3.634±0.057	5.01±0.02	24.1±3.3	9.8±0.1
HD155889AB	34.9±0.7	3.461±0.078	6.21±0.02	605.9±62.2	34.7±0.5
HD38666	33.9±0.5	3.591±0.048	4.50±0.02	7.9±1.8	5.2±0.1
HD34078	34.5±0.8	3.471±0.146	4.75±0.03	20.4±6.6	6.7±0.1
HD207538	32.0±0.5	3.611±0.048	4.70±0.04	11.6±2.0	7.1±0.1
HD36512	33.0±0.5	3.444±0.094	4.34±0.02	7.8±1.6	4.5±0.00

Table A.1: Parameter values used from the empirical sample. The first columns refers to the name in the Henry Draper catalog. These values were obtained by Holgado et al. (2020).
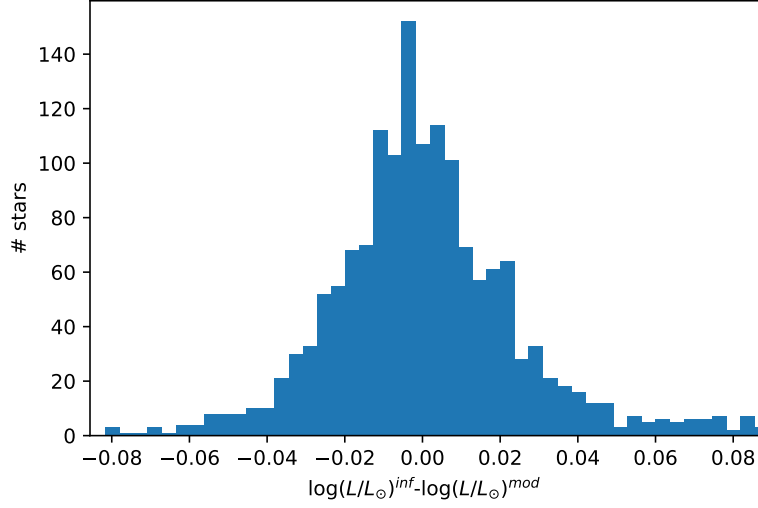
Figure A.2: *Results of the global training analysis. These histograms show the relative error for the inferred luminosity ($log(L/L_\odot)^{inf}$) and the value given by the models ($log(L/L_\odot)^{mod}$) for the $\mathcal{L}$-training. The bin size follows the Freedman-Diaconis rule (Freedman & Diaconis 1981).*
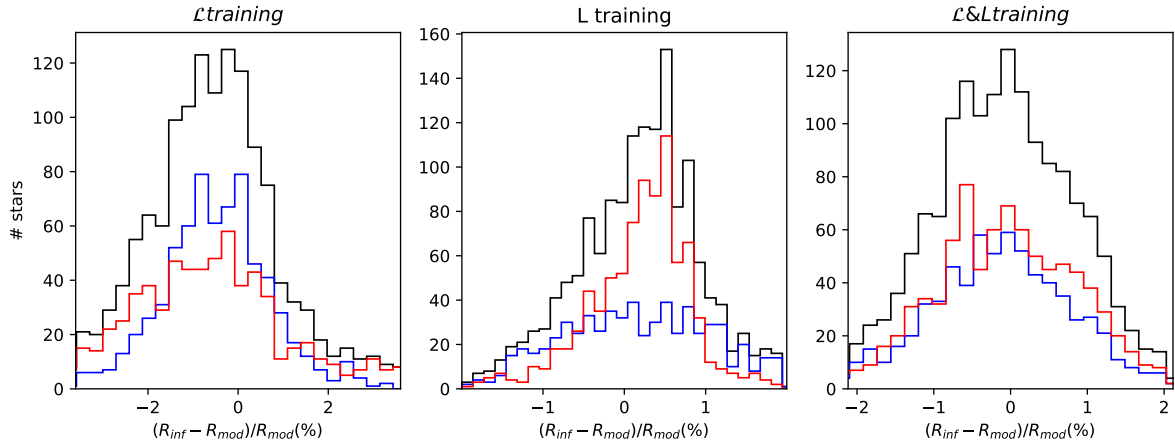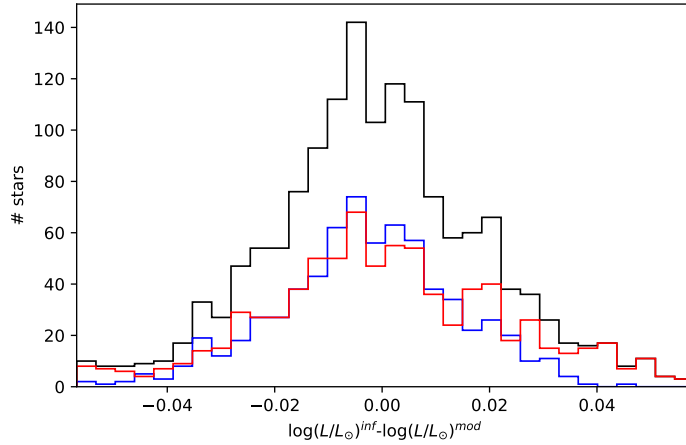

Figure A.3: *Histograms representing the distribution of intrinsic uncertainties associated with the inference of the radius using the three considered training sets. Red/blue distributions represent the case of the stars located in the upper/lower part of the sHRD (see Figure 3.1 right panel). In black results for the global sample.*
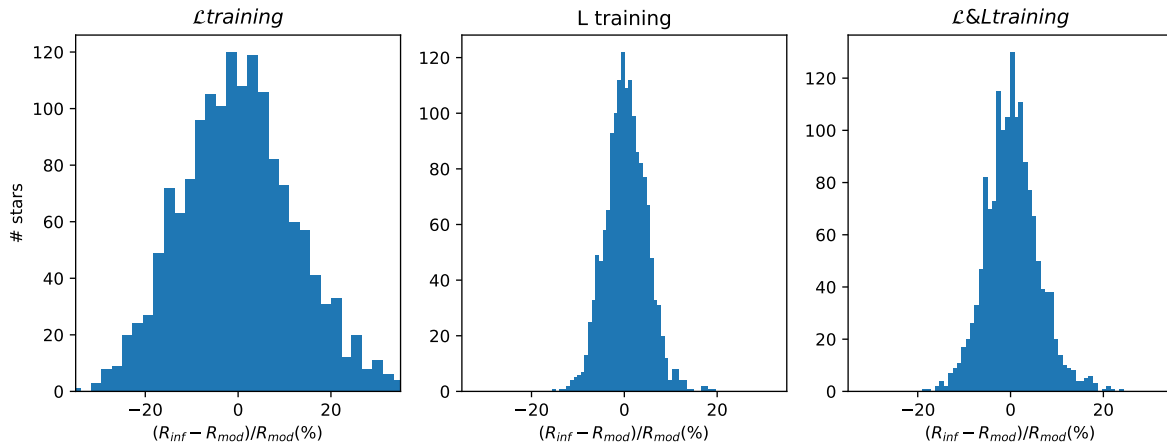
Figure A.4: *Histogram representing the distribution of intrinsic uncertainties associated with the inference of the luminosity using the $\mathcal{L}$-training. Red/blue distributions represent the case of the stars located in the upper/lower part of the sHRD (see Figure 3.1 right panel). In black results for the global sample.*

Figure A.5: *Results of the global training with added observational errors. These histograms show the relative error of the inferred radius (R_{inf}) with respect to the input value given by the models (R_{mod}) for the three types of training variable sets mentioned in Chapter 2. The bin size follows the Freedman-Diaconis rule (Freedman & Diaconis 1981).*
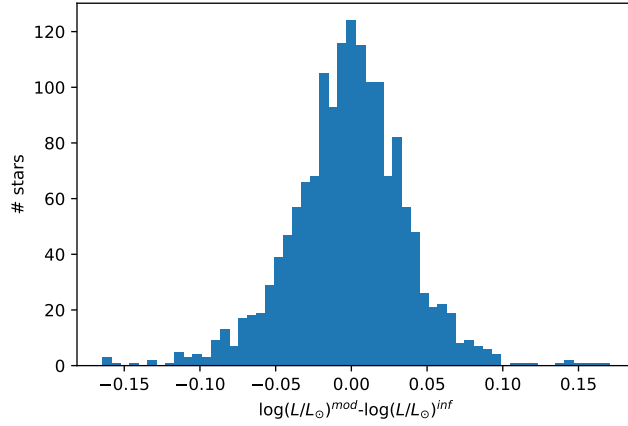
Figure A.6: *Results of the global training with added observational errors. This histogram show the relative error of the inferred luminosity ($log(L/L_\odot)^{inf}$) with respect to the input value given by the models ($log(L/L_\odot)^{mod}$) for the three types of training variable sets mentioned in Chapter 2. The bin size follows the Freedman-Diaconis rule (Freedman & Diaconis 1981).*
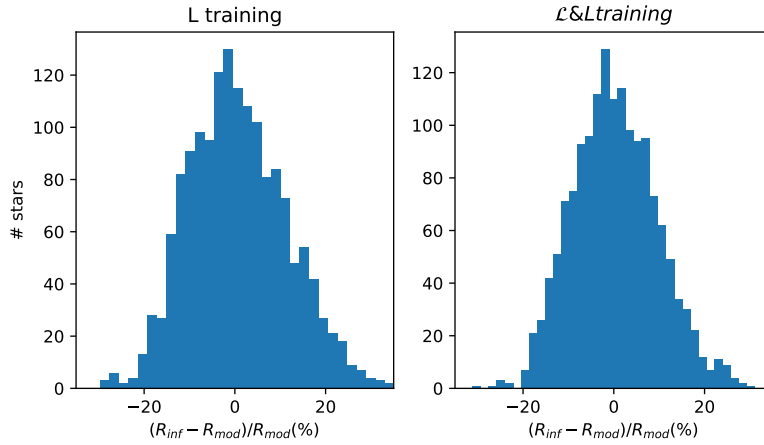

Figure A.7: *Representation of the relative error distribution of the radius for the L-training and the L&$\mathcal{L}$-training considering $\Delta log(L/L_\odot) = 0.09$. The bin size follows the Freedman-Diaconis rule (Freedman & Diaconis 1981).*
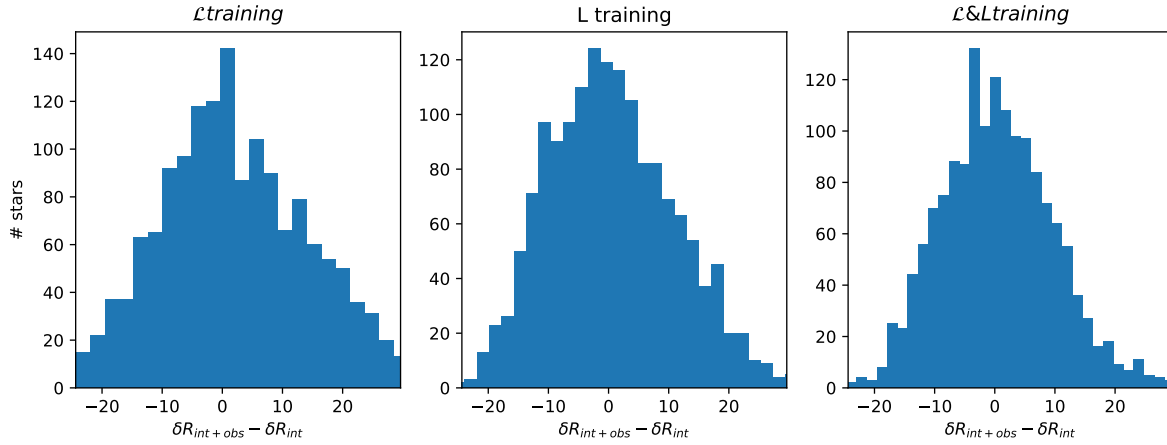
Figure A.8: *Relative error distributions of the added observational errors for the mass inference, for the three types of training variable sets mentioned in Chapter 2. The bin size follows the Freedman-Diaconis rule (Freedman & Diaconis 1981).*

Figure A.9: *Relative error distributions of the added observational errors for the mass inference, for the three types of training variable sets mentioned in Chapter 2. The bin size follows the Freedman-Diaconis rule (Freedman & Diaconis 1981).*
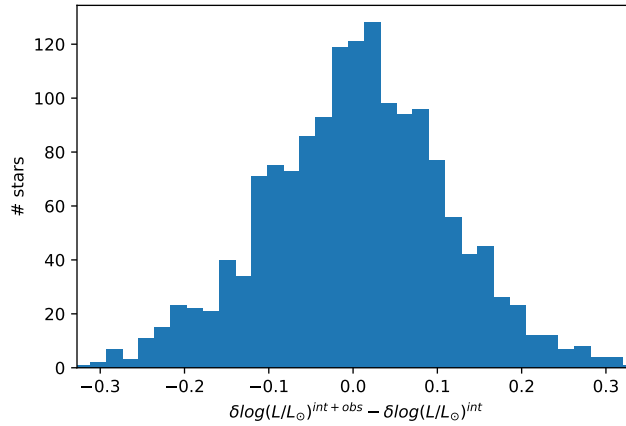
Star	T_{eff} (K)	$\log(\mathcal{L}/\mathcal{L}_\odot)$ (dex)	$\log(L/L_\odot)$ (dex)	$V sini(km/h)$	$M(M_\odot)$	$R(R_\odot)$
S1	47065	4.0389	5.860	25.688	66.30	12.840
S2	37008	4.1150	5.862	0	55.86	20.810
S3	30088	4.1162	5.572	17.597	28.60	22.560
S4	42043	3.8215	5.416	0	39.39	96.580
S5	34025	3.8898	5.349	64.020	28.84	13.650
S6	27663	3.9194	5.397	0	30.03	21.800
S7	34427	3.4699	4.787	69.232	20.78	69.800
S8	28800	3.6425	4.927	87.757	19.26	11.710

Table A.2: *Parameter values of the 8 mock stars used in the local training analysis.*