

Dissertation
submitted to the
Combined Faculty of Mathematics, Engineering and Natural Sciences
of Heidelberg University, Germany
for the degree of
Doctor of Natural Sciences

Put forward by
Kiril Maltsev
born in: Minsk (Belarus)
Oral examination: 18th December 2024

Statistical modeling of the progenitor evolution and formation of neutron stars and stellar-mass black holes

Referees:

Prof. Dr. Friedrich Röpke
Prof. Dr. Michela Mapelli

Zusammenfassung

In dieser Arbeit werden Methoden entwickelt, um effiziente und skalierbare Vorhersagemodelle für die Sternentwicklung, die Endzustände von Kernkollaps-Supernovae (CCSN) und die Gravitationswellenemission von den Überresten der Verschmelzung binärer Neutronensterne (NS) mit Hilfe von überwachtem maschinellem Lernen zu konstruieren. Mit Hilfe der Vorhersagemodelle werden die folgenden zentralen wissenschaftlichen Erkenntnisse gewonnen. Das endgültige Schicksal von massereichen Einzel- und Geber-Sternen in Doppelstern-Systemen, die einen Kollaps des Eisenkerns durchlaufen, ist nicht nur in der Struktur des SN-Vorläufers codiert, sondern bereits am Ende des Kern-Helium Brennens (CHeB) weitgehend vorherbestimmt. Das in dieser Arbeit entwickelte CCSN-Modell stimmt mit Abschätzungen der Kohlenstoff-Sauerstoff Kernmassen (M_{CO}) mehrerer beobachteter SN-Vorläufer überein, welche andere CCSN-Rezepturen, die typischerweise in schnellen Binärstern-Populationssynthese-Codes verwendet werden, nicht erklären können. Um das Problem der fehlenden Roten Überriesen teilweise mit gescheiterten SNe zu lösen, müssen Sternentwicklungsmodelle am Ende des CHeB M_{CO} -Werte und einen zentralen Massenanteil an Kohlenstoff innerhalb eines begrenzten Bereichs des Parameterraums haben. Ein NS ist der wahrscheinlichere (garantiert für $M_{\text{CO}} \lesssim 6 M_{\odot}$, sonst mit einer Wahrscheinlichkeit von 85-90%) kompakte Überrest, der nach einer erfolgreichen SN zurückbleibt, denn die Bildung eines Schwarzen Lochs über Rückfall kann unter bestimmten Bedingungen bereits vor der SN ausgeschlossen werden. Wenn zwei Neutronensterne verschmelzen, kann das Gravitationswellen-Signal des Verschmelzungsüberrests bei fortgeschrittener LIGO Sensitivität bis zu einer Leuchtkraftentfernung von etwa 12 Mpc rekonstruiert werden. Ein masseärmerer NS-Doppelstern mit einer weicheren Zustandsgleichung (EOS) kann zu einer ähnlichen Peak-Oszillationsfrequenz des Verschmelzungsüberrests führen wie ein massereicherer NS-Doppelstern mit einer steiferen EOS.

Abstract

In this thesis, methods are elaborated for constructing efficient, scalable predictive models of stellar evolution, of core collapse supernova (CCSN) outcomes and of the gravitational wave (GW) emission from binary neutron star (NS) merger remnants, with supervised machine learning techniques. Aided by the predictive models, the following central scientific findings are made. The final fates of massive single and binary-stripped stars undergoing iron core collapse are not only encoded in the pre-SN progenitor structure, but broadly pre-determined already at the end of core helium burning (CHeB). The CCSN model developed in this work is in agreement with estimates of carbon-oxygen core masses (M_{CO}) of several observed SN progenitors, which other CCSN recipes typically used in rapid binary population synthesis codes cannot explain. In order to partially address the missing red supergiant problem by failed SNe, stellar evolution models are constrained to have a M_{CO} and a central carbon mass fraction at the end of CHeB within a confined region of parameter space. A NS is the more likely (guaranteed for $M_{\text{CO}} \lesssim 6 M_{\odot}$, and at 85-90 % probability otherwise) compact remnant left behind a successful SN, while fallback black hole formation can under certain pre-SN conditions be excluded. When two NSs merge, the GW signal of the merger remnant can be reconstructed at advanced LIGO sensitivity up to a source luminosity distance of approximately 12 Mpc. A less massive NS binary at a softer equation of state (EOS) can lead to a similar peak oscillation frequency of the merger remnant as a more massive binary at a stiffer EOS.

*“Keep your eyes on the stars, and
your feet on the ground.”
— Theodore Roosevelt*

Contents

1. Introduction	3
1.1. Motivation	3
1.1.1. The need for scalable statistical models of stellar evolution and its outcomes	3
1.1.2. Gravitational wave astronomy	3
1.2. Surrogate modeling	5
1.2.1. Supervised machine learning	6
1.2.2. Model selection, training and testing	6
2. Scalable stellar evolution forecasting	9
2.1. Introduction	9
2.2. Methods	14
2.2.1. Regression problem formulation	14
2.2.2. The timescale-adapted evolutionary coordinate	15
2.2.3. Data base	16
2.2.4. Performance evaluation	20
2.3. Interpolation scheme solutions	21
2.3.1. Supervised learning model selection	22
2.3.2. Choice of loss function	23
2.3.3. Hyperparameter optimization	24
2.3.4. Hierarchical Nearest Neighbor Interpolation (HNNI)	25
2.4. Results	26
2.4.1. Prediction of classical photometric observables	27
2.4.2. Prediction of the evolutionary coordinate	33
2.4.3. Prediction of evolutionary tracks and isochrones	35
2.5. Conclusions and outlook	39
A Appendix	41
A1. Stellar evolution re-parametrization and HNNI	41
A2. HNNI in 3 dimensions or higher	43
A3. GPR	44
A4. fNN hyperparameter tuning	45
A5. Alignment problem	46
3. Explodability criteria for the neutrino-driven supernova mechanism	51
3.1. Introduction	51
3.2. Methods	55
3.2.1. 1D CCSN progenitor models	55
3.2.2. 1D CCSN explosion model	56

3.2.3.	Catalog of 3D CCSN simulation outcomes	57
3.2.4.	Supervised learning model	58
3.3.	Results	59
3.3.1.	Pre-SN explodability criterion	59
3.3.2.	Final fate landscapes of single and binary-stripped stars . . .	66
3.3.3.	Comparison with observations	82
3.4.	Discussion and outlook	90
3.5.	Conclusion	92
B	Appendix	94
B1.	The $2.5 M_{\odot}$ mass coordinate and M_{Fe}	94
B2.	Parameters of the $\log Z$ scaling model	94
B3.	Deterministic fallback model	95
4.	Gravitational wave model for neutron star merger remnants	99
4.1.	Introduction	100
4.2.	Methods	103
4.2.1.	BNS merger simulations	103
4.2.2.	K-nearest neighbors model	104
4.2.3.	Splitting of data	108
4.2.4.	Injection setup	109
4.3.	Results	110
4.3.1.	Performance of KNN models	110
4.3.2.	Execution time of the model	112
4.3.3.	Parameter estimation on simulated data	113
4.3.4.	Dependence on the EOS	115
4.4.	Conclusions	118
C	Appendix	120
C1.	KNN models	120
C2.	Best and worst predictions over the test data	122
5.	Conclusions	127
5.1.	Main method development results	127
5.2.	A cross-comparative and integral view	129
5.3.	Main scientific results	131
5.4.	Outlook	133
	Acknowledgments	139
	List of publications	141
	Bibliography	143

1. Introduction

1.1. Motivation

Stars are self-gravitating bodies that are massive enough to ignite nuclear fusion as they convert gravitational potential energy into thermal heat (e.g. [Prialnik, 1996](#)). They form through the gravitational collapse of gigantic gaseous nebula, and burn successively heavier elements as they undergo a sequence of evolutionary phases (e.g. [Kippenhahn, 1990](#)), governed by an interplay of gravity and thermodynamics, nuclear and particle quantum physics, and hydrodynamics. Stars end their lives by transformation into compact objects, or in violent supernova (SN) explosions tearing them apart. Massive stars become large, shine bright and die young.

1.1.1. The need for scalable statistical models of stellar evolution and its outcomes

The study of stars is crucial for various fields in astrophysics. For example, planets form in the protoplanetary disks around young stars. Stars make up the vast majority of baryonic matter in the Universe and are responsible for producing heavy elements (e.g. [Kippenhahn, 1990](#)). These are ejected into space through stellar winds or SN explosions, enriching the interstellar medium and driving chemical evolution (e.g. [Goswami et al., 2022](#)). Stellar radiation heats the surfaces of planets and is essential for sustaining life. Stellar feedback co-regulates the star formation rate in galaxies ([Hopkins et al., 2018](#)). The energy released by stars can heat up the surrounding gas and prevent it from collapsing to form new stars. Stars are the constituents of open and globular clusters, providing information about their formation and evolution (e.g. [Paunzen and Netopil, 2006](#)). Many such and similar astrophysical studies require efficient predictive models of stellar evolution, feedback and explosions.

Stars evolve on timescales that are orders of magnitude beyond human lifetimes. Therefore, the only way to probe stellar evolution models is to compare statistical model predictions with observations of stellar populations.

Statistical methods are necessary to model stars also because of their vast number and the variability among them. With thousands, millions or billions of stars scattered across space, a detailed simulation of each star is impossible due to the associated computational costs.

1.1.2. Gravitational wave astronomy

One example for such applications is gravitational wave astronomy. Massive stars are the progenitors of neutron stars and stellar-mass black holes (e.g. [Heger et al., 2003](#)),

and the coalescence of these compact objects gives rise to gravitational wave radiation (e.g. [Carroll, 1997](#)). Depending on the frequency range and luminosity distance, this source emission can be observed at present or will be observable with future generation detectors ([Abbott and et al., 2019](#)). Among the crucial requirements for gravitational wave astronomy are

- efficient gravitational wave models that can be used to facilitate the detection of source signals and to relate its spectral features to properties of the astrophysical source (e.g. [Meyer and Christensen, 2016](#)), and
- estimates of binary compact object merger rates ([Santoliquido et al., 2021](#)) from the two main progenitor channels: isolated binary evolution ([Broekgaarden et al., 2022](#)) and the dynamical pairing in a stellar cluster ([Mapelli et al., 2021](#)).

In this thesis, three predictive models are constructed, which all contribute to targeting these requirements:

1. The gravitational wave model traces the binary neutron star (BNS) post-merger emission over 17 ms, assuming an equal-mass NS binary over a range from the minimal NS mass at birth in a CCSN up to the threshold mass for prompt black hole (BH) formation, for two different nuclear equations of state (EOS). Given its predictive accuracy and execution time of 1-3 ms for generating a signal at sampling rates comparable to those of advanced LIGO (aLIGO), it is applicable as a matched-filtering template for the detection of signals at aLIGO sensitivity and next-generation detectors.
2. The CCSN model evaluates structural properties of a pre-SN progenitor to predict the final fate (failed or successful SN). A CCSN recipe, which parametrizes the final fate outcomes as a function of carbon-oxygen core mass M_{CO} , metallicity Z and mass transfer pre-history of binary-stripped stars, is derived from it. The CCSN recipe is applicable for rapid binary population synthesis studies, such as the prediction of binary BH, BNS and BH-NS merger rates. Both final fate prescriptions are evaluated analytically and have negligible computational cost.
3. The single star evolution model covers a mass range from red dwarves to $300 M_{\odot}$ Wolf-Rayet stars and the stellar lifetime from the zero-age-main-sequence up to the end of core helium burning. A parameter space this wide entails the progenitor evolution of neutron stars and stellar-mass black holes from the single-star channel. The predictive model achieves an accuracy that is at least one order of magnitude below typical observational uncertainties, and an efficiency of millions of point predictions within tens of seconds on a 4-core CPU. It is applicable for testing adopted stellar evolution physics against observations, for estimation of stellar parameters by iterative optimization, and in larger-scale astrophysical simulations such as stellar N -body dynamics.

Each of these is subject of a chapter in this thesis:

- Ch. 2. deals with scalable stellar evolution forecasting,
- Ch. 3. with the scalable prediction of CCSN outcomes and
- Ch. 4. with scalable gravitational wave modeling.

The three statistical models have in common that their construction is motivated by the computational costs associated with running detailed stellar evolution, core collapse and numerical relativity simulations, respectively, over a quasi-continuous parameter space in order to obtain the output predictions of the variables of interest. For example, these are necessary for population synthesis or iterative optimization. The three models have supervised machine learning (ML) as a construction means in common. In the introductory sections of each thesis chapter on the aforementioned predictive models — i.e. Sect. 2.1., Sect. 3.1. and Sect. 4.1. — respectively, the motivation for building the predictive models is covered in greater depth, and compared to alternative methods. In this thesis, supervised ML is used to fit surrogate models on data obtained by running the original (expensive) astrophysical code. A well-trained surrogate model acquires the capability to generalize the output predictions over the input parameter space, and to cast predictions at low computational cost. General aspects of this modeling framework are introduced in Sect. 1.2.. The specifics of each of the three predictive models constructed in this thesis are treated in the methods-related sections of each chapter, i.e. Sect. 2.2. and Sect. 2.3. for the stellar evolution surrogate model, Sect. 3.2. for the CCSN model and Sect. 4.2. for the GW model. The scientific and method development results are presented in Sect. 2.4., Sect. ?? and Sect. 4.3., respectively. These are reflected upon and integrated in Ch. 5., which also gives an outlook for future research that could build upon this work.

1.2. Surrogate modeling

Surrogate modeling is a technique used in scientific computing to approximate the behavior of a complex computer model F with a simpler, computationally less expensive model \hat{F} , referred to as an emulator (Gramacy, 2020). The emulator imitates the computationally expensive original model,

$$F \leftarrow \hat{F}, \quad (1.1)$$

and can therefore reliably used instead of the original to mimic its output. The process of building a surrogate model typically involves five main steps:

1. data collection,
2. regression problem formulation,
3. model selection,
4. model training and
5. model testing.

1.2.1. Supervised machine learning

Data collection involves gathering information about the system or process of interest through simulations, experiments, or observations. In context of this work, the data consists of catalogs of pre-computed simulation results using astrophysical codes that model stellar evolution, core collapse and binary neutron star mergers. This data serves as the basis for training the emulators.

The next step is the formulation of the regression problem (e.g. [Hastie et al., 2009](#)). This amounts to the choice of the target variables and of the regressor variables. This choice needs to be tailored to the science goal that is aimed to be addressed with the fitted surrogate model.

Suppose that an expensive computer experiment F produces the outcome y_i under the condition x_i for $i = 1, \dots, n$. Then, the regression problem reads

$$y_i = \hat{F}(x_i). \quad (1.2)$$

Supervised machine learning (e.g. [Hastie et al., 2009](#)) is then used to construct an emulator that predicts a target variable y_i as output, given the regressor variable x_i as its input. While the collected data discretely sample the parameter space at the grid points x_1, \dots, x_n , the training goal is to construct a predictive model that generalizes the prediction of y over a continuous parameter space.

Supervised machine learning works by training a model on a labeled dataset. This means that the training data set consists of $i = 1, \dots, n$ input-output pairs

$$D_i = (x_i, y_i) \quad (1.3)$$

which are presented to the model. During training, the surrogate model learns to map the inputs to the outputs.

1.2.2. Model selection, training and testing

A suitable model \hat{F} needs to be selected to approximate the relationship between the input and output variables of the system. Common models used in surrogate modeling include gaussian processes ([Sacks et al., 1989](#); [Rasmussen, 2004](#); [Gramacy, 2020](#)), feedforward neural networks ([Ivakhnenko and Lapa, 1967](#); [Rumelhart et al., 1985](#)), random forest models ([Ho, 1995](#)) and support vector machines ([Cortes and Vapnik, 1995](#)). The choice of the model depends on the regression problem, on the size of the data set, on the specific characteristics of the data and on the complexity of the system being modeled.

Each surrogate model has hyperparameters that configure its design. These are, for example, the number of hidden layers in a feedforward neural network. The hyperparameters need to be specified before the training. Each surrogate model also has parameters which – in contrast to the hyperparameters – are adjusted during the training.

Once a model is selected, it is trained using the collected data to learn the underlying patterns and relationships within the system. To this end, the available data is split into a training, a validation and a test data subsets (e.g. [Hastie et al.](#),

2009). The training data is used to fit a surrogate model's free parameters. Each model has free parameters that are adjusted during training to minimize the error between the predicted and actual values of the system's outputs. This comparison requires the selection of a statistical score (Gneiting, 2011) that evaluates the predictive performance of the model. One example for a statistical score is the mean squared error. The training process is realized using optimization techniques, such as gradient descent (Curry, 1944).

The validation data is used to optimize hyperparameters of the model. To this end, the surrogate model is fitted multiple times on the training data set, each time with a different hyperparameter configuration. The model performance for each hyperparameter configuration is assessed using the validation data set, based on predictive accuracy of the different surrogate models at the evaluation points in parameter space that have not been used for training. The best-fit hyperparameter configuration is chosen based on the least error score over the validation data. This process is referred to as hyperparameter optimization (e.g. Bischl et al., 2023). It can be realized by methods such as manual engineering, a grid search or Bayesian optimization (Mockus, 1982). Cross-validation (Stone, 2018) ensures that the best-fit model is not biased toward a specific data split into training and validation data subsets.

The test data is used to evaluate the performance of the surrogate model with the optimal hyperparameter choice. It assesses the fitted model's capability to generalize the prediction of the output variables over the input parameter space.

2. Scalable stellar evolution forecasting

This chapter has been published in [Maltsev et al. \(2024b\)](#). Formatting and small text edits have been made to match this thesis. I conducted the scientific work and wrote all of the text myself, with the support of the co-authors.

Abstract: Many astrophysical applications require efficient yet reliable forecasts of stellar evolution tracks. One example is population synthesis, which generates forward predictions of models for comparison with observations. The majority of state-of-the-art rapid population synthesis methods are based on analytic fitting formulae to stellar evolution tracks that are computationally cheap to sample statistically over a continuous parameter range. Running detailed stellar evolution codes, such as MESA, over wide and densely sampled parameter grids is prohibitively expensive computationally, while stellar-age based interpolation in-between sparsely sampled grid points leads to intolerably large systematic prediction errors. In this work, we provide two solutions of automated interpolation methods that find satisfactory trade-off points between cost-efficiency and accuracy. We construct a timescale-adapted evolutionary coordinate and use it in a two-step interpolation scheme that traces the evolution of stars from zero age main sequence all the way to the end of core helium burning while covering a mass range from 0.65 to 300 M_{\odot} . The feedforward neural network regression model (first solution) that we train to predict stellar surface variables can make millions of predictions, sufficiently accurate over the entire parameter space, within tens of seconds on a 4-core CPU. The hierarchical nearest neighbor interpolation algorithm (second solution) that we hard-code to the same end achieves even higher predictive accuracy, the same algorithm remains applicable to all stellar variables evolved over time, but it is two orders of magnitude slower. Our methodological framework is demonstrated to work on the *MESA Isochrones and Stellar Tracks* ([Choi et al., 2016](#)) data set, but is independent of the input stellar catalog. Finally, we discuss prospective applications and provide guidelines how to generalize our methods to higher dimensional parameter spaces.

2.1. Introduction

Several fields of astrophysics need cost-efficient and fast predictive models of stellar evolution for their deployment at scale. These include stellar population synthesis, N -body dynamics models of stellar clusters (e.g. [Kamalah et al., 2022](#)), iterative optimization-based stellar parameter estimation methods (e.g. [Bazot et al., 2012](#)), and large-scale galactic and cosmic evolution simulations (e.g. [Springel et al., 2018](#)) that require a stellar sub-grid physics.

To illustrate, the BONN STELLAR ASTROPHYSICS INTERFACE (BONNSAI; [Schneider et al., 2014](#)) is an example of a Bayesian framework that allows to test stellar evolution models, and—if the test is passed—to infer fundamental stellar model parameters given the observational data. Determination of fundamental stellar parameters that best match the observation requires costly iterative optimization procedures, such as Markov chain Monte Carlo nested sampling techniques, which need a large number of evaluations over a quasi-continuous parameter space for convergence to the best-fit model. In order to reduce systematic estimation errors, BONNSAI requires a stellar parameter grid as dense as possible.

However, there are costly computational demands arising from the traditional method of running a detailed stellar evolution code over a dense rectilinear grid in stellar parameter space: for a fixed grid spacing, the number of stellar tracks to evolve scales to the power of the dimensionality of the fundamental stellar parameter space. The most important parameters of single star evolution are age τ , initial mass M_{ini} at zero age main sequence (ZAMS), initial metallicity Z_{ini} and initial rotation velocity v_{ini} . For stars of $M_{\text{ini}} > 8 M_{\odot}$, binary interaction effects become increasingly important: 71% of all O-stars interact with a companion, and over half of them during the main sequence evolution ([Sana et al., 2012](#)). Therefore, in order to evolve massive stars, the parameter space needs be expanded to cover, in general, eight dimensions ($\tau_1, M_{\text{ini},1}, M_{\text{ini},2}, v_{\text{ini},1}, v_{\text{ini},2}, Z_{\text{ini}}, P_{\text{ini}}, \epsilon$) where P_{ini} is the initial period, ϵ the eccentricity of the binary orbit and $\tau_1 \simeq \tau_2$ to a good approximation.

MODULES FOR EXPERIMENT IN STELLAR ASTROPHYSICS (MESA; [Paxton et al., 2011](#)) is an example of a detailed one-dimensional (1D) stellar evolution code with modular structure, which allows to update the adopted physics when generating stellar evolution tracks: for instance, the equation of state, the mass loss recipe, and the opacity tables. When evolving stars numerically over a wide and densely sampled parameter grid with MESA, there are two main computational challenges: first, the computational cost associated with running the code over the large grid size, and second, numerical instabilities. To overcome the latter, substantial manual effort is required to push a simulation past failure points by reconfiguring the code, and by checking for unphysical results. The manual action mainly involves adaptation of spatial mesh refinement and time step control strategy, and of error tolerance thresholds in stellar model computation, to make sure the solvers converge over each evolutionary phase within reasonable computation time.

The problem of prohibitive computational costs has been addressed in three different ways: First, stellar evolution tracks have been approximated by analytic fitting formulae. Second, the output of detailed stellar evolution codes over a discrete parameter grid has been interpolated. Third, cost-efficient surrogate models of stellar evolution have been constructed. Below, we summarize these main approaches.

The SINGLE STAR EVOLUTION (SSE) package ([Hurley et al., 2000](#)) consists of analytic stellar evolution track formulae predicting stellar luminosity, radius and core mass as functions of age, mass and metallicity of the star. Separate formulae apply to each evolutionary phase, and the duration of each phase is estimated from physical conditions. Along with analytical expressions from stellar evolution theory, the SSE package was obtained by fitting polynomials to the set of stellar tracks by [Pols et al. \(1998\)](#). The fitting formulae method has been extended to predict evo-

lution of binary systems, to include analytical prescriptions for mass transfer, mass accretion, common-envelope evolution, collisions, supernova kicks, angular momentum loss mechanisms and tides (Hurley et al., 2002). At present, the fitting formulae are often used in connection with rapid binary population synthesis codes, for example COMPACT OBJECT MERGERS: POPULATION ASTROPHYSICS & STATISTICS (COMPAS; Riley et al., 2022), and stellar N -body dynamics codes. Though there are two main drawbacks: First, the fixed (rather than modular) input physics. And second, the limited set of predicted output variables, which—depending on the astrophysical application—may be not all the variables of interest. A re-derivation of analytic fitting formulae for a new set of stellar tracks is non-trivial (Church et al., 2009; Tanikawa et al., 2020). Overall, the analytic approach is not sustainable, since it would need be reiterated after each update in stellar input physics.

Interpolation of tracks pre-computed by a detailed code is an alternative to analytic fitting. Brott et al. (2011) interpolate stellar variables in a $(M_{\text{ini}}, v_{\text{ini}}, \tau)$ parameter space. For each stellar age, the two nearest neighbors (from above and from below) in initial mass are selected first, and then—for each of the two initial masses—the two nearest neighbors in initial rotational velocity. Values of stellar evolution variables, at each stellar age, are computed from these four neighboring grid points by a sequence of linear interpolations in the sampled parameter space. The scope of the interpolation method is restricted to the main sequence evolution of stars. Instead of stellar age, the fractional main sequence lifetime is used as interpolation variable.

Following a different approach to interpolation of stellar tracks, the METHOD OF INTERPOLATION FOR SINGLE STAR EVOLUTION code (METISSE; Agrawal et al., 2020) takes as input a discrete single-star parameter grid, and uses interpolation by a piece-wise cubic function to generate new stellar tracks in-between the sampled initial mass grid points at fixed metallicity. The parameter space covers the initial mass range from 0.5 to 50 M_{\odot} , and stars are evolved up to the late stages beyond core helium burning. Instead of stellar age, the interpolation scheme uses a uniform basis, EQUIVALENT EVOLUTIONARY POINTS (EEP; Dotter, 2016), to model evolutionary tracks. The EEP coordinate quantifies the evolutionary stage of a star based on physical conditions, derived from numerical values of evolutionary variables (e.g., depletion of central hydrogen mass fraction to a threshold value), which are readily identifiable for different evolutionary tracks. For any given stellar age, an isochrone is constructed by identifying which EEP coordinate values are valid for that age as function of M_{ini} . For each fixed EEP value, an ordered $M_{\text{ini}} - \tau$ relation is constructed over the available grid points and interpolated over. In a second step, M_{ini} is used as independent variable to obtain stellar properties by another round of interpolation. Reliable and fast stellar track interpolation with the EEP method has originally been demonstrated upon MESA ISOCHRONES AND STELLAR TRACKS (MIST; Choi et al., 2016), a catalog of stellar evolution tracks over a grid space covering the age, initial mass and initial metallicity parameters. METISSE is a more general alternative to SSE, because it may take any single star grid (at fixed initial metallicity)—produced as output of a detailed stellar evolution code—as input, i.e., it is not tied to specific input physics adopted to generate the stellar tracks.

Apart from METISSE, there are the COMBINE (Kruckow et al., 2018), SEVN (Iorio et al., 2023, in its latest version) and POSYDON (Fragos et al., 2023) population

synthesis codes that interpolate grids of detailed single or binary evolution simulations. Interpolation in COMBINE is based on the method of [Brott et al. \(2011\)](#) while in SEVN, single star evolution is divided into sub-phases analogous to the EEP method, and interpolation is performed over each sub-phase using a fractional time coordinate relative to duration of each sub-phase. Evolution of the binary companion and interaction effects are approximated using analytic fitting formulae. Since the procedure to construct the uniform EEP basis cannot be trivially automatized, the pre-processing steps to identify EEPs, to define appropriate interpolation functions and also to down-sample the stellar evolution catalog to reduce memory costs need to be re-iterated after each stellar grid update (see e.g. the TRACKCRUNCHER pre-processing modules, [Iorio et al., 2023](#), in context of SEVN).

In contrast, POSYDON interpolates output of detailed binary evolution simulations with MESA. The EEP-based interpolation method is not directly applicable to binary evolution tracks, because EEPs must be strictly ordered a priori while binary interaction, which can set on at any time, may change their order. Therefore, in POSYDON interpolation needs to be preceded by classification of binary evolution phase, over each of which separate interpolation schemes are applied.

Finally, the third way is to build a prediction-making tool that allows to replace the output of cost-intensive detailed up-to-date stellar evolution code like MESA by a cost-efficient imitation model (emulator, or surrogate) of the original. Emulation, or surrogate modeling, is a pragmatic but reliable reproduction of the output generated by an expensive computer experiment. The predictive surrogate model is constructed by training a supervised machine learning (ML) algorithm on a stellar evolution tracks data base pre-computed with the original code over a discrete parameter grid. A well-trained model will not only efficiently reproduce stellar tracks at the parameter grid points it has seen during training, but be capable of generating accurate predictions of tracks in-between the grid points, by the capability to generalize it acquired by training. Once constructed, the emulator can be used as a package to generate predictions of stellar variables of interest, instead of running the original detailed stellar evolution code like MESA over a quasi-continuous parameter range, or instead of storing the catalog data in computer memory for interpolation. Calling the emulator package saves energy costs, speeds up generation of output predictions and reduces human effort of running models over a dense grid by orders of magnitude. The speed-up is owed to the efficiency of input-to-output mapping by machine learning algorithms. The disadvantage is the introduction of prediction errors by the trained model, which reproduces stellar tracks with a finite precision. Therefore, when training machine learning models, the main task is to achieve reliable generalization over the parameter space with a prediction inaccuracy of stellar variables of interest that is tolerable for inference and astrophysical application.

Surrogate modeling of stellar evolution has yet not been explored extensively at widths of the parameter range necessary for more general applicability. [Li et al. \(2022\)](#) use Gaussian Process Regression (GPR) to emulate stellar tracks in a five-dimensional parameter space, though the initial mass range covered by the predictive models is restricted to the solar-mass neighborhood $M_{\text{ini}} \in (0.8, 1.2) M_{\odot}$, and to evolutionary sequences from the Hayashi line onward through the main sequence up to

the base of the red giant branch. GPR-based emulators have also been used, for example, for parameter space exploration of state-of-art binary population synthesis codes like COMPAS (Barrett et al., 2017; Taylor and Gerosa, 2018). Because of data set size limitation for applicability of GPR, it is not the ideal tool for emulating a large stellar model grid, and we seek for other ML based models instead. The feedforward neural network algorithm proved itself as promising in previous surrogate modeling work: Scutt et al. (2023) emulate 25 stellar output variables (classic photometric variables, asteroseismic quantities, radial and dipole mode frequencies) over a $(M_{\text{ini}}, Z_{\text{ini}})$ grid space of stars in or near the δ Scuti instability strip using neural networks, and principal component analysis to reduce the output dimension to nine. Lyttle et al. (2021) emulate five variables of red dwarfs, sun-like stars and subgiants in a five-dimensional input parameter space. While these are high dimension problems successfully addressed by neural networks, the problem settings have in common that the mass range considered is relatively narrow— $M_{\text{ini}} \in (1.3, 2.2) M_{\odot}$ and $M_{\text{ini}} \in (0.8, 1.2) M_{\odot}$, respectively—and that evolutionary sequences cover the pre-main sequence and only part of the main sequence, or main sequence and sub-giant phase, respectively.

More widely in context of stellar astrophysics, supervised machine learning has been applied to solve the inverse problem of mapping observables to models. For example, a variant of the Random Forest regression model (Bellinger et al., 2016) and invertible neural networks (Ksoll et al., 2020) have been trained to predict fundamental stellar parameters in a high dimensional parameter space given a set of observational variables. Though again, the predictive models were restricted to an initial mass range and evolutionary sequences of stars narrower (e.g., main sequence evolution of $M_{\text{ini}} \in (0.7, 1.6) M_{\odot}$ stars in Bellinger et al. 2016) than those presented in this work, where we consider an initial mass range from red dwarfs to very massive stars evolved from the zero age main sequence up to end of core helium burning.

In this work, we provide two proof-of-concept solutions of automated single star interpolation schemes over a wide parameter span, which—in contrast to the EEP-based interpolation method—do not require mapping out points of interest in stellar parameter space, because they are constructed based on a timescale-adapted evolutionary coordinate that we introduce, whose computation is easily automated. Using the latter for constructing more general interpolation models has the potential to be applicable to larger parameter spaces, such as those found in stellar binaries. The first solution we develop is a surrogate model of stellar evolution, constructed with supervised machine learning. The second is a stellar-catalog-based hierarchical nearest neighbor interpolation (HNNI) method. These find two different trade-off points between efficiency and accuracy of predictions—depending on astrophysical application, either the one or the other is preferable.

This paper is organized as follows. In Sect. 2.2., we describe the methods common to both interpolation scheme solutions that we develop: the regression problem that is addressed, the data base used for constructing predictive models, the timescale-adapted evolutionary coordinate—which is used as the primary interpolation variable—and performance scores that assess quality of predictions. Sect. 2.3. outlines how the two interpolation scheme solutions are set up. For the surrogate model, we report on the choice of loss function, on the selection of machine learning

model class, and on its hyperparameter optimization. For the interpolation-based solution, we explain how HNNI works, and how it differs from interpolation models from previous work. In Sect. 2.4., results are presented that are obtained with both the supervised machine learning and the HNNI. The paper is concluded in Sect. 2.5. with a summary of results, limitations, and an outlook on possible future developments.

2.2. Methods

In Sect. 2.2.1., we define the problem which is addressed by two different predictive frameworks (surrogate modeling of stellar evolution, and catalog-based hierarchical nearest neighbor interpolation), and motivate the two-step approach to fitting stellar evolution tracks. In Sect. 2.2.2., the timescale-adapted evolutionary coordinate is introduced, which we use to set up reliable predictive frameworks, in the two-step interpolation scheme. In Sect. 2.2.3., the methods to prepare the data base are described: a nonlinear sampling density segmentation of the initial mass parameter space, and a data augmentation routine for the core helium burning phase. This data base is used as catalog for interpolation of tracks by HNNI, and as training data for constructing surrogate models. Finally, Sect. 2.2.4. outlines how we evaluate predictive performance of our models based on error metrics.

2.2.1. Regression problem formulation

In 1D stellar evolution codes like MESA, stellar evolution is modeled as a deterministic initial value problem, and observables are predicted by cost-intensive numerical time integration of differential equations. Instead, we formulate the prediction of observables as a regression problem, which is to be addressed by supervised machine learning or by catalog-based interpolation. In a regression problem, the goal is to predict output target variables from input regressor variables, but instead of programming the rules that map the input to the output, in the surrogate modeling case the data-driven approach is used to learn the mapping. We constrain the problem to predicting three stellar surface observables, namely, log-scaled luminosity $Y_L = \log L/L_\odot$, effective temperature $Y_T = \log T_{\text{eff}}/\text{K}$, and surface gravity $Y_g = \log g/[\text{cm} \cdot \text{s}^{-2}]$. These are the target variables to be predicted for a given the input of age τ and initial mass M_{ini} of an isolated non-rotating single star, at fixed solar-like initial metallicity $Z_{\text{ini}} = Z_\odot$.

Stars evolve on different timescales, depending on the evolutionary phase they undergo, on their masses, and on other stellar parameters. Therefore, stellar track fitting across different evolutionary phases and initial masses is a temporal multiscale problem. We confirm the conclusion of Li et al. (2022) that the naive approach of training a machine learning surrogate model $f_{\text{ML}}: (\tau, M_{\text{ini}}) \mapsto Y$ to predict the observable Y , by operating directly on (scaled) age τ , does not result in accurate enough predictions of the post-main sequence evolution (see Fig. 2..11 for an illustration).

Instead, we set up a two-step interpolation scheme:

$$\begin{aligned} \text{Step 1 (age proxy fit)} \quad f_1: (\log \tau, \log M_{\text{ini}}) &\mapsto s, \\ \text{Step 2 (observables fit)} \quad f_2: (s, \log M_{\text{ini}}) &\mapsto (Y_L, Y_T, Y_g). \end{aligned}$$

Here, the evolution of stellar surface variables is modeled as function of a timescale-adapted evolutionary coordinate s —an age proxy—instead of the age τ (step 2). The transition from stellar age to the age proxy is accomplished by a second predictive model (step 1).

We find that the fits of the post-main sequence evolutionary stages resulting from this two-step interpolation scheme are orders of magnitude more accurate, as assessed by standard statistical performance scores, than the direct naive fit. We take the logarithm of initial mass values, in order to exploit the approximate mass-luminosity power law relation, which is a linear variable dependence in log-log space.

2.2.2. The timescale-adapted evolutionary coordinate

The method of using a timescale-adapted evolutionary coordinate, or age proxy, instead of the age variable for fitting stellar evolution tracks has been explored before in stellar astrophysics (e.g. [Jørgensen and Lindegren, 2005](#); [Li et al., 2022](#)). The motivation for this re-parametrization is to reduce timescale variability. Stellar age at computation step i ,

$$\tau_i = \sum_{j=1}^i \delta t_j, \quad (2.1)$$

is a monotonically increasing function which grows cumulatively at an adaptive step size δt_j after each step $j = 1, \dots, i$ of numerical time integration of the differential equations describing stellar structure and evolution. The age proxy variable

$$s_i = \sum_{j=1}^i \delta s_j \quad (2.2)$$

is constructed analogously, but here δs_j is the increment in the star’s Euclidean displacement in a diagram spanned by a set of its physical variables, obtained after the numerical time integration step $j = 1, \dots, i$. For a parametric form of δs , [Jørgensen and Lindegren \(2005\)](#) used the ansatz

$$\delta s_j = \sqrt{\left| \Delta_{j,j-1} \log \frac{L}{L_\odot} \right|^2 + \left| \Delta_{j,j-1} \log \frac{T_{\text{eff}}}{\text{K}} \right|^2} \quad (2.3)$$

where $\Delta_{j,j-1}X = X_j - X_{j-1}$. By construction, this age proxy measures the increase in Euclidean path length of a star along its evolutionary track in the Hertzsprung-Russell (HR) diagram. More recently, [Li et al. \(2022\)](#) suggested another prescription

$$\delta s_j = \left(\left| \Delta_{j,j-1} \log \frac{g}{[\text{cm s}^{-2}]} \right|^2 + \left| \Delta_{j,j-1} \log \frac{T_{\text{eff}}}{\text{K}} \right|^2 \right)^c \quad (2.4)$$

which they tailor to their problem formulation and parameter range. Their age proxy measures the displacement of the star in the Kiel diagram, to the power of a parameter c . After experimentation, they conclude that $c = 0.18$ yields the most uniform distribution of the data they train their models on. At the same time, the authors report fit inaccuracies at transition regions between consecutive evolutionary phases and over the fast ascension of the red giant branch. Over these phases—in contrast to the MS evolution—target variables change rapidly in time, and vary unsteadily even as function of the age proxy. To cure this problem, we have re-defined the timescale-adapted evolutionary coordinate by an altered prescription, whose effect is to not only smooth out transitions in-between stellar phases, but—in addition—also resolve the CHeB phase in a way that allows for reliable stellar track fitting, by keeping the resolution of variability on the same numerical age proxy scale as the previous two phases. To get there, we found as a promising approach returning to the original formulation by [Jørgensen and Lindegren \(2005\)](#) but extending it by a third variable that spans another dimension of the diagram, in which the Euclidean path length is calculated:

$$\delta\tilde{s}_j = \sqrt{\left|\Delta_{j,j-1} \log \frac{L}{L_\odot}\right|^2 + \left|\Delta_{j,j-1} \log \frac{T_{\text{eff}}}{\text{K}}\right|^2 + \left|\Delta_{j,j-1} \log \frac{\rho_c}{[\text{g} \cdot \text{cm}^{-3}]}\right|^2} \quad (2.5)$$

The motivation for introducing another variable into the computational prescription of the path length stems from the fact that during the stable CHeB, stars hardly displace in the HR diagram, although their nuclear composition and hydrodynamic properties undergo substantial changes. In order to adjust the path length prescription, we therefore sought for a suitable stellar-core-related variable. After experimental tests, we found that adding the log-scaled core density $\log \rho_c / [\text{g} \cdot \text{cm}^{-3}]$ has the desirable effect of casting the variability of all target variables of interest onto a unified numerical scale across the three consecutive phases MS, RGB, CHeB, and across the wide initial mass range that we work with.¹

We normalize the age proxy of each initial mass to the range (0, 1). The star is on the ZAMS when $s = 0$, while $s = 1$ when the stellar core has terminated helium burning (TACHeB).

2.2.3. Data base

Stellar evolution catalog: Here, we use MIST ([Choi et al., 2016](#)) as an example data set upon which we formulate and demonstrate our method, train, and validate our predictive models. However, the method we develop is general and not specific to

¹This age proxy computation prescription has the aforementioned desirable effects not only during these, but also during the pre-MS and post-CHeB phases, as shown in Fig. 2..12 in the appendix. Our age proxy construction therefore is a promising general candidate solution to the multiscale problem of stellar evolutionary track fitting beyond the evolutionary sequences considered in this work. It resolves prominent features (e.g. the Henyey MS hook, the first dredge-up, MS turnoff, the Hertzsprung gap, tip of RGB, helium flashes, blue loops, thermal pulsations on the asymptotic giant branch, and white dwarf cooling) across all evolutionary phases we tested over the wide initial mass span.

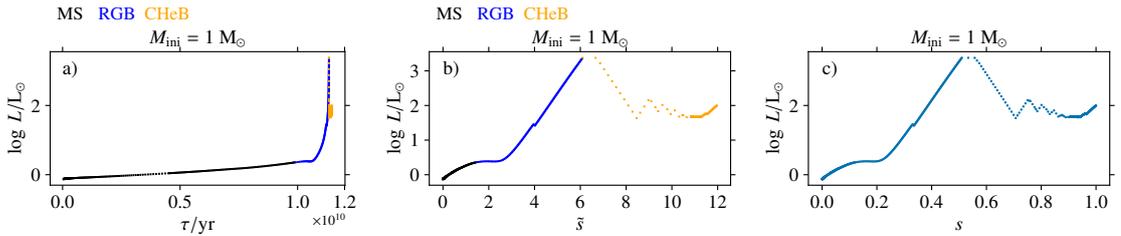


Figure 2.1: Luminosity series of a Sun-like star from the ZAMS up to TACHeB parametrized as function of stellar age τ (a) vs. of the timescale-adapted evolutionary coordinate \tilde{s} , before (b) vs. after (c) CHeB data augmentation and normalization to s . The original MIST data contains phase labels for each model, which the predictive models (the surrogate model and HNNI) do not see.

the MIST data set. We restrict the scope of ages of stars to the evolutionary sequence from ZAMS to the terminal age of core helium burning (TACHeB)², which is expected to account for $\simeq 99\%$ of stellar observations (excluding compact object sequences). The initial mass parameter range, from 0.65 to $300 M_{\odot}$, is chosen as the entire initial mass span available in the MIST data set, over which stars are evolved through all three consecutive phases main sequence (MS), red giant branch (RGB) and core helium burning (CHeB). The wide initial mass range, and at the same time the inclusion of the red giant as well as core helium burning phases have not been explored in previous work of stellar evolution surrogate modeling. We acknowledge that the two-dimensional input parameter space is small compared to the size of the eight-dimensional parameter space required for general cost-efficient binary star modeling. We see our work as a first step toward a large-scale enterprise of stellar evolution surrogate modeling and of hierarchical interpolation in high dimensional parameter space over wide parameter ranges, however, as a lay-out of basic methodology toward this end.

CHeB data augmentation: The MIST data set is generated with the MESA code, which by default outputs more stellar evolution models than what is included in the MIST data set for each M_{ini} -dependent track. The number of models per track is ~ 500 , with ~ 250 models on the MS, ~ 150 on the RGB before ignition of helium burning in the core, and ~ 100 for the CHeB phase. While the MIST data set includes phase labels for each stellar model, the predictive models that we build are not exposed to this information. All the input information they are exposed to is the value of the age (proxy) and of initial mass of the star. While in the MIST data set, the CHeB phase is the least sampled among these three, it is the phase most difficult to fit. In particular, the helium flashes of low mass stars, blue loops of upper main-sequence stars, and fast timescale dynamics of Wolf-Rayet stars during CHeB pose a challenge to fitting. To increase weight and accuracy of interpolation fits during the CHeB phase, we use local nearest neighbor 1D linear interpolation of the training data (not of the test data) along the age proxy axis (for the step 2 fit) or along the

²The end of core helium burning is determined by the condition $X_{\text{He,central}} \leq 10^{-3}$ where $X_{\text{He,central}}$ is the central helium mass fraction.

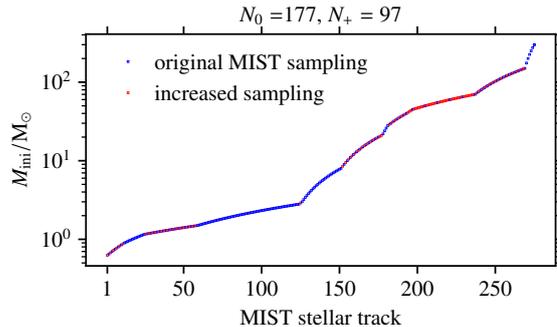


Figure 2..2: Original initial mass sampling in the MIST catalog (in blue), and the locally increased sampling (in red) that we use for training surrogate models. The stock MIST catalog contains 177 solar metallicity stellar evolution tracks within the initial mass range $(0.65, 300) M_{\odot}$. For our purposes, we expand it to 274, to achieve the desired quality of predictive accuracy necessary for a general-purpose stellar evolution emulator.

scaled age axis (for the step 1 fit) during this phase. The net effect is an artificial increase in the CHeB training data by insertion of a sample in-between each pair of age proxy neighbors. Despite simplicity of this methodological step, we find the predictive performance of our best-fit models to be boosted by around half an order of magnitude decline in the mean squared error over the validation data (to which CHeB data augmentation is not applied), after switching on CHeB data augmentation of the training data. In Fig. 2.1 the data pre-processing consisting of age proxy re-parametrization, normalization and CHeB data augmentation is illustrated upon the example of the Sun-like stellar model.

Parameter space grid sampling: A recommended standard routine for a homogeneous sampling of the parameter space that produces the data for training surrogate models is LATIN HYPERCUBE SAMPLING (LHS) (McKay et al., 1979a). LHS is an efficient alternative to random uniform and rectilinear sampling methods for homogeneity. Random sampling introduces sampling voids by consequence of statistical random clumping effects while dense rectilinear sampling is too expensive in many problem settings. However, since the stellar evolution dependence on the initial mass parameter is strongly non-linear, a homogeneous population of parameter space is not the optimal sampling scheme. We work with the pre-computed MIST data set for which a segmented parameter sampling density across the initial mass range has already been pre-determined by the makers of the catalog, based on physics-informed considerations.

In order to reach a high accuracy level of stellar track forecasts—necessary for a general-purpose stellar evolution emulator—across the entire initial mass range, we found that it is needed to locally increase the initial mass sampling. Practically, we increase initial mass sampling in those parameter space sub-regions where the fit quality is worst, while we keep the MIST stock sampling intact where the local fit accuracy is found to be satisfactory (see Fig. 2..2 and Tab. 2..1 for a summary). For

Table 2.1: Summary of the initial mass sampling density segmentation before ($\delta_0 M_{\text{ini}}$) and after (δM_{ini}) expanding the stock MIST data set.

Range (M_{ini}/M_{\odot})	$\delta_0 M_{\text{ini}}$ [M_{\odot}]	δM_{ini} [M_{\odot}]	$N_{M_{\text{ini}}}$
(0.65, 0.9)	0.05	0.025	10
(0.9, 1.16)	0.02	0.02	13
(1.16, 1.5)	0.02	0.01	34
(1.5, 2.8)	0.02	0.02	65
(2.8, 3)	0.2	0.1	2
(3, 8)	0.2	0.2	25
(8, 21)	1	0.5	26
(21, 22)	2	1	1
(22, 28)	2	2	4
(28, 40)	2	1	12
(40, 45)	5	1.25	4
(45, 70)	5	0.625	40
(70, 150)	5	2.5	32
(150, 300)	25	25	6

generating the additional stellar tracks, we use the MIST Web Interpolator³, which works by applying the EEP-based method referred to in Sect. 2.1.. Our finding is that the final sampling required to reach the predictive accuracy goal varies substantially depending on sub-region of parameter space: a least $\delta M_{\text{ini}}/M_{\odot} = 0.01$ between $M_{\text{ini}}/M_{\odot} \in (1.16, 1.5)$ and a largest $\delta M_{\text{ini}}/M_{\odot} = 25$ between $M_{\text{ini}}/M_{\odot} \in (150, 300)$. For the $M_{\text{ini}}/M_{\odot} \in (0.65, 0.9)$ interval, we double the sampling to correct for a systematic under-representation of red dwarfs in the stock MIST catalog as compared to the adjacent initial mass intervals. For the $M_{\text{ini}}/M_{\odot} \in (1.16, 1.5)$ interval, we double the sampling rate mainly because of complexity of shape changes in HR diagrams due to the helium flashes. In the interval $M_{\text{ini}}/M_{\odot} \in (1.5, 40)$, we hardly increase the sampling, except at transitions in-between neighbouring sampling segments at different rates, in order to smooth out transitions. The biggest increase in this range is within the interval $M_{\text{ini}}/M_{\odot} \in (8, 21)$. We stress that our densest sampling region (the solar neighborhood initial mass range) is the same as in Li et al. (2022), in Bellinger et al. (2016) and in Lyttle et al. (2021), while at the same time our surrogate models evolve the stars further, up to end of CHeB, and cover a much wider initial mass range. Scutt et al. (2023) adopt a sampling of $\delta M_{\text{ini}} = 0.02 M_{\odot}$ over the range $M_{\text{ini}}/M_{\odot} \in (1.3, 2.2)$, comparable to ours.

At the high mass end, the relative increase in sampling is greatest within the interval $M_{\text{ini}}/M_{\odot} \in (40, 70)$, where the increment step size $\delta M_{\text{ini}}/M_{\odot}$ was augmented from 5 to 0.625. We suspect that numerical challenges are the reason for unexpectedly sharp, peculiarly shaped changes in HR diagrams. Nevertheless, for the proof-of-concept, we assume as if MIST was a perfect data set, even if we know that it may be not.

Naturally, the denser the grid sampling, the more accurate are the forecasts of

³https://waps.cfa.harvard.edu/MIST/interp_tracks.html

surrogate models. We stress that depending on minimal performance benchmarks (as quantified by error scores) of a specific astrophysical application, the initial mass sampling required to reach that benchmark can be significantly sparser.

With the initial-mass parameter space sampling as described above, the total size N_{tot} of the data set amounts to 139016. Shuffling it, we do a uniform random split of the N_{tot} into 85 % training (N_{train}) and 15 % validation (N_{val}) data sets. To the N_{train} data, we apply CHeB data augmentation, which yields additional $N_{\text{aug}} = 32143$ samples, such that the expanded training data set is of size $N'_{\text{train}} = N_{\text{train}} + N_{\text{aug}}$. This is the final data set on which we train different classes of surrogate models (for the first solution) or which we use as the catalog for interpolation (for the second solution).

2.2.4. Performance evaluation

Validation and test data: We use two schemes to evaluate performance of predictive models: the first (model validation) based on the validation data set, and the second (model testing) based on the test data set. The validation data consists of randomly selected grid points over the input domain (initial masses and evolutionary phases of stars). It is representative, since it has similar statistical properties as the training data. In contrast, for model testing, we aim to assess the trained model’s capability to predict entire stellar tracks from ZAMS up to TACHeB for initial masses unseen during training. We choose this method of model testing since it is of main interest to obtain a predictive model that is capable of accurate interpolation over the space of fundamental stellar parameters. Only then the traditional method of running expensive simulations over densely sampled grids can be replaced by a surrogate model capable of sufficiently accurate generalization. As test data, we prepare another set of stellar tracks at 16 initial mass grid points, $\{M_{\text{ini}}^{\text{test}}/M_{\odot}\} = \{0.91, 1.51, 2.41, 4.1, 8.25, 16.25, 21.5, 31.5, 41, 51, 61, 83.75, 103.75, 155, 262.5, 295\}$, which we hold back from training. These are chosen at half of the grid step in the respective region of parameter space. This choice is motivated by the aim to test predictive accuracy at parameter space points that are farthest away from training grid points, where we likely probe the worst cases of complete stellar track predictions.⁴

Performance scores: A crucial ingredient for the optimization procedure of an automated interpolation method is a set of appropriately designed scores that quantify performance in a physically meaningful and numerically appropriate manner. Only with adequately defined quantitative performance scores, the automated interpolation scheme can be scaled up to higher dimensional fundamental stellar parameter spaces, which become too large for visual inspection based performance evaluation for comparing the predicted against the held-back test tracks.

⁴ An alternative approach to choosing test initial mass grid points is to sample the initial mass range randomly, in order to obtain the statistically likely distribution of prediction errors of stellar tracks. Since we quantitatively probe statistical error distribution already on the validation data set by appropriate statistical error scores, we opt for the half-grid step approach to probe the worst cases instead.

For model validation on the validation data set, we look at residuals for each observable independently, and at measures of overall predictive performance. A residual is a signed prediction error, $\epsilon_i = Y_i - \hat{Y}_i$, of a given prediction-label pair (\hat{Y}_i, Y_i) , and we evaluate it for each of the surface variables. We consider the following error scores that retain the physical significance of residuals: the mean residual $\bar{\epsilon} = (\epsilon_1 + \dots + \epsilon_{N_{\text{val}}})/N_{\text{val}}$, the most extremal under-prediction $\epsilon^+ = \max_{i=1, \dots, N_{\text{val}}} \{\epsilon_i\}$, and the most extremal over-prediction $\epsilon^- = \min_{i=1, \dots, N_{\text{val}}} \{\epsilon_i\}$. If ϵ^+ and ϵ^- are close enough to zero over the entire set of validation data grid points, then there is no need to further stratify the performance evaluation.⁵ Additionally, we use the following error scores to quantify overall predictive performance across the three surface variables: the Mean Squared Error (MSE) and the Mean Absolute Error (MAE). These scores are calculated from the squared residuals and from the absolute residuals, respectively, by taking the average variable by variable and over the three surface variables. We choose the MSE and the MAE, because these are standard choices for evaluating point forecasts generated by statistical learning models, but physical significance is largely lost by averaging across surface variables.

For model testing on the held-back test tracks at the 16 M_{ini} grid points stated above, we define and use the following error scores based on HR and Kiel diagrams: $L2_{\text{HR}}^+$ and $L2_{\text{K}}^+$. For a single track in HR or in Kiel diagram at a particular initial mass M_{ini} , the $L2$ score measures the cumulative deviation between predicted track and held-back track,

$$L2(M_{\text{ini}}) = \frac{1}{N(M_{\text{ini}})} \sum_{i=1}^{N(M_{\text{ini}})} * \vec{v}_i - \hat{\vec{v}}_i^2, \quad (2.6)$$

computed as the mean squared Euclidean distance in a two-dimensional plane of target variable pairs: $\vec{v}_i = (\log L_i, \log T_{\text{eff},i})$ for the HR diagram, and $\vec{v}_i = (\log g_i, \log T_{\text{eff},i})$ for the Kiel diagram. This measure agrees reasonably well with the visual assessment of how closely a predicted track aligns with the true test track. As summary measures of predictive performance on the test data, we take the maximum $L2$ measure, $L2^+ = \max \{L2(M_{\text{ini}}^{\text{test}}/M_{\odot})\}$ among the 16 initial masses of the test set, for each type of diagram, i.e. $L2_{\text{HR}}^+$ and $L2_{\text{K}}^+$.

2.3. Interpolation scheme solutions

In this section, we describe the methodology behind the development of the two solutions to cost-efficient stellar evolution forecasting over continuous parameter spaces. For construction of a stellar evolution emulator with supervised machine learning, we treat the selection of the surrogate model class in Sect. 2.3.1.. Then, we discuss loss function choice (Sect. 2.3.2.), and outline our training and hyperparameter optimization methods to obtain the best-fit model (Sect. 2.3.3.), which is a feedforward

⁵The statistical performance assessment can be further stratified by applying the scoring prescriptions not globally over the entire initial mass range and over the full evolutionary sequence, but to confined sub-regions of parameter space. For instance, $\epsilon_L^+(1.2, 4.5)_{\text{RGB}}$ is the most extremal under-prediction of log-scaled luminosity during the RGB phase within the initial mass segment $M_{\text{ini}}/M_{\odot} \in (1.2, 4.5)$.

neural network. The hierarchical nearest neighbor interpolation method is subject of Sect. 2.3.4..

2.3.1. Supervised learning model selection

There are different surrogate model class candidates available for tackling the regression problem defined in Sect. 2.2.1.. For selection of statistical learning algorithms, the following three requirements apply in our problem case: first: applicability to a large data set ($N > 150k$), second: multiple output⁶, and third: fast computational speed in forecast generation, for applicability of the surrogate model at scale. Below, we discuss a number of available options, and justify our choices.

Choice of statistical learning model: GPR has been considered the standard model choice for emulation tasks (Sacks et al., 1989). However, because of memory limitations, the default implementation of global GPR is not applicable to large training data sets. While there exist approaches to improve scalability of GPR, we do not opt for GPR-based emulators for reasons discussed in Appendix A3.. Instead, we test the performance of a number of regression models that satisfy the aforementioned constraints. After a series of manual tests, we find satisfactory starting performance with the k-Nearest Neighbors (Fix and Hodges, 1989), Random Forest (Ho, 1995), and Feedforward Neural Network (Ivakhnenko and Lapa, 1967; Rumelhart et al., 1985) regression models classes, all of which are efficient statistical learning algorithms that qualify as scalable predictive models with multiple output. Among them, in order to identify which model class is the best choice for construction of a high-accuracy surrogate model, we perform hyperparameter optimization of each of these three, to cross-compare their performance as assessed by the scores defined in Sect. 2.2.4.. We perform hyperparameter optimization of k-Nearest Neighbors (KNN) and Random Forest (RF) regression models by a grid search, with a sampling of numerical hyperparameters over a log scale, and do model selection based on 3-fold cross-validation. For the Feedforward Neural Network (ffNN) model, which has a much larger space of options for hyperparameter choices, we determine a preliminary best-fit hyperparameter configuration after training hundreds of models over a high-dimensional but coarsely sampled hyperparameter grid. We then take it as a starting configuration from which we further optimize hyperparameter selection over a series of manual experiments. The result is that a manually tuned Feedforward Neural Network (ffNN) outperforms KNN and RF models that have been optimized through grid search, as assessed by the majority of error metrics defined above (see Tab. 2..3). The KNN and RF best-fit models therefore serve us primarily as benchmarks for ffNN performance.

Deep learning models: ffNN is one out of many available deep learning architectures. We opt for a ffNN architecture because in our regression problem, the input is

⁶The multiple output condition (three target variables predicted by a single surrogate model) is motivated by pragmatic considerations: predicting a multitude of stellar variables each with a separate surrogate model requires substantially more effort, if the desired number of output variables of interest is large.

a vector of fixed dimension. To discriminate, we do not train, for instance, a recurrent neural network based architecture, which is the model class of choice if the input is a sequence of variable length; nor a convolutional neural network architecture, which is model class of choice if the input is a higher dimensional topological data array. A motivation for choosing a fNN architecture is the established theoretical result that a fNN with a number of hidden layers ≥ 1 is capable of universal function approximation (Hornik et al., 1989).

2.3.2. Choice of loss function

Choosing a loss function appropriate to the problem is a crucial step, because it defines the training goal for the supervised machine learning. During optimization of a fNN, its trainable parameters are iteratively updated, after each batch, to minimize the loss score. Choosing one error score over another is a trade-off to compromise which type of error is least tolerable against other types of errors. Common choices of scoring rules (for a more detailed reference on scoring rules for point forecast evaluation, see Gneiting, 2011) for model training as well as for point forecast evaluation are the MAE and MSE. Other choices include the Mean Squared Logarithmic Error (MSLE) and the Mean Absolute Percentage Error (MAPE). For our problem case, loss function selection was guided by the following considerations.

MAPE is not the appropriate loss function since, for instance, changes in log-scaled luminosity of massive stars in HR diagram happen on a smaller relative numerical scale than for low mass stars, and prediction errors in that range would therefore hardly be penalized. Furthermore, we do not opt for MAPE for reasons outlined in Tofallis (2015). When choosing MSLE as loss function, we observed an inefficient learning, with a too slow decline of MSE, MAE and our physical performance scores over the validation data. However, we also found neither MAE nor MSE to be optimal choices for our problem. Using MAE kept the mean averaged error scores low but admitted considerable prediction outliers. Conversely, using MSE reproduced the global shape of the distribution of values of the target variables, but predictions of stellar tracks were often not precise enough locally, and overfitting occurred at epochs much earlier than when minimizing MAE. Instead, we opt for the Huber loss (Huber, 1964), which seeks a trade-off between MAE and MSE minimization. It penalizes MSE-like for small prediction errors and MAE-like for large prediction errors, using a parameter d for the transition threshold (for a recent discussion and generalization, see Taggart, 2022):

$$L_d(Y, \hat{Y}) = \begin{cases} \frac{1}{2}(Y - \hat{Y})^2 & \text{for } Y - \hat{Y} \leq d \\ dY - \hat{Y} - \frac{1}{2}d^2 & \text{otherwise.} \end{cases} \quad (2..7)$$

During supervised learning, the Huber loss $L_d(Y, \hat{Y})$ issues a penalty for each point prediction error, given the prediction \hat{Y} by the surrogate model and the true label Y it is compared against. When training deep learning models to predict multiple output, the mean Huber loss is computed as the average across target variables, that is, over the set of labels and over multiple output predictions $\{\vec{Y}_j, \hat{\vec{Y}}_j\}_{j=1, \dots, n_b}$ that are obtained from one randomly sampled data batch of size n_b . We find our best results,

as assessed by the physically meaningful performance scores outlined in Sect. 2.2.4., with $d = 0.75$. Once a desired target value of the validation loss score is set, which comes in hand with low enough physical performance scores over the validation data, what is left is to seek a suitably configured deep learning model that reaches this target value.⁷

2.3.3. Hyperparameter optimization

There are two types of hyperparameters to optimize when constructing fNN-based emulators: architecture and learning hyperparameters. The most important architecture hyperparameters are the number of layers, the number of neurons per layer, the choice of activation function, and the kernel initialization. Typical important learning hyperparameters are the learning rate, the batch size, the choice of optimizer, and the choice of regularization method. There are three different ways to optimize hyperparameters: first, by manual fNN learning engineering; second, by automated brute-force search methods (for instance, grid or random search); third, by sophisticated search algorithms (for instance, *Bayesian optimization* or *Genetic Evolutionary Search*). We opt for manual fNN learning engineering instead of automated searches, because for deep learning models, the optimal stage when (i.e., at which epoch⁸) to stop training cannot be faithfully decided upon a priori, and requires careful consideration of numerical criteria for stopping training if models are optimized in an automated pipeline. Most reliably, it is determined a posteriori by inspection of the fluctuating training and validation data loss curve declines during run-time. Then, we continue training so long as the degree of overfitting is tolerable. We consider overfitting as tolerable so long as the validation loss—even though it may be decaying slower than the training loss at advanced learning stages (i.e. at large epoch numbers)—has yet neither reached the flattening plateau stage nor started to increase.

Best-fit model: For theoretical considerations regarding hyperparameter tuning and the selection criteria we used, refer to Appendix A4.. In practice, we found a successful hyperparameter tuning strategy (guided by Goodfellow et al., 2017) with the following configurations (see Tab. 2.2 for a summary): First, a symmetric many-layer (6 hidden layers) architecture with a moderate number of neurons per layer (128), rectified linear unit (ReLU; Hahnloser et al., 2000) activation, *Glorot* uniform (GU; Glorot and Bengio, 2010) kernel initialization, and *Layer Normalization* (LN; Ba et al., 2016) regularization after each layer. *Layer Normalization* counteracts overfitting while the 8-layer architecture with 128 neurons per hidden layer yields a large enough model capacity to prevent underfitting by over-parametrization. Second, long-term training ($\sim 70k$ epochs) at relatively small (512) batch size. Observation of the degree of fluctuation of the loss curves is a means to assess exploration of the high-dimensional trainable parameter space spanned by the biases and by the

⁷See Appendix A5. for caveats regarding choice of the loss function.

⁸One single epoch is over, once the entire training data set—presented to the network in batch subsets—has been propagated through the network.

weighted connections between neurons from neighboring layers in each backpropagation step. The small batch size (as compared to the size of N'_{train}) adds stochasticity to the learning, and thereby ensures enough exploration, which is aimed to prevent early flattening of the validation loss curve. Third, a learning rate schedule of slow exponential decay in the *Adam* optimizer (Kingma and Ba, 2014): starting with a large enough initial learning rate $\text{lr}_i = 10^{-3}$ (to accelerate the gradient descent at beginning stages of learning), and decreasing the learning rate down to a final $\text{lr}_f \sim 5 \cdot 10^{-6}$ towards end of training (in order to target global rather local minima in the value space of trainable network parameters). The slow gradual decrease is aimed to improve on subtle prediction errors.

Table 2.2: Summary of loss function choice, architecture and learning hyperparameters adopted for training our best-fit fNN model, compared to those adopted by Scutt et al. (2023).

Hyperparameter	Our choice	Scutt et al. (2023)
# of hidden layers	6	6
# of neurons per layer	128	64
Activation function	ReLU	ELU
Kernel initializer	GU	
Regularization	LN	/
Batch size	512	$6 \cdot 10^4$
Optimizer	Adam	Adam
lr schedule	Exp. decay	Fixed lr
lr range	$(10^{-3}, 5 \cdot 10^{-6})$	$7 \cdot 10^{-5}$
Loss function	Huber loss	MSE

2.3.4. Hierarchical Nearest Neighbor Interpolation (HNNI)

In this section, we present a second method to solve the problem by a HNNI scheme. Our construction of the HNNI algorithm was partly motivated by an attempt to customize the operation of the KNN algorithm to our problem setting. In KNN, the nearest neighbors are selected based on a pre-defined distance metric (for instance, Euclidean or Manhattan) over the input parameter space (which, in our case, is two-dimensional), without treating the regressor dimensions apart from one another. The key principle behind the HNNI method is to select the nearest available grid points, from above and from below, in each parameter space direction, to the location in parameter space at which the interpolation prediction is to be made, and then apply a 1D interpolation prescription subsequently in each parameter space direction according to a hierarchical order of parameters. Our method works similar to Brott et al. (2011) in that it performs a sequence of linear interpolations separately in each parameter space direction according to a hierarchical ordering of stellar variables, but different from it in that it uses a timescale-adapted evolutionary coordinate, instead of fractional age, as primary interpolation variable. We thereby show that the method is applicable not only to the MS evolution but to a sequence of evolutionary phases.

In this regard, our method is analogous to [Agrawal et al. \(2020\)](#) in that it uses an adapted evolutionary coordinate to trace the evolution of stars across phases, but we use a prescription for it that allows to automate its computation.

We prepare the data set for generating predictions with HNNI under exactly the same conditions as in the supervised machine learning case. The N'_{train} is now used as a catalog data base, upon which the hierarchical nearest neighbor interpolation is performed, instead of serving as the training data for fitting a surrogate model. The HNNI method requires continued access to the pre-computed stellar evolutionary tracks catalog. The HNNI method is applied separately to each of the three surface variables Y_L , Y_T , and Y_g , for obtaining point forecasts at unseen locations in parameter space.

As will be shown in Sect. 2.4.1., the HNNI is applicable reliably over the entire initial mass range and over all three evolutionary phases, including the transitions in between them, without the need to map out points of interest for that purpose. The level of predictive accuracy of HNNI is achieved for two main reasons. First, HNNI operates on local parameter space regions immediate to the test location at which a prediction is to be made. Predictions are calculated by an interpolation scheme that treats different dimensions apart from one another. This stands in contrast to the way fNN, RF and KNN operate. RF and fNN take the global properties of the input parameter space into account, in order to find their own rules for making local predictions. This can be a great benefit in some problem settings, but irrelevant in others. Likewise to HNNI, KNN also operates on local environments but does not take hierarchy relations among input parameters into account. Second, HNNI uses the normalized timescale-adapted evolutionary coordinate s as primary interpolation variable, without which the interpolation scheme would not produce accurate results. By virtue of using the latter, interpolation-based predictions at transitions between evolutionary phases are mostly accurate because meanwhile values of stellar log-scaled luminosity, effective temperature or core density variables change drastically. Therefore, the path length increment δs —which is computed from absolute increments in these variables—increases significantly, resulting in a higher resolution—along the age proxy axis—of the transition stages between evolutionary phases.

Given the initial mass parameter space sampling used in this work, a linear interpolator was sufficient for making accurate forecasts. More generally, for each parameter space dimension, a different (for instance, a quadratic or cubic polynomial) functional could be applied instead.

For clarity, we outline the pseudo-code of HNNI in a 3D ($s, M_{\text{ini}}, Z_{\text{ini}}$) single star parameter space in Appendix A2.. We believe that the HNNI method, in its basic principle, is applicable to those higher dimensional parameter spaces that allow for a sequential ordering of the parameters in importance of their effect on the shape of resulting stellar evolutionary tracks.

2.4. Results

In this section, the prediction results—obtained with the deep learning surrogate model and with the HNNI algorithm—are analyzed. We treat the observables fit (step 2) first (Sect. 2.4.1.), because it yields the physically meaningful outcome:

Table 2..3: Ranking of the predictive models Hierarchical Nearest Neighbor Interpolation (HNNI), feedforward Neural Network (ffNN), Random Forest (RF) and k-Nearest Neighbors (KNN) regressors, according to the performance scores outlined in Sect. 2.2.4. to assess predictive accuracy of stellar observables. The best performance is marked in bold, the worst with a ”*” tag. The manually tuned ffNN outperforms the grid search hyperparameter optimized RF and KNN models according to all scores except $\bar{\epsilon}_T$. HNNI outperforms ffNN as assessed by all scores except $\bar{\epsilon}_L$, ϵ_g^+ and ϵ_L^- .

Score	HNNI	ffNN	RF	KNN
validation data set				
$\bar{\epsilon}_L$	6.57E-05	-4.46E-05	2.612E-04*	2.506E-04
$\bar{\epsilon}_T$	-4.94E-06	1.82E-04*	-2.00E-05	-1.758E-05
$\bar{\epsilon}_g$	-9.22E-05	4.16E-04	-4.77E-04*	-4.55E-04
ϵ_L^+	0.102	0.145	0.210	0.217*
ϵ_T^+	0.014	0.032	0.093	0.095*
ϵ_g^+	0.169	0.165	1.05	1.08*
ϵ_L^-	-0.115	-0.108	-0.700	-0.721*
ϵ_T^-	-0.011	-0.016	-0.0378*	-0.0375
ϵ_g^-	-0.143	-0.191	-0.286	-0.294*
MSE	1.11E-05	2.01E-05	2.39E-04	5.79E-04*
MAE	0.00041	0.00193	0.00479*	0.00270
test data set				
L_{HR}^+	0.0166	0.0176	0.0319*	0.0237
L_{K}^+	0.0225	0.0283	0.0442*	0.0283

the prediction of stellar evolution variables and tracks. Therefore, in our two-step interpolation scheme, the observables fit needs to reach a satisfactory level of accuracy first, which can be assessed physically, before approaching the age proxy fit (step 1). Then, the performance baseline for the age proxy fit is set by the condition that the predictive accuracy of the integral two-step interpolation scheme is maintained on the same order of magnitude, as assessed by the scores. We analyze the step 1 fit in Sect. 2.4.2..

2.4.1. Prediction of classical photometric observables

Deep learning emulation

Validation data: The performance assessment on the validation data is presented in Fig. 2..3 by histograms of the residuals and by the summary statistics, defined in Sect. 2.2.4., individually for each of the three predicted surface variables. If we assume that the prediction errors of Y_L , Y_T , and Y_g were scored over the same numerical scale, then the following conclusions could be made. The mean residual, in absolute value, is largest for $\log g$ and lowest for $\log L$, while the most extremal over-prediction and under-prediction are obtained for the $\log g$ target variable. All three mean residuals take on low numerical values of order 10^{-4} or 10^{-5} . These error scores are comparable to those found with the best-fit neural network model of [Scutt et al. \(2023\)](#) ($8 \cdot 10^{-4}$

dex on $\log L$ and $2 \cdot 10^{-4}$ dex on $\log T_{\text{eff}}$) who address a similar regression problem. Since $\bar{\epsilon}_{\text{res}}$ is negative for $\log L$ but positive for $\log T_{\text{eff}}$ and $\log g$, the deep learning emulator tends to over-predict the first, but to under-predict the latter two. The most extreme prediction outliers are of the order 10^{-1} or 10^{-2} in absolute value, i.e. up to three orders of magnitude larger than the mean residuals. To better characterize the distribution of errors, we therefore compute an additional score, σ_{ϵ} , which is the standard deviation of the residuals over each target variable. It is a measure of the spread of the prediction errors around the mean residual error, which we find to be of order 10^{-3} for each of the three target variables.

Comparison to observational uncertainties: It is of interest to compare the mean residual errors on the target variables to the typical uncertainties from observations of stars. For stellar bolometric luminosity, the relative error is of order $\delta L/L \propto 0.01$ for Gaia observations of solar-like stars (Creevey, O. L. et al., 2023), which translates into $\delta \log L/L_{\odot} = \frac{\delta L}{L} \log e \propto 0.004$. For surface gravity, with $\delta \log g/[\text{cm} \cdot \text{s}^{-2}] \propto 0.1$ (see e.g. Ryabchikova et al., 2016) it is comparatively large. For effective temperature of low mass stars, the observational error is of order $\delta T_{\text{eff}}/\text{K} \propto 50 - 100$ depending on stellar class and spectral method (Ryabchikova et al., 2016). For massive stars, the observational uncertainty on the classical observables typically ranges $\delta \log L/L_{\odot} = 0.1$, $\delta T_{\text{eff}} \propto 500 - 2000 \text{ K}$ and $\delta \log g/[\text{cm} \cdot \text{s}^{-2}] \propto 0.1 - 0.2$ (Schneider et al., 2018b,a).

In sum, the mean residual errors on all three target variables are smaller than the typical observational errors on the same log-scaled quantities. In the case of ϵ_L and ϵ_g , not only these, but also the expected spreads σ_{ϵ} are smaller, by 1–3 orders of magnitude depending on statistical score. This means that the prediction errors from the emulator are greater than the observational uncertainties only when the prediction errors belong to the tail of their integral empirical histogram, which comprises cases that are statistically rare. For ϵ_T , the histogram of linear-scaled residual errors $T_{\text{eff}} - \hat{T}_{\text{eff}}$ yields a mean residual error of $\simeq 8.3 \text{ K}$, an expected spread of $\simeq 85 \text{ K}$, a worst over-prediction outlier of $\simeq 1385 \text{ K}$ and a worst under-prediction outlier of $\simeq 2885 \text{ K}$ in absolute values. The expected spread is smaller than the observational uncertainty $\delta T_{\text{eff}}/\text{K}$ but of a similar order of magnitude. Therefore, inference on effective temperature of low mass stars using the emulator is—under assumption of the aforementioned observational uncertainties—least reliable, out of the three surface variables, in a practical setting.

Test data: For model testing on the test data, in order to predict evolutionary tracks in the HR diagram, we compute the values of target variables $\log L_i$, $\log g_i$, $\log T_{\text{eff},i}$ at the evolutionary coordinate grid points $\{s_i\}_{i=1, \dots, N(M_{\text{ini}}^{\text{test}})}$ contained in the held-back series for each test initial mass $M_{\text{ini}}^{\text{test}}$. We then plot pairs of predicted target variables against one another, to obtain the predicted tracks in the HR and in the Kiel diagram, respectively. These can now be compared with the held-back tracks test data in the diagrams. As shown in Fig. 2.4, the shape of the stellar tracks is reproduced by the deep learning surrogate models across the entire initial mass range. For a closer resolution of predictive quality, Fig. 2.5 displays the best and worst prediction, respectively, of stellar evolution tracks in HR diagram at unseen

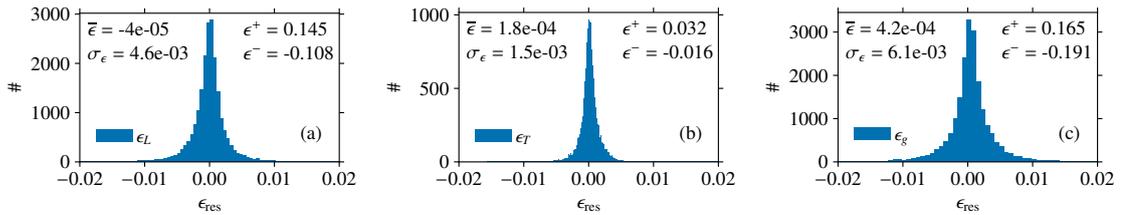


Figure 2.3: Validation data results for the ffNN-based stellar evolution emulator. The histograms and summary statistics of the residuals $\epsilon_k = Y_k - \hat{Y}_k$, over the validation data $k = 1, \dots, N_{\text{val}}$ are shown, for Y_L in panel (a), for Y_T in panel (b), and for Y_g in panel (c). We calculate the mean $\bar{\epsilon}$, the standard deviation σ_ϵ , the most extremal under-prediction ϵ^+ , and the most extremal over-prediction ϵ^- . Overall, the distribution of residuals is globally symmetric around near 0, with a sharper peak than a Gaussian, reminiscent of a Cauchy distribution.

test data initial mass grid points. The biggest deviation between predicted and held-back test stellar track is observed at the low mass end (worst fit for $M_{\text{ini}}^{\text{test}} = 0.91 M_\odot$). There are two main reasons for this. First, as low mass stars displace in the HR diagram from ZAMS up to TACHeB, they cover a larger spread in value range of log-scaled luminosity than higher mass stars, due to the stretched-out (in the HR diagram) ascension of the red giant branch. Second, the main contribution to cumulative deviation of predicted to the actual test track for low mass stars arises during the unstable core helium burning, the sequence of short-lived helium flashes. The helium flashes introduce the most prompt transition in both the $\log L$ and the $\log T_{\text{eff}}$ variables. Since these are physically uncertain from the modeling perspective, it therefore is not as important to obtain high accuracy prediction of flashes compared to other parts of the stellar evolution track. We evaluate our state-of-art worst fit as satisfactory, since the more reliable (from the modeling perspective) evolution before and after the flashes is well reproduced by the surrogate model: the evolution up to the tip of the RGB, and the stable core helium burning after electron degeneracy in the core is lifted.

HNNI

Stellar track predictions in the HR and Kiel diagram are obtained in the same way as described above for the deep learning case. Performance of the HNNI predictive model is assessed using the same data bases, procedures and metrics as the supervised machine learning models. The outcome is that over the validation and test data, HNNI even outperforms the deep learning method in accuracy of predictions, although not significantly, as is measured by the majority of statistical scores (see Tab. 2.3). Over the test data, the HNNI method yields accurate predictions of stellar evolutionary tracks across the entire initial mass range and over all three evolutionary phases, including the fast-timescale transition regions. To illustrate, Fig. 2.5 shows the best and worst fit of a stellar track in HR diagram over the test data. The HNNI and deep learning models agree on the worst fit for $M_{\text{ini}}^{\text{test}} = 0.91 M_\odot$ for reasons explained above. In the HNNI case, the worst fit is resolved at higher

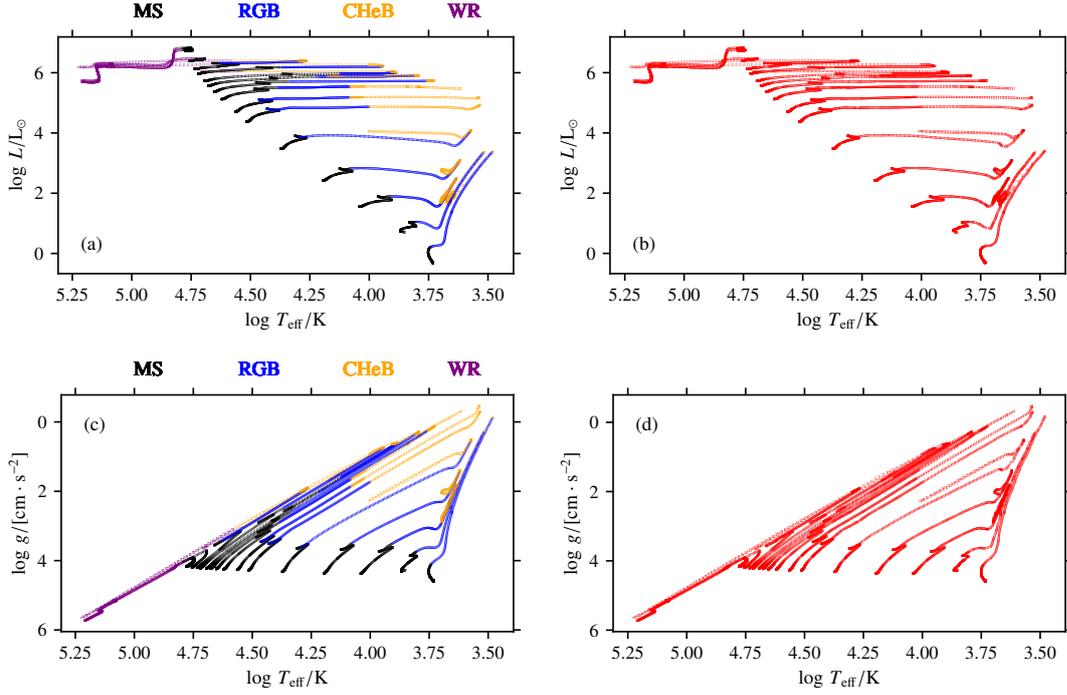


Figure 2.4: Test data results, comparing the true (left) and the fNN-predicted (right) stellar evolutionary tracks in HR (top) and Kiel (bottom) diagrams, over the entire set of our test initial masses $\{M_{\text{ini}}^{\text{test}}\}$ unseen by the predictive model during training.

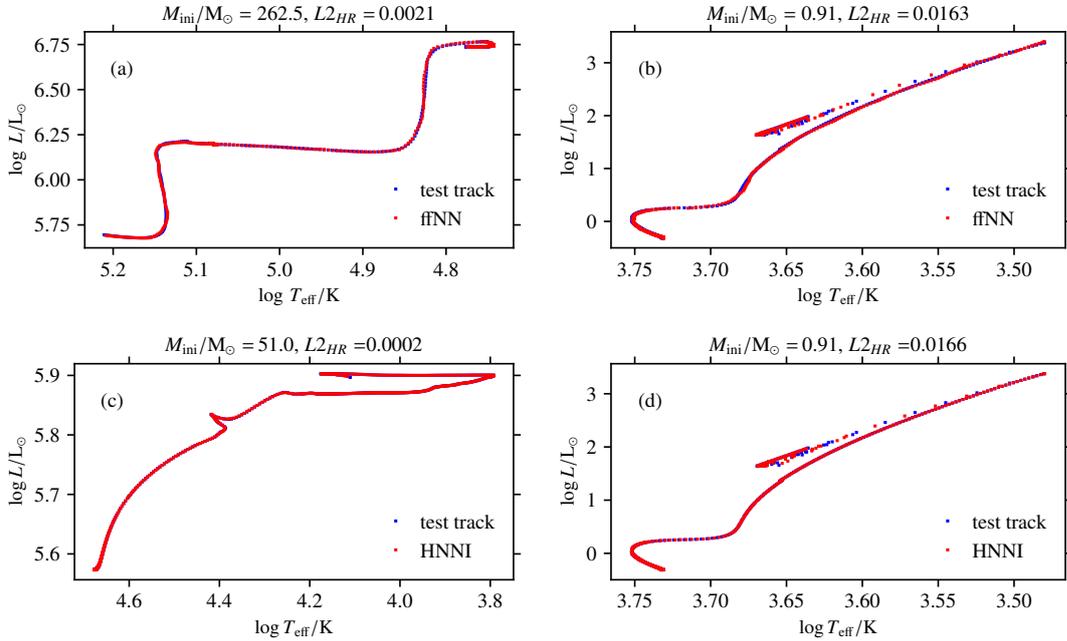


Figure 2.5: Test data results, showing the best (left) and the worst (right) predictions of stellar evolutionary tracks, as assessed by the $L2$ measure, in the HR diagram, for unseen test initial masses, by the trained fNN model (top) and by the HNNI algorithm (bottom). For comparison, the original held-back tracks are underlain.

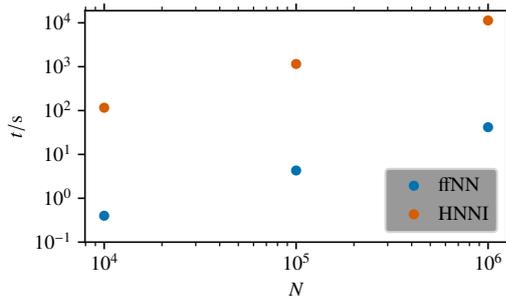


Figure 2.6: Cost-efficiency of forecast generation. Given our problem size and software implementation of HNNI (see Appendix A2. for an outline of the pseudo-code), the computing time scaling relation $t(N) \propto N$ with the number N of multiple output predictions is around 360 times larger for HNNI compared to that of the ffNN.

accuracy than in the deep learning case, with a deviation from the test track that is marginal throughout except during the helium flashes.

Furthermore, the HNNI scheme allows to predict any stellar evolution variable of interest we tested, by virtue of the same algorithmic prescription for interpolation (see Fig. 2.13 in the appendix for prediction of stellar-core related variables for unseen test data initial masses). In contrast, by the current setting, the ffNN predicts only those three surface variables which it has been trained upon, as set by the regression problem defined in Sect. 2.2.1.. In principle, a predictive framework with a large number of time-evolved variables could also be achieved with a ffNN emulator in two different ways. By the first way, the dimension of the output would need to be expanded to match the total number of stellar evolution variables of interest. For example, the values of 6 stellar variables would be produced as output of the 6 neurons in the outermost layer of the ffNN. However, optimizing such a model by a single globally defined loss score is cumbersome (for a discussion, see Appendix A5.). By the second way, a separate ffNN model with univariate output would need to be trained to predict each additional stellar variable of interest. This is the more promising approach out of the two, but requires construction of a separate hyperparameter-optimized model for each output variable.

Method comparison

The two methods for stellar evolutionary track forecasting (deep learning emulation vs. HNNI) that we develop lie at different trade-off points between cost-efficiency and accuracy of the forecasts. To summarize, the advantages of HNNI are the following:

1. The quality of predictions is reliable, with HNNI even outperforming our best-fit deep learning model.
2. All evolved stellar variables (i.e., not only $\log L_i$, $\log T_{\text{eff},i}$, $\log g_i$, whose prediction has numerically been evaluated for comparison with output of the surrogate model) are covered by the same interpolation prescription.
3. HNNI works as a sustainable out-of-the-box solution method. In contrast to the supervised machine learning approach, there is no need to re-iterate training

and optimization of a predictive model each time another stellar tracks data base is used as the catalog being accessed by the algorithm.

The disadvantages of HNNI are:

1. Continued access to the catalog data base is required which—depending on size of parameter space, sampling density and dimension of the problem—typically is of \sim GB size,
2. the computing time to generate predictions is significantly slower compared to the speed of the surrogate model. For the comparison, we have computed scaling relations on a 4-core CPU (see Fig. 2.6): on such a machine, it takes around 40 seconds to generate one million point predictions of all the three surface variables, spread randomly across the evolutionary phases and the initial mass range, with ffNN, while making the same number of predictions takes around 3 hours 13 minutes for HNNI. The computing time scales down linearly with the number of cores that are used to generate the predictions.
3. Extension to higher dimensional parameter space is not straightforward. Depending on the set of stellar parameters, a hierarchical relation may not always be identifiable. Moreover, in a high-dimensional parameter space, the required number of subsequent 1D interpolations becomes large (see the discussion in Appendix A2.). Thereby, prediction-generation is slowed down further.

In contrast, advantages of the supervised machine learning method are:

1. Prediction-making is fast, two orders of magnitude faster (in seconds) to generate than with HNNI.
2. Trained surrogate models are handy: a predictive ffNN model is of file size \sim 3 MB.
Third, the supervised machine learning approach is very general: the extension to higher dimensions does—in contrast to HNNI—neither require hierarchical ordering of regressor variables, nor faces significant increase in computing time with increasing dimension.

The disadvantages of the method are as follows:

1. Optimization of deep learning models is a more entailed task than a hard-coding adjustment of HNNI.
2. Minimizing a single global loss score during model training does not guarantee locally accurate fit results consistently over the entire parameter space (see Appendix A5. for a discussion thereof and proposed solutions).
3. The scaling of ffNN output with the number of target variables either comes under sacrifice of predictive accuracy (in the multiple output case) or implies considerably more development effort (in the single output case).

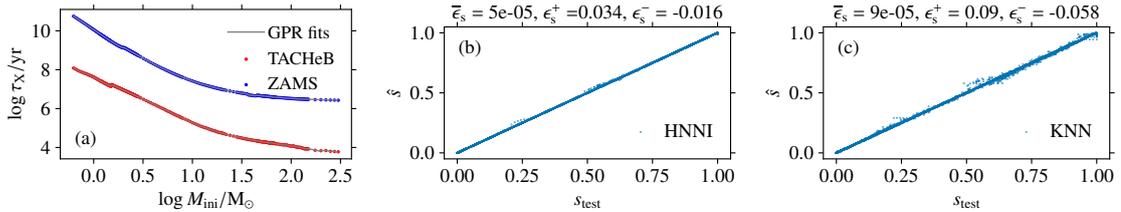


Figure 2.7: GPR fits of the $\log \tau_{\text{ZAMS}}(\log M_{\text{ini}}/M_{\odot})$ and $\log \tau_{\text{TACHeB}}(\log M_{\text{ini}}/M_{\odot})$ relations (a), and scatter plots of the age proxy predictions $\hat{s}(\log \tau, \log M_{\text{ini}})$ against the validation data s_{test} with HNNI (b) and KNN (c) methods, for performance comparison.

2.4.2. Prediction of the evolutionary coordinate

The series of age proxy values from $s = 0$ (ZAMS) to $s = 1$ (TACHeB) are not known for initial masses over which no stellar evolution tracks have been pre-computed, since s is calculated from the $\log L$, $\log T_{\text{eff}}$ and $\log \rho_c$ time series which are then not available at those initial mass grid points. Both our methods for predicting stellar evolution tracks rely on the timescale-adapted evolutionary coordinate s , which we use to re-parametrize the evolution of stars. Many astrophysical applications, however, require indication of stellar ages—for instance, drawing model isochrones into observed color-magnitude diagrams. We therefore construct another duet of interpolation methods (with HNNI and with supervised machine learning) that map the age τ onto the value of a star’s timescale-adapted evolutionary coordinate $s(\log \tau, \log M_{\text{ini}}) \in (0, 1)$ over a continuous initial mass range, in order to accomplish the two-step interpolation scheme as defined in Sect. 2.2.1..

Time counting in the MIST data set starts with the pre-MS phase. The values of ages at ZAMS, $\tau_{\text{ZAMS}}(M_{\text{ini}})$, therefore quantify its duration. Instead of τ , we use a scaled age variable $\tilde{\tau} \in (0, 1)$ for the age proxy fit with both the HNNI and the supervised ML methods:

$$\tilde{\tau}_i(M_{\text{ini}}) = \frac{\log \tau_i(M_{\text{ini}}) - \log \tau_{\text{ZAMS}}(M_{\text{ini}})}{\log \tau_{\text{TACHeB}}(M_{\text{ini}}) - \log \tau_{\text{ZAMS}}(M_{\text{ini}})}. \quad (2.8)$$

To obtain back the actual non-normalized age values (in units of years), the supply of the ZAMS $\log \tau_{\text{ZAMS}}(M_{\text{ini}})$ and the TACHeB $\log \tau_{\text{TACHeB}}(M_{\text{ini}})$ functions is needed. The $\tau_{\text{ZAMS}}(M_{\text{ini}})$ values are available from the MIST data set at the discretely sampled initial mass grid points. In order to be able to predict ZAMS and TACHeB ages of stars over a continuous M_{ini} range, we fit a Gaussian Process model to the discretely sampled catalog ZAMS and TACHeB grid points (see Fig. 2.7 a), respectively.

HNNI

The HNNI routine for the age proxy fit operates in the same way as outlined in Sect. 2.3.4., with the sole difference that the primary regressor variable now is $\tilde{\tau}$ (instead of the age proxy used in step 2), while s is itself the target variable of the fit.

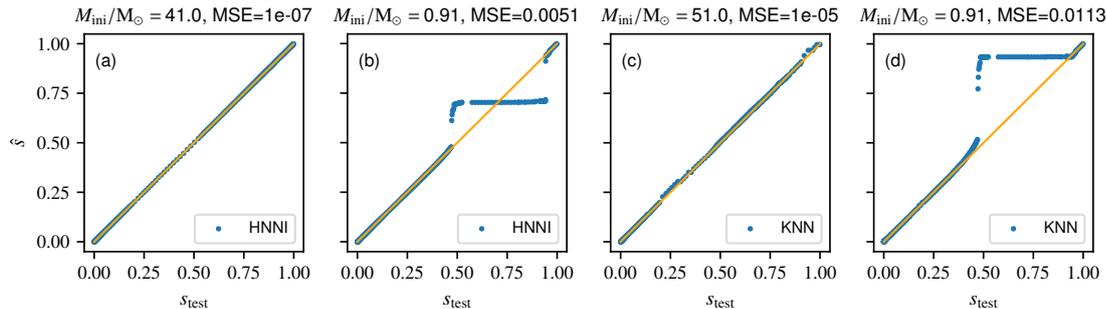


Figure 2.8: Best [(a) and (c)] and worst [(b) and (d)] fits of age proxy tracks for unseen test initial masses, with the HNNI and KNN methods, respectively.

As Fig. 2.7 b) shows, HNNI predicts the values of the age proxy reliably throughout evolution of stars from $s = 0$ up to $s = 1$ over the validation data set. The mean residual error $\bar{\epsilon}_{\text{res}}$ is of order 10^{-5} . The only clustered scatter regions off the diagonal are around $s \simeq 0.25$ and $s \simeq 0.6$, but the scatter offsets are low in amplitude. The most extremal over- and under-prediction outliers are of order 10^{-2} in absolute value. The performance evaluation based on the test data assesses the predictive accuracy of mapping the stellar age onto the timescale-adapted evolutionary coordinate over the course of the entire evolution from ZAMS up to TACHeB for unseen initial masses. HNNI predicts the value of the age proxy reliably throughout the evolution from ZAMS to TACHeB except at fast-timescale transitions, which make up little stellar lifetime but manifest themselves as sharp increases in age proxy values as function of age. Fig. 2.8 a) and b) show the best and the worst fit, respectively, assessed by the MSE metric, among all the test initial masses. The scatter plot of the best fit (for $M_{\text{ini}}^{\text{test}} = 41 M_{\odot}$) has no considerable spread, since \hat{s} aligns with s_{test} over course of the full evolution. In the scatter plot of the worst fit (for $M_{\text{ini}}^{\text{test}} = 0.91 M_{\odot}$), the local deviations of \hat{s} from s_{test} are apparent: the age proxy is first reproduced accurately up to $s \simeq 0.47$. Then a gap in the output range forms, such that the next predicted age proxy value is at $s \simeq 0.6$ and continues to be over-predicted up to $s \sim 0.7$ from where onward it transitions to an under-prediction phase. A second domain gap forms, and the predicted age proxy aligns back with its actual test data value at $s \sim 0.9$ up to end. The physical implications on the prediction of HR and Kiel diagrams in effect of the two-step interpolation scheme will be discussed in Sect. 2.4.3.. Age proxy prediction errors imply that the evolutionary state of the star is either under- or over-predicted, since a wrong evolutionary coordinate value has been assigned to a given stellar age of reference. However, as has been shown in Sect. 2.4.1., sampling target variables at homogeneously distributed δs increments (e.g. an equidistant spacing) in the step 2 scheme ensures that no significant changes in target variability will have been jumped over. In other words, artifact gap formation along curves in HR and Kiel diagrams is avoided by the fitted age proxy parametrization of stellar evolutionary tracks (step 2 fit), independent of the age proxy forecasts (step 1 fit).

KNN

Analogous to the procedure for the observables fit, we construct another solution to the age proxy fit with supervised machine learning. The reason is that we would like to obtain a more cost-efficient interpolation model than HNNI, which nevertheless is sufficiently accurate for astrophysical application purposes. The age proxy fit is a univariate regression problem distinct from the observables fit, and for which the procedure of surrogate model class selection and hyperparameter optimization needs be re-iterated. For comparing and selecting ML surrogate models, we use performance scores that are defined analogously to the performance scores for the step 2 fit (Sect. 2.2.4.), but applied to the univariate output of age proxy prediction. After a series of tests of a number of model classes, including ffNN, we obtain the best performance with the KNN algorithm. After a preliminary grid search for hyperparameter optimization of KNN, we manually fine-tune hyperparameters for best-fit results. We obtain these with two neighbors to query, a Minkowski metric, a $p = 2$ power parameter for distance calculation, the *BallTree* algorithm, distance-based weighting, and a leaf size of 300. The predictive quality is lower, but error scores are of the same order of magnitude compared to the HNNI case (see Fig. 2.7c for the summary statistics of the age proxy prediction errors over the validation data). Therefore, we evaluate the solution with the KNN algorithm as sufficiently accurate.

For performance assessment over the test data set, Fig. 2.8 c) and d) show the best and worst fits of the age proxy, respectively. HNNI and KNN agree on the worst fit to be at the low mass end, for $M_{\text{ini}}^{\text{test}} = 0.91 M_{\odot}$. Here, the KNN worst-fit has a characteristic similar to the HNNI case: the age proxy is first reproduced accurately up to $s \simeq 0.52$. Then a gap in the output range forms, such that the subsequent predicted age proxy value is at $s \sim 0.76$. The gap is larger than in the HNNI case. Hereafter, the age proxy is over-predicted, and aligns with the s_{test} values from $s \sim 0.93$ onward up to end.

2.4.3. Prediction of evolutionary tracks and isochrones

Putting consecutively together two predictive models for the age proxy and for the stellar observables, respectively, allows for the prediction of stellar evolution tracks in HR and Kiel diagrams as function of stellar age, and of isochrones showing stars of same age in the HR or Kiel diagram. In this section, we use the integral two-step interpolation scheme to predict complete stellar evolution tracks in HR and Kiel diagrams over the set of test initial masses (see Sect. 2.4.3.), and to predict stellar observables at fixed values of stellar age over a densely sampled initial mass range (see Sect. 2.4.3.).

Evolutionary tracks

For the input of age ($\log \tau$) and initial mass ($\log M_{\text{ini}}$) of the star, the value of the age proxy (\hat{s}) is predicted first by the step 1 method. \hat{s} is then used as input variable for the step 2 method, together with again the initial mass ($\log M_{\text{ini}}$). Here, we present the two-step pipeline interpolation results that are obtained with the supervised machine learning models (KNN and ffNN), which is the less accurate method than

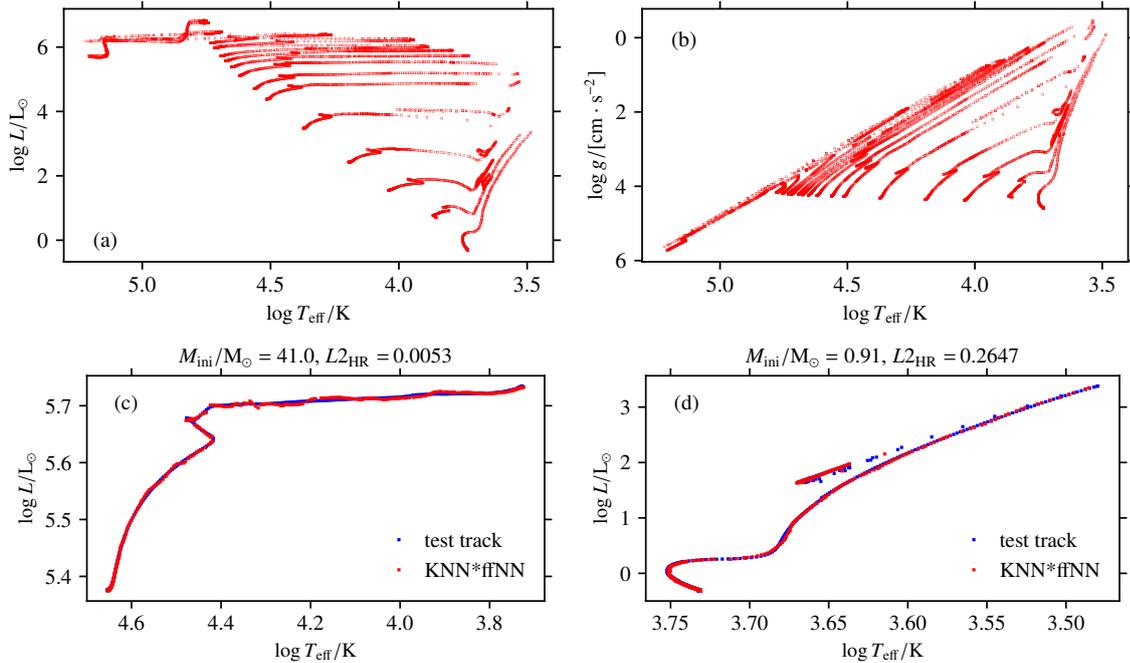


Figure 2.9: Outcome of the two-step interpolation scheme with supervised ML. Stellar tracks in the HR (top left) and Kiel diagram (top right) for unseen test initial masses are predicted as function of age τ . For comparison with the true test HR and Kiel tracks, see Fig. 2.4. For better visibility, the best- (bottom left) and worst-fit (bottom right) of tracks in the HR diagram—as assessed by the $L2$ measure—are displayed separately.

HNNI in both fitting tasks, and find that predictive quality of stellar surface variables reaches the desired accuracy level (see prediction of HR and Kiel diagrams for unseen test initial masses in Fig. 2.9). The net effect the step 1 fit errors have on the shape of predicted tracks in HR and Kiel diagrams in the two-step interpolation scheme is that step 2 based point forecasts of surface variables are incorrectly shifted along the track. If the step 1 fit by KNN accomplished a perfect $\log \tau$ -to- s mapping, then the tracks predicted by ffNN (step 2 fit) would retain the same shape as shown in Fig. 2.4 b) and d). However, as the step 1 fit introduces over- and under-prediction errors of the age proxy values, these lead to locally increased or decreased sampling densities of the age proxy axis as compared to the original unperturbed case. The under-densities along the age proxy axis result in domain gaps, over which no corresponding step 2 output (values of surface variables) is produced. These gaps form predominantly at fast-timescale transitions between evolutionary phases. As can be seen in Fig. 2.9, this applies in particular to the rear part of passage through the Hertzsprung gap and toward the late stages of CHeB (for high mass stars), to the nearing of the tip of the RGB and during the helium flashes (for low mass stars). Depending on accuracy or sampling needs of specific astrophysical applications, post-processing methods could be applied to fill the prediction gaps. The post-processing method would need to identify the domain gaps in the age proxy value range, sample the age proxy within the gap regions to obtain the prediction of observables (by the

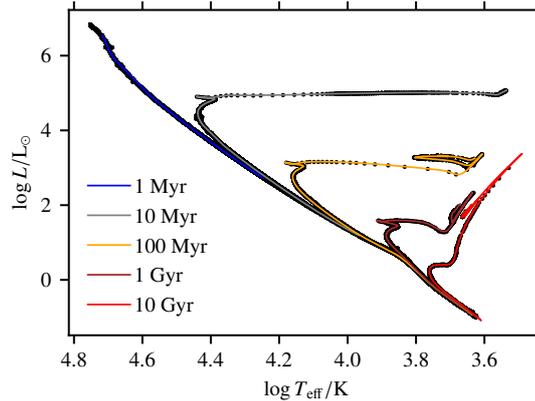


Figure 2.10: Comparison of MIST isochrones with emulator-based (GPR, KNN and fNN) predictions of stellar observables at fixed ages. The initial mass range $M_{\text{ini}} \in (0.65, 300) M_{\odot}$ is sampled over a log scale with a step size $\delta \log M_{\text{ini}}/M_{\odot} = 5 \cdot 10^{-5}$ to obtain the parameter space points at which discrete predictions are made. For the set of values of stellar isochrones, we choose a log sampling of the time axis to cover the full range of stellar ages τ . The theoretical MIST isochrones are color-marked, the emulator-based point predictions are scatter-plotted in black.

model which accomplishes the step 2 fit, which is fNN or HNNI), and then use local interpolation based methods to infer the stellar ages over the gap regions along the age proxy axis.

Isochrones

Finally, we further demonstrate consistency of our predictive models with the original MIST stellar evolution catalog by comparison of MIST isochrones with emulator-based predictions of observables at fixed stellar ages. MIST isochrones, interpolated over the parameter space using the EEP-based method discussed in Sect. 2.1., are imported from the Web interface⁹ provided by the makers of the catalog. For the emulator isochrones, we use the following multistep predictive pipeline to obtain stellar observables at fixed age:

First, for each M_{ini} value of interest, the two fitted GPR models are used to predict $\log \tau_{\text{ZAMS}}(M_{\text{ini}})$ and $\log \tau_{\text{TACHeB}}(M_{\text{ini}})$. Second, these are then used to calculate analytically the scaled age $\tilde{\tau}$ for each pair $\{M_{\text{ini}}, \tau\}$ of interest. Third, together with $\log M_{\text{ini}}$, $\tilde{\tau}$ serves as input of the trained KNN model, which predicts the corresponding value of the timescale-adapted evolutionary coordinate s . Fourth, s and $\log M_{\text{ini}}$ serve as input of the trained fNN model to predict the observables $\log L$, $\log T_{\text{eff}}$ and $\log g$, which are then plotted against one another.

Fig. 2.10 shows the outcome of the multistep predictive pipeline to obtain isochrones.¹⁰

⁹https://waps.cfa.harvard.edu/MIST/interp_isos.html

¹⁰ The value range of the imported MIST isochrones is adapted to match our problem setting: only the evolutionary sequences ZAMS–TACHeB are shown, the pre-MS and the post-CHeB evolution of stars are cut off. Note that in addition, there is an intrinsic cut-off in the MIST isochrones at the high mass end, which –in contrast to the MIST training data set we used– do not include the

For each value of stellar age, the distribution of point predictions in the HR diagrams mimics a simulated observation of stars under assumption of a log-uniform initial mass probability distribution. Therefore, most emulator-based point predictions of observables populate those regions of stellar evolution tracks over which stars evolve on the nuclear timescales. While there is some scatter of emulator-based predictions about the theoretical isochrones, most prominent at the blue end of the HR diagram and along the blue loop of the 100 Myr isochrone, we consider the overall statistical match as satisfactory.

Comparison with observations

As has been stressed in Sect. 2.1., in this work we use the MIST catalog for a proof-of-concept study to demonstrate our method of constructing accurate predictive models of stellar evolution over a width of parameter space necessary for a scalable general-purpose astrophysical applicability. We therefore proceed with the background assumption that the MIST data set is the ground truth of stellar evolution modeling. In Sect. 2.4.1. we show that the emulator-based prediction errors on all the three observables $\log L$, $\log T_{\text{eff}}$ and $\log g$ are significantly lower than typical observational uncertainties on the same variables. However, our predictive models can explain observations only as good as the original MIST models do. The important question how well MIST stellar models agree with observations, and which sources of systematic uncertainty have been identified, is addressed elsewhere, in the original paper on the catalog (Choi et al., 2016) and in follow-up studies. In this section, we provide a brief summary of their main conclusions concerning the ZAMS–TACHeB evolution at solar metallicity over the initial mass range $(0.65, 300) M_{\odot}$. It is aimed to inform interested readers about which scalable astrophysical applications of our predictive models are reasonable, and which are not, as consequence of systematic prediction errors that result from the adopted MIST input physics. Our trained machine learning models can be used for astrophysical applications in future work, and are available to corresponding authors upon reasonable request.

MS evolution, MS turn-off morphologies and red clump luminosities of low mass stars are in good agreement with MIST predictions except for those in the mass range $M_{\text{ini}} < 0.7 M_{\odot}$. MS evolution of high mass stars is reproduced well by MIST models close to ZAMS, but not the MS width at the highest masses within the tested range $M_{\text{ini}}/M_{\odot} \in (10, 80)$. The slope of model red supergiants is too shallow compared to observations, however no observed red supergiants lie in the forbidden zone cooler than the limit at the Hayashi line. Comparing the observed to predicted ratio of WC- to WN-type¹¹ stars, of WR to O-type, and of blue- to red supergiant stars allows to test mass loss, semiconvection and convective overshoot models. At Z_{\odot} , model ratios and observed ratios broadly agree on the order of magnitude, but the deviation is substantial in particular for the ratio of WC- to WN-type stars. For a more detailed analysis, the reader is referred to the original paper (Choi et al., 2016)

WR stars. Therefore, the emulator-based 1 Myr isochrone extends further to the blue part of the HR diagram than the MIST isochrones.

¹¹ WC stars are WR subtype stars that reveal helium-burning products in the outer layers, while WR stars of subtype WN reveal hydrogen-burning products.

and references therein.

2.5. Conclusions and outlook

We develop two method classes for interpolation of stellar evolution tracks over an initial mass range from red dwarfs to very massive stars, evolved from the zero age main sequence up to end of core helium burning. The two interpolation methods are, first, a surrogate model of stellar evolution constructed with supervised machine learning, and second, a catalog-based hard-coded hierarchical nearest neighbors interpolation algorithm. Both of these invoke a two-step interpolation procedure that makes use of a timescale-adapted evolutionary coordinate s (age proxy) that we introduce to re-parametrize the evolution of stars. This re-parametrization reduces the timescale variability of evolutionary variables, and thereby allows for more accurate predictions across timescale-separated evolutionary phases.

For the predictive two-step pipeline constructed with supervised machine learning, we optimize a K-Nearest Neighbors model to predict the age proxy for the input of (scaled) stellar age $\tilde{\tau}$ and initial mass M_{ini} . The predicted age proxy value, together with initial mass, is then used as input by a hyperparameter-tuned Feedforward Neural Network model to produce the multiple output prediction of the log-scaled surface variables luminosity $\log L$, effective temperature $\log T_{\text{eff}}$, and surface gravity $\log g$. These predictions allow tracing the evolution of stars in the HR and Kiel diagrams over the dominant duration of their lifetimes.

For the predictive two-step pipeline constructed with HNNI, we use the same syntax in the algorithmic prescription for both the age proxy prediction and for the prediction of observables. It operates by selecting the two nearest neighbors, from above and from below, in each parameter space direction, and by then performing a sequence of linear interpolations according to a hierarchical ordering of parameters. Depending on the astrophysical application, one method is preferable over the other. The supervised machine learning approach is more cost-efficient (by two orders of magnitude in seconds) but more difficult to develop. The hard-coded HNNI is more accurate by one order of magnitude on the MAE and by two orders of magnitude on $\bar{\epsilon}_T$, while all other error scores are on the same scale), but less handy, since continued access to the stellar evolution catalog is required.

With a wide initial mass range and with a sequence of evolutionary phases from zero age main sequence up to the end of core helium burning, astrophysical application of our models is of interest, first, in context of rapid single star population synthesis. The second promising application prospect is the incorporation of stellar evolution emulators as stellar microphysics sub-grid models in large-scale stellar N -body dynamics or galactic evolution simulations. The third application prospect is the usage of our interpolation methods to infer fundamental stellar parameters (given multiple observables of a single star) or the initial mass function (given the observation of a stellar population). The latter astrophysical application prospect could follow the DALEK (Kerzendorf et al., 2021) working example in context of spectral modeling of Type Ia supernovae. DALEK is a deep learning based emulator of the output of the TARDIS (Kerzendorf and Sim, 2014) radiative transfer code. A variant thereof has been used in a Bayesian framework, where it represents the output of TARDIS, to

infer supernova progenitor parameters from the observation of its spectrum (O’Brien et al., 2021). The variant has been trained from scratch on a training data set which has been generated over a reasonably constrained parameter range given likely properties of the progenitor system. Reliable inference on the parameters of the progenitor system without the emulator, with the traditional grid-based methods instead, is impossible—it would take thousands of years of clock time to evaluate the high-dimensional parameter space by the classical Bayesian inference approach of running TARDIS models at all those parameter space grid points as selected by iterative optimization that typically requires millions of evaluations.

Sampling a stellar evolution track as function of the age proxy instead of the stellar age, for instance at equidistant δs increments, facilitates the adequate resolution of all significant changes in the stellar output variables. This applies not only to the ZAMS–TACHeB sequence, but also to the pre-MS and post-CHeB evolution (up to white dwarf cooling for low mass stars).

For generalization of our methods to a higher-dimensional space of fundamental parameters, additional considerations need to be taken into account. Sampling of a high-dimensional parameter space to generate the grid data needs to be efficient: sparse enough to keep the computational expenses low but dense enough to maintain the predictive accuracy satisfactory locally across all directions in parameter space. The MIST single star grid space sampling density distribution, which we used to construct our models, has been decided upon by the makers of the catalog, based on physical insight from domain expertise. We have expanded the data set in parameter ranges of interest based on inspection of local fit results by the surrogate model, to locally improve predictions where needed, by supplying more training data in those regions. An alternative approach to determining the optimal parameter space sampling goes by using Active Learning (AL; Settles, 2009). By pre-defined heuristics, decision-making with AL is automated, and therefore better adapted to high dimensional parameter spaces for finding an optimized distribution of grid points. In the context of stellar astrophysics, Rocha et al. (2022) apply AL in a case study involving the mapping of initial binary star parameters to the final orbital period, and show that it can be used to reduce the training data grid size.

For stellar parameter spaces greater than those tested here, we recommend using HNNI as the predictive interpolation model as far as it is applicable given computational cost constraints. The HNNI method generalizes to higher dimensions: for clarity, we have provided the recipe for a 3D ($s, M_{\text{ini}}, Z_{\text{ini}}$) formulation of HNNI in order to show the systematic of its dimensional extension. In the case that either the HNNI method we developed will break down or be computationally too inefficient in the high dimensional parameter space (given the impractically large cumulative number of 1D interpolations to make), we recommend using supervised machine learning, in particular deep learning, to train univariate surrogate models of stellar evolution on segments of the initial mass parameter space. For training deep learning models, we have provided basic guidance on selection of feedforward neural network architecture and learning hyperparameters and on the choice of the loss function. Finally, we have found a successful training strategy that, in its basic design, could—since it has been adjusted to data base specifics of a stellar evolution catalog—continue to produce satisfactory fit results when trained on data in a higher-dimensional parameter

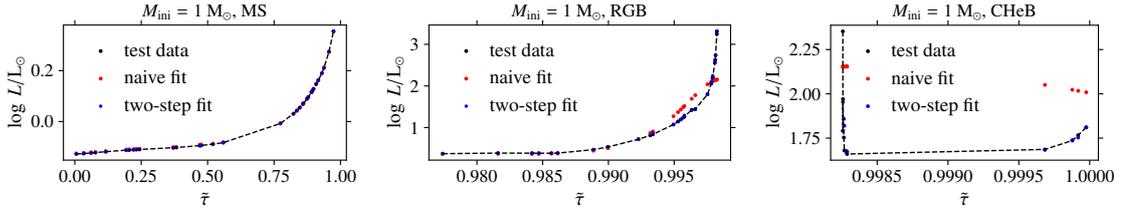


Figure 2.11: Comparison of naive vs. two-step fit of stellar evolution time series, upon the 1D test case of predicting log-scaled luminosity of a Sun-like star over the MS, RGB and CHeB phases.

space.

A Appendix

A1. Stellar evolution re-parametrization and HNNI

Figures 2.11–2.13 provide additional materials that support the use of a timescale-adapted evolutionary coordinate and of the HNNI method for interpolation of stellar evolution tracks.

Fig. 2.11 demonstrates the general suitability of our age proxy prescription for resolution of variability in stellar tracks over a wide span of sequential evolutionary phases and across the initial mass range.

Fig. 2.12 compares the predictive quality of the two-step fitting approach with the naive direct fit, when applied to the test case of modeling the log-scaled luminosity time series of a Sun-like star from ZAMS up to TACHeB. We make an 85:15% train-test split of the data and use the scaled age variable $\tilde{\tau}$ defined in Sect. 2.4.2. as regressor variable. For the naive fit, we train a GPR model on the log-scaled luminosity training data, and use it to predict the test data. For the two-step approach, we first train a KNN model to predict the normalized age proxy s based on the training data. Second, we use the age proxy prediction as regressor variable when training another GPR model to predict the log-scaled luminosity training data. To compare the outcomes, we plot the prediction of the naive fit and the one resulting from the two-step pipeline separately for each evolutionary phase MS, RGB and CHeB. For better discrimination of the test data stellar track, neighboring test data points are connected by piecewise linear dashed lines over each phase. The MS evolution data is accurately predicted by both methods, and so is the sub-giant and early red giant phase. The naive fit loses out to the two-step fit during the later stages: the ascension of the RGB and throughout the CHeB phase.

Fig. 2.13 shows that HNNI, by virtue of the same algorithmic prescription, yields accurate forecasts of not only the surface variables $\log L$, $\log T_{\text{eff}}$ and $\log g$ —which have been evaluated in the main part of the paper—but also of all other stellar variables we tested. The shape of the tracks in the $(\log \rho_c, \log T_c)$ diagram is represented well by HNNI, except at fast-timescale transitions during the helium flashes of low mass stars.

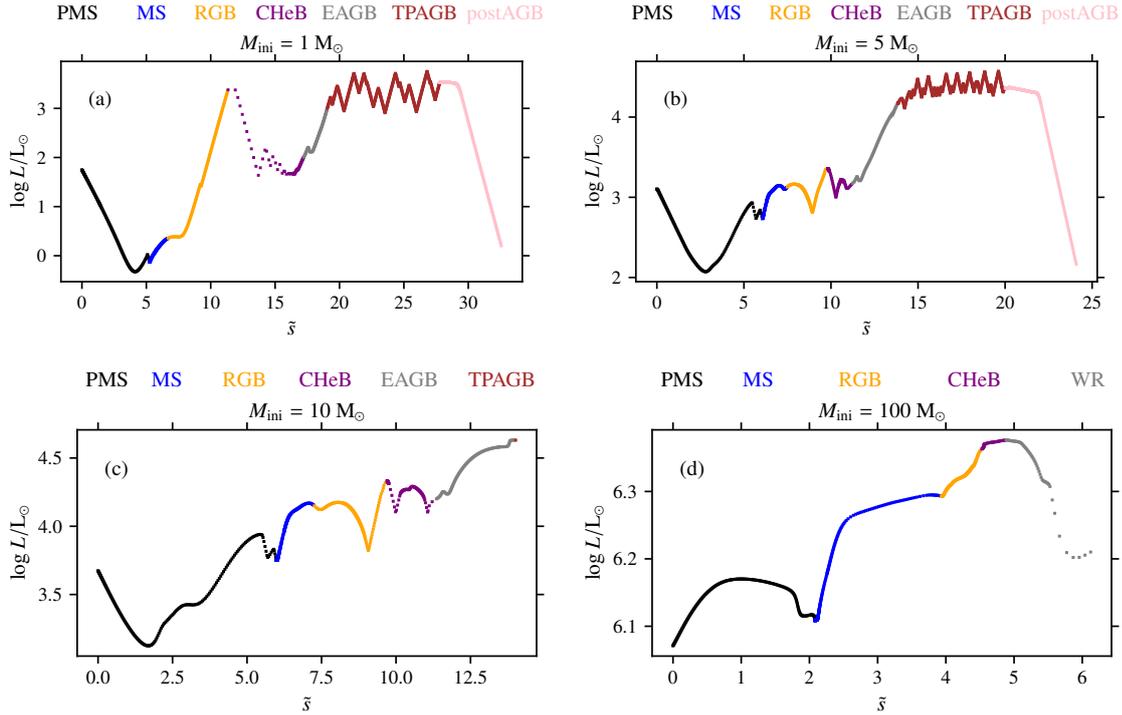


Figure 2.12: Parametrization of stellar evolution as function of the timescale-adapted evolutionary coordinate over phases beyond the ZAMS–TACHeB sequence. Stars across the initial mass range are evolved from the pre-MS up to post-CHeB evolution (post-asymptotic giant branch evolution for low mass stars, and onset of core carbon burning for massive stars) as function of (unnormalized) timescale-adapted evolutionary coordinate \tilde{s} . The color marking denotes the evolutionary phases pre-main sequence (PMS), main sequence (MS), red giant branch (RGB), core helium burning (CHeB), early asymptotic giant branch (EAGB), thermally pulsating asymptotic giant branch (TPAGB), post asymptotic giant branch (postAGB), and Wolf-Rayet (WR) phase.

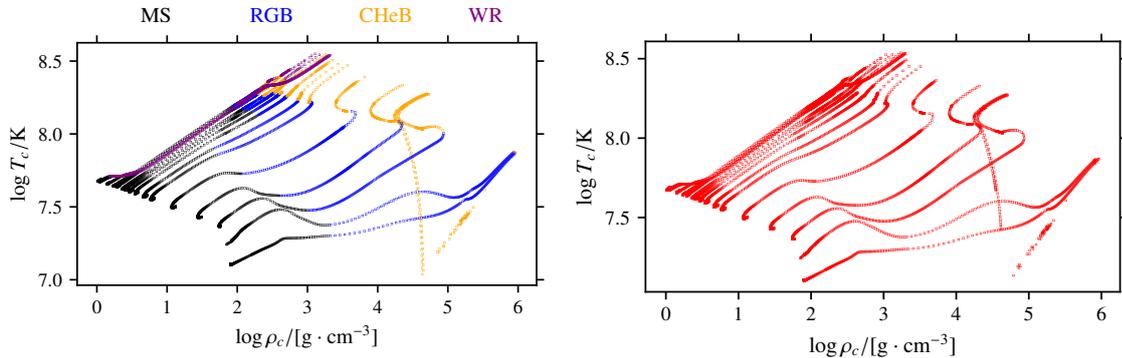


Figure 2.13: Prediction of stellar evolutionary tracks in the core density-temperature diagram over the test data, with the HNNI method (right), and the true held-back tracks (left) for comparison.

A2. HNNI in 3 dimensions or higher

Here, we assume that the stellar parameter space is spanned by a third dimension—the initial metallicity Z_{ini} . The resulting set of parameters $(s, M_{\text{ini}}, Z_{\text{ini}})$ allows for a hierarchical ordering: Always first is the age proxy axis s . Always second is the initial mass axis M_{ini} . Third and least significant out of the three, is the initial metallicity axis Z_{ini} , whose effect on the shape of stellar evolutionary tracks—at fixed initial mass—results in minor corrections.

The pseudo-code below provides the recipe for a numerical response to the following inquiry: What is the value of target variable $Y_j = Y(s_j)$ at test location $(Z_{\text{ini}}^{\text{test}}, M_{\text{ini}}^{\text{test}}, s_j)$ in parameter space?

Y is any evolved stellar variable: for example, $Y = \log L/L_{\odot}$. For full generality, we assume that neither $M_{\text{ini}}^{\text{test}}$ nor $Z_{\text{ini}}^{\text{test}}$ is contained in the catalog grid database spanned by $\{M_{\text{ini}}, Z_{\text{ini}}\}_{\text{cat}}$.

Below, the linear interpolation equation $y(x) = y_1 + \frac{y_2 - y_1}{x_2 - x_1}(x - x_1)$ is referred to by its parameters: $y(x) \leftarrow y_2, y_1, x_2, x_1$. We assume that, in the catalog, a similar initial mass grid sampling density is available for each initial metallicity.

Pseudo-code:

1. Determine the nearest neighbors $Z_{\text{ini}}^+, Z_{\text{ini}}^- \in \{Z_{\text{ini}}\}_{\text{cat}}$ to the test initial metallicity value $Z_{\text{ini}}^{\text{test}}$ from above and from below, respectively.
2. From the set of available initial masses $\{M_{\text{ini}}\}(Z_{\text{ini}}^+)$ and $\{M_{\text{ini}}\}(Z_{\text{ini}}^-)$ contained in the catalog at these two metallicities, determine the nearest neighbors to $M_{\text{ini}}^{\text{test}}$ from above and from below, respectively:
$$M_+^+, M_+^- \in \{M_{\text{ini}}\}(Z_{\text{ini}}^+),$$

$$M_-^+, M_-^- \in \{M_{\text{ini}}\}(Z_{\text{ini}}^-).$$
3. From the age proxy series available in the catalog at each of these four initial mass grid points, find the nearest neighbors to s_j along the age proxy axis that satisfy:
$$s_{+, \text{max}}^+, s_{+, \text{min}}^+ \in \{s_i\}(M_+^+) \text{ with } s_{+, \text{min}}^+ < s_j < s_{+, \text{max}}^+,$$

$$s_{+, \text{max}}^-, s_{+, \text{min}}^- \in \{s_i\}(M_+^-) \text{ with } s_{+, \text{min}}^- < s_j < s_{+, \text{max}}^-,$$

$$s_{-, \text{max}}^+, s_{-, \text{min}}^+ \in \{s_i\}(M_-^+) \text{ with } s_{-, \text{min}}^+ < s_j < s_{-, \text{max}}^+,$$

$$s_{-, \text{max}}^-, s_{-, \text{min}}^- \in \{s_i\}(M_-^-) \text{ with } s_{-, \text{min}}^- < s_j < s_{-, \text{max}}^-.$$
4. Interpolate along the age proxy axis to find
$$Y_+^+(s_j) \leftarrow Y_+^+(s_{+, \text{max}}^+), Y_+^+(s_{+, \text{min}}^+), s_{+, \text{max}}^+, s_{+, \text{min}}^+,$$

$$Y_+^-(s_j) \leftarrow Y_+^-(s_{+, \text{max}}^-), Y_+^-(s_{+, \text{min}}^-), s_{+, \text{max}}^-, s_{+, \text{min}}^-,$$

$$Y_-^+(s_j) \leftarrow Y_-^+(s_{-, \text{max}}^+), Y_-^+(s_{-, \text{min}}^+), s_{-, \text{max}}^+, s_{-, \text{min}}^+,$$

$$Y_-^-(s_j) \leftarrow Y_-^-(s_{-, \text{max}}^-), Y_-^-(s_{-, \text{min}}^-), s_{-, \text{max}}^-, s_{-, \text{min}}^-.$$
5. Interpolate along the log-scaled initial mass axis to find
$$Y_+(s_j) \leftarrow Y_+^+(s_j), Y_+^-(s_j), \log(M_+^+/M_{\odot}), \log(M_+^-/M_{\odot}),$$

$$Y_-(s_j) \leftarrow Y_-^+(s_j), Y_-^-(s_j), \log(M_-^+/M_{\odot}), \log(M_-^-/M_{\odot}).$$
6. Interpolate along the log-scaled initial metallicity axis to find
$$Y(s_j) \leftarrow Y_+(s_j), Y_-(s_j), \log(Z_{\text{ini}}^+/Z_{\odot}), \log(Z_{\text{ini}}^-/Z_{\odot}).$$

Generalization: The HNNI method is extended analogously to higher-dimensional parameter spaces. The number n of hierarchical 1D interpolations to perform and number k of neighboring grid points to query scale as follows with dimensionality of the parameter space:

- 1D (s): $n = 1, k = 2$
- 2D (s, M_{ini}): $n = (1 + 1) + 1 = 3, k = (2 + 2) + 2 = 6$
- 3D ($s, M_{\text{ini}}, Z_{\text{ini}}$): $n = (3 + 3) + 1 = 7, k = (6 + 6) + 2 = 14$
- 4D ($s, M_{\text{ini}}, Z_{\text{ini}}, v_{\text{ini}}$): $n = (7 + 7) + 1 = 15, k = (14 + 14) + 2 = 30$.

We believe that our HNNI method is generalizable to even higher parameter space dimensions but have not verified this hypothesis. For a binary system composed of two non-rotating stars of the same initial metallicity, we expect the following hierarchical ordering of variables to yield accurate interpolation results:

$$5\text{D}: (s_1, M_{\text{ini},1}, M_{\text{ini},2}, P_{\text{ini}}, \epsilon)$$

with $n = (15 + 15) + 1 = 31$, and $k = (30 + 30) + 2 = 62$.

A3. GPR

Ever since pioneering work by [Sacks et al. \(1989\)](#), GPR has been considered standard method choice for emulation tasks because of flexibility of the fitting model and regulatory effect of the Gaussian assumption (for a detailed discussion of application of GPR to computer model calibration, see [Kennedy and O’Hagan, 2001](#)). In general, GPR becomes increasingly time-prohibitive and computationally expensive as the size of training data grows. Particularly, GPR involves the Cholesky factorization and inversion of the covariance matrix, which are computationally costly operations for a large data set.

The literature on GPR includes multiple approaches to improve scalability ([Liu et al., 2020](#)), which can be broadly classified into global approximation and local approximation of the GPR. While approaches to global approximation tend to focus on methods based on sparse kernels ([Kaufman et al., 2008](#)) and approximate likelihoods ([Varin et al., 2011](#)), approaches to local approximation center around inference and prediction on local subsets of data, such as moving-window GPR ([Van Tooren and Haas, 1993](#); [Ni et al., 2012](#)).

Following the local approximation approach, [Li et al. \(2022\)](#) solved the forward problem of stellar evolutionary track forecasts for given fundamental input parameters with separate GPR models that each cover a subspace of the narrow but five dimensional stellar parameter space. However, using a separate GPR on subspaces is likely to ignore the potential dependence across them, which in turn can lead to suboptimal predictions. We expect this to become problematic upon extension of the input space, when exploring parameter spaces of binary star systems. For these reasons, we investigate machine learning models that can be trained on the full data set more closely.

A4. fNN hyperparameter tuning

Architecture design: There are a number of relevant theoretical considerations that guided our approach to fNN architecture design. The main role of the activation function is to introduce non-linearities into the information processing pipeline of the neural network. We adopt the standard recommendation of choosing the ReLU activation function, and instead focus on tuning the model capacity.¹² There are approaches to tuning the model capacity based on complexity of the learning task (Achille et al., 2021), which can be estimated using the Kolmogorov complexity measure (Kolmogorov, 1963). But its estimation is more a theoretical, less a practical enterprise, due to its non-trivial computation. Instead, there is a body of theoretical hint suggesting that over-parametrization of the deep learning model is required in order to overcome an inherent bias of learning simple (rather than complex) input-output mapping rules (Dingle et al., 2018; Nichani et al., 2020). Model capacity ought to be chosen large enough to prevent underfitting, though not overly large to prevent overfitting. The model capacity, once fixed, can be built up in two contrasting ways: by few hidden layers and many neurons per hidden layer, or by many hidden layers and few neurons per hidden layer. An incentive towards the first approach is the success of GPR as emulation method: A first-order Taylor approximation to the output of a wide network, initialized with independent and identically distributed weights and trained for a large number of epochs, approximates the predictions of a GPR model, and the selection of activation functions corresponds to a particular kernel (the neural tangent kernel) in GPR (Jacot et al., 2018; Lee et al., 2019). The other approach to building up a fixed model capacity, by choosing a higher number of layers, can—on the other hand—be more valuable than increasing the width. For example, Eldan and Shamir (2016) show that approximating certain functions requires an exponentially higher number of neurons in a wide network configuration to achieve the same accuracy as that of a deeper network, and the result holds irrespective of the choice of activation function. We tested both approaches on our problem and found the best result by building up model capacity through a many-layer architecture with a moderate number of neurons.

When training deep learning models, we tested dropout, batch normalization and layer normalization as regularization techniques, in order to push the validation loss further down past stagnation phases.

Selection criteria: We performed empirical tests of manually designed hyperparameter (HP) combinations, and then applied selection criteria to decide whether or not to train the configured model up to end. The main HP that we varied were the number of hidden layers, the number of neurons per layer (assuming a symmetric network architecture), learning rate schedule parameters and the batch size. For each HP combination, we evaluated the loss curve decline during run-time of learning and applied the following selection criteria at the $\{500, 10^3, 2 \cdot 10^3, 5 \cdot 10^3\}$ epoch stages: First, the speed of learning (judged upon by cross-comparison of validation

¹²The model capacity corresponds to the number of free trainable parameters. In a fully connected fNN without regularization layers, the model capacity is given by the total number of weighted connections between neural nodes plus the total number of biases in the network.

loss scores at the aforementioned epoch stages for different HP combinations, and estimation of the slopes). Second, the degree of overfit (judged upon by visual assessment of departure of the validation from the training loss curve with increasing epoch number). Subsequently, we manually adjusted the HP choice for the next series of empirical tests, informed by performance of HP combinations from the previous trials. Promising models (with fast validation loss curve decline and tolerable overfit over long training periods) were trained until the validation loss curve either flattened out or started to oscillate over epoch scales of order $5 - 10k$. Out of those promising models, the 'best' deep learning model was selected as the one that had the least error scores on the validation data. This procedure was iterated until we trained a surrogate model that attained a threshold value of the validation loss, with the lowest error scores among all the deep learning models we tested over a series of generations. Training our best-fit fNN model lasted around 8 hours on a Nvidia RTX 3060 GPU machine.

Randomness and reproducibility: A trained deep learning model is the outcome of a stochastic computer experiment. In order to obtain reproducible results, the random seed needs to be fixed twice: First, before the train-test split of the total data set $N_{\text{total}} = N_{\text{train}} + N_{\text{val}}$. Second, before initializing the fNN kernel at model compilation.

A5. Alignment problem

Globally defined loss functions: What remains an issue when building stellar evolution surrogate models with supervised machine learning models to approach the regression problem we formulated, is what we refer to as the Alignment Problem (AP): our expectation of the surrogate model's predictive capability (characterized by locally accurate performance over all three target variables, across all three evolutionary phases, and across the entire initial mass range) does not align with the formalized numerical condition (characterized by minimization of one single global error score) that is optimized during training of machine learning models. We find that none of the standard loss function choices optimally match our problem setting.¹³ The reason for the AP is that the evolution of stars, traced in the HR diagram, does neither happen over the same absolute nor relative numerical scale range for different initial masses. At the same time, our expectation of the surrogate model's

¹³ MSE is the average squared residual, where the squared penalization incentivizes to avoid large absolute residuals in model training. Clearly, this behavior is globally desirable for stellar evolutionary track fitting, but leads to too much leniency when a surface variable does not vary much over a star's lifetime. Then, residuals would be small compared to the global range, but large as perceived in HR or Kiel diagrams for a given initial mass. MAE is the average absolute residual, where penalization is linear, and the behavior is reversed in comparison to the MSE. Common scale-free measures, such as MSLE and MAPE, essentially penalize multiplicative errors. MSLE evaluates squared penalties on a log scale (that is, squared log ratios), and MAPE is the average ratio of the absolute residual over the actual value of predicted target variable. In a nutshell, both of these measures tolerate larger absolute residuals as the observed value increases, but we require the opposite for luminosity, which tends stay in a smaller range for tracks at overall high values of luminosity (see top-left panel of Fig. 2..9).

performance is that it yields locally accurate fit results for all three predicted surface variables. Instead, the surrogate model learns by minimizing a globally defined error score, which means by improving to reproduce the overall global shape of the three-dimensional hypersurface of the target variables over the two-dimensional input parameter space. While training deep learning models, we encountered cases when the statistical MSE scores on training and validation data decreased further (i.e., no overfit in the statistical learning sense of the term), but our physical performance scores, which are locally defined, worsened. In essence, this means that the surrogate model continued to learn, but not that what we appreciate. It may happen that the emulator will have improved predictive capability globally, as assessed by the global loss score, by substantial gains in predictive accuracy in those parameter space regions where the accuracy was already good enough according to our physical performance metrics, though at the sacrifice of losing predictive accuracy in other parameter space regions admitted by statistical fluctuations. That latter loss in local predictive accuracy, however, may manifest itself in a decrease of physical performance scores over the target variables, adverse to expectations. However, this performance loss is not considered problematic by the surrogate model based on the global error score that too little penalizes the prediction errors in relevant parameter regions of concern.

The AP is only partially addressed by choosing a ffNN model class (which minimizes the loss of—not the global data set in a single step, but of—a sequence of randomly selected data batches¹⁴), by choosing the Huber loss score (which seeks a trade-off between MSE and MAE minimization), and by locally increasing the initial mass parameter space sampling (which, statistically, increases the importance of specific parameter space regions by the increased amount of data for that region). If supervised ML is to be applied in high dimensional parameter space for stellar track fitting, this issue requires adequate resolution.

Solution approaches: For future extensions of our surrogate modeling method to wide high dimensional parameter spaces, we propose the following approaches:

- (i) To train a separate ffNN model for each target variable, instead of training a single ffNN model with multiple output.
- (ii) To segment the initial mass parameter space into parts, and train a different ffNN model on each segment of the initial mass range.
- (iii) To tailor a loss function (parameterized by input variables, in particular the initial mass) to account for differences in numerical scale range over which

¹⁴This point is best understood by comparison of ffNN optimization to that of another statistical learning algorithm. For instance, a RF model is optimized in a single step: a loss score (such as MSE) is minimized after the complete data set is fed into the RF by the bagging technique that distributes the input data onto the individual decision trees. A RF forecast is an ensemble forecast from an ensemble of decision trees, each of which receives a random split of data samples. For this subset of training data (which differ from one tree to another), the decision tree finds its own hierarchically conditioned numerical rules during the supervised machine learning. However, during training the minimization of the loss score happens globally, not locally for each subset of the total training data set. In contrast, ffNN minimizes the loss score for each batch (a small, randomly selected chunk of the total training data set) and repeatedly over a large number of iterations.

stellar variables change across the initial mass range, across evolutionary phases and—in the case of multiple output—across different target variables.

Approaches (i) and (ii) are the most common and straightforward. [Li et al. \(2022\)](#) employed both of them when modeling stars by a set of global GPR models. Here, approach (ii) facilitated a splitting of the total training data set of size $\sim 300k$ into subsets of size $\sim 20k$, which is their stated limit of computational feasibility for applying global GPR models on a training subset. With approaches (i) and (ii), the learning task is simplified and that can lead to more accurate individual interpolations.

However, all three approaches to solving the AP, which can be employed individually or in concert, do have their drawbacks: For approaches (i) and (ii), many separate models need to be trained, and the capability to capture dependence structures is impaired. Approach (iii) is the only solution that leads to truly multivariate predictions, but is the most difficult to realize. If the training target loss does not account for inter-variable dependence, then any loss-based training lacks the required guidance. The first step in accounting for dependence is accounting for variability and covariance. Already that initial step is a challenge since the range of a single stellar surface variable over a star’s lifetime can be drastically different depending on initial mass alone. Possibly, creating a suitable multivariate loss function, tailored to stellar evolution tracks, is similarly complex as the multivariate interpolation task itself.

For future research towards extending stellar evolution emulators to wide high dimensional parameter spaces, we therefore believe that a good starting approach is to build a separate deep learning model for each target variable, and to segment the initial mass range into parts, in order to train sets of univariate deep learning models on each initial mass segment, in an otherwise high dimensional parameter space.

3. Explodability criteria for the neutrino-driven supernova mechanism

*This chapter is based on a paper manuscript that is to-be-submitted as [Maltsev et al. \(2024a\)](#) for publication in the *A&A* journal. I conducted the scientific work and wrote all of the text myself, with the support of the co-authors.*

Abstract: Massive stars undergoing iron core collapse at the end of their evolution either explode and leave a neutron star behind, or explode but form a black hole by fallback, or do not explode and form a black hole directly. The outcomes of 3D core collapse supernova (CCSN) simulations depend on the structure of the stellar progenitor, on the nuclear equation of state and on the treatment of the complex physical ingredients such as neutrino physics and transport, turbulent convection and other instabilities. While CCSNe are inherently multi-dimensional phenomena, due to computational cost constraints, 3D simulations are prohibitive over the wide parameter space of stellar progenitors as necessary for population synthesis. In this work, we formulate criteria that allow us to predict the final fates of stars by evaluating multiple explodability proxies derived from the stellar structure at the onset of core collapse. These criteria are derived from the outcomes of a semi-analytic supernova model evaluated over a set of stellar models consisting of ~ 3900 heterogeneous stellar progenitors (single stars, binary-stripped stars and accretor stars), over which they achieve an accuracy of 99%. The criteria are validated over 27 state-of-art 3D core collapse simulation outcomes from two different groups, over which the accuracy is 89%. We further find that all explodability proxies relevant for our pre-SN based criterion have two distinct peaks with an intervening valley as a function of the carbon-oxygen (CO) core mass M_{CO} . The CO core masses of explodability peaks shift systematically with metallicity Z and with the timing of hydrogen-rich envelope removal of binary-stripped stars. We identify critical values in M_{CO} that define windows over which black holes form directly. The outcome is a CCSN recipe based on M_{CO} and Z , applicable for rapid binary population synthesis studies. We find that our explodability formalism is consistent with observations of Type IIP and Type IIb/Ib supernova progenitors and partially addresses the missing red supergiant problem by direct black hole formation. Moreover, it seems to be in agreement with observations of inferred Type IIn and Type Ic progenitors that other supernova recipes used in binary population synthesis codes cannot explain.

3.1. Introduction

Asymmetry-induced turbulence in both the progenitor structure at the onset of and during the collapse as well as other multi-dimensional effects can be decisive to shock

revival by the delayed neutrino-driven perturbation-aided SN mechanism (Müller, 2020). Therefore, the appropriate modeling of CCSNe¹ is addressed by 3D simulations. However, 3D CCSN simulations are notoriously computationally expensive due to the spatial and temporal scales that need to be resolved and the complexity of the physical processes involved (Janka et al., 2007; Burrows and Vartanyan, 2021). For population synthesis over hundreds and thousands of stars, only 1D codes of stellar evolution up to core collapse, and only 1D CCSN codes are computationally feasible. However, in most spherically symmetric CCSN simulations, successful explosions do not occur except for the lightest progenitors (Kitaura et al., 2006; Janka et al., 2008), and stars therefore need pushed to explode, e.g., by artificially enhancing neutrino heating (O’Connor and Ott, 2010) or the proto-neutron star contraction dynamics (Ugliano et al., 2012) using some appropriate calibration.

The degree of difficulty to explode a star by neutrino energy transport has been linked to the steepness of density decline outside the iron core (Burrows and Lattimer, 1987). Based on large samples of 1D CCSN simulations, different explodability proxies, derived from the pre-SN structure, have been introduced to predict whether iron core collapse will result in a successful neutrino-driven supernova explosion or not, which we summarize below. O’Connor and Ott (2011) introduced the compactness parameter

$$\xi_M = \frac{M/M_\odot}{R(M)/1000 \text{ km}}, \quad (3.1)$$

where R is the radius of the Lagrangian mass shell enclosing baryonic mass M in the pre-supernova star, as a first-order criterion for predicting final fates at the moment of core bounce. When choosing the mass coordinate $M = 2.5 M_\odot$, they found that setting the condition $\xi_{2.5} > 0.45$ for failed SNe agreed with CCSN outcomes predicted with the GR1D SN model (O’Connor and Ott, 2010) over a sample of over 700 single-star progenitors. It has further been shown that compactness at bounce and at the pre-SN stage is comparable, and that it distinguishes whether the steep density stratification at the interface between the silicon core and the silicon/oxygen burning shells that contribute to neutrino heating inside the gain region and revival of the stalled shock is inside or outside $2.5 M_\odot$ (Sukhbold and Woosley, 2014). Based on 2D and 3D CCSN simulations with approximate neutrino transport, Horiuchi et al. (2014) concluded that a lower threshold of $\xi_{2.5} > 0.2$ better represents the final fate outcomes in their sample and is in line with observational constraints. Many other works confirmed the following gross trend, based on different sets of stellar progenitors and various 1D SN codes: statistically, stars with lower $\xi_{2.5}$ are more likely to explode (e.g. O’Connor and Ott, 2013; Nakamura et al., 2015; Sukhbold et al., 2018; Schneider et al., 2021). With the PROMETHEUS-HOTBATH 1D hydrodynamics code that parametrizes the contraction of the proto-neutron star (PNS) which is excised from the computational domain, Ugliano et al. (2012) obtained successful SNe for $\xi_{2.5} < 0.15$, failed SNe for $\xi_{2.5} > 0.35$, and their co-existence for values in-between. Using a semi-analytical approach of comparing the neutrino luminosity obtained from GR1D simulations to the estimated critical neutrino luminosity required for shock

¹In this work, we consider the neutrino-driven engine and not alternative SN mechanisms, such as those driven by jittering jets (Papish and Soker, 2011), magnetars (Kasen and Bildsten, 2010) or collapsars (Woosley and Bloom, 2006).

revival, [Pejcha and Thompson \(2015\)](#) found a similar degeneracy in the final fate outcomes over a large set of pre-SN progenitors, even when optimizing the choice of M for ξ_M . [Ertl et al. \(2016\)](#) made the choice of the mass coordinate dependent on entropy (dimensionless entropy per nucleon number of $s = 4$) and suggested to use a two-parameter criterion to classify explodability. To this end, they introduced the normalized mass

$$M_4 = m(s = 4)/M_\odot, \quad (3..2)$$

which is located where the entropy abruptly rises and the density declines. It typically defines the mass shell of the proto-neutron star (PNS) in whose vicinity the shock is revived in case of a successful SN explosion. Matter falling onto the stalled shock at a rate \dot{M} exerts ram pressure and thereby obstructs shock expansion. Smaller values of the radial gradient at the mass coordinate M_4 ,

$$\mu_4 = \left. \frac{dm/M_\odot}{dr/1000\text{km}} \right|_{s=4}, \quad (3..3)$$

imply a steeper density jump, such that at core collapse, lower-density matter arrives at the neutrino-sphere and the ram pressure is reduced ([Sukhbold et al., 2018](#)). The hot accretion mantle pushes onto the PNS, giving rise to an accretion luminosity that is maintained by persistent mass accretion. For neutrino luminosities above a critical threshold $L_{\nu,\text{crit}} \propto G M_{\text{PNS}} \dot{M} / R_{\text{PNS}}$ that depends on the mass accretion rate as well as on the PNS mass $M_{\text{PNS}} \propto M_4$ and radius R_{PNS} , the neutrino heating triggers the onset of an explosion by shock runaway expansion ([Burrows and Goshy, 1993](#)). μ_4 , when divided by the free-fall timescale and multiplied by the radius up to which M_4 is enclosed, is proportional to \dot{M} . Since R_{PNS} was found to be similar across different progenitors and only weakly time-dependent in late-time explosions, [Ertl et al. \(2016\)](#) argued based on analytical considerations that the product $\mu_4 M_4$ is a proxy for $L_{\nu,\text{crit}}$. To separate exploding from non-exploding pre-SN progenitors, they suggested to draw a separation line in the $(\mu_4 M_4, \mu_4)$ plane that compares two competing forces onto the CCSN outcome: failed SNe are favoured by a high density outside of the iron core (i.e., a large μ_4) and by weak accretion luminosities (i.e., a small $\mu_4 M_4$). With this explodability criterion (E16), [Ertl et al. \(2016\)](#) achieve an accuracy of $\simeq 97\%$ over a heterogeneous set of 621 massive single star progenitors, when exploding them with a 1D CCSN model that updates the 1D approach of [Ugliano et al. \(2012\)](#) by a reduced neutrino wind power. While a two parameter-criterion is more appropriate to model explodability (see e.g. [Heger et al., 2023](#)) a separation line in the $(\mu_4 M_4, \mu_4)$ plane has not been shown to be a reliable criterion to discriminate the outcomes of other 1D SN codes, such as those that model the effects of turbulence in 1D by a modified mixing-length theory approach², STIR ([Couch et al., 2020](#); [Boccioli and Fragione, 2024](#)).

E16 assesses the shock revival conditions at one instant of time – at the onset of iron core infall. It therefore does not capture the temporal-dynamical nature of how the heating and accretion conditions compare as the collapse proceeds. One of the original motivations for the formulation of a semi-analytic supernova model (M16

²However, see [Müller \(2019\)](#) for a critical assessment of this approach to incorporating the effects of turbulence.

Müller et al., 2016) was to supply a set of ordinary differential equations that takes the dynamical evolution of these competing effects into account. M16 draws together theoretical insight, scaling relations and analytical approaches from previous work into a unified semi-analytic framework that models the neutrino-driven perturbation-aided SN mechanism. It takes as input the density, chemical composition, binding energy, sound speed and entropy profiles from the pre-SN progenitor. To assess whether the stalled shock is revived or not, it adopts the earlier established criterion that stars are expected to explode if the mass advection timescale is greater than that of neutrino heating (Janka and Keil, 1998):

$$\tau_{\text{adv}} > \tau_{\text{heat}}. \quad (3.4)$$

After a successful shock revival and SN launch, the mass accretion onto the PNS is continued while mass is ejected. M16 models the effects due to the competing co-existence of outflows and downflows during the explosion phase using dissipation terms. A fallback BH is formed if the predicted initial explosion energy is insufficient to unbind the envelope or if the mass gaining PNS surpasses the maximal equation-of-state dependent baryonic NS mass $M_{\text{NS,bary}}^{\text{max}}$. When the downflows resume, the final explosion energy, the compact remnant mass, total ejecta mass and nickel mass nucleosynthesis yield are obtained from a set of equations modeling explosive nuclear burning during shock propagation up to breakout.

The M16 model outcomes have been compared to those of the $\xi_{2.5}$ -based and E16 criteria over the same set of pre-SN progenitors (e.g., Sukhbold et al., 2018; Schneider et al., 2023; Takahashi et al., 2023; Aguilera-Dena et al., 2023), but so far no simple pre-SN criterion to the M16 outcomes based on a reduced set of pre-SN variables parameters has been formulated. In this work, we evaluate M16, with a parameter calibration from Schneider et al. (2021) over a large sample of $\simeq 3900$ single, binary-stripped and accretor star 1D models, in order to formulate a method to classify explodability of massive stars based on multiple diagnostic pre-SN structural parameters that are used in concert: $\xi_{2.5}$, the E16 parameters ($\mu_4 M_4, \mu_4$), the carbon-oxygen mass M_{CO} and central specific entropy s_c . We further find that the CCSN outcome not only is encoded in the pre-SN stellar progenitor structure, but – given a fixed model for the late burning phases – already pre-determined at the end of core helium burning. This has far-reaching consequences in particular for population synthesis. Since 1D CCSN codes such as M16 require as input the entire progenitor structure profiles at pre-SN stage, they cannot be applied for binary population synthesis (BPS) for two reasons. First, rapid and detailed BPS codes such as POSYDON (Fragos et al., 2023) and BPASS (Byrne et al., 2022) impose an evolutionary cut-off at – if not before – central neon ignition. Second, rapid BPS codes such as COMPAS (Riley et al., 2022) and BINARY C (Izzard and Jermyn, 2023) do not keep track of entire stellar structure profiles up to the cut-off point, and evolve only global parameters. At the same time, in BPS codes that are used to – for example – predict binary black hole merger rates, the CCSN model is one of the most principal components of the predictive pipeline. Rapid codes use recipes such as those introduced in Fryer et al. (2012), Mapelli et al. (2020) or Mandel and Müller (2020) for predicting the final fate that are all based on M_{CO} . In this work, we use our pre-SN explodability criterion to construct a predictive framework that bridges the gap between the evolutionary

cut-off imposed by rapid BPS codes and the CCSN outcomes as predicted by the M16 model.

This paper is structured as follows. In Sect. 3.2.1. we refer to the set of single, binary-stripped and accretor star CCSN progenitor models, in Sect. 3.2.2. the parameter choice of the M16 CCSN model, and in Sect. 3.2.3. the catalog of 3D CCSN simulation outcomes considered in this work. In Sect. 3.2.4., the supervised learning model is described, which we use to map out CCSN outcomes as a function of progenitor M_{CO} . The pre-SN explodability criterion telling apart successful and failed SNe is formulated in Sect. 3.3.1.. It is compared against 3D CCSN simulation outcomes, and extended by a scheme predicting the remnant type (NS or fallback BH) left behind a successful SN explosion. The final fate landscape of single and binary-stripped stars is presented in Sect. 3.3.2.. It is derived from the trends of explodability proxies relevant for the pre-SN criterion with M_{CO} and Z , while taking into account differences between single and binary-stripped star progenitors. We use these results to formulate a CCSN recipe applicable for rapid BPS studies, and compare it to previous recipes. In Sect. 3.3.3. the prediction of CCSN outcomes is compared against observations that constrain the CCSN progenitor properties. The results are discussed and an outlook given in Sect. 3.4.. Conclusions are drawn in Sect. 3.5..

3.2. Methods

3.2.1. 1D CCSN progenitor models

We consider the following set of stellar evolution models:

- 127 single and binary-stripped stars [Case A, Case Be, Case Bl, Case C] of variable zero-age-main-sequence mass $M_{\text{ZAMS}} \in (11, 90) M_{\odot}$ at solar³ metallicity $Z = Z_{\odot}$ (S21) from [Schneider et al. \(2021\)](#),
- 121 single and binary-stripped stars [Case A, Case Be, Case Bl, Case C] of variable $M_{\text{ZAMS}} \in (11, 90) M_{\odot}$ at $Z = Z_{\odot}/10$ (S23) from [Schneider et al. \(2023\)](#),
- 570 accretor stars [Case Ae⁴, Case Al, Case Be, Case Bl, Case C] of variable $M_{\text{ZAMS}} \in (11, 90) M_{\odot}$ and fraction $f \in (0.1, 2)$ of ZAMS mass accreted on the thermal timescale at $Z = Z_{\odot}$ (S24) from [Schneider et al. \(2024\)](#),
- 169 single stars of variable $M_{\text{ZAMS}} \in (9, 70) M_{\odot}$ and convective core overshoot parameter $\alpha_{\text{ov}}/H_P \in (0.05, 5)$ at $Z = Z_{\odot}$ (T24) from [Temaj et al. \(2024\)](#),
- 2910 single stars of variable $M_{\text{ZAMS}} \in (9, 45) M_{\odot}$ at $Z = Z_{\odot}$ (H16) from [Müller et al. \(2016\)](#).

All 3897 stellar models were evolved from ZAMS up to onset of iron-core infall. The first four data sets, S21, S23, S24 and T24 have in common that for the advanced

³As solar metallicity, we assume $Z_{\odot} = 0.001432$ following [Asplund et al. \(2009\)](#)

⁴The distinction between case Ae and case Al is made depending on whether the mass transfer occurs before (Ae) or after (Al) the mid-MS.

burning phases, the same input physics is assumed, and that these are evolved using the MESA (Paxton et al., 2011) code. H16 adopts a different input physics for the main and the advanced burning phases, and the stellar models are evolved using the KEPLER (Weaver et al., 1978; Heger and Woosley, 2010) code.

The classification of binary-stripped stars in S21 and S23 based on the mass transfer (MT) history follows the nomenclature summarized in Table 3.1. The stripped stars are modeled effectively as single stars, with a prescription for removal of the hydrogen-rich envelope over 10% of the thermal timescale. For details on the effective modeling approach, see Schneider et al. (2021).

Table 3.1: Classification of binary-stripped star models based on the mass transfer history of the donor star. Stable MT leading to removal of the hydrogen-rich envelope takes place during the main sequence (MS) evolution for case A systems, between terminal age main sequence (TAMS) and terminal age core helium burning (TACHeB) for case B systems. In early case B (Be), the envelope of the star being stripped is still radiative, while it is convective for late case B (Bl). In case C systems, the hydrogen-rich envelope is removed after TACHeB.

MT class	timing of h-rich envelope removal
single	none
case A	during MS
case Be	between TAMS and TACHeB, radiative envelope
case Bl	between TAMS and TACHeB, convective envelope
case C	after TACHeB

3.2.2. 1D CCSN explosion model

CCSN outcomes predicted by the M16 code depend on in total six explicitly specifiable free parameters. The accretion efficiency $\zeta = 0.8$, the cooling timescale $\tau_{1.5}/s = 1.2$ of a $1.5 M_{\odot}$ PNS and the mass outflow fraction $\alpha_{\text{out}} = 0.5$ are left with the default values from Müller et al. (2016). For this work, we adopt a lower shock compression ratio value, $\beta = 3.3$ (default: 4.0), a greater shock expansion parameter due to turbulent stresses, $\alpha_{\text{turb}} = 1.22$ (default: 1.18), and lower the maximal gravitational NS mass to $M_{\text{NS,grav}}^{\text{max}} = 2.0 M_{\odot}$ (default: $2.05 M_{\odot}$). In Schneider et al. (2021), these values were tuned manually in order to get an average explosion energy of Type IIP SNe in the range 0.5 – 1.0 B for consistency with observations. For single stars, this choice of parameters preserves the highly skewed shape of the explosion energy distribution landscape as is also obtained with the default parameters but admits explosion energies above 2 B, which is the maximal value over the H16 progenitors when using the M16 model with default parameters. The explosion energy distribution from the customized parameter choice has a longer tail, i.e. an extension toward greater explosion energies up to $\simeq 3$ B for single stars at $Z = Z_{\odot}$ (see Fig. 3.18). The predicted chirp mass landscape obtained from this parameter choice is consistent with the current LIGO-Virgo-KAGRA observations of gravitational-wave sources (Schneider et al., 2023).

3.2.3. Catalog of 3D CCSN simulation outcomes

We compare the final fate predictions using our pre-SN criterion to those of state-of-art 3D CCSN simulations from two different groups: 1) the Garching group, over single-star progenitors from [Sukhbold et al. \(2018\)](#), and 2) the Monash group, over single-star progenitors and stripped-star progenitors from catalogs other than those used in this work. Unless otherwise specified, the pre-SN progenitors were non-rotating. The model naming acronyms encode metallicity ('s' for solar, 'z' for Population III), ZAMS mass (e.g., 9.5 for $M_{\text{ZAMS}} = 9.5 M_{\odot}$) and binary-stripped status ('he' for helium stars at $Z = Z_{\odot}$) of the progenitors. In the Monash group archive, 3D CCSN simulations have been run over 21 single and binary-stripped star pre-SN progenitors. Simulations have been carried out either over spherically symmetric progenitors with or without magnetic fields, or with a perturbation introduced into the progenitor stratifications as initial conditions, obtained from simulations of pre-collapse oxygen burning.

Over the low-mass ($M_{\text{CO}} < 2 M_{\odot}$) progenitors he2.8, he3.5, s9.5, z9.6, s11.5 and s11.8 and the high-mass ($M_{\text{CO}} > 8 M_{\odot}$) progenitors y20, z40, z85 and z100 and the rapidly rotating m39, CCSN simulations used spherically symmetric stratifications and did not include magnetic fields. With the exception of z100, all of these exploded. In the intermediate M_{CO} range mostly in-between the aforementioned low-mass and high-mass models, perturbations were introduced into the spherically symmetric stratifications of the progenitors he3, z12, s12.5 and s14.07, and these models exploded in 3D simulations without the enhancing effect of magnetic fields. 3D CCSN simulations with magnetic fields over the spherically symmetric stratifications of s14, s15s7b2 and of the very slowly rotating s16.9 lead to a non-explosion (s14) and an explosion (s15s7b2, s16.9). The effect of perturbations and of magnetic fields was studied systematically upon the pre-SN progenitors s18 and the slowly rotating m15b2. The latter did not explode when starting from spherically symmetric progenitor stratifications and not including magnetic fields, but exploded once magnetic fields were introduced. The same outcomes were observed for s18. In addition, a 3D CCSN simulation over perturbed s18 stratifications lead to an explosion even without magnetic field effects.

Multi-dimensional effects enhance shock revival due to turbulent stresses, increased advection timescale and the increased heating efficiency compared to the axisymmetric treatment, which are coupled to each other by feedback processes ([Müller and Janka, 2015](#)). One approach to taking the shock revival enhancing effects into account in the 1D modeling is to modify equations for hydrostatic structure and jump conditions at the shock by scaling up the shock radius $r_{\text{sh}} \rightarrow \alpha_{\text{turb}} \cdot r_{\text{sh}}$ by a factor α_{turb} that is set by the root-mean-square averaged turbulent Mach number in the gain region ([Müller and Janka, 2015](#)). One of the origins for strong seed perturbations are oxygen-neon shell mergers, convective burning in the silicon burning phase or pulsations before the iron core forms and surpasses its effective Chandrasekhar mass ([Timmes et al., 1996](#)). [Müller et al. \(2016\)](#) suggest to consider α_{turb} within a range between 1.08 and 1.28 depending on progenitor properties and due to theoretical uncertainty about the reasons for the turbulent motion boosts behind the shock. Further on, they suggest that explosion-enhancing multi-D effects are switched off at

$\alpha_{\text{turb}} \simeq 0.86$ rather than at $\alpha_{\text{turb}} = 1$ due to renormalization procedures. Since in our 1D CCSN model, we keep the value $\alpha_{\text{turb}} = 1.22 > 0.86$ fixed, we implicitly assume the presence of seed perturbation or other shock-revival enhancing multi-dimensional effects (such as magnetic fields).

Therefore, in our comparison of pre-SN progenitor properties to outcomes of 3D CCSN simulations, we use the outcomes of 3D simulations with perturbations or magnetic fields wherever applicable. For the lighter and heavier progenitors, these are not available, but since the spherically symmetric stratifications already explode (except for z100), it is to be expected that the same progenitors would still explode if perturbations are introduced. This distinction becomes relevant in particular for m15b2 and s18. For the models s14 (non-exploding), s16.9 (exploding) and s24 (exploding), 3D CCSN simulations have been carried out only with magnetic fields.

3.2.4. Supervised learning model

Since the stellar structure profiles are not available beyond the stellar parameter grid nodes over which massive stars have been evolved up to pre-SN stage, it is not possible to directly predict final fates over a quasi-continuous parameter space of stellar progenitors as necessary for BPS studies. To overcome this gap, in this work, we directly relate stellar parameters to CCSN outcomes. To this end, we first formulate a pre-SN explodability criterion based on scalar quantities derived from the stellar structure at the onset of core collapse. This criterion coincides with the final fate outcomes issued by the M16 model and therefore can be reliably used as its surrogate. In a second step, in order to obtain final fate landscape forecasts, we evaluate the pre-SN criterion over a quasi-continuous parameter space in M_{CO} using fitted models of the scalar proxies necessary for evaluation of the criterion.

As a fitting model, we use Gaussian Process Regression (GPR; Rasmussen, 2004). Ever since pioneering the work by Sacks et al. (1989), GPR has been a standard method for emulation tasks because of its flexibility, smoothness and regulatory effect of the Gaussianity assumption (for a detailed discussion of application of GPR to computer model calibration, see Kennedy and O’Hagan, 2001). The data-driven supervised learning approach is used to fit GPR models. If n is the size of the training data set, then GPR interprets the output data as a random sample drawn from an n -dimensional multivariate normal (MVN)

$$\mathcal{N}_n(\mu, \Sigma) = \frac{\exp\left(-\frac{1}{2}(\mathbf{X} - \mu)^T \Sigma^{-1} (\mathbf{X} - \mu)\right)}{\sqrt{(2\pi)^n |\Sigma|}} \quad (3.5)$$

which has the mean vector μ and the functional form of the $n \times n$ covariance matrix $\Sigma_{i,j} = \text{Cov}[x_i, x_j]$ as free choices that need to be pre-set before the training.

While μ is typically set to the zero vector, different kernel functions $k(x_i, x_j)$ are available⁵ that specify Σ . In this work we adapt the kernel choice to each fitting

⁵For a selection of kernel models, see e.g. https://scikit-learn.org/stable/modules/generated/sklearn.gaussian_process.GaussianProcessRegressor.html

task. We often find best performance with the Matérn kernel

$$k(x_i, x_j) = \frac{1}{\Gamma(\nu)2^{\nu-1}} \left(\frac{\sqrt{2\nu}}{l} d(x_i, x_j) \right)^\nu K_\nu \left(\frac{\sqrt{2\nu}}{l} d(x_i, x_j) \right), \quad (3.6)$$

where $d(x_i, x_j)$ is the Euclidean distance between x_i and x_j , K_ν the modified Bessel function, and Γ the Gamma function. The Matérn kernel has free parameters ν and l . ν controls the smoothness of the approximated function, and l is the length scale of the kernel. The training task is to optimize the choice of the kernel parameters by returning a probability distribution over their values, such that the resulting MVN has maximum likelihood over the training data.

Once the GPR model is fitted, predictions over evaluation points are made by conditioning over the training data. and their prediction intervals are obtained from marginalization. For more technical detail, refer to [Rasmussen \(2004\)](#).

3.3. Results

3.3.1. Pre-SN explodability criterion

Is the CCSN outcome as predicted by M16 universally (i.e., over single, binary-stripped and accretor stars having very different pre-SN properties) encoded already at the pre-SN stage? We find that the final fate cannot be discriminated using a single explodability proxy, be it $\xi_{2.5}$, central specific entropy s_c , iron core mass M_{Fe} , nor when using two of them with calibrated threshold values (e.g. $\xi_{2.5}$ and s_c), nor with a separation line in the $(\mu_4 M_4, \mu_4)$ plane (see Fig. 3.1 and 3.2). However, there are upper (lower) limits in $\xi_{2.5}$, s_c and M_{CO} , beyond (below) which only implusions (explosions) occur (see Table 3.2 and the red demarcation lines in the aforementioned figures). For the remaining pre-SN progenitor models, we find that a modified E16 criterion discriminates the CCSN outcomes. To summarize, the following set of conditions replicates the CCSN outcomes of M16 with an accuracy of $\simeq 99.3\%$ over the super-set consisting of 3897 heterogeneous pre-SN progenitor models (see Fig. 3.1 and Fig. 3.2):

Pre-SN distinction between failed and successful SNe: If either $\xi_{2.5}$, or s_c or M_{CO} are above an upper threshold value, the SN outcome is a failed SN. If, conversely, any of these proxies is below a lower threshold value, the outcome is a successful SN launch. For values in the intermediate (“overlap”) region or in the rare case of conflicting final fate assignments using these variables, the final fate is decided in the $(\mu_4 M_4, \mu_4)$ plane: if the $\mu_4 M_4$ coordinate is critically low (large), the outcome is a successful (failed) SN, while for intermediate values a separation line tells the exploding and imploding stars apart. The critical values are summarized in Tab. 3.2. The separation line is set by the parameters $(k_1, k_2) = (0.005, 0.421)$ and yields the following condition for a failed SN, over $\mu_4 M_4 \in (0.247, 0.438)$:

$$\mu_4 < k_1 + k_2 \cdot \mu_4 M_4. \quad (3.7)$$

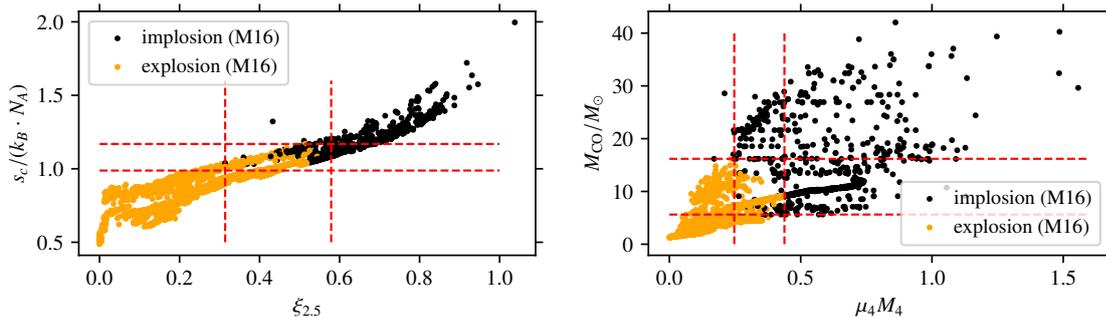


Figure 3.1: Final fate assignment in the $(\xi_{2.5}, s_c)$ and $(\mu_4 M_4, M_{CO})$ planes, based on CCSN outcomes predicted by M16 over the set of S21, S23, S24, T24 and H16 pre-SN progenitors. The red lines indicate the critical lower and upper thresholds of each expodability proxy considered, which are summarized in Table 3.2. The fates of those collapsing stars that hitherto have not been assigned, are mapped out by the separation line in the $(\mu_4 M_4, \mu_4)$ plane (see Fig. 3.2).

Stellar models that fulfill this condition, i.e. those found *below* the separation line in the $(\mu_4 M_4, \mu_4)$ plane, are found to implode; otherwise, they explode. Note that according to E16, the final fate outcomes are reversed: the non-exploding models are found *above* the separation line. We interpret the reversed (within a subrange of $\mu_4 M_4$) E16 criterion as follows:

Since $\mu_4 M_4$ is a proxy for the required neutrino luminosity $L_{\nu, \text{crit}}$ for runaway shock expansion, for $\mu_4 M_4 < (\mu_4 M_4)^{\text{min}}$, $L_{\nu, \text{crit}}$ is critically low, making it easy to explode the progenitor. Conversely, for $\mu_4 M_4 > (\mu_4 M_4)^{\text{max}}$, $L_{\nu, \text{crit}}$ is critically high, making it difficult to explode the star. In the intermediate range $(\mu_4 M_4)^{\text{min}} \leq \mu_4 M_4 \leq (\mu_4 M_4)^{\text{max}}$, a large μ_4 compared to $k_1 + k_2 \cdot \mu_4 M_4 \propto \mu_4 M_4$ implies a comparatively small PNS mass $M_4 = \mu_4 M_4 / \mu_4$, which makes it easier to explode the progenitor for the following reasons. The mass accretion onto the gain region has two adverse effects: on the one hand, it will increase the PNS mass, and on the other hand, it supplies fresh material into the gain region that is neutrino-heated. With a comparatively small M_4 , the first effect is less dramatic (since more mass needs to be accreted for surpassing the limiting PNS mass for stability) while the second effect is enhanced (because $L_{\nu, \text{crit}}$ is comparatively low. With a comparatively large μ_4 , it then is easier to reach the critical neutrino luminosity for shock revival than the threshold mass for PNS collapse.

Out of 3987 progenitor models, 2685 ($\simeq 69\%$) explode. The pre-SN explodability criterion does not depend on evolutionary pre-history of the progenitors, since it is constructed based on a heterogeneous set that contains single-stars, binary-stripped stars and accretor stars. The most essential part of our criterion is the one based on the $(\mu_4 M_4, \mu_4)$ plane. Using it alone leads to an accuracy in the final fate assignment of $\simeq 98\%$. When adding the critical values in $\xi_{2.5}$, s_c and M_{CO} , the accuracy rises to 99%.

There are two main ideas advocated by our approach to parametrizing explodability. First, there is no sharp boundary in $\xi_{2.5}$ or any other explodability proxies that would allow to map out explodability based on a single threshold value alone.

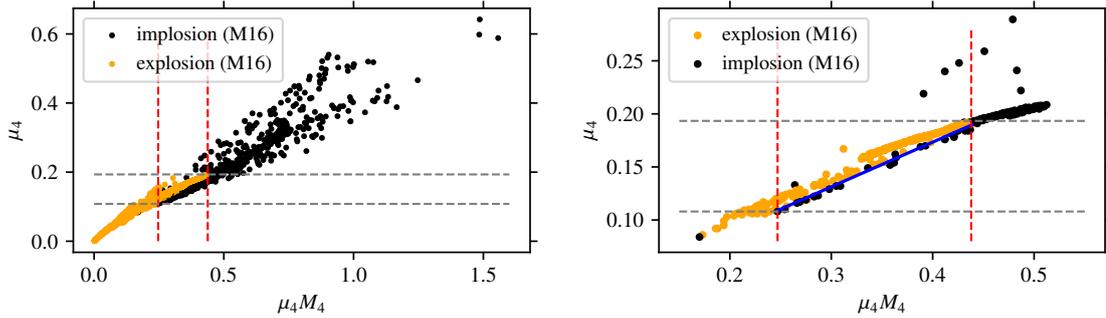


Figure 3.2: Final fate distribution in the $(\mu_4 M_4, \mu_4)$ plane over the entire set of S21, S23, S24, T24 and H16 pre-SN progenitors (left panel), and over a subset of progenitors that fall into the overlap region with degenerate final fate outcomes, as their $\xi_{2.5}$, s_c and M_{CO} are neither super-critical nor sub-critical to assign a final fate forecast based on $\xi_{2.5}$, s_c or M_{CO} (right panel). Stars from this overlap region explode if $\mu_4 M_4$ is below the lower threshold, and implode if it is greater than the upper threshold. For $\mu_4 M_4$ values in-between, a separation line (in blue) discriminates the SN outcomes.

Table 3.2: Upper and lower threshold values in explodability proxies that pre-determine CCSN outcomes as predicted by M16. Below the lower threshold value, only SNe occur. Above the upper threshold value, only implosions occur. Over the value range in-between, the final fate outcome is degenerate (co-existence of successful and failed SN). Final fate outcomes in this intermediate range are mapped out using Eq. 3.7 with coefficients $(k_1, k_2) = (0.005, 0.421)$, which determine the separation line between exploding and non-exploding stars in the $(\mu_4 M_4, \mu_4)$ plane (see Fig. 3.2).

variable X	lower threshold X^{\min}	upper threshold X^{\max}
$\xi_{2.5}$	0.314	0.579
$s_c/[N_A \cdot k_B]$	0.988	1.169
M_{CO}/M_{\odot}	5.6	16.2
$\mu_4 M_4$	0.247	0.438

Looking at one single explodability proxy, such as $\xi_{2.5}$, is thus insufficient to make faithful final fate forecasts. Neither can a globally valid segmentation line be constructed in the $(\mu_4 M_4, \mu_4)$ plane for discriminating the CCSN outcomes. Instead, there are relatively broad value ranges in $\xi_{2.5}$, s_c and in other pre-SN variables, over which both successful and failed SN outcome are the outcomes (see Figs. 3.1 and 3.2). Second, the explodability proxies $\xi_{2.5}$, s_c and the E16 parameters $(\mu_4, \mu_4 M_4)$ are *not* equivalent in their significance for the final fate outcome. While these can in some cases follow similar trends (e.g. a large $\xi_{2.5}$ accompanied by a large s_c), they may carry complimentary information about explodability (e.g. a critically low s_c for a successful SN, but a moderate $\xi_{2.5}$ within the degenerate range over which failed and successful SNe co-exist). To assign the final fate of a CCSN at the pre-SN stage,

insight from several explodability proxies needs be drawn together. Our criterion states that it is sufficient to evaluate $\xi_{2.5}$, s_c , μ_4 , $\mu_4 M_4$ and M_{CO} . These proxies probe the pre-SN progenitor at four different mass coordinates: the innermost region of the stellar core at constant entropy, the mass coordinate at which typically the PNS is enclosed once it gets formed, the mass coordinate $2.5M_{\odot}$ which for many stars⁶ is at the interface between the silicon core and the silicon-enriched oxygen layers, and the carbon-oxygen rich layers much further out.

Comparison with 3D CCSN simulation outcomes

For our validation of the pre-SN criterion against 3D CCSN simulation outcomes, we take into account 6 exploding and 2 non-exploding models from the Garching archive, as well as 17 exploding and 2 non-exploding models from the Monash archive⁷. Fig. 3.3 compares the outcomes of 3D CCSN simulations to their pre-SN properties necessary for diagnosing a final fate using our explodability criterion. We first analyze the models that explode in 3D. As in the 1D case, most of them are scattered in the region of critically low $\xi_{2.5}$ and critically low s_c . There are a few exceptions:

- the stripped-star model y20 represented by the “left-hand triangle” symbol has a conflicting final fate prediction issued by $\xi_{2.5}$ (to explode) versus by s_c (to implode). Its $M_{\text{CO}} \simeq 8.2 M_{\odot}$ is within the overlap region of degenerate final fate outcomes. It is predicted to explode – in agreement with the 3D outcome – because of a critically low $\mu_4 M_4$ value.
- The stripped-star model m39 represented by the “right-hand triangle” symbol is within the overlap region in the $(\xi_{2.5}, s_c)$ plane, but is predicted to implode because of a critically large $M_{\text{CO}} \simeq 21 M_{\odot}$, in disagreement with the 3D outcome. Interestingly, if the M_{CO} condition is ignored, according to the final fate assignment in the $(\mu_4 M_4, \mu_4)$ plane, the model is predicted to explode as it lies above the separation line in the $(\mu_4 M_4, \mu_4)$ plane, on the verge of the transition to the regime of having a critically large $\mu_4 M_4$ value for an implosion. We conclude that since stars with M_{CO} as high as $21 M_{\odot}$ are found to explode in 3D simulations, the final fate assignment based on M_{CO} alone is not faithful. We speculate that the progenitor has had a high (at least $X_C > 0.25$) central carbon mass fraction X_C core carbon burning onward, which enhances the probability for an explosion. If we lift the M_{CO} -based condition for final fate assignment in our pre-SN criterion, we still achieve a predictive accuracy of 98.5 % over the super-set of pre-SN progenitors.
- The models represented by the rhombus and square have a critically large compactness and $\mu_4 M_4$ coordinate for an implosion according to our pre-SN criterion, but explode nevertheless in the 3D simulations. These are the Population

⁶For many stars, the mass coordinate at $2.5 M_{\odot}$ is large enough to be located outside the iron core M_{Fe} , and small enough to be within the mass accreting region. However, the compactness criterion becomes inadequate for stars massive enough to develop iron cores heavier than $2.5 M_{\odot}$ (see Fig 3.16 in the appendix).

⁷We plan to compare our explodability criterion with 3D CCSN simulation outcomes from other groups in follow-up work.

III star models z85 and z40, having an M_{CO} of around 31 and 13 solar masses, respectively. Their $(\mu_4 M_4, \mu_4)$ coordinates reach beyond the region sampled by our set of models. We suspect that the reason for the discrepancy in the final fate outcome predictions are effects of the altered nuclear reactions during iron core collapse compared to stars initially with metals, and conclude that our pre-SN criterion breaks down when it comes to Population III stars, which were not part of the set of stellar models based on which we constructed the pre-SN criterion. We aim to introduce possible fixes to our pre-SN criterion for stellar progenitors with $(\mu_4 M_4, \mu_4)$ values beyond the region sampled by our models and for better applicability to Population III stars in future work.

All models that are non-exploding in the 3D simulation are predicted to implode by the pre-SN criterion, except for the spherically symmetric progenitor model s14 from the Monash group archive. Whether or not s14 explodes in a 3D simulation if asymmetry is introduced into the progenitor stratifications has not been tested.

The final fate predictions of the non-exploding models with the pre-SN criterion have the following origins:

- the non-exploding u75 model (represented by a brown circle) has critically large $\xi_{2.5}$, s_c , M_{CO} and $\mu_4 M_4$ values for an implosion,
- the non-exploding s40 model (represented by a brown circle) has its $\xi_{2.5}$, s_c and M_{CO} within the overlap region, but it is predicted to implode given its critically large $\mu_4 M_4$ coordinate.
- the Population III star model z100 represented by the “upward” triangle symbol has a conflicting final fate prediction based on s_c (explosion) and M_{CO} (implosion), but its $\mu_4 M_4$ coordinate is critically large for an implosion.

When using our pre-SN criterion for prediction-making, the value of the $\mu_4 M_4$ coordinate is found to play the decisive role for the explosion of y20, for the implosion of s40 and z100, and – together with μ_4 – for the explosion of m39. The number of false final fate assignments with our pre-SN criterion over a set of $8+19=27$ progenitors is 4 when the M_{CO} -based condition is included and 3 when it is excluded. The resulting overall accuracy over 3D CCSN simulation outcomes is 85% and 89%, respectively.

Three out of the four prediction errors issued by the pre-SN criterion compared to the 3D outcomes are false negatives for an explosion, i.e. the 3D simulations are even more optimistic about explosions than our pre-SN criterion is. Note that when considering the 3D CCSN simulations on their own, bypassing the M16 1D CCSN model outcomes over the large set of progenitor models, the same predictive accuracy is also achieved by setting a critical $\xi_{2.5}^{\text{crit}} \simeq 0.37 - 0.39$, or by using our $(\mu_4 M_4, \mu_4)$ plane based prescription only. Compared to the M16 model outcomes, the $\xi_{2.5}^{\text{crit}} > 0.38$ criterion for direct BH formation would predict substantially more implosion outcomes (compare Fig. 3.1), and have an agreement with M16 in only 86% over the sample of 3897 1D progenitors. With the $(\mu_4 M_4, \mu_4)$ -only pre-SN criterion applied to our set of SN progenitor models, the agreement with the M16 outcomes is as high as 98.3%, but results in even more explosion outcomes. While the threshold value in $\xi_{2.5}^{\text{crit}}$ is tightly constrained by 3D CCSN simulation outcomes, the final fate

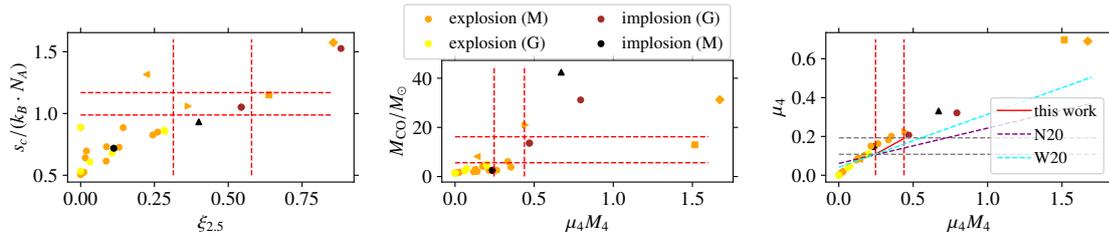


Figure 3.3: Comparison of the final fate assignment using our pre-SN explodability criterion to 3D CCSN simulation outcomes performed in the Garching (G) and Monash (M) groups. The dashed red lines in the two-parameter planes spanned by $(\xi_{2.5}, s_c)$, $(M_{\text{CO}}, \mu_4 M_4)$ and $(\mu_4 M_4, \mu_4)$ planes indicate the critical lower and upper thresholds of each explodability proxy involved in our pre-SN criterion. The fates of collapsing stars for $\mu_4 M_4$ in-between the lower and the upper threshold are mapped out by the reversed E16 criterion (right panel): it predicts explosions if the progenitor is found above the red solid separation line, and implosions if below. The final fate assignment in the $(\mu_4 M_4, \mu_4)$ plane is compared to the standard E16 criterion, with calibrations from Ertl et al. (2020) for the updated W20 and N20 engines. The final fates (success or failure of shock revival at the end of the 3D CCSN simulation) are color-marked in yellow/brown for the Garching, and in orange/black for the Monash group. Specific progenitors that are referenced in the main text are represented by symbols other than circles.

outcomes in 3D CCSN simulations in the $(\mu_4 M_4, \mu_4)$ plane can be segmented in ways different from our scheme while retaining the same level of consistency over the sparsely sampled $(\mu_4 M_4, \mu_4)$ plane given the low number of 3D CCSN simulations available. However, as is evident from Fig. 3.3, no single separation line with BH forming models above and exploding models below the line can be drawn to segregate the final fate outcomes. To illustrate this, the separation lines with (k_1, k_2) calibrations from Ertl et al. (2020), assuming the updated N20 and W20 engines⁸, respectively, in 1D CCSN simulations with P-HOTB are shown in the right panel. These are consistent with explosions at low $\mu_4 M_4$ and low μ_4 but evidently over-predict BH formation compared to the 3D outcomes.

BH formation by fallback

If shock revival is successful, outflows and downflows onto the PNS co-exist until the layers surrounding the PNS are unbound. During this process the PNS accretes matter. If the mass gain lifts its mass beyond the limit for hydrostatic stability or if the explosion is not energetic enough to unbind the envelope, the PNS undergoes accretion-induced collapse to form a fallback BH. If not, the compact remnant becomes a NS. The M16 CCSN model not only predicts the final fate (successful or

⁸The excision of the PNS core introduces free parameters that regulate the neutrino-emission evolution and settling of the hot accretion mantle above the PNS. These are constrained to reproduce the explosion energy, nickel mass, total neutrino energy loss and duration of the neutrino signal of SN 1987A. However, different parameter choices satisfy these observational constraints, and CCSN outcomes still change depending on their calibration (Ertl et al., 2016).

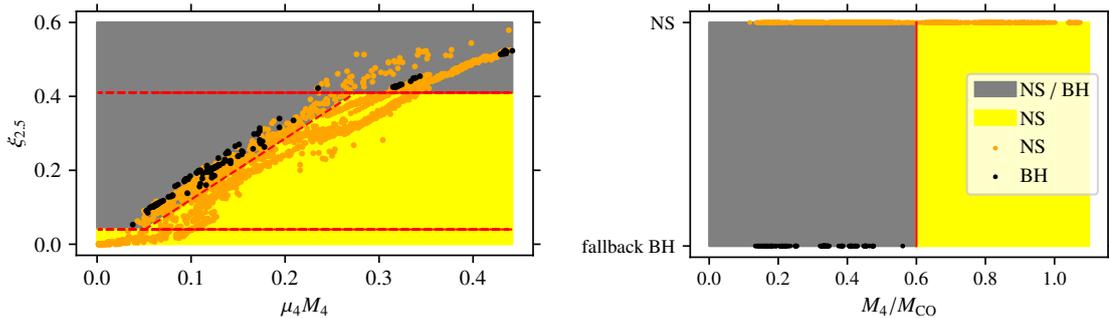


Figure 3.4: Prerequisites for black hole formation by fallback in a successful CCSN explosion. Left panel: the exploding pre-SN progenitors are projected onto a plane spanned by $(\xi_{2.5}, \mu_4 M_4)$. The minimal $\xi_{2.5} = 0.04$ value for fallback BH formation to occur is indicated by a red horizontal line. The linear $\xi_{2.5}$ -to- $\mu_4 M_4$ fit model that sets a minimal $\xi_{2.5}$ value for fallback to occur depending on $\mu_4 M_4$ is represented by a diagonal dashed red line. This constraint applies to progenitors with $\xi_{2.5} < 0.41$ (threshold indicated by horizontal line). For progenitors with a greater $\xi_{2.5}$, fallback may occur regardless of $\mu_4 M_4$. Right panel: the ratio M_4/M_{CO} discriminates the compact remnant type. In both panels, the gray-shaded background indicates the parameter subspace that permits both compact remnant types (NS and fallback BH) while the yellow-shaded background indicates the subspace that excludes BHs and guarantees NS formation.

failed SN) but also the remnant type (NS or fallback BH) for the case of a successful SN. Out of 2685 exploding pre-SN progenitors in our sample, only 167 ($\simeq 6\%$) leave fallback BHs behind. Based on pre-SN properties alone, we do not find a scheme that could discriminate the remnant type (NS versus fallback BH) in a deterministic way. Instead, we identify empirically prerequisites placed on pre-SN variables that need to be fulfilled for fallback BH formation to occur with a certain probability:

- Fallback BH formation occurs only if the compactness is not critically low. At

$$\xi_{2.5} < 0.04, \quad (3.8)$$

only NSs are left behind, since the silicon-oxygen layers (typically found at the $2.5 M_{\odot}$ mass coordinate surrounding the infalling iron core) are then too loosely bound for substantial fallback onto the PNS.

- Fallback BH formation may occur if the dimensionless $\xi_{2.5}$ parameter is large compared to the dimensionless $\mu_4 M_4$ coordinate, which scales with the accretion luminosity. Under these governing conditions, a typical outcome will be that the kinetic explosion energy imparted to the gain region by neutrino heating does not overcompensate its comparatively strong gravitational binding energy. If the opposite holds, that is when

$$\xi_{2.5} < a \cdot \mu_4 M_4 + b \quad (3.9)$$

with $(a, b) = (1.75, -0.044)$, only NSs are left behind. This condition remains valid for $\xi_{2.5} \in (0.04, 0.41)$. For

$$\xi_{2.5} \geq 0.41, \quad (3.10)$$

the outer layers are tightly bound, and fallback may occur regardless of the accretion luminosity.

- Fallback BH formation occurs only if M_{CO} is large compared to the PNS mass (which correlates with the M_4 coordinate). These two masses allow to compare two adverse effects on the final fate. The PNS mass is an important proxy for the neutrino heating conditions, since it sets the PNS surface temperature. M_{CO} is a proxy for the gravitational binding energy of the entire carbon-oxygen core and for the ram pressure exerted onto the revived shock as it propagates outward. If

$$M_4/M_{\text{CO}} > 0.6 \quad (3.11)$$

is satisfied, only NSs are left behind.

Fig. 3.4 summarizes these pre-requisites for BH formation by fallback. We find that among the exploding progenitors that satisfy these conditions, the fraction of SN explosions that lead to fallback BHs is $f = 0.15$, while NS formation is guaranteed if these conditions are not satisfied.

3.3.2. Final fate landscapes of single and binary-stripped stars

Explodability dependence on M_{CO} and Z

Motivation: Previous works have shown that for the same M_{CO} , pre-SN profile structures of massive stars differ, depending on metallicity (e.g. [Limongi and Chieffi, 2018](#); [Schneider et al., 2023](#)) and on whether the star evolved in isolation or underwent binary mass transfer (e.g. [Brown et al., 2001](#); [Wellstein and Langer, 1999](#); [Schneider et al., 2021](#); [Laplace et al., 2021](#)). In spite of this, M_{CO} -based CCSN recipes that are typically used in rapid BPS codes do not make such a distinction. Here we construct predictive models that bridge M_{CO} to the final fate outcomes in a class-specific way, i.e. with a distinct treatment of single and binary-stripped stars, and by taking Z dependence into account. To facilitate the mapping, we assume a particular model of the late burning stages, [Schneider et al. \(2021\)](#), that evolves stars from the end of core helium burning (CHeB) up to the onset of iron core collapse. We then extract the scalar quantities used in our pre-SN explodability criterion from the catalogs S21 and S23, and fit these as a function of M_{CO} . The fits of the explodability proxies are then used to assign final fates for M_{CO} values in-between the $M_{\text{CO}}^{\text{min}}$ and $M_{\text{CO}}^{\text{max}}$ thresholds⁹ for having explosion and implosion outcomes only, i.e. over the range

⁹With the 3D CCSN simulations of the stripped star m39 and Population III star z85 progenitors, there are two counter-examples of explosions at $M_{\text{CO}} > M_{\text{CO}}^{\text{max}} = 16.2 M_{\odot}$. Nevertheless, for the construction of a M_{CO} -based CCSN recipe for rapid BPS codes, we keep these limiting values for the following reason. The stellar models behind m39 and z85 have different input physics from the single and stripped star models S21, S23, S24 and T24. Over these, the M_{CO} axis is sampled up to above $40 M_{\odot}$ and we do not encounter explosions at such large CO core masses (see Fig. 3.10).

where we find their co-existence, when applying the M16 SN code to the pre-SN progenitors from S21 and S23.

Bimodality: We find that not only the explodability proxies (such as $\xi_{2.5}$) of single stars (e.g. Sukhbold et al., 2018; Limongi and Chieffi, 2018; Patton and Sukhbold, 2020; Schneider et al., 2021, 2023; Takahashi et al., 2023) as well as the $\xi_{2.5}$ and s_c parameters of binary-stripped stars (Schneider et al., 2021), but that all the explodability proxies relevant for our criterion introduced in Sect. 3.3.1. follow structured bimodal trends – characterized globally by two peaks and a valley in-between – with M_{CO} , over the value range in-between $M_{\text{CO}}^{\text{min}}$ and $M_{\text{CO}}^{\text{max}}$. This is shown in Fig. 3.5: typically, the explodability proxy rises up to a peak at M_{CO} values between around 6.1 and 8.4 M_{\odot} , depending on the MT class and on Z , which for most proxies eventually reaches the respective threshold value for BH formation by direct collapse, as predicted by the pre-SN criterion. With further increasing M_{CO} , the values of the proxies first decrease and then increase again. At $M_{\text{CO}} > M_{\text{CO}}^{\text{max}}$, even though the explodability variable such as s_c may oscillate down to below its upper threshold (e.g., s_c^{max}), this does not change the final fate assignment due to the critically large M_{CO} value itself.

The reason why the dependence of explodability on M_{CO} is non-monotonic is that M_{CO} is a proxy for the interior core structure at the onset of collapse as set by the evolution through the late burning phases. The pioneering work of Sukhbold and Woosley (2014) studied the physical origin of explodability properties on the nuclear burning conditions during the burning stages past core helium exhaustion, and how these are set by M_{CO} . M_{CO} controls the core temperature T_c and density ρ_c and thereby also the neutrino luminosity leaking out of the core. At greater helium core masses (which are proportional to the CO core masses), T_c and ρ_c during CHeB are greater, and the $^{12}\text{C}(\alpha, \gamma)^{16}\text{O}$ reaction sets in earlier. This leads to a general trend of having a lower X_C at the end of CHeB with increasing M_{CO} . The non-monotonicity in the final fate dependence on M_{CO} has been linked to episodes of enhanced versus weakened neutrino-emission during core carbon and neon burning phases (Brown et al., 2001; Sukhbold and Woosley, 2014), which in turn are primarily set by M_{CO} and X_C (Chieffi and Limongi, 2020; Schneider et al., 2023; Laplace et al., 2024). Carbon and neon burning becoming neutrino-dominated leads to an increase in $\xi_{2.5}$, and – by taking away thermal energy from the core – marks the transition from a convective to radiative burning core, and changes the number as well as size of carbon burning shells (Sukhbold and Woosley, 2014). However recent work by Laplace et al. (2024) made the case that this transition is not the cause for the changes in the explodability patterns, and identified the mechanisms explaining the formation of the peaks in $\xi_{2.5}$. When the temperature and density conditions (set by M_{CO}) and the amount of nuclear fuel (which, in the case of carbon burning, is given by X_C at carbon ignition) are such that the central burning source is strongly neutrino-dominated, the core contraction increases, leading to a large fuel-free core and – ultimately – to an increase in the iron core mass M_{Fe} and $\xi_{2.5}$. For even more neutrino-dominated cases at higher masses and lower initial fuel abundance, the next nuclear burning episode ignites early, which counters the core contraction and leads to a drop in M_{Fe} and $\xi_{2.5}$. Ultimately, these mechanisms lead to the formation of

characteristic $\xi_{2.5}$ peaks when carbon and neon burning become neutrino-dominated. We discuss how these findings relate to our predictive framework in Sect. 3.4..

Prediction of final fate landscapes: With the fitted models for each quantity relevant for our pre-SN criterion and the critical values from Sect. 3.3.1., we sample – for each MT class – M_{CO} in increments of $\delta M_{\text{CO}}/M_{\odot} = 0.1$ within the range $M_{\text{CO}}/M_{\odot} \in (5.6, 16.2)$ that has co-existence of failed and successful SN outcomes, and evaluate the fitted GPR models of $\xi_{2.5}$, s_c , $\mu_4 M_4$ and μ_4 . If any of the $\xi_{2.5}$, s_c or $\mu_4 M_4$ are above (below) the upper (lower) threshold — i.e., above $\xi_{2.5}^{\text{max}}$, s_c^{max} or $\mu_4 M_4^{\text{max}}$ (or below $\xi_{2.5}^{\text{min}}$, s_c^{min} or $(\mu_4 M_4)^{\text{min}}$) — the outcome at the corresponding M_{CO} is a failed (successful) SN. If not yet classified given these threshold values, the μ_4 fit model is used to discriminate the final fate outcome based on the separation line in the $(\mu_4 M_4, \mu_4)$ plane set by eq. (3..7). Tables 3..3 and 3..4 summarize the results at the two metallicities Z_{\odot} and $Z_{\odot}/10$.

Table 3..3: Boundary values in M_{CO} demarcating non-explosive BH formation by direct collapse at $Z = Z_{\odot}$. Stars are expected to explode for $M_{\text{CO}} < M_{\text{CO}}^{(1)}$ and for $M_{\text{CO}}^{(2)} < M_{\text{CO}} < M_{\text{CO}}^{(3)}$. Conversely, stars are expected to implode, forming BHs without a supernova, between $M_{\text{CO}}^{(1)}$ and $M_{\text{CO}}^{(2)}$, and again from $M_{\text{CO}}^{(3)}$ up to the pair-instability mass gap.

	$M_{\text{CO}}^{(1)}/M_{\odot}$	$M_{\text{CO}}^{(2)}/M_{\odot}$	$M_{\text{CO}}^{(3)}/M_{\odot}$
single	6.6	7.2	12.9
case C	6.6	7.1	13.2
case Bl	7.7	8.3	15.2
case Be	7.8	8.3	15.3
case A	7.4	8.4	15.4

Table 3..4: Same as Table 3..3, but for $Z = Z_{\odot}/10$.

	$M_{\text{CO}}^{(1)}/M_{\odot}$	$M_{\text{CO}}^{(2)}/M_{\odot}$	$M_{\text{CO}}^{(3)}/M_{\odot}$
single	6.1	6.6	12.9
case C	6.3	7.1	12.3
case Bl	7.0	7.9	14.0
case Be	6.9	7.9	13.5
case A	6.9	7.4	13.7

The following structural pattern of explodability dependence on M_{CO} is preserved:

- At low $M_{\text{CO}} < M_{\text{CO}}^{(1)}$, the outcomes are successful SNe.
- As M_{CO} increases, a narrow direct collapse BH formation window arises for $M_{\text{CO}} \in (M_{\text{CO}}^{(1)}, M_{\text{CO}}^{(2)})$.
- It is followed by successful SNe for $M_{\text{CO}}^{(2)} < M_{\text{CO}} < M_{\text{CO}}^{(3)}$.

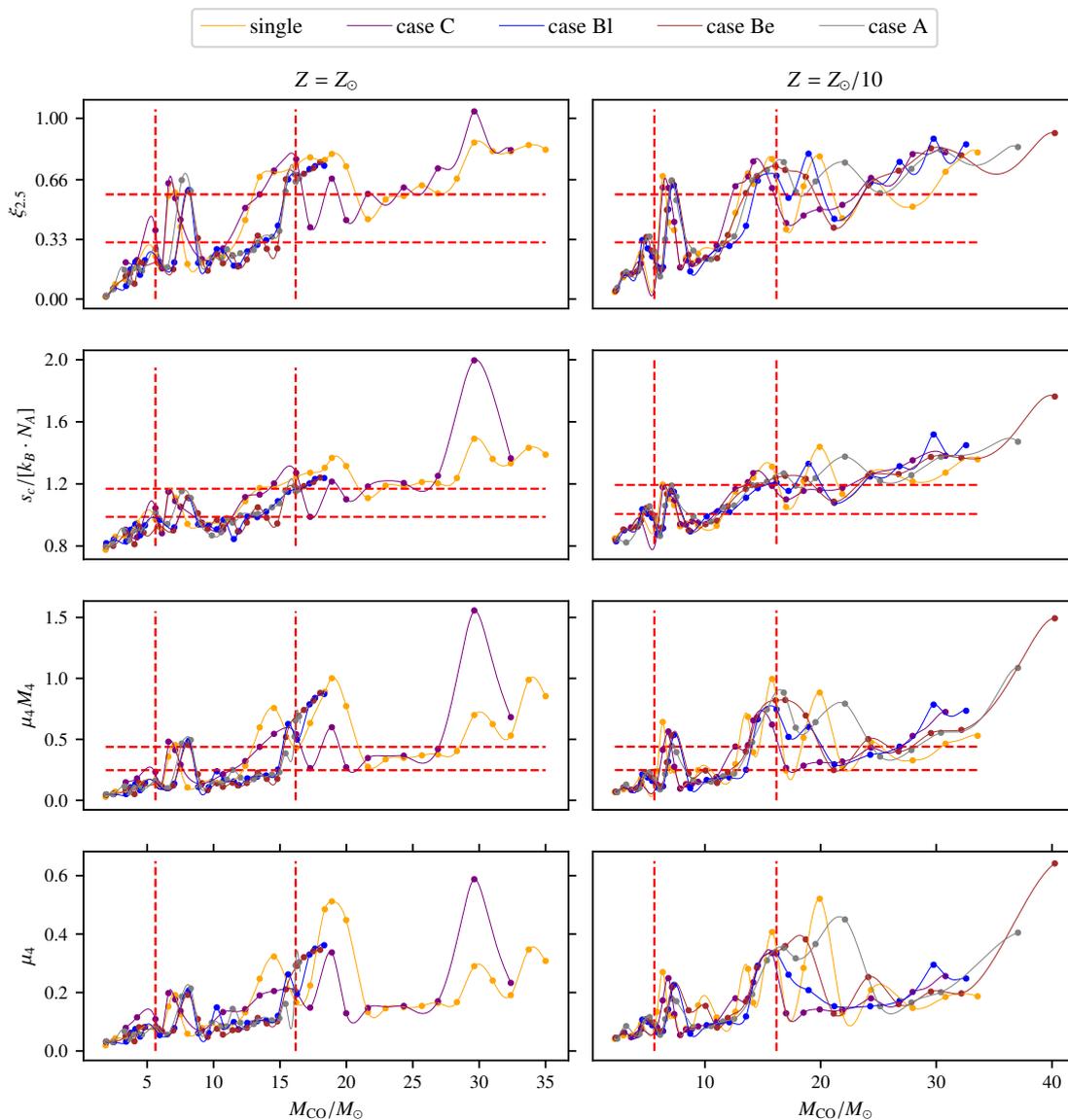


Figure 3.5: Dependence of the explodability parameters $\xi_{2.5}$, s_c , $\mu_4 M_4$ and μ_4 of single and binary-stripped stars (case A, case Be, case B1 and case C donors) on M_{CO} at $Z = Z_{\odot}$ and at $Z = Z_{\odot}/10$. To obtain these, the data points (circles) from S21 and S23 have been fitted using GPR models (solid lines). The red vertical lines indicate the lower (upper) threshold in M_{CO} below (beyond) which only explosions (implosions) occur. The red horizontal lines indicate the lower (upper) threshold in $\xi_{2.5}$, $\mu_4 M_4$ and s_c , respectively, below (beyond) which only explosions (implosions) occur.

- For critically large $M_{\text{CO}} \geq M_{\text{CO}}^{(3)}$ a plateau of BH formation outcomes by direct collapse sets in.

What changes with Z and MT pre-history are boundaries $M_{\text{CO}}^{(1)}$, $M_{\text{CO}}^{(2)}$ and $M_{\text{CO}}^{(3)}$.

Differences in the M_{CO} dependence of final fates among single and stripped stars: Comparing the final fates of single and binary-stripped stars at $Z = Z_{\odot}$ and at $Z = Z_{\odot}/10$ over the M_{CO} range (see Table 3..3, Table 3..4 and Fig. 3..5), the three most important conclusions are the following:

1. The boundaries $M_{\text{CO}}^{(1)}$, $M_{\text{CO}}^{(2)}$ and $M_{\text{CO}}^{(3)}$ shift systematically toward lower values as Z decreases from Z_{\odot} to $Z_{\odot}/10$.
2. The need to discriminate between single and stripped star pre-SN progenitors is apparent: for example, at $Z = Z_{\odot}$, the BH formation window by direct collapse is within $(6.6, 7.1) M_{\odot}$ for case C donors, while it is $(7.4, 8.4) M_{\odot}$ for case A donors, i.e. these do not even overlap.
3. The critical M_{CO} values of single stars / case C donors differ more substantially from those of case A/B donors.

We further observe the following trends:

- the direct¹⁰ BH formation window widens for case C, case B1 and case Be donors, as Z decreases.
- At Z_{\odot} , $M_{\text{CO}}^{(1)}$ and $M_{\text{CO}}^{(2)}$ of single stars and case C stripped stars are almost the same, however they differ at $Z_{\odot}/10$. As Z decreases from Z_{\odot} to $Z_{\odot}/10$, $M_{\text{CO}}^{(3)}$ of case C stars lowers by more than a solar mass. This stands in contrast to single stars, whose $M_{\text{CO}}^{(3)}$ remains unaffected.
- The direct BH formation window of case A/B donors shifts toward larger M_{CO} values by $0.4 - 1.3 M_{\odot}$ compared to single stars / case C donors, depending on Z and on the mass transfer case. At Z_{\odot} , the plateau of direct BH formation outcomes sets in at significantly larger $M_{\text{CO}}^{(3)}$ (by $2 M_{\odot}$ or more) in case A/B donors compared to single stars / case C donors.
- $M_{\text{CO}}^{(3)}$ is greater, the earlier the hydrogen-rich envelope is removed.

¹⁰To shorten the terminology, in the following we use the terms "direct BH" and "fallback BH". For direct BH formation, failure of shock revival and implosion of the core without a supernova (failed supernova) is assumed. For fallback BH formation, success of shock revival is assumed, which launches a supernova but leaves a BH behind once the PNS collapses following the accretion of fallback material. This classification allows to decouple the question of a star's final fate from that of its compact remnant mass, which is an additional degree of freedom associated with uncertainties. A scheme to determine the compact remnant mass inevitably needs to assume whether most of the hydrogen-rich envelope or only the helium core fall into the BH formed by direct collapse, and what fraction of the ejecta mass falls back onto the PNS when a fallback BH is formed. Such a scheme is not subject of this work.

- The direct BH formation boundaries of case Be and case Bl systems are almost the same at Z_\odot and at $Z_\odot/10$, except that $M_{\text{CO}}^{(3)}$ is greater by half a solar mass for case Bl compared to case Be donors at $Z = Z_\odot/10$.

Among the stellar types considered, the direct BH formation window is widest for case A donors at Z_\odot (of $1 M_\odot$ width), whereas it is (together with that of single stars) the narrowest at $Z_\odot/10$ (of $0.5 M_\odot$ width). Direct BH formation windows of all stellar types taken together cover a range from 6.1 to $8.4 M_\odot$.

S21 versus H16 single star models: We further study the systematics behind differences in final fate outcomes based on a comparison of H16 to S21, which adopt a different input physics for the advanced burning phases. H16 comprises a densely sampled grid of single-star models over $M_{\text{CO}} \in (2, 15) M_\odot$. It therefore is ideal to study phenomenologically the distribution of remnant types (NS, direct BH, fallback BH). Fig. 3.6 shows their distribution over M_{CO} . In contrast to the S21 set of single-star models, there is no window at intermediate M_{CO} values (i.e., around the first peak in $\xi_{2.5}$) over which BHs are predicted to form. The first peak in $\xi_{2.5}$ is not large enough to surpass $\xi_{2.5}^{\text{max}}$ for having direct BH formation guaranteed. With the exception of one model at $M_{\text{CO}} \simeq 2.8 M_\odot$, all pre-SN progenitors up to a threshold value $M_{\text{CO}} \simeq 9 M_\odot$ explode, and BHs form by direct collapse for values beyond. The plateau of direct collapse outcomes sets in at a much greater value of $M_{\text{CO}}^{(3)} = 12.9 M_\odot$ in S21. In H16, $\xi_{2.5}$ peaks at $M_{\text{CO}} \simeq 4.8 M_\odot$, with a $\xi_{2.5}$ value substantially lower and the peak position in M_{CO} shifted to a lower value by roughly $2 M_\odot$ compared to S21. Few fallback BHs form near the $\xi_{2.5}$ peak and close to the transition to the plateau of direct BH outcomes. In Fig. 3.7, we further compare the M_{ZAMS} -to- M_{CO} relations for the H16 and the S21 single stars. Single star models starting from the same M_{ZAMS} can result in remarkably different M_{CO} values at the end of CHeB. M_{CO} and M_{ZAMS} can therefore not be used interchangeably for parametrizing explodability.

Guaranteed NS formation at large M_{CO}

After having mapped out the final fates of single and binary-stripped stars in Sect. 3.3.3., as a next step, we investigate the type of the compact remnant left behind a successful SN, which is either a NS or a fallback BH. Two important conclusions can be drawn from the remnant type distribution over the densely sampled exploding models in the H16 set (see Fig. 3.6):

1. Fallback BH formation is not randomly distributed over the M_{CO} range but is clustered, leaving regions of NS formation in-between the clusters.
2. Fallback BH remnants co-exist with NS remnants over clustered M_{CO} ranges.

The M16 model with parameter choice from S21 predicts NS formation at large $M_{\text{CO}} > 8 M_\odot$ over both H16 and S21, which CCSN recipes such as those introduced in Fryer et al. (2012) and Mandel and Müller (2020) do not allow for.

We inquire whether there is systematics with M_{CO} behind the clustered occurrence of the rarer fallback BH formation outcomes in CCSN explosions. To this end, we evaluate our probabilistic criterion defined in Sect. 3.3.1. for fallback BH formation

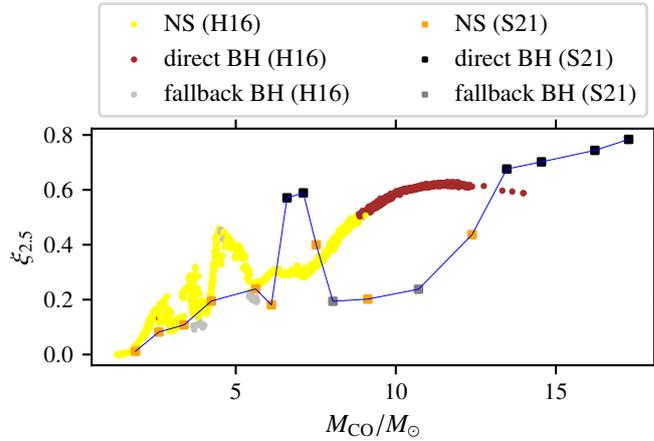


Figure 3.6: The CCSN outcomes over the Z_{\odot} single star pre-SN progenitor models S21 are compared to those over the Z_{\odot} single star pre-SN progenitor models H16. Peaks in the compactness are shifted systematically towards higher M_{CO} in S21 compared to H16. The compactness peaks are at $\simeq 5 M_{\odot}$ in H16 but at $\simeq 7 M_{\odot}$ in S21, and at $\simeq 9 M_{\odot}$ in H16 but at $\simeq 13 M_{\odot}$ in S21.

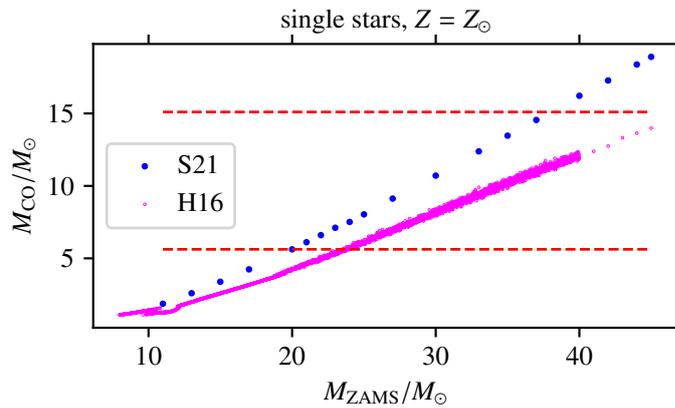


Figure 3.7: Comparison of $M_{\text{CO}}(M_{\text{ZAMS}})$ relations of S21 to H16 for single stars at $Z = Z_{\odot}$ over a ZAMS mass range up to $45 M_{\odot}$. For the same ZAMS mass, the H16 models result in a substantially lower M_{CO} compared to S21. The red horizontal lines indicate the M_{CO}^{\min} and M_{CO}^{\max} boundaries.

based on the pre-SN variables $\xi_{2.5}$, $\mu_4 M_4$, M_4 and M_{CO} . In a successful SN explosion, when the criterion for fallback BH formation based on these variables is not satisfied, the compact remnant is guaranteed to be a NS. The variables $\xi_{2.5}$, $\mu_4 M_4$ and $M_4 = \mu_4 M_4 / \mu_4$ all show bimodal trends with M_{CO} and sharply decrease for $M_{\text{CO}} > M_{\text{CO}}^{(2)}$. For probability of the occurrence of fallback according to our pre-SN based criterion, it is the difference in the slopes at which these quantities decrease (increase) compared to one another what decides the remnant type. Tables 3.5 and 3.6 summarize the widest intervals δM_{CO} for which NS formation is guaranteed at $Z = Z_{\odot}$ and at $Z = Z_{\odot}/10$, over the range $M_{\text{CO}}^{(2)} < M_{\text{CO}} < M_{\text{CO}}^{(3)}$.

In our sample of the S21 and S23 stellar models, we only find pre-SN progenitors that leave fallback BHs behind for $M_{\text{CO}}^{(2)} < M_{\text{CO}} < M_{\text{CO}}^{(3)}$ (see right panel in Fig. 3.10). Stellar models that fulfill the criterion form fallback BHs with a frequency of 0.15, which we interpret as a probability $P = 0.15$. The probability has an objective and a subjective origin: first, we expect that the map from M_{CO} to the remnant type is only partially bijective (i.e. allowing for co-existence of NSs and fallback BHs), and second, we are ignorant of the precise location and width of the window over which fallback BHs are expected to be constrained to cluster.

Table 3.5: Critical values in M_{CO} for NS formation at $Z = Z_{\odot}$ from single and binary-stripped star SN progenitors¹². $M_{\text{CO}}^{\text{NS},1}$ and $M_{\text{CO}}^{\text{NS},2}$ indicate the width of the windows over which stars are expected to explode and to leave NSs behind. Outside these intervals, the compact remnant for exploding stars with $M_{\text{CO}}^{(2)} < M_{\text{CO}} < M_{\text{CO}}^{(3)}$ is either a NS or a fallback BH.

	$M_{\text{CO}}^{\text{NS},1}/M_{\odot}$	$M_{\text{CO}}^{\text{NS},2}/M_{\odot}$
single	9	10.2
case C	9.6	10.7
case B	9.9	10.3
case A	11.1	12.1

Table 3.6: Same as Table 3.5, but at $Z = Z_{\odot}/10$.

	$M_{\text{CO}}^{\text{NS},1}/M_{\odot}$	$M_{\text{CO}}^{\text{NS},2}/M_{\odot}$
single	7.4	11
case C	8.9	9.5
case B	9.3	10.3
case A	10.4	11.1

For the same MT pre-history class, the guaranteed NS formation windows in M_{CO} shift systematically toward larger values as Z increases. For the same Z , the earlier MT sets in, the larger is the value range in CO core mass over which NSs are predicted to form.

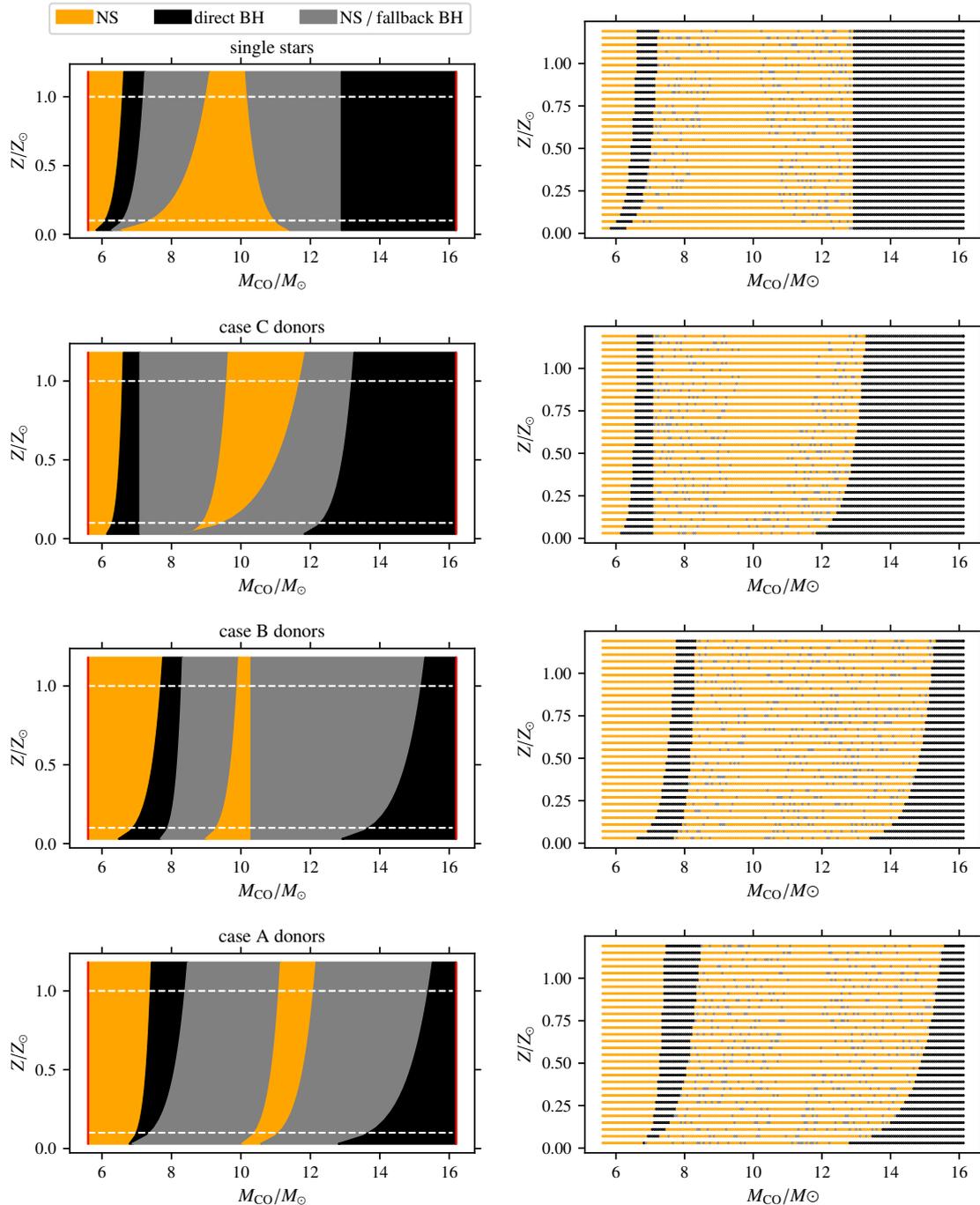


Figure 3.8: Left panels: Regions of BH formation by direct collapse (in black), of NS formation (in orange) and of the co-existence of fallback BH and NS remnants (in grey) in the stellar parameter space spanned by M_{CO} and Z for each MT pre-history class (none, case C, case B and case A). The M_{CO} range is limited to within $M_{\text{CO}} \in (M_{\text{CO}}^{\min}, M_{\text{CO}}^{\max})$ (borderlines in red), outside of which only implosions and explosions, respectively, are predicted to occur. Right panels: Statistical realization of our probabilistic fallback BH formation recipe within the M_{CO} intervals as generalized by the $M_{\text{CO}}^{\text{NS},i} \propto \log(Z/Z_{\odot})$ model for $i = 1, 2$. The boundaries demarcate windows inside of which NS formation is guaranteed. Outside of these, fallback BH formation is predicted to occur at a probability $P = 0.15$ over $M_{\text{CO}}^{(2)} < M_{\text{CO}} < M_{\text{CO}}^{(3)}$.

CCSN recipe for rapid BPS

We construct a CCSN recipe that retains the distinction between single and binary-stripped stars and a dependence on Z . To get a first-order estimate of how BH formation boundaries we derived in Sect. 3.3.3. and 3.3.2. for Z_\odot and $Z_\odot/10$ depend on Z more generally, we assume a linear model in $\log Z$ for scaling the boundaries

$$M_{\text{CO}}^{(i)}(Z)/M_\odot = a_i + b_i \cdot \log Z/Z_\odot \quad (3..12)$$

for $i = 1, 2, 3$ and each MT pre-history class.

For each MT case, we determine a separate set of parameters (a_i, b_i) that fit the model defined by eq. (3..12) to connect the two known critical value points $(M_{\text{CO}}^{(i)}(Z = Z_\odot), M_{\text{CO}}^{(i)}(Z = Z_\odot/10))$ for $i = 1, 2, 3$ at $Z = Z_\odot$ and at $Z = Z_\odot/10$, respectively, by a curve (see Appendix B2. for explicit formulae how these are calculated analytically). The same procedure is applied to the fallback BH formation boundaries set by $M_{\text{CO}}^{(\text{NS},1)}$ and $M_{\text{CO}}^{(\text{NS},2)}$ in-between $M_{\text{CO}}^{(2)}$ and $M_{\text{CO}}^{(3)}$. Since we do not encounter BH formation by fallback in our S21 and S23 samples for $M_{\text{CO}} < M_{\text{CO}}^{(1)}$, we infer that these are statistically insignificant and – for the CCSN recipe – assume that only NSs form. Fig. 3..8 shows the outcome for single and binary-stripped stars. The prediction of the occurrence of fallback is limited by both systematic uncertainties about CCSN outcomes and the restricted modeling approach using global parameters given at the time of evolutionary cut-off alone. With the exception of single stars and case C donors at Z_\odot , the guaranteed NS formation windows are rather narrow. We therefore construct a second, even simpler probabilistic fallback model for rapid BPS, which we designate as model B. For model B, we assume a uniform probability of 10% for the occurrence of fallback BH formation in-between $M_{\text{CO}}^{(2)}$ and $M_{\text{CO}}^{(3)}$. This probability is assumed given the relative frequencies for the occurrence of fallback BH formation from successfully exploding stellar models:

- 8.5% over the entire set of pre-SN progenitors (S21, S23, S24, T24, H16),
- 11.5 % over single and binary stripped stars S21, S23, T24, for which the same physics for the late burning stages has been assumed.

These frequencies are similar despite differences in adopted stellar evolution physics and resulting pre-SN properties. We thus coarse-grain over these differences and assume a value of 10%.

The resulting final fate landscape of single and binary-stripped stars shows a recurrent pattern: a direct BH formation window is followed by an interval permitting co-existence of fallback BHs and NSs, then by a NS formation window at even higher M_{CO} , which again is followed by an interval with co-existence of fallback BH and NS outcomes, until the plateau of BH formation outcomes is reached.

This remnant type distribution follows directly from the bimodal dependence of explodability proxies $\mu_4 M_4$, μ_4 , $\xi_{2.5}$ and s_c on M_{CO} (see Fig. 3..5), and has its shifts in critical M_{CO} value ranges depending on Z and MT pre-history.

Finally, with the fitted parameters of the assumed scaling law given by eq. (3..12), we obtain a CCSN recipe that predicts final fates (failed or successful SN) and remnant type (NS or BH) for given M_{CO} , Z and assignment of MT pre-history class.

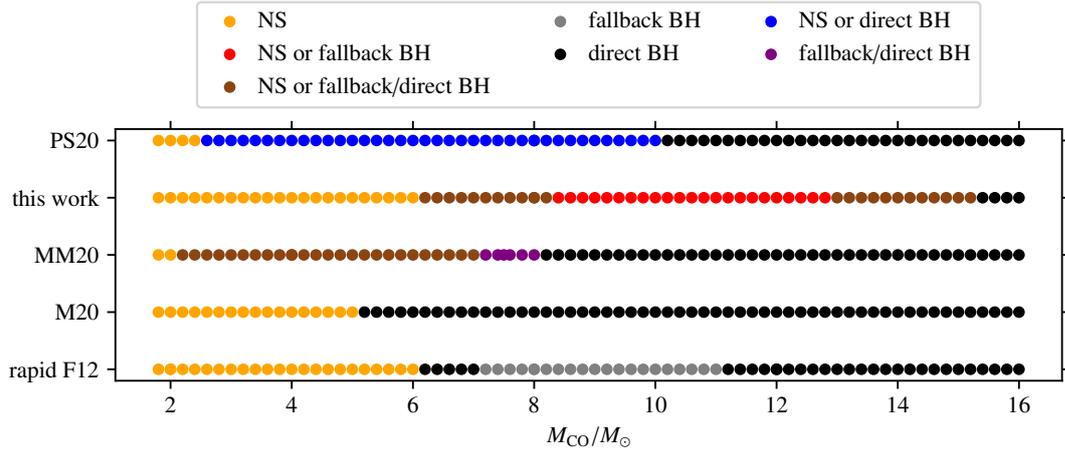


Figure 3..9: Cross-comparison of CCSN outcomes predicted by M_{CO} -based CCSN recipes for binary population synthesis.

The source code behind its construction is applicable for rapid BPS studies and available upon reasonable request to the corresponding author. Usage of this CCSN recipe beyond $Z \in (Z_{\odot}/10, Z_{\odot})$ is extrapolation. The recipe breaks down at a cut-off $Z_{\text{cut-off}} < Z_{\odot}/10$ presumably higher than zero (Population III stars). The CCSN recipe and the pre-SN explodability criterion formulated in Sect. 3.3.1. are based on the same CCSN model. However, the former, i.e. the final fate parametrization scheme as function of M_{CO} loses the universality property, since the critical values in M_{CO} become dependent on adopted physics of the late burning stages that evolve the star carbon burning onward up to the pre-SN stage, and since we find different critical M_{CO} value ranges depending on stellar MT pre-history and metallicity.

Comparison with other M_{CO} -based CCSN recipes

The following more general conclusions can be drawn from the final fate landscapes of single and binary-stripped stars at $Z \geq Z_{\odot}/10$ that result from our predictive framework, regardless of Z and MT pre-history:

- for $M_{\text{CO}}/M_{\odot} < 6.1$, only NSs form,
- for $M_{\text{CO}}/M_{\odot} \in (6.1, 15.4)$, NS, direct BH, and fallback BH remnants co-exist,
- for $M_{\text{CO}}/M_{\odot} \in (8.4, 12.4)$, direct BH formation is excluded and a successful SN explosion guaranteed, leaving NSs or fallback BHs behind, and
- for $M_{\text{CO}}/M_{\odot} > 15.4$, only direct BHs form.

This final fate parametrization using M_{CO} differs substantially from others that have commonly been used in binary population synthesis codes and that adopt a similar approach of predicting CCSN outcomes based on M_{CO} . In Fig. 3..9 we compare the CCSN outcomes predicted by our recipe to those by Mandel and Müller (2020) (MM20), Mapelli et al. (2020) (M20), Fryer et al. (2012) (F12) and Patton

and Sukhbold (2020) (PS20). In what follows, we summarize each of these CCSN recipes and discuss in more detail how these compare to ours.

MM20: In MM20, the remnant mass and type is assigned using probabilistic formulae based on calibrated threshold values M_i , with $i_1, \dots, 4$, in M_{CO} . Core collapse is predicted to result in a successful SN that leaves only NSs behind for $M_{\text{CO}} \leq M_1$. BHs (formed either by fallback or directly) and NSs co-exist for $M_1 < M_{\text{CO}} \leq M_3$. For $M_3 < M_{\text{CO}} \leq M_4$, BH formation, either directly or by fallback, is guaranteed. For $M_{\text{CO}} > M_4$, only direct BHs form. The default threshold values are $M_1 = 2 M_{\odot}$, $M_3 = 7 M_{\odot}$ and $M_4 = 8 M_{\odot}$. MM20 is used as CCSN recipe in COMPAS, for example.

While both MM20 and our CCSN recipe are constructed based on outcomes of the M16 SN codes, there are two main differences. The first concerns the parameter choice for the M16 SN code: in MM20 the default settings from M16 are used, except for a higher accretion efficiency ($\zeta = 0.8$) and a different calculation of the final mass cut in case of BH formation by fallback in a successful CCSN explosion. The second difference concerns the pre-SN progenitor models: MM20 is based on the H16 single star models at $Z = Z_{\odot}$. A randomized scheme is used that follows gross trends of compact remnant mass M_{rem} with M_{CO} .

M20: According to M20, there is no co-existence of BHs and NSs for the same M_{CO} . Only NS form up to a critical CO core mass $M_{\text{CO}}^{\text{crit}}$, and only direct BHs form for more massive CO cores.¹³ M20 differs from our framework in three principal regards. First, the stellar progenitors considered in M20 are the single star models from Limongi and Chieffi (2018) evolved from ZAMS up to the onset of iron core infall over a parameter space spanned by M_{ZAMS} , Z and initial rotation v_{ini} . Second, as pre-SN explodability criterion, $\xi_{2.5}$ is used in M20. Third, in order to relate CCSN outcomes to M_{CO} , in M20 the $\xi_{2.5}$ values over the sampled parameter space in $(M_{\text{ZAMS}}, Z, v_{\text{ini}})$ are stacked together, and then fitted as function of M_{CO} with a monotonically increasing parametric power-law model. The critical compactness value $\xi_{2.5}^{\text{crit}}$ for BH formation is a free parameter in M20. Setting the threshold for BH formation to $\xi_{2.5}^{\text{crit}} = 0.3$ as is suggested in M20, results in $M_{\text{CO}}^{\text{crit}} = 4.4 M_{\odot}$. With greater threshold values $\xi_{2.5}^{\text{crit}} \in [0.32, 0.33, 0.37, 0.45]$, the resulting critical CO core masses are $M_{\text{CO}}^{\text{crit}}/M_{\odot} \in [4.8, 5, 6.1, 11]$.

M20 and MM20 both are based on stellar evolution tracks that exhibit a weak compactness peak ($\xi_{2.5}^{\text{max}} < 0.45$) at intermediate $M_{\text{CO}} < 7 M_{\odot}$, which does not reach the upper threshold for direct BH formation (see Fig. 1 in Mapelli et al. (2020) for M20, and Fig. 3..6 for MM20), according to our pre-SN based criterion. The CCSN recipes are indirectly testable by comparison to observations of compact remnant masses. M20 and MM20 both do not predict the presence of a BH-BH (“upper”) mass gap, because direct BH formation outcomes are not interrupted after having set in at sufficiently large M_{CO} . Our CCSN recipe is compatible with a BH-BH mass gap, since there is a SN window between $M_{\text{CO}} \in (8.4, 12.4) M_{\odot}$ independent of MT pre-history and $Z > Z_{\odot}/10$, over which the expected outcome is a NS or a fallback

¹³In context of the BPS code SEVN (Iorio et al., 2023), a fallback BH formation window is inserted over M_{CO} values in-between the NS and direct BH outcomes.

BH of lower mass than one that would arise from a direct BH at the same M_{CO} . The default M20 model (without fallback) predicts a NS-BH (“lower”) mass gap because of a sharp transition between remnant types at a critical $M_{\text{CO}}^{\text{crit}}$ and because BH progenitors with lower M_{CO} are weakly affected by stellar winds (Mapelli et al., 2020). MM20 does not predict a lower mass gap, since over the $M_{\text{CO}} \in (2, 7) M_{\odot}$ range, direct BH, fallback BH and NSs co-exist.

F12: In Fryer et al. (2012), a recipe is formulated for computing compact remnant masses. The original recipe uses parametric fits to model remnant masses as function of ZAMS mass and metallicity of single-stars. It is constructed based on hydrodynamical simulations and stellar progenitors at two reference metallicities (solar and Population III) from Woosley et al. (2002). The original recipe is then reformulated as a function of M_{CO} and final mass, to account for differences in outcomes due to different assumptions about wind mass loss and binary mass transfer. It is assumed that the amount of fallback onto the PNS (of fixed mass $\simeq 1 M_{\odot}$) depends on the timing of the explosion (a “fast-convection” explosion, if it happens within 250 ms after core bounce; a “delayed-convection” explosion otherwise), since the accretion rate of the infalling material decreases with time and therefore also the total kinetic energy stored in the convective region between the PNS and the base of the shock. The fast-convection (“rapid”) explosion model and the delayed-convection (“delayed”) explosion models do not explicitly distinguish the remnant type. However, the remnant type can be distinguished implicitly in the rapid model, since it predicts a lower mass gap. According to the rapid F12, there is no co-existence of NSs and BHs for the same M_{CO} . For $M_{\text{CO}} < 6 M_{\odot}$, only NSs form. For $M_{\text{CO}} \in (6, 7) M_{\odot}$ and $M_{\text{CO}} > 11 M_{\odot}$, only direct BHs form. For $M_{\text{CO}} \in (7, 11) M_{\odot}$, stars are predicted to explode and leave only fallback BHs behind, with a fallback mass fraction that increases with M_{CO} .

Comparison to 3D CCSN simulation outcomes: The MM20, M20 and F12 recipes are less consistent with 3D CCSN simulation outcomes than the new recipe presented in this work. For example, the explosion (in the 3D simulation) of the $M_{\text{CO}} = 6.07 M_{\odot}$ single star $Z = Z_{\odot}$ model s24 from the Monash group is consistent with our recipe but not with rapid F12. The explosion in 3D of the high $M_{\text{CO}} = 8.2 M_{\odot}$ but low $\xi_{2.5} = 0.22$ binary-stripped star y20 is not consistent with the default upper mass limit for explosions (M_4) in MM20, and with the $\xi_{2.5}$ -to- M_{CO} relation assumed in M20. The explosion in the 3D simulation of Population III star z40 at $M_{\text{CO}} = 12.9 M_{\odot}$ is consistent with our CCSN recipe (when extrapolated to $Z = 0$), but not with any of the others mentioned.

PS20: In PS20, the final fate dependence on the starting point in (M_{CO}, X_C) plane at zero age core carbon burning has been studied systematically. The final fates are predicted by applying E16¹⁴ to the pre-SN profiles that are obtained after evolving bare CO cores through the late burning phases over a densely sampled grid in the

¹⁴E16 is the default criterion suggested by PS20. However, other explodability criteria can equally be extracted from the pre-SN profiles, such as the $\xi_{2.5}$ -based final fate determination.

(M_{CO}, X_C) parameter space. The motivation for evolving bare CO cores is that after the end of CHeB, the evolution of the core and that of the envelope are largely decoupled. The envelope re-structures itself on the thermal timescale and has little time to re-adjust to the core whose evolution speeds up after core helium burning due to the enhanced neutrino losses - it only takes a few thousands of years from carbon ignition up to collapse. Final fates are looked up given $(M_{\text{CO}}$ and $X_C)$, and the grid base interpolated over to get predictions at arbitrary values of interest within the grid boundaries. When assuming E16 as explodability criterion, the gross trend is that at high X_C and low M_{CO} , explosions dominate, while it is implusions at low X_C and high M_{CO} . At the grid boundaries in M_{CO} , the most frequent outcomes are explosions over all X_C at the low-mass end of $M_{\text{CO}} = 2.5 M_{\odot}$ and implusions over the majority of X_C at the high-mass end of $M_{\text{CO}} = 10 M_{\odot}$ (see Fig. 3.12). For population synthesis purposes, it therefore has been suggested to assume that all stars with $M_{\text{CO}} < 2.5 M_{\odot}$ explode and all stars with $M_{\text{CO}} > 10 M_{\odot}$ implode (Patton et al., 2022). Implusions and explosions may in principle co-exist over the entire range from 2.5 to 10 M_{\odot} , provided that X_C is suitably chosen.

Explodability dependence on X_C

Given a fixed M_{CO} , we inferred that the final fate of a given pre-SN progenitor may differ, depending on its Z and its MT pre-history. We find systematics behind these differences by taking into account X_C at the end of core helium burning. Fig. 3.10 shows the X_C -to- M_{CO} dependence of single and binary-stripped stars at Z_{\odot} and at $Z_{\odot}/10$, respectively. For the same M_{CO} and MT pre-history, as Z increases, X_C increases. This is because with greater Z , the helium core is less massive due to its reaction to stronger wind mass loss from the envelope. In a less massive helium core, the core temperature is lower and the $^{12}\text{C}(\alpha, \gamma)^{16}\text{O}$ reaction sets in later during core helium burning. This leaves more carbon in the core at core helium exhaustion. With a higher X_C for the same M_{CO} , more nuclear fuel is available during the relatively long-lasting carbon burning phase. A higher X_C shifts the peaks in explodability proxies such as $\xi_{2.5}$ toward larger M_{CO} values (Patton and Sukhbold, 2020; Chieffi and Limongi, 2020). At a critically low X_C , the dependence of explodability proxies on M_{CO} flattens and the peak structure vanishes altogether.

Single stars and case C donors follow similar tracks in the (M_{CO}, X_C) plane at Z_{\odot} and at $Z_{\odot}/10$, respectively. The same applies to the $X_C(M_{\text{CO}})$ tracks of case A and case B donors. The differences are noteworthy only at $M_{\text{CO}} < 2 M_{\odot}$ and at $M_{\text{CO}} > 15 M_{\odot}$. For the same M_{CO} and Z , Case A/B donors have a higher X_C compared to single stars/case C donors. Effects of case A/B MT change the core structure of the donor because before core carbon burning, the core and envelope evolution are not yet decoupled. For the same metallicity, earlier removal of the hydrogen-rich envelope leads to a lower helium core mass, as a response to the mass loss. A less massive helium core leads to a cooler core temperature during CHeB. This results in a greater X_C , since the $^{12}\text{C}(\alpha, \gamma)^{16}\text{O}$ reaction that uses up carbon sets in at hotter temperatures than the Triple- α reaction that produces it.

The starting points in (M_{CO}, X_C) plane set the neutrino burning conditions for the late burning stages, which pre-determine the explodability of the pre-SN progenitors.

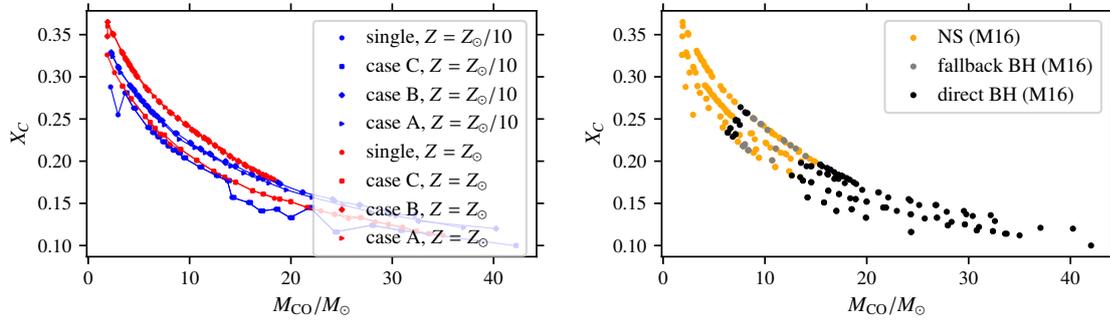


Figure 3.10: Left panel: $X_C(M_{CO})$ tracks of single and binary-stripped stars (case C, case B and case A donors) at Z_\odot and at $Z_\odot/10$, respectively. Right panel: CCSN outcomes of single and binary-stripped stars at Z_\odot and at $Z_\odot/10$ in the (M_{CO}, X_C) plane.

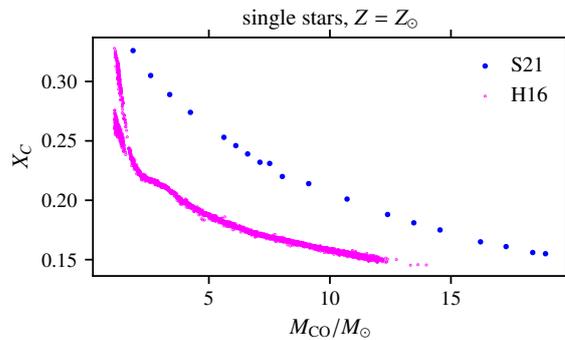


Figure 3.11: The lower X_C abundance in the H16 set compared to the S21 set, for the same M_{CO} mass. It explains the shift of the compactness peaks toward lower values in H16 compared to S21, and the resulting differences in the CCSN outcomes over the Z_\odot single star models (see Fig. 3.6).

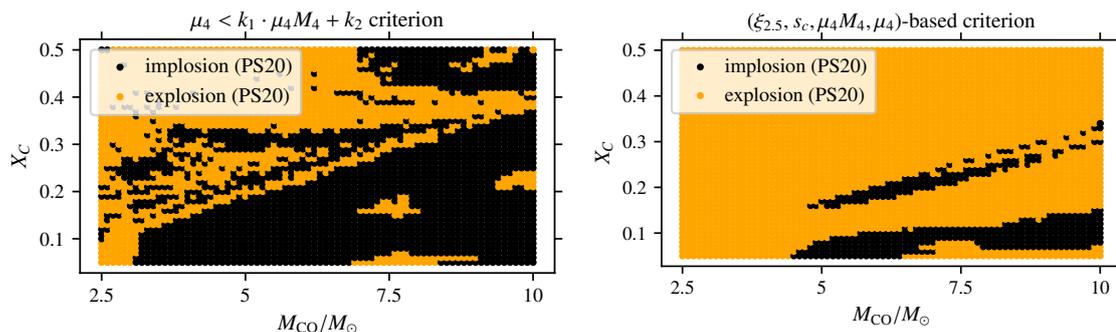


Figure 3.12: Final fate landscapes resulting from E16 (left panel) and from our pre-SN based explodability criteria (right panel), when applied to the pre-SN profiles over the PS20 grid in the (M_{CO}, X_C) plane. The parameters (k_1, k_2) of the separation line for the E16 criterion are calibrated to the updated W20 engine from [Ertl et al. \(2020\)](#). The explodability criterion introduced in this work is more optimistic about successful explosions than E16, and leads to a segmented final fate landscape. The latter features two islands of direct BH formation over a similar M_{CO} range but different value ranges in X_C . In contrast to the landscape resulting from the E16 criterion, in the one resulting from our criterion the plateau of BH formation outcomes is not yet in sight at $M_{\text{CO}} = 10 M_{\odot}$.

The systematic shifts (see right panel of Fig. 3.10) of direct BH formation boundaries in M_{CO} with increasing Z and when comparing those of single star/case C systems to those of case A/B windows can thus be reverted to differences in X_C for the same M_{CO} .

More generally, differences in the adopted stellar evolution physics up to the end of CHeB will manifest themselves in differences in X_C -to- M_{CO} relations (e.g. [Chieffi and Limongi, 2020](#); [Patton and Sukhbold, 2020](#); [Schneider et al., 2021](#); [Temaj et al., 2024](#)). For example, for the same model (S21) of the late burning phases, increased core overshooting smoothly shifts the $X_C(M_{\text{CO}})$ tracks for the same X_C towards larger M_{CO} values (see Fig. 3 in [Temaj et al. \(2024\)](#)). Likewise, the difference in the predicted final fate landscape of the H16 versus the S21 single stars as a function of M_{CO} can be traced back to a lower X_C for the same M_{CO} in H16 compared to S21 (see Fig. 3.11).

While we relate the differences in CCSN outcomes among single and binary-stripped stars and with variable Z to a higher X_C in binary-stripped stars compared to single stars and to a higher X_C with increasing Z , rapid BPS codes do not track the X_C variable. BPS codes that do keep track the X_C variable can use the PS20 grid to look up the final fate. When the pre-SN models over the PS20 grid are assigned a final fate by using the E16 criterion, the final fate is sensitive to the location in (M_{CO}, X_C) plane, featuring a landscape that has explosion islands in implosion-dominated regions and vice-versa (see left panel of Fig. 3.12). Note that while E16 fails to reproduce the 3D CCSN simulation outcomes (see right panel of Fig. 3.3), it is the pre-SN criterion that has recently been used in BPASS ([Patton et al., 2022](#)) and POSYDON ([Fragos et al., 2023](#)) BPS codes over the PS20 grid. To highlight the

differences between our¹⁵ and the E16 pre-SN criteria, we compute the resulting final fate landscapes over the same grid of PS20 pre-SN progenitors (see Fig. 3.12). As is evident from Fig. 3.12, applying E16 over PS20 yields an implosion-dominated final fate landscape at the grid boundary $M_{\text{CO}} = 10 M_{\odot}$ for most X_C values. In contrast, our pre-SN criterion over PS20 results in a decay of the direct BH formation stripe at $M_{\text{CO}} \simeq 9 M_{\odot}$ and a final fate landscape that is dominated by explosions except at low X_C . The parameter space width in M_{CO} is not wide enough to assess when and whether at all – according to our pre-SN criterion – the final fate landscape becomes implosion-dominated regardless of X_C as M_{CO} increases beyond $10 M_{\odot}$. Over the S21 and S23 tracks in (M_{CO}, X_C) , no explosions are found for $M_{\text{CO}} > 15.4 M_{\odot}$, however this statement is valid only for X_C roughly within 0.1 and 0.2.

We defer the development of a final fate look-up table over an extended parameter space in M_{CO} with arbitrary starting points in (M_{CO}, X_C) plane to future work. In order to make final fate predictions over the (M_{CO}, X_C) parameter space with our explodability criterion, the $(\xi_{2.5}, s_c, \mu_4 M_4, \mu_4)$ variables necessary for its evaluation are readily available in the PS20 data base and can be interpolated over in the (M_{CO}, X_C) for $M_{\text{CO}} \leq 10 M_{\odot}$ plane in the same way as is done at present with the $(\mu_4 M_4, \mu_4)$ parameters for evaluating E16. For $M_{\text{CO}} > 10 M_{\odot}$, our CCSN recipe for rapid BPS (see Sect. 3.3.2.) can be used to look up $M_{\text{CO}}^{(3)}$ of single and binary-stripped stars.

3.3.3. Comparison with observations

In what follows, our CCSN recipe as well as the aforementioned ones (MM20, M20, F12, PS20) are tested against CCSN observations that allow to estimate the M_{CO} of pre-SN progenitors.

Type IIP SN progenitors and the missing RSG problem

Type IIP SNe are the most frequent type of observed transients in a volume-limited sample, and result from CCSNe of massive stars with hydrogen-rich envelopes. In a few dozen of cases, observational missions have directly imaged the explosion site years before the transient detection in nearby galaxies, using space- or ground-based telescopes. Pre-explosion imaging allowed to estimate photometric properties of Type IIP SN progenitors, in particular their effective temperature and bolometric luminosity, and confirm the expectation that these are red supergiants (RSGs) (Smartt, 2015). During the advanced burning phases, the bolometric luminosity of RSGs changes little, and is primarily set by the core mass. The CO core mass sets the inner temperature and density stratifications, and thereby also determines the burning rate of helium in the shell surrounding the CO core. This leads to a dependence of bolometric luminosity of pre-SN RSGs on the CO core mass. RSGs not only preserve a hydrogen-rich envelope up to collapse, but also retain a nearly constant CO core

¹⁵For our pre-SN criterion, we lift the M_{CO} -based condition for the distinction between failed and successful SNe and assign final fates based on the variables $\xi_{2.5}, s_c, \mu_4 M_4$ and μ_4 .

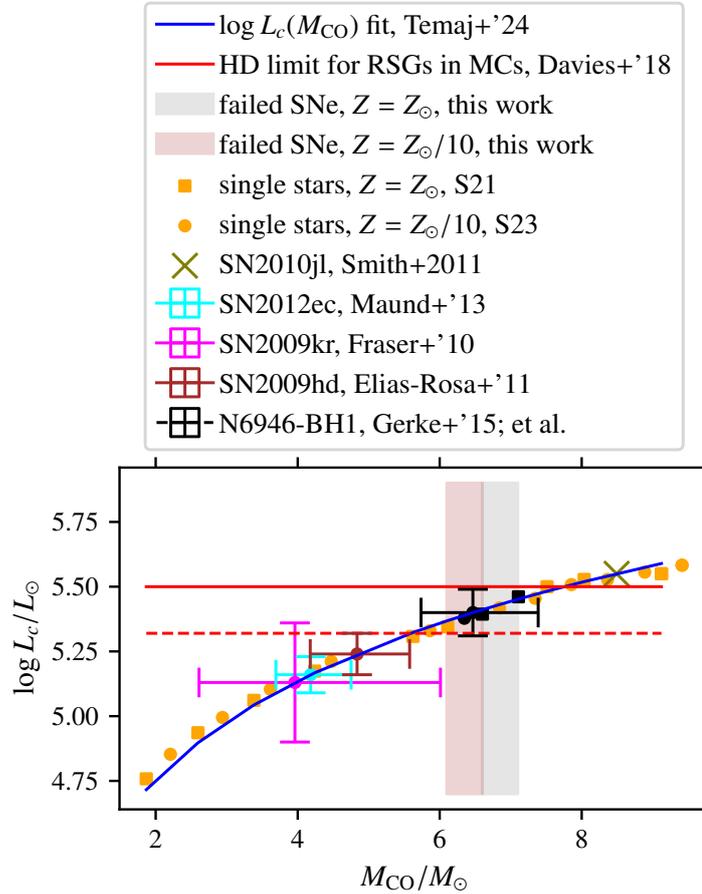


Figure 3.13: Inferred M_{CO} values (within uncertainty bounds) of the most luminous Type IIP SN progenitors SN2012ec, SN2009kr and SN2009hd, of the failed SN candidate N6946-BH1 and of the progenitor of the Type IIn SN2019jl, assuming that all these systems stem from the single star progenitor channel. Inference is performed using a parametric scaling law $\log L_{\text{TACCB}}(M_{\text{CO}})$ that relates bolometric luminosity $\log L_{\text{TACCB}}$ at terminal age core carbon burning to M_{CO} . Observations are compared with the direct BH formation windows predicted by our CCSN recipe model as function of Z , and the Humphreys-Davidson limit for Red Supergiants in the Magellanic Clouds, that together with the upper bound on the SN2009hd progenitor defines the value range in $\log L_{\text{TACCB}}$ (and M_{CO}) over which RSGs are missing out as observed Type IIP SN progenitors.

mass after formation at the end of CHeB¹⁶. In the present work, we use the empirical formula given by eq. (6) in [Temaj et al. \(2024\)](#) to estimate M_{CO} of the most luminous observed Type IIP pre-SN progenitors. To this end, we invert it in order to express the estimated CO core mass \hat{M}_{CO} ,

$$\hat{M}_{\text{CO}}/M_{\odot} = 10^{(\log L_{\text{pre-SN,obs}}/L_{\odot} - 4.372)/1.268} \quad (3.13)$$

as function of observed pre-SN bolometric luminosity $\log L_{\text{pre-SN,obs}}$.

According to [Davies and Beasor \(2018\)](#), the three most luminous observed Type IIP pre-SN progenitor are those of SN2009hd ([Elias-Rosa et al., 2011](#)), SN2012ec ([Maund et al., 2013](#)) and SN2009kr ([Fraser et al., 2010](#)). With its $\log L_{\text{pre-SN,obs}}/L_{\odot} = 5.24 \pm 0.08$, the progenitor of SN2009hd is estimated to have $\hat{M}_{\text{CO}} = 4.84_{-0.66}^{+0.76} M_{\odot}$. The Humphreys-Davidson (HD) limit of the most luminous RSGs that have been observed in the Magellanic Clouds is $\log L/L_{\odot} \simeq 5.5$ ([Davies et al., 2018](#)). The lack of observed Type IIP SN progenitors at higher luminosities up to the HD limit defines the problem of missing RSGs as Type IIP progenitors ([Smartt, 2009](#)). The luminosity of the brightest, (most likely) helium burning stars that define the HD limit is lower than that of pre-SN stars. Therefore, $\log L/L_{\odot} \simeq 5.5$ is in fact a lower bound on the upper boundary to the luminosity range over which RSGs are missing out as Type IIP SN progenitors. Assuming that the lower boundary is set by the upper limit $\log L/L_{\odot} = 5.32$ on SN2009hd, using eq. (3.13), the missing RSG problem can be reformulated as the lack of RSG progenitors of Type IIP SNe over $M_{\text{CO}} \in (5.6, 7.8) M_{\odot}$. One of the solution proposals is that RSGs over this range do not explode and instead collapse to form BHs quietly ([Smartt, 2009, 2015](#)). We test whether this hypothesis is compatible with our predictive models. Since our M_{CO} -based CCSN recipe for single stars predicts explosions for $M_{\text{CO}} < 6.1 M_{\odot}$ at $Z > Z_{\odot}/10$, the three most luminous Type IIP SNe cited in this work and observations of all fainter ones are all consistent with our predictive model (see Fig. 3.13). The integral BH formation window of single stars over all $Z > Z_{\odot}/10$ predicted by our model is contained within the value range in M_{CO} over which RSGs are indeed found to be missing. However, this failed SN window does not entail the intervals $\log L/L_{\odot} \in (5.32, 5.37)$ and $\log L/L_{\odot} \in (5.45, 5.5)$. Within our framework, the direct BH formation window over M_{CO} is of width $\simeq 0.4 - 0.5 M_{\odot}$ as it gradually slides from $(6.1, 6.6) M_{\odot}$ to $(6.6, 7.2) M_{\odot}$ while Z increases from $Z_{\odot}/10$ to Z_{\odot} . For example, this means that a star of $M_{\text{CO}} = 6.5 M_{\odot}$ is predicted to implode at $Z = Z_{\odot}/10$ but to explode at $Z = Z_{\odot}$. If RSGs happen to co-exist at variable Z over the $M_{\text{CO}} \in (6.1, 7.2) M_{\odot}$ range, then Type IIP are expected to be suppressed but observable. In other words, direct BH formation can be part of solution to the missing RSG problem by a suppressed SN rate, but there must be other physical reasons on top to add to it, in particular over the intervals $\log L/L_{\odot} \in (5.32, 5.37)$ and $\log L/L_{\odot} \in (5.45, 5.5)$.

The source N6946-BH1 is a failed SN candidate ([Gerke et al., 2015](#)), whose bolometric luminosity $\log L/L_{\odot} = 5.40 \pm 0.09$ imaged before disappearing in the optical ([Adams et al., 2017](#)) is within the luminosity range over which RSGs are missing out as SN progenitors. Assuming that it is indeed a failed SN¹⁷, it is consistent with the

¹⁶The CO core mass at the end of CHeB and at the onset of iron core infall has negligible variation not only for RSGs but over all the S21 and S23 progenitors (except at $M_{\text{CO}} > 18 M_{\odot}$).

¹⁷Based on observations of a luminous infrared source at the same sky location, recent work has

direct BH formation interval in M_{CO} predicted by our CCSN recipe. However, the observation is not constrained enough to confirm it. With $\log L/L_{\odot} = 5.4$, assuming that the progenitor is a RSG, $\dot{M}_{\text{CO}} \simeq 6.5 M_{\odot}$, which lies within the span of direct BH formation windows from $Z > Z_{\odot}/10$ up to slightly sub-solar. A higher $\log L/L_{\odot}$ of N6946-BH1 favours an implosion at higher Z , a lower one favours an implosion at lower Z . The value ranges $\log L/L_{\odot} > 5.45$ or $\log L/L_{\odot} < 5.36$ would be too bright or too faint to be inside any of our direct BH formation windows in M_{CO} between Z_{\odot} and $Z_{\odot}/10$.

Since MM20 admits failed and successful SNe up to $M_{\text{CO}} = 8 M_{\odot}$, it is consistent with observations of the most luminous Type IIP SN progenitors. It can partially address the missing RSG problem by direct BH formation because of an increasing probability for direct BH formation with increasing M_{CO} . In order to account for M_{CO} of the progenitor of SN2009hd, the compactness threshold in M20 needs to be placed at $\xi_{2.5} \geq 0.323$, which implies $M_{\text{CO}}^{\text{crit}} \geq 4.85 M_{\odot}$. Assuming that all stars with $M_{\text{CO}} > M_{\text{CO}}^{\text{crit}}$ implode, M20 can address the missing RSG problem by BH formation entirely. F12 is consistent with the most luminous Type IIP SN progenitors and can partially explain the missing RSG problem by BH formation over a similar range ($M_{\text{CO}}/M_{\odot} \in (6, 7)$) as our CCSN recipe. In contrast to the Z -dependence of the BH forming windows in M_{CO} of our CCSN recipe, F12 does not predict co-existence of failed and successful SNe over the $M_{\text{CO}}/M_{\odot} \in (6, 7)$ range, and therefore predicts the absence of Type IIP SNe over this M_{CO} range, while our model only predicts their suppression. M20 and MM20 are consistent with N6946-BH1 constituting a failed SN for any value within $\log L/L_{\odot} = 5.40 \pm 0.09$, whereas – similar to the constraint valid for our CCSN recipe – F12 requires it to be within $\log L/L_{\odot} = 5.40 \pm 0.04$.

PS20 coupled to E16 can explain the most luminous Type IIP SNe and the missing RSG problem by failed SNe provided that stellar models of RSGs at variable Z have $X_C(M_{\text{CO}})$ relations suitably low for passing through explosion/implosion sites.

When assuming PS20 as model for the late burning stages but using our pre-SN explodability criterion instead of E16 to map out final fates in the (M_{CO}, X_C) plane, tighter constraints are posed on stellar evolution models up to end of CHeB. In order to “land” on the implosion stripe at intermediate M_{CO} values in-between the M_{CO} range over which RSGs are found to be missing (see Fig. 3.14), stellar models need to have specific X_C values over M_{CO} intervals (and thereby $\log L_{\text{pre-SN}}/L_{\odot}$ value ranges) of interest.

Type IIn SN progenitors:

Type IIL SN progenitors: In the few cases of SNe IIL progenitor identification, no progenitor luminosity greater than the most luminous Type IIP of $\log L/L_{\odot} = 5.24 \pm 0.08$ has been estimated.

Type IIn SN progenitor channels and observations: In contrast, Type IIn SN progenitors as bright as $\log L_{\text{pre-SN}}/L_{\odot} > 6$ have been observed (Gal-Yam et al., 2007; Boian and Groh, 2018; Kankare, E. et al., 2015). Type IIn SNe are distinctive by

questioned this hypothesis and made other progenitor scenarios, such as a stellar merger event, plausible. (Beasor et al., 2024)

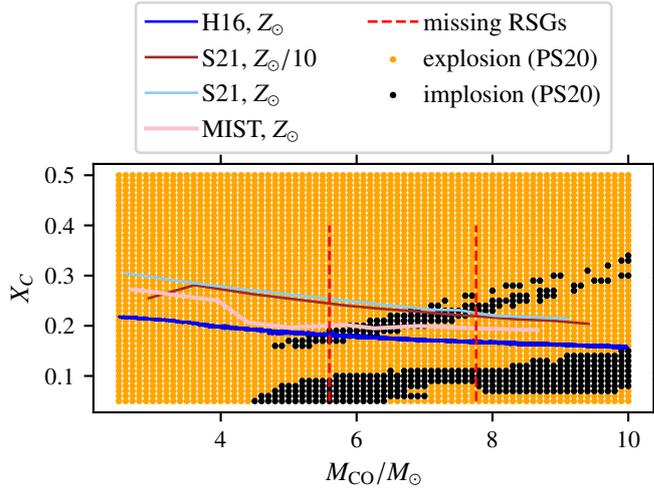


Figure 3.14: Final fates (failed or successful SNe) in (M_{CO}, X_C) plane, when applying the pre-SN explodability criterion introduced in this work to the pre-SN progenitors from PS20, and the constraints it poses on stellar evolution models for partially addressing the missing RSG problem by direct BH formation. The M_{CO} range over which RSGs are found to be missing is delineated in red. The H16, S21 and MESA ISOCHRONES AND STELLAR TRACKS (MIST; Choi et al., 2016) single star models lead to different $X_C(M_{\text{CO}})$ tracks through this plane.

narrow, bright multi-component hydrogen Balmer lines in the spectrum. These lines are attributed to interaction of the supernova with the circumstellar medium, which may have been formed by episodes of enhanced mass loss from the pre-SN progenitor. Progenitor channels of Type IIn SNe may be single stars, but also accretor stars and stellar merger products. The latter two categories are the more likely ones, given their explosion sites in the HR diagram (Justham et al., 2014; Schneider et al., 2024). In the accretor star scenario, a binary system is subject to stable mass transfer, wherein the accretor star accretes mass from the hydrogen-rich envelope of the donor star, and then explodes to produce a hydrogen-rich transient. The large $\log L$ values do not necessarily imply large M_{CO} for accretor stars, because luminosity in that case also is contributed by hydrogen-rich envelope mass and hydrogen shell burning luminosity, not only the helium burning shell whose burning conditions are largely set by M_{CO} .

Single star progenitor channel for Type IIn: However, not all Type IIn need to have accretor star or stellar merger progenitors. In the single star scenario, the progenitor is expected to have gone through a luminous blue variable (LBV) phase of enhanced mass loss outbursts, which however did not shed away the entire hydrogen-rich envelope by the time the explosion sets in. We explore consequences of the hypothesis that the Type IIn SN2010jl (Smith et al., 2011) is such a case. Its comparatively faint progenitor is inferred to have a bolometric luminosity of $\log L/L_{\odot} = 5.55$, and the observational data is consistent with a progenitor that has gone through a LBV phase. We find that the $\log L_{\text{TACCB}}(M_{\text{CO}})$ scaling law remains reliably applicable up to $\log L/L_{\odot} \simeq 5.7$, even though the stellar models then are not anymore

RSGs. From eq. (3.13) follows that assuming that the progenitor is a single star, its carbon-oxygen core is estimated to weigh $\hat{M}_{\text{CO}} = 8.5 M_{\odot}$. This value is inside the region over which our CCSN recipe predicts explosions of single and stripped stars independent of Z , and the same applies to F12. MM20 is challenged to explain a single star progenitor channel of SN2010jl, since it allows for explosions only up to $M_{\text{CO}} < 8 M_{\odot}$. Consistency requires lifting the M_4 threshold value up to which MM20 allows for the occurrence of fallback BHs. M20 cannot explain the missing RSG problem by direct BH formation and SN2010jl by the single star progenitor channel at the same time. Lifting $\xi_{2.5}^{\text{crit}}$ to a greater value to explain the progenitor SN2010jl results in loss of explanatory power over the missing RSG problem.

Type IIb/Ib SN progenitors:

The progenitors of stripped-envelope SNe (SESNe) are considered to either be massive single stars that experienced strong mass loss or donor stars that evolved through a binary MT phase. At the time of explosion, the progenitor could be a blue supergiant, a cool supergiant or a Wolf-Rayet (WR) star. Only five progenitors of Type IIb and two progenitors of Type Ib SNe have been imaged directly (Gilkis and Arcavi, 2022). The most luminous progenitors are those of the Type IIb SN1993J (Aldering et al., 1994), estimated to have $\log L/L_{\odot} = 5.1 \pm 0.3$, of the Type IIb SN2016gkg (Arcavi et al., 2017), estimated to have $\log L/L_{\odot} = 4.99 \pm 0.32$, and of the Type Ib SN2019yvr (Kilpatrick et al., 2021), estimated to have $\log L/L_{\odot} = 5.3 \pm 0.2$. All three estimates have been revised in Gilkis and Arcavi (2022), which assesses the most luminous progenitor source to be that of SN2016gkg with $\log L/L_{\odot} = 5.28 \pm 0.16$.

We explore the consequences of the hypothesis that the progenitors of these systems are case B donors that explode after having lost all or most of their hydrogen-rich envelope. This progenitor channel is supported by comparison of observations to detailed stellar evolution models (Yoon et al., 2017). For inferring M_{CO} of the progenitors, we find that the $\log L_{\text{TACCB}}(M_{\text{CO}})$ scaling law is applicable also to case B donors in S21 and S23 independent of Z up to $\log L_{\text{TACCB}}/L_{\odot} \simeq 5.2$. At greater $\log L_{\text{TACCB}}$, dependence on Z emerges: eq. (3.13) predicts a M_{CO} value lower than the actual stellar models at $Z_{\odot}/10$. Further more, the progenitor M_{CO} is greater the greater is Z (see Fig. 3.15). At $\log L_{\text{TACCB}} > 5.2 L_{\odot}$, the scaling law therefore provides a lower limit on the progenitor M_{CO} .

Within the observational uncertainty bounds, the progenitor M_{CO} inferred from eq. (3.13) is admitted to be $> 6 M_{\odot}$ in four out of in total six reference luminosity estimates. In addition, a progenitor metallicity closer to Z_{\odot} than to $Z_{\odot}/10$ makes a greater progenitor M_{CO} than the one inferred from eq. (3.13) more likely. Compared to the most luminous Type IIP SN progenitors detected, those of Type IIb/Ib admit greater progenitor luminosities and therefore greater M_{CO} . This is consistent with our CCSN recipe, which predicts the BH formation windows in case B systems to be shifted toward greater M_{CO} values compared to single stars. MM20 can explain the progenitor observations since it allows for explosions for $M_{\text{CO}} < 8 M_{\odot}$. M20 can explain the observations provided that a high enough $M_{\text{CO}}^{\text{crit}}$ is set. However, the observations challenge the F12 criterion, which predicts direct BH formation universally for single and stripped stars satisfying $6 < M_{\text{CO}}/M_{\odot} < 7$. One or several out of the four

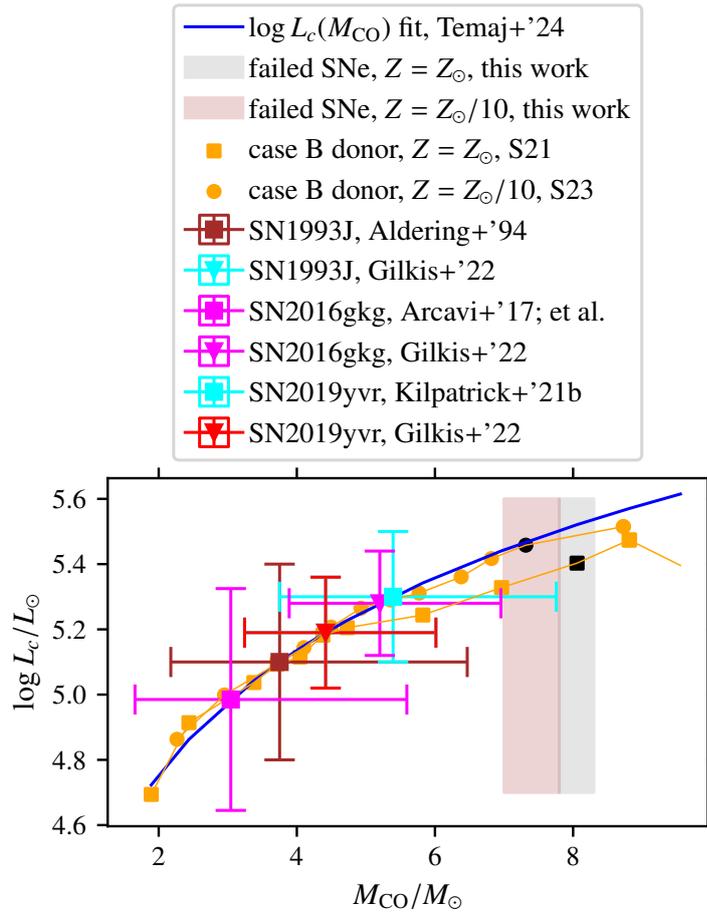


Figure 3.15: Inferred M_{CO} (within uncertainty bounds) of the most luminous Type IIb and Type Ib pre-SN progenitors. The parametric scaling law $\log L_{\text{TACCB}}(M_{\text{CO}})$ is applicable at arbitrary Z up to $\log L_{\text{TACCB}} < 5.2 L_{\odot}$, and provides a lower limit to M_{CO} for SN progenitor luminosities beyond. Observations are compared with the direct BH formation windows for case B donors predicted by our CCSN recipe as function of Z .

M_{CO} estimates that permit $M_{\text{CO}} > 6 M_{\odot}$ potentially constitute a counter-example to F12. The large observational uncertainties on the SESN progenitor luminosities do not allow for a decisive statement.

Type Ic SN progenitors:

The spectra of Type Ic SNe lack both hydrogen and helium lines, and the progenitors of these transients are WR stars. In the single-star progenitor channel, these form after a period of enhanced wind mass loss that removes all or most of the helium-rich envelope, which requires a higher Z for strong enough winds. In the binary star progenitor channel, a carbon-oxygen star can be formed by helium-rich envelope removal through the combined effect of mass loss by Roche lobe overflow to an acceptor star and winds, and therefore is not limited to higher Z . Since the helium-rich envelope is then mostly lost by the time the iron core collapses, the final pre-SN

progenitor mass M_{final} is bounded from above by M_{CO} at the end of CHeB, A naive way to place a lower bound on its value is to assume that

$$M_{\text{CO}} \geq M_{\text{final}} \geq M_{\text{rem}} + M_{\text{ej}} \quad (3.14)$$

where M_{ej} is the ejecta mass and M_{rem} the compact remnant mass. Since M_{ej} of Type Ic SNe can be deduced from the light curve, eq. (3.14) places constraints on the progenitor M_{CO} . Within the context of this framework, the spectroscopically normal nickel-rich Type Ic SN 2011bm (Valenti et al., 2012) poses a challenging test case to CCSN recipes. Its ejecta mass is estimated to be $M_{\text{ej}} \in (7, 17) M_{\odot}$. In order to account for this transient, M20 needs to lift $M_{\text{CO}}^{\text{crit}}$ to $\geq 8.2 M_{\odot}$ when assuming that a NS of mass $M_{\text{rem}} = 1.2 M_{\odot}$ is formed. Since MM20 admits SNe leaving fallback BHs of mass $M_{\text{rem}} > 2 M_{\odot}$ behind over $M_{\text{CO}} \in (7, 8) M_{\odot}$, it cannot explain SN2011bm unless it assumes a greater threshold value $M_4 > 9 M_{\odot}$.

The PS20 grid coupled to E16 admits explosions up to $M_{\text{CO}} \leq 10 M_{\odot}$ for suitable values of X_C at the end of CHeB. Assuming that the compact remnant is a NS of mass $M_{\text{rem}} = 1.2 M_{\odot}$, it is consistent with $M_{\text{ej}} < 11.2 M_{\odot}$.

For $M_{\text{CO}} \in (7, 11) M_{\odot}$, F12 expects SN explosions and fallback BH remnants. In order to satisfy eq. (3.14) with a fallback BH of mass $M_{\text{rem}} > 2 M_{\odot}$ and $M_{\text{ej}} \geq 7 M_{\odot}$, $M_{\text{CO}} > 9 M_{\odot}$ needs to be assumed. However a CCSN explosion at $M_{\text{CO}} > 9 M_{\odot}$ of a Type Ic progenitor of pre-SN mass $M_{\text{final}} \simeq M_{\text{CO}}$ leaving a fallback BH remnant of mass $M_{\text{rem}} < M_{\text{CO}} - 7 M_{\odot}$ is not consistent with the compact remnant mass calculation formalism of the ‘‘fast-convection’’ explosion model. It yields $M_{\text{rem}} = 7.272 M_{\odot}$ for a SN progenitor of $M_{\text{CO}} = 9 M_{\odot}$ and $M_{\text{rem}} = 10.76 M_{\odot}$ for $M_{\text{CO}} = 10.9 M_{\odot}$, since the fallback mass fraction is predicted to increase with M_{CO} . A distinctive feature of our CCSN recipe is that it guarantees explosions for $M_{\text{CO}} \in (8.4, 12.4) M_{\odot}$ independent of Z and binarity, admits explosions up to $M_{\text{CO}} = 15.4 M_{\odot}$ provided that Z is high enough and removal of the hydrogen-rich envelope sets in early enough, and expects NSs to be the more frequent remnant type than fallback BHs at larger M_{CO} values past the direct BH formation window. It therefore is consistent with $M_{\text{ej}} \in (7, 16.6) M_{\odot}$ so long as no further constraint is put on progenitor Z and MT pre-history.

SN remnants:

The CO core masses of SN progenitors can also be constrained by nebular line spectroscopy of SN remnants. After explosive nuclear burning, the ejecta mass of Type Ic SNe is mostly composed of oxygen and iron group elements. The nebular line ratio $[\text{OI}/\text{CaII}]$ is an indicator of the oxygen mass M_O released in the SN explosion according to the calibrated scaling law

$$\log[\text{OI}/\text{CaII}] = 0.9 \cdot \log(M_O/M_{\odot}) + 0.03 \quad (3.15)$$

that is applicable to SESNe of Type IIb/Ib and Type Ic/Ic-BL (Fang and Maeda, 2023; Fang et al., 2022). The greatest values of Type IIb/Ib are $\log[\text{OI}/\text{CaII}] \simeq 0.5$ while those of Type Ic/Ic-BL reach out up to $\log[\text{OI}/\text{CaII}] \simeq 0.7$ (Taddia et al., 2019; Pellegrino et al., 2022; Fang et al., 2022). These imply $M_O \leq 4.29 M_{\odot}$ and

$M_O \leq 6.23 M_\odot$, respectively. The oxygen mass M_O takes up a significant fraction X_O of the total SN ejecta,

$$M_O = X_O \cdot M_{\text{ej}} \quad (3.16)$$

that depends on the progenitor. According to radiation-hydrodynamics calculations with SNEC (Morozova et al., 2015), the mass fraction is found to be $X_O < 0.5$ for progenitor CO cores up to $M_{\text{CO}} < 6.6 M_\odot$ (which result in oxygen ejecta masses up to $M_O < 3.1 M_\odot$) and to increase to greater fractions for progenitor CO core masses somewhat beyond (Fang and Maeda, 2023). In the case of SN2011bm, the estimates are $M_O \in (5, 10) M_\odot$ (Valenti et al., 2012) while $M_{\text{ej}} \in (7, 17)$, which implies $X_O \simeq 0.6 - 0.7$. Starting from these reference X_O values within the range $M_O \in (5, 10)$, if we assume that $X_O = 0.7$ places a lower bound on M_{ej} for $M_O = 6.23 M_\odot$, then eq. (3.14) implies progenitor $M_{\text{CO}} \geq 8.9 M_\odot$ for the most oxygen-rich Type Ic SN explosions. While the rapid F12 and the PS20 table coupled to E16 can account for this observation, it challenges the M20 and MM20 recipes. In order to achieve compatibility with these observations, $M_{\text{CO}}^{\text{crit}}$ in M20 needs to be lifted accordingly, which again implies that M20 loses its power to explain the missing RSG problem by failed SNe. MM20 needs to lift M_4 by at least $1 M_\odot$ for consistency with observations.

3.4. Discussion and outlook

Pre-SN structure dependence on core evolution through the advanced burning phases: The findings from Laplace et al. (2024), which build the theoretical background for the the M_{CO} -based parametrization of explodability established by this work, are discussed in relation to the CCSN recipe below.

- At low $M_{\text{CO}} < M_{\text{CO}}^{(1)}$, X_C is comparatively high and the core temperature T_c comparatively low. Neutrino cooling is inefficient and the core carbon burning phase either radiation-dominated or weakly neutrino-dominated. The convective burning leads to an expanded core, and — due to the large amount of fuel X_C — the burning front does not move outward far in mass coordinate. Ultimately, these result in a lower core density and low iron core mass at the onset of collapse. The explodability is therefore high.
- At $M_{\text{CO}}^{(2)} \geq M_{\text{CO}} \geq M_{\text{CO}}^{(1)}$, less fuel X_C is available and the neutrino losses are greater. These lead to a neutrino-dominated core carbon burning phase. The core cools and turns radiative. The decreasing amount of fuel and the neutrino cooling accelerate core contraction and the outward progression of the burning front. The burning front moves further out in mass coordinate but stays below the effective Chandrasekhar mass. With partial degeneracy support, the core burns almost all of the X_C fuel in the convective regions before ignition of radiation-dominated neon burning. After core neon burning, the burning front quickly burns the former convective region, moving out far in mass coordinate. This leads to the growth of a large and dense fuel-free core. The explodability is therefore low.
- At greater $M_{\text{CO}} > M_{\text{CO}}^{(2)}$, due to high T_c and low X_C , core carbon burning is even more neutrino-dominated. The burning phase proceeds faster, and the

core contraction is even more accelerated. The burning front moves further out in mass coordinate, until it exceeds the effective Chandrasekhar mass. The contraction leads to an early core neon ignition. This next (radiation-dominated) burning stage suppresses nuclear burning at the front above, preventing it from moving far outward. Ultimately, this results in a low-mass iron core and high explodability.

- At high $M_{\text{CO}} \geq M_{\text{CO}}^{(3)}$, it is the central neon burning phase that becomes neutrino-dominated. The neutrino cooling leads to a quickly contracting radiative core, and accelerated progression of the burning front above. The burning front again moves further out but stays below the effective Chandrasekhar mass. With partial degeneracy support, the core burns most of the neon fuel before ignition of the radiation-dominated oxygen burning. Central burning of the large oxygen core leads to an enhanced growth of the silicon-rich core, as the burning front moves out in mass coordinate. These lead to a heavy iron core and low explodability.

The role of s_c for gravitational collapse: The relevance of the pre-SN variables used in our criterion for explodability is discussed in detail elsewhere, except for s_c (although see Schneider et al. (2021); Laplace et al. (2024)). The PNS is supported by repulsive nuclear forces, degeneracy pressure, thermal and other effects. The central specific entropy measures the thermal energy content inside the hot inner electron-degenerate core. With a greater s_c , the thermal effect add more to the mechanical support of the core, and therefore increase the gravitational mass $M_{\text{PNS,grav}}$ of the PNS. More generally, the maximal hydrostatic equilibrium mass for a PNS against collapse is EOS-dependent. From the empirically found monotonically increasing trend of $M_{\text{NS,grav}}$ with s_c (Temaj et al., 2024) follows that if an EOS-dependent upper NS mass limit $M_{\text{NS,grav}}^{\text{max}}$ exists beyond which it collapses, then accordingly there must also be a s_c^{max} for NS stability. da Silva Schneider et al. (2020) relates the maximal gravitational mass $M_{\text{grav}}^{\text{max}}$ of the PNS to the “most common specific entropy” \tilde{s} within the PNS (as defined therein), due to temperature contributions to the degeneracy-dominated EOS, and identifies an ordering with $\xi_{2.5}$ of the PNS pathway in the $(M_{\text{grav}}^{\text{PNS}}, \tilde{s})$ plane up to collapse. Not only the PNS mass, but also s_c increase, until the PNS loses hydrostatic stability and begins to collapse. Therefore the gross trend is that with greater s_c in the inner core as a starting condition, it is more difficult to retain a stable NS during the collapse phase, and the star more likely will form a BH.

How appropriate is our parameter choice for the M16 SN model? In future work, we aim to investigate in greater detail how well the semi-analytic SN code M16 agrees with state-of-art 3D simulations beyond solely the final fate prediction (successful or failed SN). Instead, we aim to compare the computed variables by the 1D SN code – such as the solutions to the gain radius, to neutrino luminosity and to the shock velocity – to those actually obtained in state-of-art 3D simulations. It is questionable, for example, whether the same values of β and α_{turb} are appropriate for all progenitors. After potential updates to adopted physics in the M16 SN code,

sophistication in choice of its free parameters, and re-evaluation over a large set of pre-SN progenitors, consequently also the pre-SN explodability criteria will need to be re-formulated to account for the differences. Given that our current formulation of the explodability criteria already are broadly in agreement with 3D CCSN outcomes mainly due to the approach we chose to segment the final fate landscape in the $(\mu_4 M_4, \mu_4)$ plane, we anticipate that it is mainly the critical value choice and slope of the separation line that will need be adjusted to account for the final fate outcome differences. In future work, we intend to construct a probabilistic scheme that predicts the CCSN outcome at the pre-SN stage, which takes into account stochastic effects due to magnetohydrodynamics and turbulence.

3.5. Conclusion

We have formulated explodability criteria that probe the pre-SN progenitor at four different mass coordinates based on the scalar variables $\xi_{2.5}$, s_c , μ_4 , $\mu_4 M_4$, and M_{CO} . Our pre-SN criterion is more optimistic about successful CCSN explosions by the neutrino-driven perturbation-aided mechanism than both E16 – based on $(\mu_4 M_4, \mu_4)$ with E20 parameters for the separation line – and the default compactness ($\xi_{2.5} > 0.45$) criterion. Our explodability formalism is calibrated to outcomes of the semi-analytic SN code (with parameter choice from S21) over a heterogeneous set of $\simeq 3900$ single, binary-stripped and accretor star pre-SN progenitors, over which it achieves a predictive accuracy of $> 99\%$. Enhancement of shock revival by multi-D effects is implicitly assumed by the pre-SN criterion, and in the original M16 code modeled with the parameter α_{turb} that scales up the shock radius expansion.

The pre-SN criterion is in broad agreement with 3D CCSN simulation outcomes of the Garching and the Monash group (in total, 27 simulations) this is a very sloppy attribution that you can perhaps make in a talk, but not in a paper, where proper citations are needed., achieving an accuracy of 85 % or 89 % depending on whether the more restrictive M_{CO} -based condition is included into the criterion or not) provided that for progenitors at intermediate M_{CO} , asymmetry is introduced as starting condition of the 3D simulation or magnetic field effects are taken into account during the explosion These enhancement effects can be decisive to the final fate outcome. Since a separation line in $(\mu_4 M_4, \mu_4)$ plane that would segregate the exploding from the non-exploding models cannot be drawn for any (k_1, k_2) choice, E16 cannot explain the sample of 3D CCSN simulation outcomes considered in this work. When placing a compactness-based single-parameter condition $\xi_{2.5} > 0.38$ for failed SNe, the same predictive accuracy over the 3D CCSN simulation outcomes is achieved, but the agreement with the M16 model outcomes over the large sample of pre-SN progenitors drops down to 86 %, over-predicting BH formation.

We also formulate a probabilistic model for BH formation by fallback in a successful SN explosion after shock revival. It excludes BH formation by fallback, if $\mu_4 M_4$ is large compared to $\xi_{2.5}$, if $\xi_{2.5}$ is critically low or if M_4 large compared to M_{CO} . Otherwise, we find that fallback BH formation occurs at a frequency of 0.15 over exploding models, i.e. NSs are the several times more likely CCSN outcome than fallback BHs. The fallback model is a particularly uncertain part of our predictive framework for several reasons. First, the deterministic criterion is not robust against

different parameter choices of the M16 model. Second, it has not been validated against 3D CCSN simulation outcomes. Third, it has not been compared against observations of fallback BH formation that constrain the progenitor properties.

We further find that all explodability proxies of single and binary-stripped star pre-SN progenitors, which are evaluated by our criterion, show bimodal trends as a function of M_{CO} within the wide range ($M_{\text{CO}} \in (5.1, 15.4) M_{\odot}$), over which explosions and implosions co-exist. We relate explodability proxies to M_{CO} over the S21 and S23 models of single and binary-stripped stars at different Z using supervised machine learning methods. The bimodality in explodability dependence on M_{CO} allows us to predict CCSN outcomes already at the end of core helium burning, when most stellar lifetime is over. Aided by the fitted ML models, we map out the windows in M_{CO} over which we predict direct BH formation to occur, depending on over which M_{CO} ranges the critical values in explodability proxies for a failed SN are reached. The bimodality in explodability dependence on M_{CO} allows us to predict CCSN outcomes already at the end of core helium burning, when most stellar lifetime is over. Our prediction is that direct BH formation outcomes are not spread randomly over a wider M_{CO} range, but occur within windows that shift in location and width that depend on MT pre-history and metallicity of the pre-SN progenitor. At higher Z , the direct BH formation windows of single and binary-stripped stars shift systematically to higher M_{CO} ranges. At any Z , the direct BH formation windows of stripped stars occur at greater M_{CO} values than those of single stars. These shifts can be reverted to differences in their central carbon mass fractions X_{C} at the end of core helium burning. Envelope mass loss by stellar winds (enhanced by a greater Z) or by stable MT to a companion star results in lower-mass helium cores. Lower-mass helium cores imply a lower core temperature and later ignition of the $^{12}\text{C}(\alpha, \gamma)^{16}\text{O}$ reaction that uses up carbon. This leaves more carbon behind at core helium exhaustion. Over the same M_{CO} range, a greater X_{C} shifts the peaks in explodability proxies toward larger M_{CO} values. Based on these results, we construct a CCSN recipe that allows to predict the final fate of a star, when M_{CO} and Z are given and a MT pre-history class (single stars, case C, case B and case A donor stars) is assigned. Our M_{CO} -based recipe is applicable in rapid BPS codes, which only keep track of global parameters and impose an evolutionary cut-off at – if not earlier – central neon ignition, well before the pre-SN stage is reached. Our proposed recipe is a first, preliminary version of a predictive model of CCSN outcomes, which — for a fixed pre-SN explodability criterion — will need be updated as adopted physics for stellar evolution up to the end of core helium burning, through the late burning stages, and for binary mass transfer are changed.

We have compared our CCSN recipe to competing recipes that are also based on M_{CO} and typically used in BPS codes: the rapid F12 criterion, the PS20 “look-up” table coupled to the E16 criterion, the $\xi_{2.5}$ -criterion based M20 recipe, and the stochastic MM20 recipe to the M16 1D SN code based on H16 single star pre-SN progenitor models. The aforementioned recipes were tested against observations that constrain the M_{CO} of Type IIP, Type IIn, Type IIb/Ib and Type Ic SN progenitors. We find that M20 needs to lift $M_{\text{CO}}^{\text{crit}}$ to $> 8.5 M_{\odot}$ in order to be consistent with Type IIn SN2010jl progenitor imaging, to $> 8.2 M_{\odot}$ in order to be consistent with the lower bound on the Type Ic SN2011bm ejecta mass, and to $M_{\text{CO}}^{\text{crit}} > 8.9 M_{\odot}$

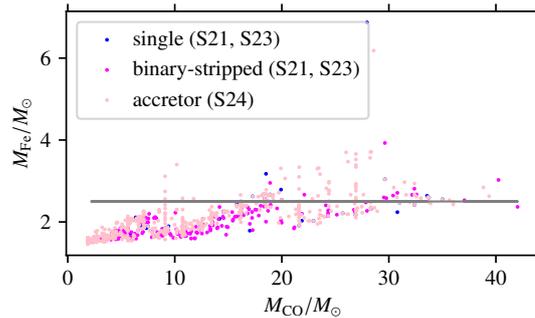


Figure 3.16: Trends of iron core mass M_{Fe} with carbon-oxygen core mass M_{CO} for single, binary-stripped and accretor star pre-SN progenitors. For some stars, the $M = 2.5 M_{\odot}$ mass coordinate used for ξ_{M} (grey solid line) is found inside the iron core.

in order to be consistent with the most oxygen-rich Type Ic SNe that have been observed to date. However, M20 then loses the explanatory power for the missing RSG problem, which we inferred to be the lack of Type IIP SN progenitors over the $M_{\text{CO}} \in (5.6, 7.8) M_{\odot}$ range, by failed SNe. Accordingly, for the MM20 recipe, the threshold value for guaranteed direct BH formation needs to be lifted to $M_4 > 8.9 M_{\odot}$ in order to be consistent with the aforementioned transients. F12, MM20 and our recipe can partially explain the missing RSG problem by direct BH formation. The fixed direct BH formation window of F12, between 6 and 7 M_{\odot} , is challenged by direct imaging of the progenitors of Type IIb/Ib SNe, which are consistent with progenitor $M_{\text{CO}} > 6 M_{\odot}$ within the observational uncertainty bounds. F12 cannot explain the lower bound on the ejecta mass of Type Ic SN2011bm, since at this progenitor M_{CO} range, this recipe predicts a fallback BH rather than the less massive NS remnant.

Since binary-stripped stars are the expected progenitors of both components in a binary BH (BBH) mergers, we expect that our CCSN recipe will result in a suppression of the predicted BBH merger rates compared to previous estimates. We aim to explore implications of our explodability formalism for gravitational wave astronomy in subsequent work.

B Appendix

B1. The $2.5 M_{\odot}$ mass coordinate and M_{Fe}

Fig. 3.16 illustrates that the $\xi_{2.5}$ criterion does not for all pre-SN progenitors probe the density of the mass-accretion regions outside the iron core.

B2. Parameters of the log Z scaling model

The linear model $f(x) = a + b \cdot x$ has two free parameters a and b , and we use pairs of data points $(M_{\text{CO}}^{\text{crit},i}(Z_{\odot}), Z_{\odot})$ and $(M_{\text{CO}}^{\text{crit},i}(Z_{\odot}/10), Z_{\odot}/10)$ to determine these analytically. This means that the curve for the critical boundary $M_{\text{CO}}^{(i)}(Z)$ is given

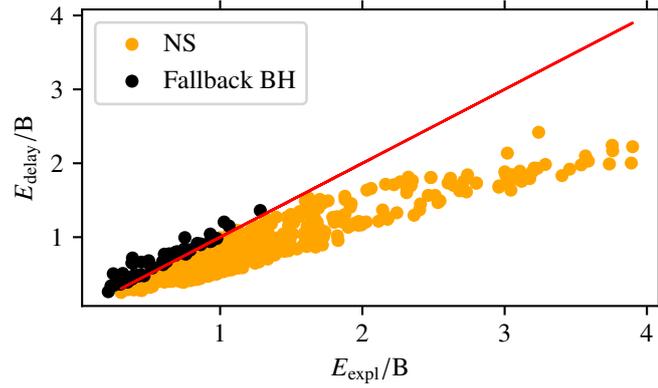


Figure 3.17: Determination of the remnant type (NS or fallback BH) in a successful CCSN explosion. By comparison of the explosion energy E_{expl} and the delay energy E_{delay} , the remnant type can be discriminated deterministically. The criterion $E_{\text{delay}} > E_{\text{expl}}$ for BH formation by fallback holds exactly over the S21, S23, S24 and T24 pre-SN progenitors (evaluated in the panel), and approximately over H16.

by

$$M_{\text{CO}}^{(i)}(Z) = a + \log Z/Z_{\odot} \cdot b$$

and the linear model's free parameters are given by

- $b = a_1 - a_2$ and
- $a = a_1$
- $(a_1, a_2) = (M_{\text{CO}}^{(i)}(Z_{\odot}), M_{\text{CO}}^{(i)}(Z_{\odot}/10))$

are the critical CO core mass values for BH formation by direct collapse at Z_{\odot} and at $Z_{\odot}/10$ are given in Tables 3.3 and 3.4. The same formalism is applied to the guaranteed NS formation windows that delineate boundaries of non-zero fallback BH probability, listed in Tables 3.5 and 3.6.

B3. Deterministic fallback model

The condition for BH formation from Müller and Janka (2015) applies during the explosion phase: if the diagnostic energy E_{diag} happens to become negative, then the gravitational binding energy of the matter enclosed by the mass shell contained within the shock is greater than the kinetic energy of the explosion. In this case, a BH forms by fallback of ejected matter onto the PNS. With the set of M16 model parameters adopted in Schneider et al. (2021), we find that remnant type can be predicted by comparing two characteristic energies during the explosion phase: the final explosion energy E_{expl} of the shock at breakout, and the energy term E_{delay} which – within the limitations of a 1D formulation – by construction accounts for the co-existence of outflows and downflows in the region surrounding the PNS during the explosion phase. E_{delay} is an auxiliary variable defined implicitly from eq. (42)

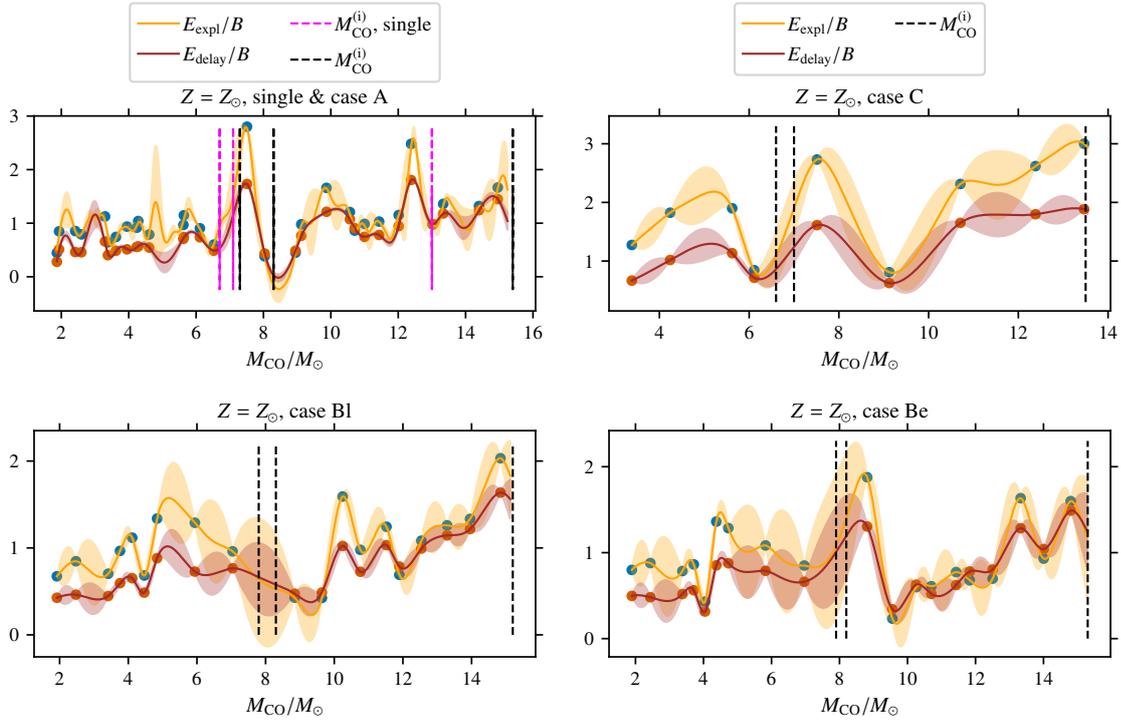


Figure 3.18: BH formation by fallback in successful CCSN explosions, as predicted by the deterministic criterion $E_{\text{delay}} - E_{\text{expl}} > 0$, at $Z = Z_{\odot}$. Non-exploding single and binary-stripped stars are removed from the sample, and the estimated direct BH formation boundaries are indicated by black dashed lines. The explosion energies E_{expl} and delay energies E_{delay} are fitted as a function of M_{CO} using GPR regressors. The GPR regressors are consistent with the hypothesis that fallback BH formation is not randomly distributed but occurs over windows in M_{CO} , and that the width and position of the windows varies with MT class. With the systematic uncertainties given the wide, overlapping prediction intervals on the relevant variables, the deterministic fallback BH formation window predictions are unreliable.

and (43) in Müller et al. (2016), which are used to calculate the evolution of E_{diag} as the revived shock moves outward in mass shell. The fallback condition reads

$$E_{\text{delay}} > E_{\text{expl}} \quad (3.17)$$

where E_{expl} is the final kinetic energy of the explosion. The criterion achieves an accuracy of 100% over the S21, S23, S24 and T24 stellar models (shown in Fig. 3.17), and 93% over the H16 models, and is sensitive to our particular parameter choice for M16.

E_{delay} is expected to correlate well with the initial explosion energy E_{ini} at shock revival, say $E_{\text{delay}} = \eta \cdot E_{\text{ini}}$, where η is a parameter not too far below 1. η measures the amount of energy that is dissipated during shock propagation. Physically one expects fallback to be determined primarily by the ratio of initial explosion energy to envelope binding energy E_{bind} , with a drastic increase of fallback once E_{bind} gets close to the diagnostic energy E_{diag} . Fallback BH formation is expected to occur, as soon as $E_{\text{bind}}/E_{\text{diag}}$ crosses some threshold. To zeroth approximation, $E_{\text{expl}} = E_{\text{ini}} + E_{\text{delay}} - E_{\text{bind}}$. The threshold is of the order of the binding energy at the mass cut in the “weak shock” regime defined in Mandel and Müller (2020), in which case $E_{\text{expl}} \ll E_{\text{ini}}$. So $E_{\text{expl}} - E_{\text{delay}} < 0$ essentially means just $E_{\text{expl}}/E_{\text{ini}} < \eta$, i.e. the explosion has lost a substantial fraction of its initial energy.

As shown exemplarily in Fig. 3.18, single and binary-stripped stars of different MT classes show complex, oscillatory patterns in the dependence of the E_{expl} and E_{diag} variables on M_{CO} . In principle, windows in M_{CO} can be mapped out over which eq. 3.17 is satisfied as a function of MT pre-history and Z . However, in contrast to H16, the sampling of the M_{CO} axes in S21 and S23 is too sparse for drawing faithful conclusions. The prediction intervals of our fitted GPR models overlap over wide ranges in M_{CO} and therefore do not allow for confident deterministic prediction of the remnant type as a function of M_{CO} .

4. Gravitational wave model for neutron star merger remnants

*This chapter has been submitted as [Soultanis et al. \(2024\)](#) for publication in the *Phys. Rev. D* journal. I contribute as the second author and have conducted substantial parts to the scientific work. Formatting and small text edits have been made to match this thesis.*

Abstract: We present a time-domain model for the gravitational waves emitted by equal-mass binary neutron star merger remnants i.e., quasi-stable hypermassive neutron stars, for a fixed equation of state. We construct a large set of numerical relativity simulations for a consistent with current constraints, totaling 157 equal-mass binary neutron star merger configurations. The gravitational-wave model is constructed using the supervised learning method of K-nearest neighbor regression. As a first step toward developing a general model with supervised learning methods that accounts for the dependencies on equation of state and the binary masses of the system, we explore the impact of the size of the dataset on the model. We assess the accuracy of the model for a varied dataset size and number density in total binary mass. Specifically, we consider five training sets of $\{20, 40, 60, 80, 100\}$ simulations uniformly distributed in total binary mass. We evaluate the resulting models in terms of faithfulness using a test set of 30 additional simulations that are not used during training and which are equidistantly spaced in total binary mass. The models achieve faithfulness with maximum values in the range of 0.980 to 0.995. We assess our models simulating signals observed by the three-detector network of Advanced LIGO-Virgo. We find that all models with training sets of size equal to or larger than 40 achieve an unbiased measurement of the main gravitational-wave frequency. We confirm that our results do not depend qualitatively on the choice of the (fixed) equation of state. We conclude that training sets, with a minimum size of 40 simulations, or a number density of approximately 11 simulations per $0.1 M_{\odot}$ of total binary mass, suffice for the construction of faithful templates for the post-merger signal for a single equation of state and equal-mass binaries, and lead to mean faithfulness values of $\mathcal{F} \simeq 0.95$. Our model being based on only one fixed equation of state represents only a first step towards a method that is fully applicable for gravitational-wave parameter estimation. However, our findings are encouraging since we show that our supervised learning model built on a set of simulations for a fixed equation of state successfully recovers the main gravitational-wave features of a simulated signal obtained using another equation of state. This may indicate that the extension of this model to an arbitrary equation of state may actually be achieved with a manageable set of simulations.

4.1. Introduction

Two detections of gravitational-waves (GWs) from sources identified as binary neutron star (BNS) mergers, GW170817 (Abbott et al., 2017a) and GW190425 (Abbott et al., 2020), have been reported already, and many more observations are anticipated in the next years (Abbott et al., 2018b). The GWs emitted by BNS mergers provide a new channel for probing the properties of high-density matter in the interior of neutron stars. Already GW170817 has led to stringent constraints on the so far incompletely known neutron star equation of state (EOS) (Abbott et al., 2017a; Abbott et al., 2019, 2018a), by the measurement of the tidal deformation of the binary components of the inspiral phase (see Chatziioannou (2020); Dietrich et al. (2021a) for reviews). Further constraints on the EOS could be derived from the electromagnetic counterparts of GW170817 or by combined measurements, such as those discussed in Margalit and Metzger (2017); Bauswein et al. (2017); Shibata et al. (2017); Radice et al. (2018); Ruiz et al. (2018); Rezzolla et al. (2018); Capano et al. (2020,?); Dietrich et al. (2020); Landry et al. (2020); Dietrich et al. (2020); Bauswein et al. (2021); Raaijmakers et al. (2021); Legred et al. (2021); Pang et al. (2021); Nicholl et al. (2021); Huth et al. (2022); Brandes et al. (2023) and the references therein. In addition, the increased number of detections expected in the near future would improve those constraints (Del Pozzo et al., 2013; Chatziioannou et al., 2015; Lackey and Wade, 2015; Hernandez Vivanco et al., 2019; Chatziioannou and Han, 2020; Criswell et al., 2023). The post-merger GW signal in GW170817 could not be detected because the sensitivity of the Advanced LIGO Aasi et al. (2015) and Advanced Virgo Acernese et al. (2015) detectors was not sufficient (Abbott et al., 2017a,b; Abbott et al., 2019). However, with upgraded second-generation (Abbott et al., 2020), future next-generation (Abbott et al., 2017; Maggiore et al., 2020; Reitze et al., 2019; Srivastava et al., 2022), or dedicated high-frequency detectors (Martynov et al., 2019; Ackley et al., 2020; Ganapathy et al., 2021; Page et al., 2021; Sarin and Lasky, 2021a), such detections are likely to be achieved in the next years.

In the post-merger phase, if the total binary mass is lower than a threshold mass (M_{thres}) Hotokezaka et al. (2011); Bauswein et al. (2013, 2021); Tootle et al. (2021) for prompt black hole formation, the remnant is a quasi-stable rapidly rotating neutron star that is supported against gravitational collapse by differential rotation and thermal pressure. This object is typically referred to as a *hypermassive neutron star* (HMNS) in the literature. The remnant oscillates in various fluid (quasi) oscillation modes and emits GWs in the range of 2-4 kHz. The most prominent feature of the post-merger GW emission is associated with the fundamental quadrupolar oscillation mode (see Zhuge et al. (1996); Shibata (2005); Shibata et al. (2005); Oechslin and Janka (2007); Stergioulas et al. (2011); Bauswein and Janka (2012); Bauswein et al. (2012); Hotokezaka et al. (2013); Takami et al. (2015); Bernuzzi et al. (2015); Clark et al. (2016); Bauswein et al. (2016); Foucart et al. (2016); Dietrich et al. (2017); Maione et al. (2017); Liebling et al. (2021); Dudi et al. (2022)). The frequency of this mode, denoted as f_{peak} or f_2 , depends on the EOS (see Stergioulas et al. (2011); Bauswein and Stergioulas (2015); Paschalidis and Stergioulas (2017); Dietrich et al. (2021a); Bernuzzi (2020)). As a result, a measurement of the GWs emitted in the post-merger phase would provide substantial and independent con-

straints to the EOS, see [Bauswein and Stergioulas \(2019\)](#); [Baiotti \(2019\)](#); [Friedman and Stergioulas \(2020\)](#); [Bernuzzi \(2020\)](#); [Dietrich et al. \(2021b\)](#); [Sarin and Lasky \(2021b\)](#) and references therein.

The detection of GWs emitted from BNS mergers relies on robust and efficient data analysis techniques. One approach is to employ matched filtering search methods that rely on post-merger GW models. To achieve this, reliable models in the time or frequency domain for the post-merger GW emission which are informed by numerical relativity simulations, also called templates, are required. Several parametric models have been introduced already, either in the frequency domain ([Messenger et al., 2014](#); [Tsang et al., 2019](#); [Puecher et al., 2023](#); [Breschi et al., 2022](#)) or in the time domain ([Hotokezaka et al. \(2013\)](#); [Bauswein et al. \(2016\)](#); [Bose et al. \(2018\)](#); [Yang et al. \(2018\)](#); [Breschi et al. \(2019\)](#); [Easter et al. \(2020\)](#); [Soultanis et al. \(2022\)](#); Ω). An alternative method employs morphology-independent models ([Chatziioannou et al., 2017](#); [Torres-Rivas et al., 2019](#); [Wijngaarden et al., 2022](#); [Tringali et al., 2023](#); [Criswell et al., 2023](#); [Miravet-Tenés et al., 2023](#); [Sasli et al., 2023](#)).

In this work, we focus on the construction of models informed by numerical relativity simulations. [Whittaker et al. \(2022\)](#) introduced a time domain model employing a conditional variational auto-encoder (CVAE) using simulations of BNS mergers in numerical relativity. The CVAE is a deep-learning method where two sets of variables are employed for the interpolation. The first set of variables characterizes the system. These variables can be the total mass, binary mass ratio, initial spins, etc. The second set of variables, that is, the latent variables, are not known a priori but the CVAE algorithm learns to use them during training. Latent variables include the EOS, discretization errors, physics-modeling choices, etc. The authors generated an extended data set using a stochastic model and estimated that training such CVAE models would require approximately 10^4 signals. [Clark et al. \(2016\)](#) introduced a frequency domain model for the amplitude and phase of the spectra, based on principal component analysis for a set of numerical relativity simulations. [Easter et al. \(2019\)](#) introduced a hierarchical frequency domain model, for the amplitude of spectra, that trains on existing numerical-relativity post-merger simulations. They employed GW spectra aligned in frequency, as in ([Clark et al., 2016](#)), and fitted the amplitudes with a linear model. In [Pesios et al. \(2024\)](#), the amplitude of the post-merger spectrum was predicted using two different types of regression models in the frequency domain, multi-linear regression (MLR) and artificial neural network (ANN) regression, which were trained on a total of 87 simulations spread among 14 different EOS. With both regression methods, high fitting factors were achieved, when the method was calibrated to reduce the uncertainties inherent in empirical relations for the main post-merger frequency.

One of the main challenges in the construction of post-merger models informed by numerical relativity simulations is that they can be computationally costly, and thus it is difficult to create a sufficiently large library of simulations. Furthermore, general-purpose GW models must take into account the various degrees of freedom that influence the properties of the GW signals, such as the total binary mass $M_{\text{tot}} = m_1 + m_2$, the binary mass ratio $q = m_1/m_2$ ¹, the EOS model, intrinsic spin, etc.

¹ m_1, m_2 correspond to the individual gravitational masses, at infinite separation, of the two companion neutron stars.

which significantly increase the size of the required data set of simulations. In this work, we vary the total binary mass M_{tot} and fix the EOS model and the mass ratio to $q = 1$. We intend to study the other degrees of freedom in future work. The current study allows to understand the minimum requirement for using supervised learning techniques and thus to assess the performance of such methods.

We introduce a model for post-merger GWs employing the supervised learning method of K-nearest neighbor (KNN) regression. In this method, the assigned value at the interpolation point is computed based on the mean of the values of its K-nearest neighbors. Weights can be assigned to each point in the local neighborhood using arbitrary functions of distance in the parameter space. We consider a sequence of equal-mass binaries with increasing total binary mass M_{tot} for a fixed EOS model. We build a large library of simulations for a fixed EOS model, consisting of a total of 157 binaries uniformly distributed with respect to the total binary mass of the system using a smooth particle hydrodynamics (SPH) code (Oechslin et al., 2002; Oechslin, R. et al., 2007; Bauswein et al., 2010b) (see Sec. 4.2.). We split the data set into three subsets used for training, validation, and testing, and optimize models with training sets of varied sizes. We assess the KNN models, i.e., models constructed with KNN regression, in terms of a noise weighted inner product between two signals, the faithfulness (\mathcal{F})². In addition, we carry out injections, that mimic real detection scenarios, and enable us to evaluate our models. In those, a simulated signal with specific binary source parameters is inserted into a detector network and then reconstructed by the model using parameter estimation methods. We inject the simulated signals from the test set configurations and consider the Advanced LIGO-Virgo network. We find that KNN models with training sets of size equal or larger than 40 exhibit convergence, and accurately reconstruct the dominant f_{peak} frequency of the injected signals. The models achieve high values of \mathcal{F} with mean values of approximately 0.95 and maximum values in the range of 0.992 to 0.995, for simulated signals of the test set.

Furthermore, we study the execution time of templates generated with the KNN method. We find that the KNN models achieve execution times of approximately a few milliseconds for sampling rates comparable to those employed by Advanced LIGO.

Finally, we explore the generality of our findings by constructing another KNN model for a fixed EOS using an alternative EOS model. We show that a training set with 40 simulations suffices to build a reliable time-domain model that achieves high mean values of \mathcal{F} in that case too. In Pesios et al. (2024) the same order of magnitude of number of simulations was used to create MLR and ANN-based regression models to predict the amplitude in the frequency domain.

This paper is structured as follows: In Sec. 4.2. we describe the methods employed for the different stages of this work, i.e., the simulation tool and the parameters of the simulated systems, the employed supervised learning method, and the numerical setup of the injection studies. In Sec. 4.3. we present our results. We evaluate the performance of the various KNN models in terms of the faithfulness \mathcal{F} achieved between the model predictions and the simulations of the test and validation sets.

²A perfect match between two signals corresponds to the value of $\mathcal{F} = 1$.

We present information regarding the speed of such type of models. We discuss the results of the injection studies. Finally, we construct a KNN model using a different EOS model and assess the effect of the EOS on the accuracy of the template and the necessary sampling density to achieve high performance. We draw conclusions in Sec. 4.4..

4.2. Methods

In this section, we discuss the methods employed in this work. We describe the numerical setup of the BNS simulations. We describe the main concepts of the KNN scheme, the preparation of data, and how the optimization of the KNN model and the hyper-parameter tuning is carried out. We discuss the scheme for splitting the original GW data set into subsets that are used for training, validation, and testing of the KNN model. We present the complete form of the model. Finally, we describe the setup of our injection studies.

4.2.1. BNS merger simulations

We perform numerical relativity BNS merger simulations using a general relativistic SPH code (Oechslin et al., 2002; Oechslin, R. et al., 2007; Bauswein et al., 2010b). The spacetime is evolved under the conformal flatness condition (CFC) (Isenberg and Nester, 1980; Wilson et al., 1996). In this version of the code we use the Wendland kernel function (Schaback and Wendland, 2006; Rosswog, 2015).

We simulate a sequence of 157 equal-mass binaries with increasing binary mass. We employ the APR4 (Akmal et al., 1998) EOS model. For this EOS model, the thermal effects are approximated by an ideal gas pressure component with $\Gamma_{\text{th}} = 1.75$ (see e.g. Bauswein et al. (2010a)). The total binary masses of the sequence of simulations range from $2.4 M_{\odot}$ to $2.76 M_{\odot}$. The simulations are uniformly distributed with respect to total binary mass M_{tot} (see Fig. 4.8). Because the threshold mass M_{thres} for prompt black hole formation is $2.825 M_{\odot}$ (Bauswein et al., 2021) for this EOS model, the chosen M_{tot} range results in remnants that do not undergo a delayed collapse for at least a few tens of milliseconds. From every simulation, we obtain the $+$ and \times polarizations of the GW signal, denoted by $h_{+}(t; M_{\text{tot}})$ and $h_{\times}(t; M_{\text{tot}})$, for an observer at the rotational axis of the binary. We then interpolate the signal with a sampling rate of $dt \approx 0.005$ ms (or 200 kHz).

Furthermore, we construct a second mass-sequence of equal mass binaries using the SFHX EOS model (Steiner et al., 2013). The SFHX EOS model provides a consistent treatment of thermal effects. We simulate a total of 99 binaries with M_{tot} ranging from $2.4 M_{\odot}$ to $2.8 M_{\odot}$. With this data set, we construct another KNN model for a fixed EOS model, i.e. the SFHX EOS, and discuss the agreement between the results obtained using the two different EOS models obtained from the same M_{tot} sampling density of the training data (see Sec. 4.3.4.).

To extract the f_{peak} frequency, for every simulation, we construct the complex signal defined as $h(t) = h_{+}(t) - ih_{\times}(t)$. We then compute the Fourier Transform of the post-merger GW signal using the implementation of Numpy (Harris et al., 2020). We use the Tukey window function with a roll-off parameter of 0.05. Then we obtain

the characteristic frequency-domain GW strain defined as $h_{\text{char}}(f) = \tilde{h}(f) \cdot f$, where $\tilde{h}(f)$ is the Fourier Transform of $h(t)$. We identify f_{peak} , i.e., the frequency of the quadrupolar oscillation mode, as the frequency corresponding to the most dominant peak in the GW spectrum. Finally, a second-order polynomial fit is performed in the local region of this peak to determine the exact f_{peak} value at which the maximum occurs.

4.2.2. K-nearest neighbors model

K-nearest neighbor regression is a supervised learning method that allows for non-parametric fitting. In this work, we use KNN regression to describe the emission of GWs in the post-merger phase using training sets of different sizes (number of simulations). The goal is to predict GW signals of unseen configurations. As is typically done in other supervised learning methods, we train the model using the strain data from the training set. In addition, we aim to estimate the size of the dataset that is required for building accurate post-merger models. In this regard, the KNN regression is a simple yet robust method that allows us to explore concepts of supervised learning regression without dealing with the complexity of more sophisticated schemes such as deep learning approaches. Furthermore, a relatively smooth transition between the GW spectra with respect to M_{tot} is expected as in our model the EOS is fixed. Thus, regression schemes that use information from neighbors may be appropriate to use. For our analysis we use the KNN implementation provided by SciKit (Pedregosa et al., 2011).

Definitions

The overlap (\mathcal{O}) is a widely used metric that quantifies the match between two signals. Perfect agreement corresponds to an \mathcal{O} value of 1. The overlap \mathcal{O} is defined as

$$\mathcal{O} \equiv \frac{\langle h_1(t), h_2(t) \rangle}{\sqrt{\langle h_1(t), h_1(t) \rangle \langle h_2(t), h_2(t) \rangle}}. \quad (4.1)$$

The noise-weighted inner product $\langle h_1(t), h_2(t) \rangle$ between the signals $h_1(t)$ and $h_2(t)$ is given by

$$\langle h_1(t), h_2(t) \rangle \equiv 4\text{Re} \int_0^\infty df \frac{\tilde{h}_1(f) \cdot \tilde{h}_2^*(f)}{S_h(f)}, \quad (4.2)$$

where the one sided noise spectral density of the detector $S_h(f)$ of Advanced LIGO at design sensitivity (Aasi et al., 2015), and the Fourier transforms $\tilde{h}_1(f)$, $\tilde{h}_2(f)$ are used.

The maximized overlap \mathcal{O} with respect to initial phase ϕ_0 and merger time t_0 defines the faithfulness \mathcal{F} (or match) between two signals. It reads

$$\mathcal{F} \equiv \max_{\phi_0, t_0} \frac{\langle h_1(t), h_2(t) \rangle}{\sqrt{\langle h_1(t), h_1(t) \rangle \langle h_2(t), h_2(t) \rangle}}. \quad (4.3)$$

In what follows, \mathcal{O} and \mathcal{F} are calculated only for the post-merger signals ($t > t_0$) while the inspiral is not used (see below). We consider a frequency band, for Eq (4.7), from 10 Hz to 5.0 kHz.

We assess how well the model reproduces the simulated signals in terms of faithfulness \mathcal{F} , i.e., the maximized in phase and time overlap, between the simulations and the model’s predictions for the same value of M_{tot} (see Sec. 4.3.1. and Sec. 4.3.4.). For the injections (see Sec. 4.3.3.), that correspond to the realistic parameter estimation scenario, all parameters of the model are varied (including M_{tot}). In this regard, the overlap \mathcal{O} distribution shows how well the reconstructed signal matches the injected signal.

K-nearest neighbors regression

We summarize the general concepts of the KNN regression (see [James et al. \(2023\)](#)). We consider a point \vec{X}_0 in the N-dimensional regressor parameter space of the interpolation problem and the corresponding target variable prediction denoted by \vec{Y}_0 . The algorithm first identifies the K points of the training set that are closest to \vec{X}_0 , denoted by \mathcal{N}_0 . The prediction \vec{Y}_0 is then estimated using the average (or weighted average) of all points in \mathcal{N}_0 . The prediction estimate is given by the general form

$$\vec{Y}_0 = \frac{1}{K} \sum_{\vec{X}_i \in \mathcal{N}_0} \frac{w_i \cdot \vec{Y}_i}{W}, \quad \text{where } W = \sum_{\vec{X}_i \in \mathcal{N}_0} w_i, \quad (4.4)$$

where \vec{Y}_i are the observations i.e., the strain data in our case, of the training set, and w_i are assigned weights between \vec{X}_0 and the points in \mathcal{N}_0 .

The weights can be chosen to be equal for all neighbors or according to arbitrary criteria. For example, an inverse proportionality of the form $w_i = 1/d_i$ where d_i is the distance between \vec{X}_0 and the neighbors \vec{X}_i in \mathcal{N}_0 is commonly used. Furthermore, the metric that defines the calculation of distances between points in the N-dimensional space need to be chosen before the training.

The number of neighbors K , the weighting criterion, or the metric which determines how distances in the parameter space are measured, are not known a priori and are considered tuning parameters, commonly referred as *hyper-parameters* (see SciKit ([Pedregosa et al., 2011](#)) and references therein). The hyper-parameters affect the accuracy and performance of the KNN regression model. For this reason, we use the validation data to determine the optimal set of hyper-parameters. This stage is called *hyper-parameter tuning*.

In order to determine the optimal set of hyper-parameters, for each training set, we evaluate the accuracy of the KNN regression predictions for the simulations of the validation sets by considering a global quantity that is representative of the overall performance. We opt for the mean value of faithfulness \mathcal{F} , defined in Eq. (4.8),

between the signals of the validation set and the corresponding predictions of the various KNN models. A possible caveat of this choice of global quantity may be that the late part of the signal, where the amplitude is low, is penalized less than the early part. For each training set, we iterate over all combinations of hyper-parameters predefined over a grid, and store the set of hyper-parameters that maximize the mean

\mathcal{F} over the validation data. For each N_{train} , we obtain a different hyperparameter configuration of the best-fit KNN model as scored by the mean \mathcal{F} .

In our case, the input features, i.e., the space where the vectors \vec{X}_i live, of the KNN model are the total binary mass M_{tot} and the simulation coordinate time t expressed as an array with sampling rate $f_s = 1/dt$. The vectors \vec{Y}_i are + polarization simulated signals $h_+(t; M_{\text{tot}})$ expressed as arrays of the same length as t . The output of the KNN model is a GW signal for the + polarization.

The overlap (\mathcal{O}) is a widely used metric that quantifies the match between two signals. Perfect agreement corresponds to an \mathcal{O} value of 1. The overlap \mathcal{O} is defined as

$$\mathcal{O} \equiv \frac{\langle h_1(t), h_2(t) \rangle}{\sqrt{\langle h_1(t), h_1(t) \rangle \langle h_2(t), h_2(t) \rangle}}. \quad (4.5)$$

The noise-weighted inner product $\langle h_1(t), h_2(t) \rangle$ between the signals $h_1(t)$ and $h_2(t)$ is given by

$$\langle h_1(t), h_2(t) \rangle \equiv 4\text{Re} \int_0^\infty df \frac{\tilde{h}_1(f) \cdot \tilde{h}_2^*(f)}{S_h(f)}, \quad (4.6)$$

where the one sided noise spectral density of the detector $S_h(f)$ of Advanced LIGO at design sensitivity, and the Fourier transforms $\tilde{h}_1(f)$, $\tilde{h}_2(f)$ are used.

The noise-weighted inner product $\langle h_1(t), h_2(t) \rangle$ between the signals $h_1(t)$ and $h_2(t)$ is given by

$$\langle h_1(t), h_2(t) \rangle \equiv 4\text{Re} \int_0^\infty df \frac{\tilde{h}_1(f) \cdot \tilde{h}_2^*(f)}{S_h(f)}, \quad (4.7)$$

where the Fourier transforms $\tilde{h}_1(f)$, $\tilde{h}_2(f)$ are normalized by the one sided noise spectral density of the detector $S_h(f)$ of Advanced LIGO at design sensitivity

The maximized overlap \mathcal{O} with respect to initial phase ϕ_0 and merger time t_0 defines the faithfulness \mathcal{F} (or match) between two signals. It reads

$$\mathcal{F} \equiv \max_{\phi_0, t_0} \frac{\langle h_1(t), h_2(t) \rangle}{\sqrt{\langle h_1(t), h_1(t) \rangle \langle h_2(t), h_2(t) \rangle}}. \quad (4.8)$$

In what follows, \mathcal{O} and \mathcal{F} are calculated only for the post-merger phase signals ($t > t_0$) while the inspiral is not used (see below). We consider a frequency band, for Eq (4.5), Eq. (4.8), from 10 Hz to 5.0 kHz. We assess how well the model reproduces the simulated signals in terms of faithfulness \mathcal{F} , i.e., the maximized agreement in phase and time overlap, between the simulations and the model's predictions for the same value of M_{tot} (see Sec. 4.3.1. and Sec. 4.3.4.). For the injections (see Sec. 4.3.3.), that correspond to the realistic parameter estimation scenario, all parameters of the model are varied (including M_{tot}). In this regard, the overlap \mathcal{O} distribution shows how well the reconstructed signal matches the injected signal.

The model

We find that the KNN algorithm performs best when the input GW signals used in training are aligned in phase. In order to align the signals in phase, we express the GW signal in the complex form

$$h(t) = h_+(t) + i h_\times(t) \quad (4.9)$$

$$= |h(t)| \cdot e^{+i\phi(t)}, \quad (4.10)$$

where the amplitude $|h(t)|$, and phase $\phi(t)$ are defined as

$$|h(t)| = \sqrt{h_+^2(t) + h_\times^2(t)}, \quad (4.11)$$

$$\phi(t) = \tan^{-1} \left(\frac{h_\times(t)}{h_+(t)} \right). \quad (4.12)$$

The merger time t_{merge} is the time at which $|h(t)|$ reaches its maximum value, and the phase during the merger time is given by $\phi_{\text{merge}} = \phi(t_{\text{merge}})$. We then introduce a phase shift $\phi \rightarrow \phi - \phi_{\text{merge}}$, and reconstruct the two polarizations as

$$h_+(t) = \text{Re}(h(t)), \quad (4.13)$$

$$h_\times(t) = \text{Im}(h(t)). \quad (4.14)$$

We also introduce a time shift $t \rightarrow t - t_{\text{merge}}$ so that $t = 0$ corresponds to the

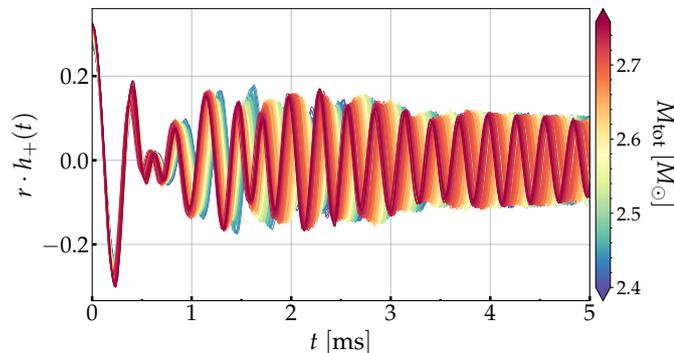


Figure 4.1: Aligned GW signals for the complete set of 157 simulations. The colors indicate the different total binary mass. The first 5 ms of the GW signals are shown, while the total duration of the signals is 17 ms.

merger time. Figure 4.1 shows the aligned signals $h_+(t)$ for the complete set of simulations where the color code indicates the total binary mass M_{tot} . Interestingly, we find remarkable agreement in the morphology of the aligned signals in the first 1 ms. In the early phase, up to 1 ms, the signal is simple, as it exhibits almost no dependence on the total binary mass M_{tot} . The dependence on M_{tot} becomes apparent as the different signals start to fall out of phase. This is because the characteristic frequencies of the GW signal scale proportional to the total binary mass of the system. Phase alignment simplifies the first 1 ms of the post-merger

signal and ensures that it is better modeled by the KNN model than in the case without phase alignment.

Finally, we discuss the general form of our model in the binary frame, which is numerically equivalent to a face-on observer. The antenna response functions of the detectors for a source with arbitrary position and orientation in the sky can be computed using the detector projection tensors (Maggiore, 2000, 2007) which introduce several parameters: a) the right ascension (r.a.) and the declination (δ), which describe the location of the source in the sky; b) the inclination angle of the binary’s orbital plane (ι); and c) the polarization angle (ψ), which is an angle in the plane perpendicular to the line of sight that defines the $+$ and \times polarizations. In order for the KNN model’s predictions to describe general GW signals, we also introduce an appropriate phase shift of the form $e^{i\phi_0}$. This phase shift is typically used in every detection search scenario, since GW signals have arbitrary initial phases depending on the position of the detector and arrival of the signal (see Sec. 4.3.3.). In order to account for the arbitrary arrival time of the GW signal we also introduce a time shift parameter t_0 which corresponds to the start of the post-merger signal, or equivalently to the merger time. The model reads

$$h_+(t) = \text{KNN}(t - t_0; M_{\text{tot}}) e^{i\phi_0}, \quad (4.15)$$

$$h_\times(t) = h_+(t - t_0) e^{i\frac{\pi}{2}}, \quad (4.16)$$

where KNN corresponds to the KNN model, and t_0 is the starting time of the post-merger signal.

Throughout the training and validation of the KNN model, we impose a cut-off at $t = 17$ ms of post-merger signal. The KNN model produces longer post-merger signals than the cut-off time. The extrapolated signal depends on the data close to the cut-off region.

4.2.3. Splitting of data

We split the APR4 data set (157 simulations) into subsets used for *training*, *validation*, and *testing* of the models. The test set contains $N_{\text{test}} = 30$ simulations that we set aside and use for testing. The remaining data are split into training and validation sets.

The training set includes the simulations used to obtain KNN models for given sets of hyper-parameters. The validation set is used for the optimization of the KNN models, that is the hyper-parameter tuning, where the set of hyper-parameters that maximizes the mean \mathcal{F} , is determined. As we want to quantify the number of simulations required to build accurate models, we vary the size N_{train} of the training set (the number of simulations for training) in the range of $N_{\text{train}} = \{20, 40, 60, 80, 100\}$. The size of the validation set is thus $N_{\text{val}} = N_{\text{tot}} - N_{\text{test}} - N_{\text{train}}$ where N_{tot} is the total number of simulations. This split thus results in a total of five KNN models labeled by the size of their training set and the EOS, e.g. $\text{KNN}_{\text{tr}=X}^{\text{EOS}}$ where X denotes the size of the training set.

The hold-out test set provides an independent way to assess the performance of KNN models (with varying N_{train}) under the same conditions. The simulations of the test set are distributed in an approximate equidistant manner with respect to M_{tot} .

We employ the Latin Hypercube Sampler (LHS) (McKay et al., 1979b) to draw N_{train} simulations from a uniform distribution to construct the training set. LHS is a method for generating a homogeneously distributed sequence of parameter values using a randomized sampling scheme with memory. The advantage of LHS over the memoryless uniform random sampler is that it prevents formation of sampling gaps. These arise for the uniform random sampler as a statistical effect in consequence of grid point clustering, when the sample size is small (McKay et al., 1979b). The validation set contains the remaining simulations.

4.2.4. Injection setup

We describe the setup for the injection study (see Sec. 4.3.3.). We perform injections using Bilby (Ashton et al., 2019; Romero-Shaw et al., 2020), an inference code implemented in Python, which provides an infrastructure for GW parameter estimation. In this work, we use the Dynesty sampler (Speagle, 2020; Kuposov et al., 2023; Skilling, 2004, 2006; Higson et al., 2019), implemented in Bilby³ (see Ashton et al. (2019) and the references therein for other sampler options), and consider 1000 live points. For the auxiliary parameters of the Dynesty sampler, i.e., the number of autocorrelation lengths in the Markov-chain Monte Carlo algorithm and the number of walks taken by the live points, we choose values 100 and 3, respectively. We do not find a significant difference in the outcome of the injections using slightly different values. We evaluate the performance of the KNN models by comparing the source parameters of the reconstructed signals to those of the injected signals. We carry out several injections with varied luminosity distance (D_L) of the source, and sample the parameters M_{tot} , ϕ_0 , t_0 , D_L . We use the posterior probabilities of the model's parameters and convert them to posterior probabilities of quantities connected to the the post-merger signal (see Sec. 4.3.3.). In this way, we examine the posterior probabilities of the dominant frequency peak, f_{peak} , in the GW spectra and compare them to the values of the injected simulated signals. In addition, we discuss the posterior probabilities of the overlap \mathcal{O} values between the reconstructed and the injected signals.

We inject a GW signal for a binary configuration from the test set with a total mass of $2.656 M_\odot$ in the source frame and vary the luminosity distance D_L of the source from 1 Mpc to 40 Mpc. We use the three-detector Advanced LIGO Hanford, Livingston, and Advanced Virgo network, at design sensitivity. For the sky location of the source, we choose the combination of r.a. and δ which corresponds to optimal orientation. We consider a face-on binary, i.e., $\iota = 0$. We do not include a noise realization in our analysis. We employ uniform priors for the intrinsic parameters M_{tot} , ϕ_0 , t_0 in the range of $[2.4 M_\odot, 2.8 M_\odot]$, $[0.0, 2\pi]$, $[t_0^{\text{inject}} - 5 \text{ ms}, t_0^{\text{inject}} + 5 \text{ ms}]$, respectively⁴. For D_L , we employ a logarithmic uniform prior in the range of $[0.48 \text{ Mpc}, 48 \text{ Mpc}]$. We use the sampling rate of Advanced LIGO, i.e., $f_s = 16384 \text{ Hz}$, and consider a signal of total duration of 1 second (see Chatziioannou et al. (2017)).

³The package version 2.1.1. is used.

⁴ t_0^{inject} represents the injected t_0 value.

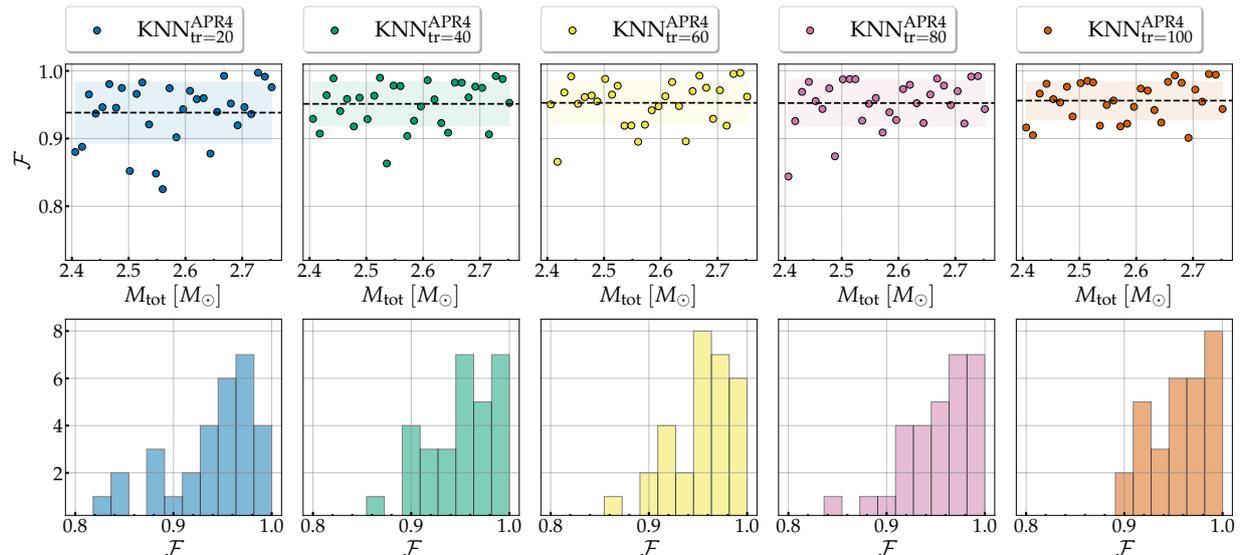


Figure 4.2: Upper panel: Faithfulness \mathcal{F} as a function of the total binary mass M_{tot} for the simulations of the test set and the predictions of the different KNN models, where training sets of sizes $N_{\text{train}} = \{20, 40, 60, 80, 100\}$ were employed. Black dashed line indicates the mean value. Shaded areas indicate the scatter of the data defined by the standard deviation. Bottom panel: Histograms of the aforementioned \mathcal{F} .

4.3. Results

In this section, we discuss the predictive accuracy of the KNN models with varied N_{train} and their overall performance. We compute the faithfulness \mathcal{F} between the KNN predicted signals and those of the validation and test sets. We also discuss the execution time of the KNN models. We perform injection studies in which GW simulated signals from the test set are reconstructed using the KNN models. We investigate how accurately the characteristics of the GW spectra, such as the f_{peak} frequency, can be reproduced using the different KNN models. Finally, we explore the generality of our results by considering a KNN template constructed using an other EOS model.

4.3.1. Performance of KNN models

We assess the quality of the fits by calculating the values of \mathcal{F} between the simulated signals of the test set and the signal predictions obtained from the KNN models. We consider a known starting time of the signal of $t = 0$. Because the simulated GW signals have arbitrary initial phases, we perform phase maximization with the SciPy curve-fit routine (Virtanen et al., 2020), that employs a non-linear least squares algorithm (Branch et al., 1999; Byrd et al., 1988). This procedure is repeated for each of the GW signals of the test set. The test set is a unique collection of configurations equidistantly distributed with respect to M_{tot} , and thus allows us to compare the different KNN models on the same GW signals. In Appendix C1., we present

the distribution of the faithfulness \mathcal{F} , computed using the above procedure, for the simulations of the validation set that were used during the hyper-parameter tuning of the KNN models.

We present the values of \mathcal{F} between the predictions of the various KNN models, for each N_{train} , and the 30 simulations of the test set in Fig. 4.2. The upper row shows \mathcal{F} as a function of the total binary mass M_{tot} , and the lower row shows the distribution of these \mathcal{F} values. We find very good performance for most KNN models with almost no dependence on the size of the training set. More specifically, a minimum of $N_{\text{train}} = 40$ simulations used for training is sufficient to achieve high \mathcal{F} , and larger values of N_{train} do not significantly enhance the overall performance of the model.

As shown in Fig. 4.2, the mean value of \mathcal{F} (indicated by a black dashed line) for most KNN models, i.e. $\text{KNN}_{\text{tr}=40}^{\text{APR4}}$ to $\text{KNN}_{\text{tr}=100}^{\text{APR4}}$, is approximately constant at 0.95, and only marginally increases with N_{train} . These mean values are comparable to those of [Easter et al. \(2020\)](#), where \mathcal{F} in the range of 0.92 to roughly 0.97 are reported for 9 simulated signals. The mean value of the model $\text{KNN}_{\text{tr}=20}^{\text{APR4}}$, $\mathcal{F} = 0.938$, is lower than that of the other models, however, it is still well above 0.90. We also report the standard deviation, with respect to the mean value, for each KNN model (indicated by the shaded areas). The scatter is roughly constant for most KNN models except for the low-resolution $\text{KNN}_{\text{tr}=20}^{\text{APR4}}$ model where it is slightly larger than the rest. This may not be surprising since the overall performance of the KNN models with $N_{\text{train}} \geq 40$ converges (see Fig. 4.2). Furthermore, the best reproduced configurations achieve significantly larger \mathcal{F} values, close to perfect match ($\mathcal{F} = 1$), considering that the test set consists of signals that are not used in training or validation. We do not find a clear trend in \mathcal{F} as a function of M_{tot} (see Fig. 4.2) for any of the KNN models.

The lower row of Fig. 4.2 shows the distribution (histograms) of \mathcal{F} for the configurations of the test set. The KNN models with large N_{train} exhibit a large number of counts at the last two bins with the highest \mathcal{F} (in range of approximately 0.96 to 1.0). In particular, the $\text{KNN}_{\text{tr}=100}^{\text{APR4}}$ model corresponds to no counts below the histogram bin with center at 0.90 (with minimum \mathcal{F} at 0.901). Quantitatively, the difference between the models $\text{KNN}_{\text{tr}=40}^{\text{APR4}}$ to $\text{KNN}_{\text{tr}=80}^{\text{APR4}}$ is negligible with respect to the accuracy of the predictions. In practice, the model $\text{KNN}_{\text{tr}=40}^{\text{APR4}}$ is preferable, as it is computationally less expensive to construct a training set of simulations of size $N_{\text{train}} = 40$ compared to $N_{\text{train}} = 100$.

In addition, we examine the outliers with the lowest \mathcal{F} . As shown in Fig. 4.2, for most KNN models ($N_{\text{train}} \geq 40$), the lowest values of \mathcal{F} are roughly around 0.85. The mean value of \mathcal{F} (see above) represents the average error assigned to a prediction of the KNN models.

Figure 4.3 displays the GW signal obtained from the $\text{KNN}_{\text{tr}=100}^{\text{APR4}}$ model compared to the simulated signal from the test set for the configuration with $M_{\text{tot}} = 2.728 M_{\odot}$. This configuration corresponds to the best performing prediction, with $\mathcal{F} = 0.995$, of the test set. The agreement between these signals is remarkable, as indicated by the high \mathcal{F} . This large \mathcal{F} is a consequence of our large data set as every configuration of the test set has close neighbors with respect to M_{tot} . In this particular case, the test simulation with $M_{\text{tot}} = 2.728 M_{\odot}$, has closest neighbors in M_{tot} from both sides, $M_{\text{tot}} = 2.732 M_{\text{tot}}$ and $M_{\text{tot}} = 2.726 M_{\text{tot}}$. In Appendix C2., we present the five best

and five worst predictions for the simulations of the test set for the KNN models with $N_{\text{train}} = 40, 100$ (see Fig. 4.11 and Fig. 4.12). Overall, we find very good agreement between the initial parts of the two signals (first 10 ms) which is important for achieving large values of \mathcal{F} . Approximately in the first 5-7 ms, the signal also contains information on the secondary frequency components of the spectrum, apart from the dominant component, and thus it is important for resolving those.

In the later parts of the signals, on average, all KNN models, regardless of N_{train} , tend to overestimate the decay of the signal (see also Fig. 4.12). This effect may be a consequence of the chosen global metric, i.e., mean \mathcal{F} , used for the selection of the best-performing KNN models during the hyper-parameter tuning. As the late part of the signal is low in amplitude compared to the early part, leniency in fitting performance of the late part is not penalized as much as that of the early part. We also note that configurations neighbouring in M_{tot} may exhibit small stochastic fluctuations in the decay times of their respective signals, which are triggered by numerical errors, and thus the KNN algorithm cannot resolve those trends over local neighborhoods. Although this systematic trend should be further investigated in the future, we find this effect to be negligible, since the \mathcal{F} of the configurations of the test set are already high.

In summary, the various KNN models reproduce exceptionally well the GW signals of the test set with high precision (\mathcal{F} close to 1) in the whole M_{tot} range. Our results suggest that KNN models with training sets of size $N_{\text{train}} \geq 40$ lead to accurate predictions for test set GW signals with large values of the \mathcal{F} , i.e., mean values roughly around 0.95. The mean of \mathcal{F} appears to converge as the performance of the various KNN models is qualitatively the same for $N_{\text{train}} \geq 40$. However, it should be noted that the \mathcal{F} is a global metric for the match between two signals, and thus depends on the total length of the signals.

4.3.2. Execution time of the model

Furthermore, we examine the average execution time of the KNN models, i.e., the average time required to generate a signal with a duration of 17 ms. As the goal is to use post-merger templates based on a KNN method (or other supervised learning methods) in real parameter estimation efforts, the execution time of the model must be reasonably short. This is because in inference, the model can potentially be called thousands of times.

The speed of the model is largely dominated by the sampling rate used for the generated GW signal. As the sampling rate decreases (increases), the coordinate time t array becomes less dense (more dense), and thus the KNN model generates fewer (more) data points. This results in a shorter (longer) execution time for a call as fewer (more) data points need to be computed. To evaluate the speed of the model, we generate signals for every configuration of the test set and vary the sampling rate $f_s = 1/dt$. For each sampling rate we repeat the procedure a thousand times and store the mean value of the execution times.

Figure 4.4 displays the average execution time for a single call for all KNN models. This corroborates the above statement that the execution time is an increasing function of the sampling rate. For reference, the black dashed line depicts the sam-

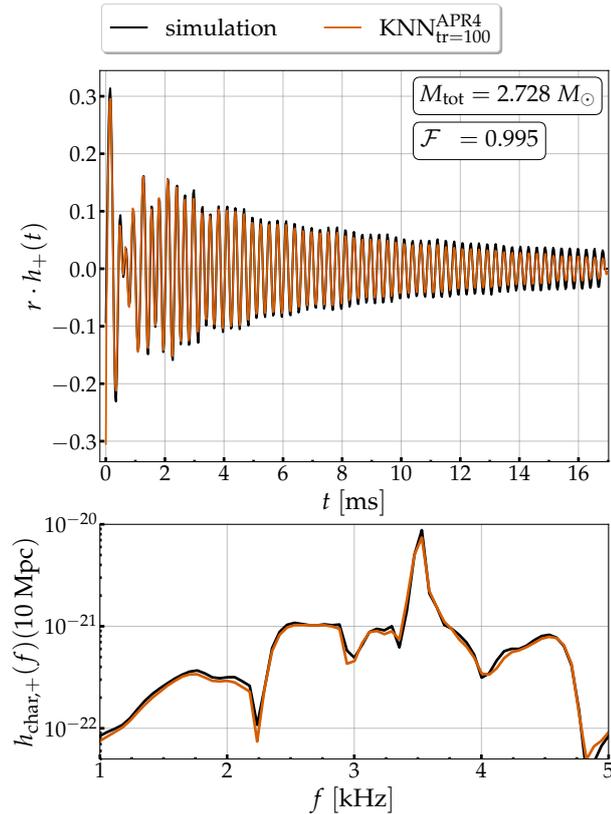


Figure 4.3: Upper panel: Comparison between the time-domain GW signal for a simulation from the test set (black) and the respective $\text{KNN}_{\text{tr}=100}^{\text{APR4}}$ model’s prediction (orange). Bottom panel: The same comparison but for the effective strain $h_{\text{char}}(f)$.

pling rate for Advanced LIGO (aLIGO) data, $f_s = 16384$ Hz. The execution time for all KNN models ranges from 1 – 3 ms for sampling rates comparable to those of aLIGO. We note that the execution time reaches values of ~ 10 ms for unrealistically high sampling rates. This result is very encouraging, as it shows that templates constructed with supervised learning schemes can be employed for parameter estimation, which requires million of waveform evaluations.

Interestingly, as shown in Fig. 4.4, all KNN models exhibit approximately identical execution time curves, implying that the size of the training set does not influence the speed of the templates. This may be explained by the fact that for all KNN models, the *number of neighbors* hyper-parameter (determined during hyper-parameter tuning) is approximately equal ($\sim 8 - 14$). This is encouraging since a full EOS-dependent model will obviously rely on a large training set.

4.3.3. Parameter estimation on simulated data

We discuss the results of our injection study for the configuration of the test set with the total binary mass $M_{\text{tot}} = 2.656 M_{\odot}$ and varied luminosity distance (see Sec. 4.2.4.). The three-detector network of Advanced LIGO and Advanced Virgo at design sensitivity is employed. For each injection, we sample the multi-dimensional

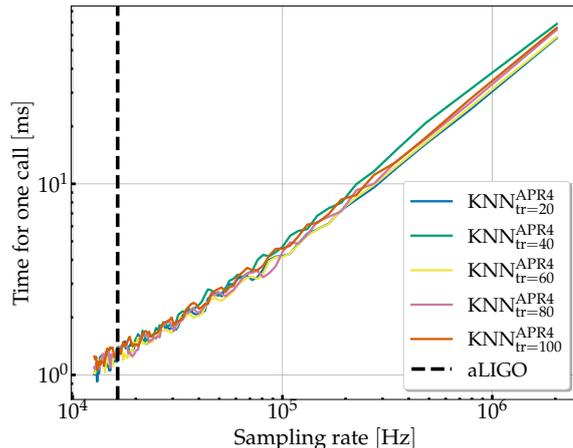


Figure 4.4: Average execution time for one call of the KNN models as function of the sampling rate. Dashed black line indicates the sampling rate of Advanced LIGO.

posterior for M_{tot} , ϕ_0 , t_0 , and D_L , from which the distributions of f_{peak} and \mathcal{O} are derived. To achieve this, we iterate through all samples $(M_{\text{tot}}, \phi_0, t_0, D_L)_i$ of the posterior distribution and generate GW signals $h_+(t; M_{\text{tot},i}, \phi_{0,i}, t_{0,i}, D_{L,i})$ using the KNN models. We then take the Fourier Transform of the GW signals, extract f_{peak} , as described in Sec. 4.2., for all samples, and thus obtain the posterior distribution of f_{peak} . Finally, we derive the posterior distribution of \mathcal{O} by computing the match between each reconstructed GW sample and the GW signal that corresponds to the injected parameters $(M_{\text{tot}}, \phi_0, t_0, D_L)$.

Figure 4.5 shows the 90% credible interval for f_{peak} (left) and the corresponding \mathcal{O} (right) as a function of the luminosity distance of the injected signal for the various KNN models. The black dashed lines correspond to the f_{peak} and \mathcal{F} of the injected simulated signal. As the luminosity distance of the injected signal decreases, the signal-to-noise ratio S/N (Cutler and Flanagan, 1994) increases and thus the ability to reconstruct the signal is enhanced. This can be seen in Fig. 4.5 where the uncertainties in \mathcal{O} , and also in f_{peak} , become smaller at low D_L values.

We find remarkable agreement between the KNN models with $N_{\text{train}} \geq 40$, as all of these models successfully reconstruct the injected signal and achieve high overlaps of approximately $\mathcal{O} \simeq 0.969 - 0.992$. As D_L decreases, the posterior distributions of f_{peak} and \mathcal{O} become narrower, thus exhibiting reconstructed values of f_{peak} and \mathcal{O} with smaller uncertainty at 90% credible intervals. The f_{peak} medians, indicated by the solid curves in Fig. 4.5, approach the corresponding true value of the injected signal. As expected, we observe a drop of the \mathcal{O} as a function of D_L . At a luminosity distance of roughly $D_L = 12$ Mpc, the median \mathcal{O} continues to be relatively high, i.e., $\mathcal{O} \simeq 0.80$, but drops below 0.60 at approximately $D_L = 14$ Mpc. The 90% uncertainty in the reconstructed f_{peak} only becomes large at $D_L \simeq 12$ Mpc.

The $\text{KNN}_{\text{tr}=20}^{\text{APR4}}$ model successfully reconstructs the injected signal. However, it reaches lower median \mathcal{O} values than the other models. At low luminosity distances, the $\text{KNN}_{\text{tr}=20}^{\text{APR4}}$ model achieves a median \mathcal{O} of $\simeq 0.95$. This trend is in agreement with the results discussed in Sec. 4.3.1., i.e., the mean \mathcal{F} for the $\text{KNN}_{\text{tr}=20}^{\text{APR4}}$ model is lower than those of the other models, $N_{\text{train}} \geq 40$, as depicted in Fig. 4.10.

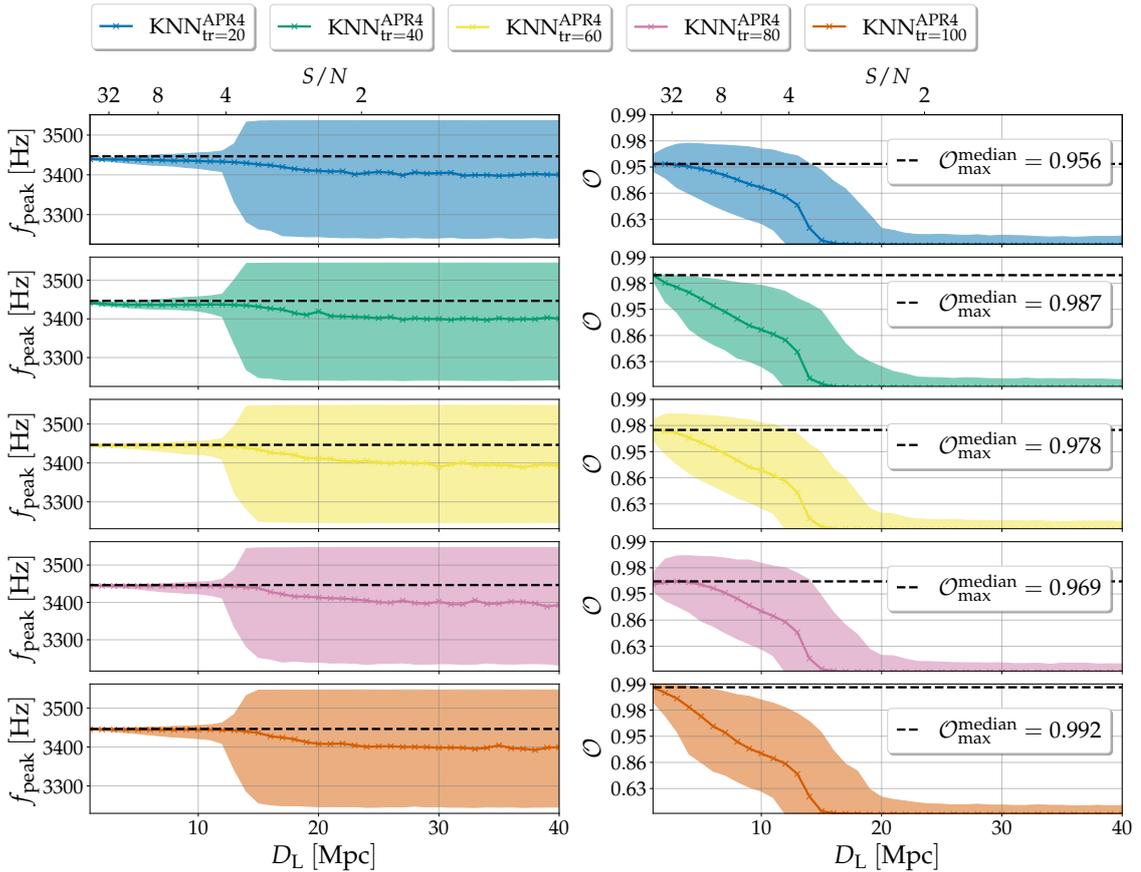


Figure 4.5: Left panel: Posterior of f_{peak} for the simulation of the test set with $M_{\text{tot}} = 2.656 M_{\odot}$ using the KNN models as a function of the luminosity distance D_L of the simulated signal. Black dashed lines indicate the f_{peak} value of the injected signal. Colored curves show the medians, while shaded areas correspond to the 90% credible intervals of the posterior probabilities. Right panel: As in the left panel, but for the corresponding \mathcal{O} values. Black dashed lines indicate the maximum median overlap \mathcal{O} of the injected signal. A three-detector network of Advanced LIGO and Advanced Virgo at design sensitivity is considered. The upper horizontal axis in both panels displays the corresponding S/N .

The injection study provides additional information on the reconstructive capacity of the KNN models under realistic conditions. We observe very good agreement between the models with $N_{\text{train}} \geq 40$, and thus conclude that a training set of 40 simulations suffices for the construction of robust templates.

4.3.4. Dependence on the EOS

As discussed in Sec. 4.3., a training set with a minimum size of $N_{\text{train}} = 40$ suffices to build a robust model that leads to large values of \mathcal{F} and successful injection studies for a fixed EOS model and equal-mass binaries. To further test our findings, we repeat our analysis for the construction of a KNN model using the SFHX EOS.

We only consider the case where $N_{\text{train}} = 40$, since we previously determined that such a size is sufficient to achieve convergence. As before, we split the data set into training ($N_{\text{train}} = 40$), validation ($N_{\text{val}} = 29$), and test ($N_{\text{test}} = 30$) sets. The validation sets between the APR4 and SFHX KNN models have different sizes since the total number of simulations for each EOS model is different. Despite these differences and inconsistencies, the comparison is still meaningful as it enables us to examine the robustness of the predictions, for models constructed with KNN regression, in terms of unknown GW signals, from a test set of simulations, covering a significant binary mass range. In the following, we denote the KNN models with training set of size $N_{\text{train}} = 40$ for the SFHX and APR4 EOS models as $\text{KNN}_{\text{tr}=40}^{\text{APR4}}$ and $\text{KNN}_{\text{tr}=40}^{\text{SFHX}}$, respectively.

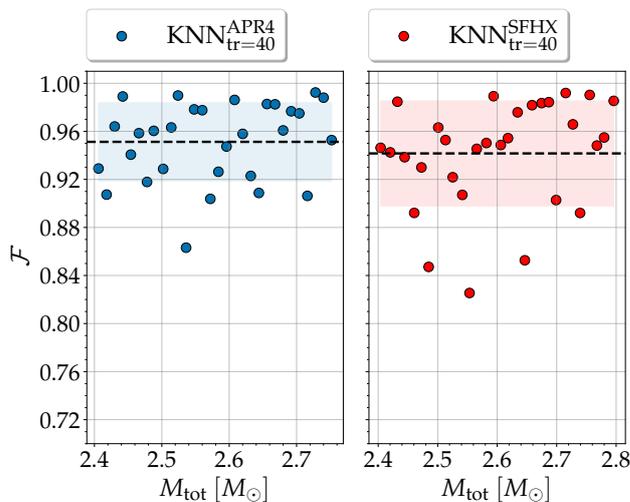


Figure 4.6: Faithfulness \mathcal{F} as a function of M_{tot} for the $\text{KNN}_{\text{tr}=40}^{\text{APR4}}$ (blue) and $\text{KNN}_{\text{tr}=40}^{\text{SFHX}}$ (red) KNN models constructed using a training set of size $N_{\text{train}} = 40$.

Figure 4.6 shows the \mathcal{F} values as a function of the total binary mass M_{tot} for the unseen GW signals of the test sets of the APR4 and SFHX EOS models. It is evident from Fig. 4.6 that the general behavior of the data is qualitatively similar in both cases. Both KNN models exhibit high values of \mathcal{F} where for most of the GW signals the \mathcal{F} values range roughly between 0.90 and 1.0, with some outliers that still have $\mathcal{F} > 0.80$. The mean value of \mathcal{F} for the $\text{KNN}_{\text{tr}=40}^{\text{SFHX}}$ model is slightly lower ($\simeq 1\%$ difference). Both values of \mathcal{F} are high, considering that these signals are not used for training or validation. Furthermore, the $\text{KNN}_{\text{tr}=40}^{\text{SFHX}}$ predictions with the highest \mathcal{F} reach values close to 1 (perfect match) as observed for $\text{KNN}_{\text{tr}=40}^{\text{APR4}}$. The $\text{KNN}_{\text{tr}=40}^{\text{SFHX}}$ model exhibits slightly larger scatter (and more outliers) compared to the $\text{KNN}_{\text{tr}=40}^{\text{APR4}}$ case. However, the lowest values of \mathcal{F} for the $\text{KNN}_{\text{tr}=40}^{\text{APR4}}$ and $\text{KNN}_{\text{tr}=40}^{\text{SFHX}}$ models exhibit comparable values, i.e., larger than 0.85 and 0.82, respectively.

The good agreement between the two KNN models suggests that our results are robust in terms of the \mathcal{F} values achieved, and in fact a training set with a minimum size of 40 simulations may be sufficient to construct accurate templates. We also note that, for practical use, the size N_{train} should be converted to number density defined

EOS	N_{train}	$n \left[\frac{\text{sims.}}{M_{\odot}} \right]$	$N_{\text{per } 0.1M_{\odot}}$	$[M_{\text{tot}}^{\text{min}}, M_{\text{tot}}^{\text{max}}]$	mean \mathcal{F}
APR4	20	55.87	6	2.40, 2.76	0.938
APR4	40	111.73	11	2.40, 2.76	0.951
APR4	60	167.60	17	2.40, 2.76	0.953
APR4	80	223.46	22	2.40, 2.76	0.953
APR4	100	279.33	28	2.40, 2.76	0.956
SFHX	40	100.00	10	2.40, 2.80	0.941

Table 4.1: Number densities and number of simulations per $0.1M_{\odot}$ for all KNN models considered in this work.

as

$$n = \frac{N_{\text{train}}}{M_{\text{tot}}^{\text{max}} - M_{\text{tot}}^{\text{min}}}, \quad (4.17)$$

where $M_{\text{tot}}^{\text{max}}$ ($M_{\text{tot}}^{\text{min}}$) refers to the maximum (minimum) total binary mass of the respective data set. Table 4.1 reports the number densities and the corresponding number of simulations per $0.1 M_{\odot}$, for all KNN models used in this work. For both EOS models, we considered comparable M_{tot} ranges, and thus, the choice of N_{train} corresponds roughly to the same number density. For a stiffer EOS model, where binary configurations with higher M_{tot} can occur, the corresponding N_{train} value may need to be adjusted.

We conduct an additional test to determine whether the KNN models discussed in this work, constructed for single EOSs, can be used to reconstruct general signals, i.e., those originating from simulations where other EOSs are employed. We inject a signal from the SFHX test set, with $M_{\text{tot}} = 2.796 M_{\odot}$, and use the $\text{KNN}_{\text{tr}=40}^{\text{APR4}}$ as reconstruction model. This configuration is chosen so that the characteristic frequencies of its GW spectrum are in the frequency regime of the $\text{KNN}_{\text{tr}=40}^{\text{APR4}}$ model. Both EOS models represent relatively soft EOSs, so there is overlap in the frequency range of their respective f_{peak} as a function of mass. We consider a luminosity distance of $D_L = 5$ Mpc and sample $M_{\text{tot}}, \phi_0, t_0, D_L$ using the setup described in Sec. 4.3.3.. We find that the $\text{KNN}_{\text{tr}=40}^{\text{APR4}}$ model is capable of reconstructing the injected SFHX signal. Figure 4.7 displays the GW spectra and time-domain signals for the injected signal and the reconstructed signal, which maximizes the likelihood. The agreement is impressive, achieving a median overlap of $\mathcal{O} = 0.776$. The spectra match well, particularly for the value of f_{peak} , and there is also relatively good agreement in the subdominant feature, at roughly 2.5 kHz. Therefore, $\text{KNN}_{\text{tr}=40}^{\text{APR4}}$ manages to capture the dominant and approximately the subdominant frequency peaks.

The injected luminosity distance of the source is not reconstructed accurately. This is because the injected signal corresponds to a system with $M_{\text{tot}} = 2.796 M_{\odot}$, while the total mass for the $\text{KNN}_{\text{tr}=40}^{\text{APR4}}$ model that best reconstructs the injected signal is $M_{\text{tot}} = 2.41 M_{\odot}$, which is a much lighter binary, resulting in weaker excitation of the quadrupolar mode.

This result suggests that the effect of the EOS may not be as important as expected, and that the construction of EOS-dependent post-merger models, i.e., a model that is suitable for any EOS, may be accomplished by a moderate number

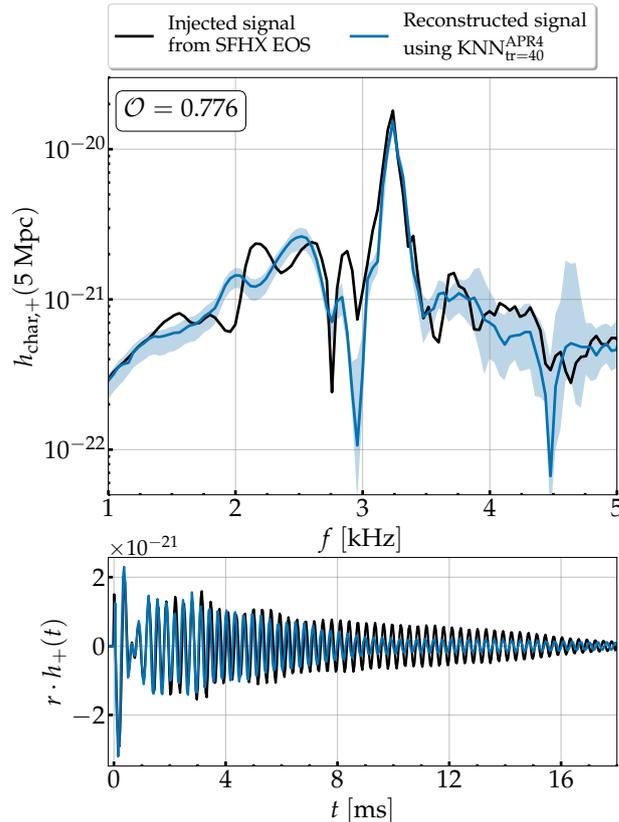


Figure 4.7: Reconstructed signal (blue) using the $\text{KNN}_{\text{tr}=40}^{\text{APR4}}$ model for the injected signal (black) of the simulation from the SFHX test set. Blue shaded areas represent the 90% credible intervals of the posterior probability of the GW spectra.

of simulations. In addition, it may be possible to introduce modifications to extend the capabilities of the models we presented in this work. In addition, it is possible that even without fully general EOS-dependent models, modifications may exist to take advantage of the capabilities of models such as those discussed in this work. For example, shifts in the frequency-domain may be employed to ensure that KNN models, constructed for a single EOS, cover a wider range of f_{peak} values that would allow approximate reconstruction of signals from stiffer or softer EOS models than those discussed here.

4.4. Conclusions

In this work, we construct accurate models for the post-merger GW emission from BNS merger remnants. We use a supervised learning method, i.e., KNN regression, and create a large library of numerical relativity simulations, for a single EOS model (APR4), which includes 157 equal-mass binaries. We split the entire data set into subsets used for training, validation, and testing. To determine the optimal size of the data set required for the construction of reliable templates, we vary the size of the training set for values in $N_{\text{train}} = \{20, 40, 60, 80, 100\}$ and optimize the corresponding KNN models.

We evaluate the performance of the KNN models in terms of the distribution of the faithfulness \mathcal{F} values for the simulations of the test set, but we also monitor those of the validation set. The test set is a holdout set of 30 simulations covering the entire M_{tot} range that enables us to compare the various KNN models. We further explore the robustness of the KNN models by performing injection studies using the three-detector network of Advanced LIGO and Advanced Virgo with a varied luminosity distance of the source. We find that the KNN models achieve high values of \mathcal{F} for completely unseen GW signals from the test set. The predictions with the highest \mathcal{F} , reach values ranging from 0.980 to 0.995, which is remarkably high, considering these signals are not used during the optimization of the KNN models. We find that KNN models with $N_{\text{train}} \geq 40$, exhibit convergence with mean \mathcal{F} approximately equal to 0.95, and standard deviation approximately around 0.032, for the simulations of the holdout test set (see Fig. 4.10).

Furthermore, all KNN models successfully reconstruct the injected signals. Using the posterior probabilities of the M_{tot} , ϕ_0 , t_0 , and D_L , we derive the distributions of f_{peak} and \mathcal{O} , and show that the ability to reconstruct f_{peak} scales with the luminosity distance, as expected. Models with $N_{\text{train}} \geq 40$ exhibit convergence as the medians of f_{peak} approach the injected value. The median \mathcal{O} values, for models with $N_{\text{train}} \geq 40$, are in the range of 0.969 to 0.992, while the $\text{KNN}_{\text{tr}=20}^{\text{APR4}}$ model achieves lower values.

All KNN models, for sampling rates comparable to those employed by Advanced LIGO-Virgo, exhibit execution times that range from 1 to 3 ms, and only slightly increase for unrealistically large sampling rates. Thus, such models can indeed be employed in real detection searches, that require million of waveform evaluations, with future detectors.

Furthermore, we verify our findings using an additional sequence of equal-mass binaries for the SFHX EOS. For this data set, the mean \mathcal{F} and standard deviation for a holdout test set of 30 simulations is 0.941 and 0.0437, respectively. This large value of mean \mathcal{F} qualitatively agrees with the values for the KNN models of the APR4 data set. This further supports the argument that a size of $N_{\text{train}} = 40$, which corresponds to a similar number density of simulations per unit of mass for these two EOS models, suffices for the construction of accurate GW templates.

We show that the APR4 KNN models may potentially be used to reconstruct the dominant component of GW signals from different EOS models. To accomplish this, we perform an injection with a signal from the SFHX data set and utilize the $\text{KNN}_{\text{tr}=40}^{\text{APR4}}$ model as a reconstruction tool. We find that the $\text{KNN}_{\text{tr}=40}^{\text{APR4}}$ model, provided a sufficiently large S/N , accurately reconstructs the f_{peak} component of the GW spectrum and achieves an overlap of $\mathcal{O} = 0.776$. This result indicates that the impact of the EOS might be less significant than expected, and thus the construction of post-merger models that cover all possible EOSs, may be accomplished using a moderate number of simulations. However, it is clear that this test is limited to only a particular region in the parameter space of possible EOS models. The GW spectra of these binary configurations exhibit similarities that may not be present for other EOS models.

Finally, we conclude that a minimum size of $N_{\text{train}} = 40$ simulations used for training, or equivalently a number density of approximately 11 simulations per $0.1 M_{\odot}$ of total binary mass, is sufficient (mean $\mathcal{F} \simeq 0.95$) for the construction of reliable

and accurate GW models for fixed EOS and binary mass ratio of $q = 1$. For models with training sets of $N_{\text{train}} \geq 40$, the mean values of \mathcal{F} for the simulations of the test set converge to the value $\simeq 0.95$. In those models, the maximum median \mathcal{O} values are in the range of 0.969 to 0.992, while the $\text{KNN}_{\text{tr}=20}^{\text{APR4}}$ model achieves lower \mathcal{O} values.

In this work, we have addressed the first of a series of questions toward the construction of a general purpose EOS-dependent GW model for the post-merger emission based on supervised learning methods. We systematically investigated the minimum size of the data set required for developing accurate templates, using the supervised learning method of KNN regression, for a sequence of equal-mass binaries and a fixed EOS model. Our findings are based on this particular method, however, the KNN regression may be indicative of other types of regression methods. However, it is evident that the models discussed in this work need to be extended in order to be applicable to realistic scenarios where the EOS is only partially known. The intrinsic parameters have to be adjusted to include the effects of the EOS and binary mass ratio. In this regard, varying the additional degrees of freedom would significantly increase the number of simulations needed, as discussed in [Whittaker et al. \(2022\)](#). A potential caveat of any model based on simulations is the fact that the simulated GW signal depends on the employed resolution and may be affected by missing physics in the simulations or the finite number of EOS models. However, the first few milliseconds of the post-merger phase, in which most of the GW radiation occurs, may be more reliable than the later evolution, i.e., of the order of tens of milliseconds. In addition, the fact that an APR model successfully recovers a SFHX signal may indicate that these issues are less relevant in practice.

In future work, we plan to investigate the behavior of the remaining degrees of freedom, such as the EOS model, or the binary mass ratio q , and construct templates for the post-merger GW emission that consider those quantities as intrinsic parameters. The effect of the binary mass ratio is likely less dominant in comparison to that of the EOS model. This is because the GW spectra for unequal-mass configurations exhibit similarities to those of equal-mass systems, while the range of characteristic frequencies of the GW spectra are not expected to differ significantly. However, the mechanisms that explain some of the subdominant frequency peaks, such as f_{spiral} (see [Bauswein and Stergioulas \(2015\)](#)), in the GW spectrum depend on the mass ratio, and thus the amplitudes of those may change for binaries with low values of q . The EOS model affects the dynamics of the post-merger system, and thus the characteristic frequencies of the GW spectra. To obtain faithful templates, that achieve large \mathcal{O} values, the model must accurately describe the dominant and subdominant frequency components of the GW spectrum. For this reason, the model must incorporate EOSs, with different degrees of softness and stiffness.

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

C Appendix

C1. KNN models

In this section, we provide additional information regarding the KNN models. We

show histograms of the mass distribution for simulated binary configurations. We present the mean and standard deviation of \mathcal{F} for the simulations of the test set for the various training set sizes employed in this work. Furthermore, we discuss the \mathcal{F} values of the simulations of the validation sets.

Data set of simulations

Figure 4.8 shows the histograms of the simulated binaries for the APR4 and SFHX EOS models. The complete set consists of simulations uniformly distributed with respect to total binary mass. The test set includes simulations distributed in an equidistant manner.

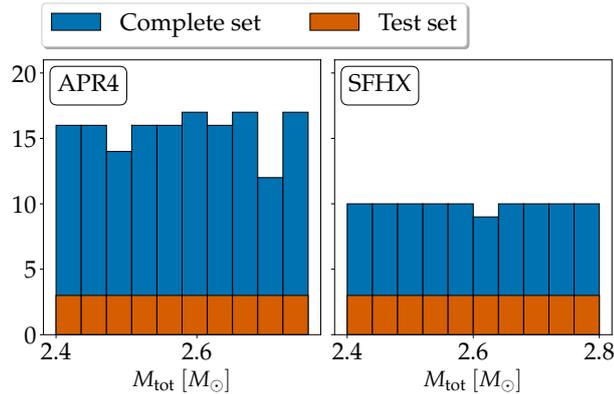


Figure 4.8: Histograms for the simulated binaries for the APR4 (left) and SFHX (right) EOS models. Blue histograms indicate the simulations of the complete set. Orange histograms indicate the configurations of the test set.

Validation data

The validation sets have different sizes (for the various KNN models) that are supplementary to the training sets. The validation sets are used during the grid search for the hyper-parameter tuning to determine the optimal hyper-parameters of the KNN models. Figure 4.9 shows the \mathcal{F} as a function of the total binary mass M_{tot} for the configurations of the validation sets. As in the case of the test set comparison, we find that the mean value of the \mathcal{F} exhibits an increase as the size of the training set increases. I think the mean consistently goes up as you increase the training set. For the $\text{KNN}_{\text{tr}=20}^{\text{APR4}}$ model the mean value slightly drops, however, it still corresponds to a large value of 0.931. The scatter of the data around the mean, expressed by the standard deviation, also decreases as the size N_{train} of the training set increases, and thus the resolution of the KNN model increases.

In addition, all KNN models exhibit some outliers that decrease as N_{train} increases. These outliers tend to occur at the corners of the parameter space (low-mass and high-mass configurations). This is a well-known characteristic of the KNN schemes as these methods rely on neighbors to make accurate predictions, and thus configurations at the edge of the parameter space have more one-sided neighbors. This effect is possibly minimized in the test data, as we did not include the two simulations at the edges

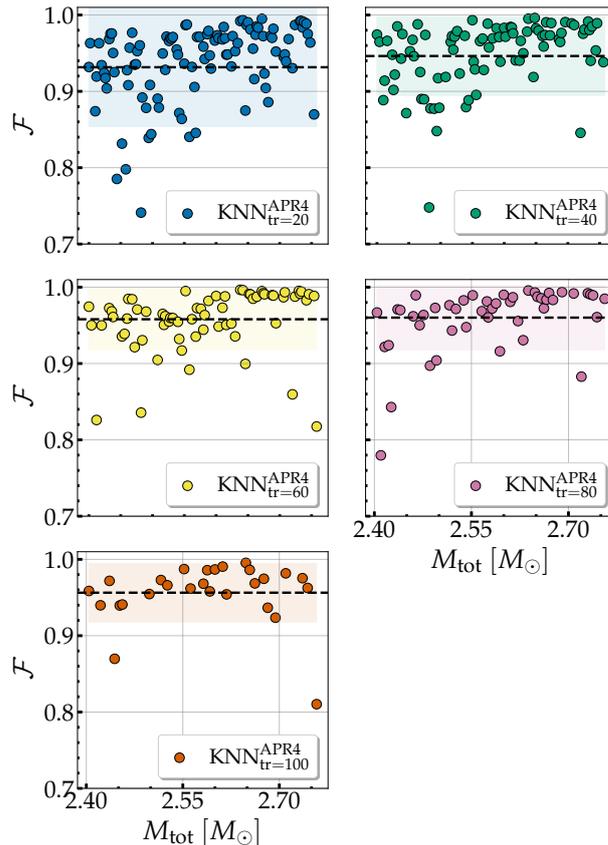


Figure 4.9: \mathcal{F} s for the signals of the validation set, which is used for hyperparameter tuning, as a function of total binary mass M_{tot} for the all KNN models.

(minimum and maximum M_{tot}) and also used a uniform distribution of configurations covering the entire range. We also observe a slight preference for outliers to occur at low M_{tot} , especially for the $\text{KNN}_{\text{tr}=20}^{\text{APR4}}$ model.

Test data

Figure 4.10 displays the mean and standard deviation for the KNN models. There is a small systematic increase in the mean \mathcal{F} as N_{train} increases. Furthermore, the standard deviation decreases as N_{train} increases.

C2. Best and worst predictions over the test data

In this section, we present the predictions of the $\text{KNN}_{\text{tr}=40}^{\text{APR4}}$ and $\text{KNN}_{\text{tr}=100}^{\text{APR4}}$ KNN models for the configurations of the test set that lead to the highest and lowest values of \mathcal{F} . Figures 4.11 and 4.12 show the predicted GW signals in the time domain in comparison to the corresponding simulated signals. This comparison enables us to better understand how \mathcal{F} , which is a global metric for the match between two signals, is affected by the different segments of the signal. For example, fits with high \mathcal{F} show a good match over the whole duration of the signal. In contrast, fits with low

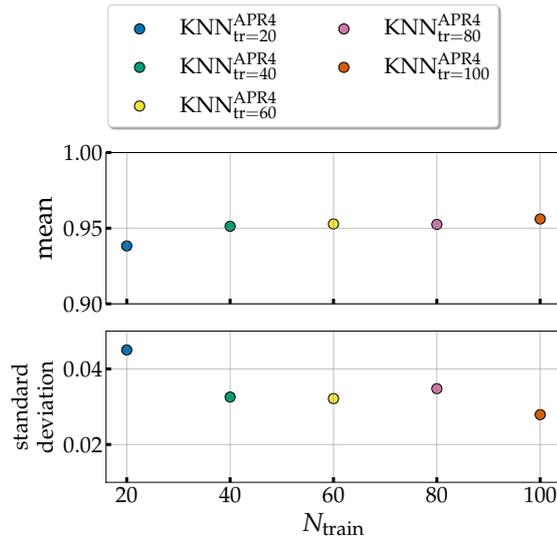


Figure 4.10: Upper panel: Mean \mathcal{F} as a function of the size of the training set N_{train} . Bottom panel: Standard deviation as a function of the size of the training set N_{train} .

\mathcal{F} exhibit certain periods within the signal where there is an apparent mismatch between the oscillations that reduce the overall faithfulness \mathcal{F} .

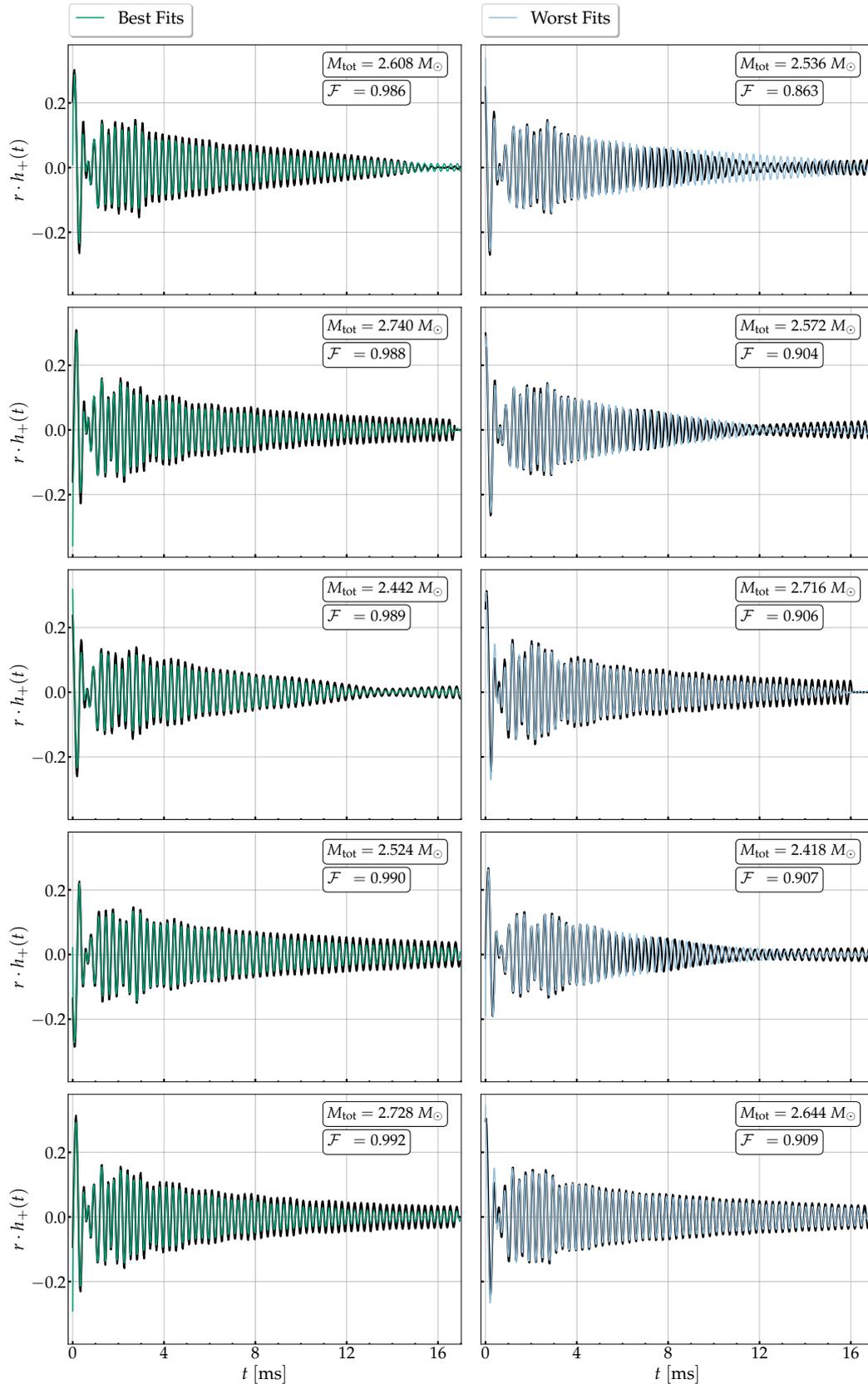


Figure 4.11: Comparison between the predictions of the $\text{KNN}_{\text{tr}=40}^{\text{APR4}}$ model that lead to the highest (green) and lowest (cyan) \mathcal{F} values and the simulated GW signals (black) for the binary configurations of the test set.

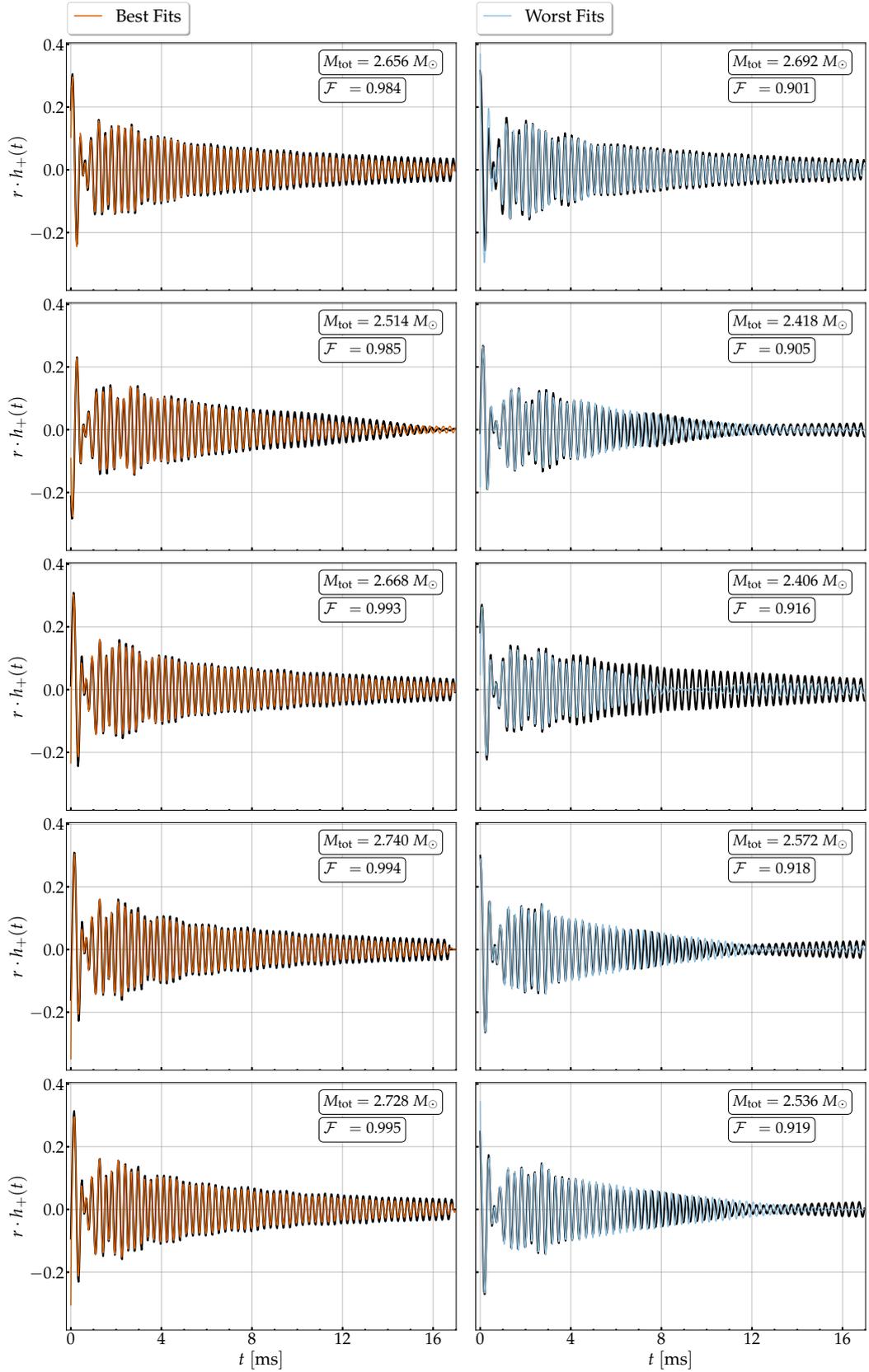


Figure 4.12: Comparison between the predictions of the $\text{KNN}_{\text{tr}=100}^{\text{APR4}}$ model that lead to the highest (orange) and lowest (cyan) \mathcal{F} values and the simulated GW signals (black) for the binary configurations of the test set.

5. Conclusions

There are method development results and scientific results as yields of this PhD research. I summarize and discuss these, and give an outlook on possible follow-up work.

5.1. Main method development results

In what follows, the main method development related outcomes of the three research projects conducted as parts of this thesis are summarized. These are ordered by subject domain:

Stellar evolution (Ch. 2.):

- Stellar evolution modeling across different evolutionary phases is a temporal multi-scale problem. A timescale-adapted evolutionary coordinate has been defined for re-parametrization of stellar evolution from zero age main sequence (ZAMS) up to the end of core helium burning (Sect. 2.2.2.) and beyond (Sect. A1.). Its primary purpose was a simplification of the fitting task by surrogate models, which yielded substantially more accurate predictions over the post-MS evolution (Sect. A1.). It can however be used independent of the fitting task, in particular to sample and visualize variability of stellar variables across scale ranges and across evolutionary phases.
- An efficient general-purpose stellar evolution surrogate model was constructed, covering a ZAMS mass range from $0.65 M_{\odot}$ red dwarves to $300 M_{\odot}$ Wolf-Rayet stars, and stellar lifetimes from ZAMS up to the end of CHeB (Sect. 2.3.3.). It predicts evolutionary tracks and isochrones in the Hertzsprung-Russell and Kiel diagrams (Sect. 2.4.3.). The model achieves a predictive accuracy over classical photometric variables that is at least one order of magnitude lower than typical observational uncertainties, and an efficiency of 1 million point predictions per 40 seconds on a 4-core CPU hardware (Sect. 2.4.1.).
- The hierarchical nearest-neighbor interpolation algorithm for interpolating stellar evolution tracks (Sect. 2.3.4., A2.) was developed, as a more accurate but less efficient (Sect. 2.4.1.) alternative to the deep learning based surrogate model.

CCSNe (Ch. 3.):

- An adjustable scheme was formulated to predict outcomes of iron core collapse (failed or successful SN) at the pre-SN stage based on multiple pre-SN variables

(Sect. 3.3.1.). The scheme is adjustable in the sense that the threshold values for each of the five pre-SN variables involved ($\xi_{2.5}$, s_c , M_{CO} , $\mu_4 M_4$ and μ_4) can easily be modified to incorporate future changes in the SN code that is used to compute the CCSN outcomes.

- A CCSN recipe was constructed (Sect. 3.3.2.) that predicts the remnant type (neutron star, fallback black hole or direct collapse black hole) as a function of M_{CO} and Z , and distinguishes between single and binary-stripped stars of different mass transfer pre-history (case A, case B and case C donors), making it applicable for rapid binary population synthesis studies. The methodology behind the construction of the CCSN recipe is sustainable, because – with future updates to the single and binary-stripped star evolutionary models – it is straightforward to re-derive the critical values $M_{\text{CO}}^{(1)}$, $M_{\text{CO}}^{(2)}$ and $M_{\text{CO}}^{(3)}$ (Sect. 3.3.3.) for direct BH formation from the fits of the explodability proxies. A distinctive feature of our CCSN recipe is that it guarantees explosions for $M_{\text{CO}} \in (8.4, 12.4) M_{\odot}$, independent of Z and binarity, with a more likely NS than fallback BH remnant. Moreover, it admits CCSN explosions at progenitor M_{CO} as high as $15.4 M_{\odot}$.
- The missing red supergiant (RSG) problem (Smartt, 2009) is usually discussed in terms of ranges in ZAMS mass of RSG stellar models that result in pre-SN bolometric luminosities over which RSGs are observed to miss out as Type IIP SN progenitors. Instead, I have re-formulated the missing RSG problem in terms of M_{CO} masses (Sect. ??): there is a lack in Type IIP SN progenitor observations over the range $M_{\text{CO}} \in (5.6, 7.8) M_{\odot}$. The formation of the CO core at the end of CHeB is an important anchor point in the stellar evolution which directly relates to the pre-SN bolometric luminosity, and different M_{CO} core masses can be obtained for the same ZAMS mass as the adopted physics up to the end of core helium burning is changed (Fig. 3.7 in Sect. 3.3.2.).

Gravitational waves (Ch. 4.):

- A gravitational wave template, predicting (equal mass) binary neutron star post-merger emission in time domain as function of the source parameters total binary mass M_{tot} and two different fixed EOSs (Sect. 4.2.2.), was constructed. The M_{tot} parameter space spans the entire range from twice the minimal NS mass assumed at birth in a CCSN ($1.2 M_{\odot}$) to the EOS-dependent threshold mass M_{thres} for prompt black hole formation. The model achieves mean faithfulness scores $\mathcal{F} \geq 0.938$ (Sect. 4.3.1.) and an execution time of 1-3 ms (Sect. 4.3.2.) for sampling rates comparable to those of aLIGO, making it applicable for future online searches of source signals.
- The minimal parameter space sampling density in total binary mass M_{tot} , required for reliable reconstruction of the equation of state dependent peak oscillation frequency from a colored noise background (Sect. 4.3.1. and 4.3.3.), was determined to be 11 simulations per $0.1 M_{\odot}$. This sampling density result gives a first-order estimate of the number of numerical relativity simulations

required to extend the parameter space of the GW template model by covering a greater number of EOSs over the full total binary mass range.

5.2. A cross-comparative and integral view

In this section, the three research projects are considered in their integrity, to emphasize how they compare and relate to one other.

Table 5.1: Cross-comparison of the three research projects conducted in this thesis. It outlines the evolutionary part modeled by emulator-based techniques, the references to the adopted physics behind the data set used to fit the models, the supervised learning models used to address the fitting task, the simplifying step introduced into the formulation of the regression problem, the predicted observables and the long-term research goal associated with each of the projects.

Evolutionary part		Adopted physics
(1)	ZAMS - end of CHeB	Choi et al. (2016)
(2)	End of CHeB - iron core collapse	Schneider et al. (2021, 2023) Müller et al. (2016)
(3)	BNS post-merger phase	Soultanis et al. (2024)
Predictive model	Simplifying step	
(1)	ffNN	timescale-adapted evolutionary coordinate
(2)	GPR	pre-SN explodability criterion
(3)	KNN	phase alignment
Observables	Long-term research goal	
(1)	$\log L$, $\log T_{\text{eff}}$, $\log g$	test of stellar evolution model / parameter inference
(2)	SNe	Re-estimation of CO merger and CCSN rates
(3)	GWs	Constraints on the nuclear EOS

Table 5.1 compares the three projects systematically. Each of the predictive models constructed in this work addresses a different evolutionary part, which relates to its consecutive part:

1. The stellar evolution model (Sect. 2.4.1.) predicts evolutionary tracks from ZAMS up to the end of CHeB. To this end, an intermediate step is introduced to map stellar age onto the timescale-adapted evolutionary coordinate, which is used instead of stellar age to trace the evolution of stars.
2. The fits of the pre-SN variables as a function of M_{CO} (Sect. 3.3.3.) bridge the evolution from the end of core helium burning through the late burning phases up to the onset of iron core collapse. These are then used for final fate and remnant type prediction over single and binary-stripped star progenitors using the pre-SN explodability criterion.

3. Granted that two neutron stars get formed through one of the stellar binary evolutionary channels, and that these are found close enough to merge within the Hubble time (these two conditions are assumed, their statistical modeling is out of scope of this thesis), the gravitational wave model traces the BNS merger remnant evolution over the entire total binary mass range up to prompt collapse.

An advantage of the integral predictive framework elaborated in this thesis is its modular structure: when the adopted physics behind each evolutionary part – be it (1), (2) or (3) – is replaced, the predictive surrogate models need to be re-fitted on an updated data base independent of the other evolutionary parts.

For each of the fitting tasks, a different surrogate model class was found to achieve best performance:

- a feedforward neural network (ffNN) was most powerful in generalizing the prediction of stellar evolution tracks,
- gaussian process regression (GPR) yielded the most accurate interpolation of the pre-SN variables, and
- k -nearest neighbor regression (KNN) was the most appropriate model choice for resolving local variability structures in the gravitational wave parameter space.

Their hyperparameters were optimized using different techniques (manual tuning, randomized search and grid search under k -fold cross-validation), and the performance metrics (see Sect. 2.2.4. for scoring the prediction of stellar evolution, and Sect. 4.2.2. for scoring the prediction of gravitational waves; the mean squared error was used to score the prediction of pre-SN variables) were tailored to the application context.

In all of the three projects, a simplifying step was introduced when formulating the regression problem. This step enhanced the accuracy of the fits compared to the brute-force approach, by altering the distribution of data in the parameter space or by reducing the dimensionality of the regression problem. The introduction of the timescale-adapted evolutionary coordinate addressed the temporal multi-scale problem of stellar evolution modeling across phases. The alignment of the gravitational waves in-phase over the first few ms (Sect. 4.2.2.) allowed to accurately fit the most relevant part of the gravitational wave signal, which has the largest strain amplitude and which carries information about the peak oscillation frequency. The formulation of the pre-SN explodability criterion to the M16 SN code made it sufficient to predict the five explodability proxies – which are derived from the pre-SN progenitor structure, and required for evaluation of the criterion – instead of predicting entire pre-SN stellar structure profiles – needed as input for running the M16 1D SN code – as a function of M_{CO} , in order to obtain the resulting CCSN outcomes.

All three research projects lead to the prediction of observables: classical photometric variables are predicted by the stellar evolution emulator (Ch.2.), SNe by the explodability formalism (Ch. 3.) and gravitational waves by the template model (Ch.4.). These immediately connect to the long-term research goal associated with

the construction of the surrogate models. The stellar evolution emulator (in the way it is currently set up) allows to test the adopted physics behind the catalog upon which it was trained, by comparison of model predictions to observations of stellar populations. The emulator can also be used for iterative optimization to infer the stellar age and the ZAMS mass of a star (given a single observation), or to estimate the age of a stellar population (by comparison of model isochrones to population observations). The construction of the M_{CO} -based CCSN recipe was tailored for application in rapid binary population synthesis codes. Its integration in these codes will allow to re-assess the CCSN rates and binary compact object merger rates (BH-BH, BH-NS, NS-NS).

5.3. Main scientific results

The research project reported in Ch. 2. was a proof-of-concept study, in which the elaboration of a sustainable methodology for fitting stellar evolution tracks over a large stellar parameter space was put into the foreground. Since binary neutron star post-merger gravitational waves have too high frequencies for detection with current interferometry, the template constructed in the research project reported in Ch. 4. cannot be used for detection of gravitational waves at present day. However, scientific conclusions were drawn regarding their detectability and the dependence of the gravitational wave signal morphology on the source parameters. These are presented in [Soultanis et al. \(2024\)](#). The main scientific findings were made in the research project associated with Ch. 3., which is based on the to-be-submitted paper [Maltsev et al. \(2024a\)](#). The main results are summarized below:

Detectability of GWs from BNS merger remnants and dependence of the GW morphology on source parameters (Ch. 4.):

- A gravitational wave (GW) from a neutron star merger remnant at one EOS can be characterized by the same main peak oscillation frequency and similar sub-dominant frequency peaks as a gravitational wave from a NS merger remnant at another EOS (Sect. 4.3.4.). This is the case if the NS binary of the stiffer EOS has a higher total mass M_{tot} compared to the softer EOS. It is therefore difficult to disentangle the effect of M_{tot} and of the EOS on the morphology of the GW signal, and to use the detection of GWs from the NS post-merger phase to constrain competing models on the nuclear EOS. These two reference GW signals will, however, differ in strain amplitude, since a lighter binary results in a weaker excitation of the quadrupolar mode. For a fixed signal-to-noise ratio of the detecting interferometry, this will therefore lead to differences in luminosity distance estimates: the GW from the stiffer EOS and the more massive binary will be predicted to have originated closer to Earth than the GW from the softer EOS and the less massive binary.
- While the less pronounced effect of the EOS on the GW morphology than expected makes it more difficult to select EOS models upon detection of a source GW, it makes it easier to detect one (Sect. 4.3.4.). A detection of a

GW from a BNS merger remnant is therefore possible even if the EOS model is inaccurate, further away from the real conditions that govern the actual astrophysical source.

- BNS post-merger GW signals can be reconstructed reliably at advanced LIGO sensitivity, i.e. at a high overlap and narrow posterior probability distributions on f_{peak} , if the source is close enough (a luminosity distance of $D_L \simeq 12$ Mpc (Sect. 4.3.3.).

Pre-determination of the final fate and compact remnant type at the onset of iron core infall:

- The explodability formalism, which was developed based on outcomes of the M16 CCSN model (Sect. 3.2.2.) over a heterogeneous super-set of $\simeq 3900$ pre-SN progenitors (Sec. 3.2.1.) is in broad agreement with 3D CCSN simulation outcomes from the Garching and the Monash groups (Sec. 3.3.1.). The formalism thereby improves previous approaches to predict the outcome of iron core collapse, by using multiple explodability proxies – derived from the pre-SN progenitor structure – in concert. These are the compactness parameter $\xi_{2.5}$, the central specific entropy s_c , the carbon-oxygen core mass M_{CO} and the E16 parameters μ_4 and $\mu_4 M_4$ that scale with the mass accretion rate and the accretion luminosity, respectively. To address the question whether a given pre-SN progenitor will explode or not, the explodability criterion can be used to make a final fate prediction at high confidence, without the need to run expensive 3D CCSN simulations in order to get the answer.
- Black hole formation by fallback in a successful SN explosion is rarer than NS formation ($< 13\%$ of exploding progenitors from the super-set of stellar models over M_{CO} values in-between peaks in $\xi_{2.5}$, and none otherwise), and can be excluded at pre-SN stage if $\xi_{2.5}$ is small compared to $\mu_4 M_4$, and M_4 large compared to M_{CO} , respectively (Sect. 3.3.1.).

Explodability systematics in the (M_{CO}, X_C) plane at the end of CHeB:

- Outcomes of CCSN simulations not only are encoded in the pre-SN progenitor structure, but – provided that the model for the late burning phases is fixed – pre-determined already at the end of core helium burning (Sect. 3.3.3.). Differences in CCSN outcomes for stars of the same M_{CO} but different Z , of the same M_{CO} and Z but different mass transfer pre-history, or of the same M_{CO} , Z and mass transfer pre-history but different input physics such as convective core overshooting, can all be traced back to differences in the central carbon mass fraction X_C at the end of core helium burning (see Fig. 3.10 and 3.6, and Sect. 3.3.2.).
- When applying the pre-SN explodability criterion introduced in Sect. 3.3.1. to a densely sampled model grid of bare carbon-oxygen cores (Patton and Sukhbold, 2020), evolved from the zero age core carbon burning up to onset of iron core infall over a parameter space spanned by (M_{CO}, X_C) , the resulting final fate

landscape is radically different: it features a less scattered distribution of CCSN outcomes and is more optimistic about successful SNe than the one resulting from the standard E16 criterion (Fig. 3.12, Sect. 3.3.2.). E16 is currently widely used for binary population synthesis, while it does not agree with 3D CCSN simulation outcomes, in contrast to our formalism (Fig. 3.3).

Constraints on stellar evolution models and on CCSN recipes from SN observations:

- The explodability systematics in the aforementioned (M_{CO}, X_C) plane, featuring only two “islands” of failed SN outcomes for $M_{\text{CO}} \leq 10 M_{\odot}$ constrains stellar evolution models to have specific value ranges in M_{CO} and X_C in order to land on these at the end of core helium burning, for partially addressing the missing red supergiant problem (Smartt, 2009) by direct BH formation (see Fig. 3.14 and Sect. 3.3.3.). This implies, first, that adopted physics such as the $^{12}\text{C}(\alpha, \gamma)^{16}\text{O}$ rate during core helium burning, the wind mass loss scheme and/or convective core overshooting must be suitably configured to yield the appropriate X_C values for resulting in failed SN outcomes over the M_{CO} interval of interest. This constraint may serve as an indirect test of stellar evolution models. The second implication is that failed SNe cannot be the only reason for the lack of RSGs as Type IIP progenitors: other physical mechanisms need be invoked on top for its explanation.
- The CCSN recipe introduced in Sect. 3.3.2. is in better agreement with observations that constrain CCSN progenitors than other M_{CO} -based recipes Mapelli et al. (2020), Mandel and Müller (2020) and Fryer et al. (2012) that are widely used in binary population synthesis codes. The CCSN recipe is consistent with the most luminous Type IIP, IIL, IIb and Ib progenitor (direct imaging) observations (Sect. 3.3.3.) while at the same time partially addressing the missing red supergiant problem by failed SNe. Furthermore, it explains the Type IIn SN2010jl (Smith et al., 2011), the Type Ic SN2011bm (Valenti et al., 2012) and the most oxygen-rich SN remnants (Fang et al., 2022), which other recipes are challenged to explain (Sect. 3.3.3.).

5.4. Outlook

There are several directions for extending the research conducted in this thesis further by building upon the aforementioned results. These are again ordered by research project.

Stellar evolution (Ch. 2.): The stellar evolution emulator – in its current version – can be used, as already anticipated, for the following three astrophysical applications:

1. a test of the input physics adopted for the Choi et al. (2016) stellar evolution models, which are emulated by the surrogate model constructed in Ch. 2., by comparison against population observations,

2. age determination of a stellar population, such as an open cluster, by comparison of isochrones predicted by the emulator to population observations in the Hertzsprung-Russell and Kiel diagrams, and
3. estimation of the age and the ZAMS mass of a single star, by optimization techniques that iteratively evaluate the emulator at different points in parameter space,

provided that the astrophysical sources can be constrained to have a metallicity close to solar, Z_{\odot} . An obvious way to extend the current version of the stellar evolution surrogate model is to enlarge the regressor parameter space to include the ZAMS metallicity Z_{ini} . Following Sect. 2.2.1., the two-step interpolation scheme then reads

Step 1 (age proxy fit) $f_1: (\log \tau, \log M_{\text{ini}}, \log Z_{\text{ini}}) \mapsto s$,

Step 2 (observables fit) $f_2: (s, \log M_{\text{ini}}, \log Z_{\text{ini}}) \mapsto (Y_L, Y_T, Y_g)$.

Similarly, other observables could be added to the set of target variables predicted by the emulator, such as surface element abundances. To this end, in order to avoid the alignment problem (Sect. A5.), I have argued that training multiple emulators that each predict a single target variable is a better approach than training a single emulator predicting multiple target variables. These two extensions will not only increase the realm of applicability of the emulator for the aforementioned science cases, but also for the other ones mentioned in Sect. 2.1.:

- Population synthesis codes,
- stellar N -body dynamics codes and
- large-scale galactic evolution simulations

all are astrophysical applications that could benefit from a stellar evolution surrogate model, as an alternative to SSE (Hurley et al., 2000) or to the simpler approaches that integrate effects of stellar evolution over a larger scale. These astrophysical applications require adaptation of the emulator set-up compared to the one presented in this work. In order to replace SSE in population synthesis codes that are based on SSE (and on BSE (Hurley et al., 2002)) – such as COMPAS and BINARY C – by a stellar evolution surrogate model that is based on a more recent stellar evolution catalog, the surrogate model needs to predict the same set of variables over the same span of stellar parameter space as is currently predicted by SSE. In stellar N -body dynamics codes, the target variables of interest are those that are relevant for the dynamical evolution, such as present-day mass, radius and rotation. In large-scale galactic evolution simulations, the target variables of interest are those that contribute to stellar feedback and chemical evolution, such as bolometric luminosity, stellar wind rates, compositional variables and nucleosynthesis yields. The advantage of the methodological framework elaborated in this thesis is that the same stellar evolution surrogate modeling method can be applied to an arbitrary choice of the target variable of interest.

CCSNe (Ch. 3.): The pre-SN explodability criterion formulated in this work has been tested against 3D CCSN simulations performed in the Garching (Germany) and Monash (Australia) research groups (Sect. 3.3.1.). An insightful follow-up study would be the test of the criterion against the 3D CCSN simulation outcomes obtained at some of the other few research centers around the world that perform these simulations: the Stockholm (Sweden), Fukuoka (Japan), Princeton and Oak Ridge (USA) groups.

As already discussed in Sect. 3.4., a profound step forward would be the comparison of 3D CCSN simulations with the M16 model not only in regard of the final fate (successful or failed SN), but also in regard of the modeled core collapse physics. This research requires running the M16 code over a few selected pre-SN progenitors (such a non-exploding, an exploding high-mass and an exploding low-mass progenitor) over which 3D CCSN simulations have been performed, and comparing the predicted size of the gain region, the shock radius evolution, the core temperature conditions, the explosive nuclear burning, the PNS mass growth and other processes modeled in 1D and 3D. The expected outcome is that the free M16 1D model parameters will need to be tuned for an improved agreement with the 3D CCSN simulation, presumably with parameter choices that depend on the progenitor. On the other hand, insight from 3D simulations of the ultimate stages between the end of the core silicon burning phases and the onset of iron core collapse can be used to improve the estimation of the α_{turb} parameter. As stressed in Müller and Janka (2015), the root-mean-square averaged turbulent Mach number informs the value choice for α_{turb} . It needs be greater, the greater are the perturbations to spherically symmetric pre-SN progenitor stratifications, induced by late phase oxygen-neon shell mergers, convective silicon burning and other effects. Again, presumably a progenitor-dependent choice of α_{turb} will be the most appropriate modeling assumption for M16 that better agrees with 3D simulations. After sophistication in the choice of the M16 model parameters, the M16 code needs to be re-run over a large sample of pre-SN progenitors (such as those considered in Sect. 3.2.1.) to re-evaluate the final fate outcomes. Any substantial differences therein will need to be reflected in a re-formulation of the pre-SN explodability criterion. In simplest way, this would go by retaining the scheme proposed in Sect. 3.3.1. and updating the critical values in the pre-SN variables $\xi_{2.5}$, s_c , $\mu_4 M_4$ and fit parameters (k_1, k_2).

A M_{CO} -based CCSN recipe in rapid BPS needs to assume a model for the advanced burning phases that bridges evolution from the cut-off point (such as central carbon ignition) to the pre-SN stage (and to the final fate). While the Schneider et al. (2021) model was assumed for the recipe constructed in Sect. 3.3.2., differences in the adopted physics for the late burning phases will lead to different pre-SN variables for the same progenitor (such as, for example, a single star of $M_{\text{CO}} = 6 M_{\odot}$ at $Z = Z_{\odot}$). For a fixed pre-SN explodability criterion, the differences in the pre-SN variables will result in shifts of the critical values $M_{\text{CO}}^{(1)}$, $M_{\text{CO}}^{(2)}$ and $M_{\text{CO}}^{(3)}$ that delineate the direct collapse threshold values.

As already mentioned, a work-in-progress is the implementation of the current version of the M_{CO} -based CCSN recipe into a rapid BPS code, and its use to re-estimate BBH merger rates. Since the recipe developed in this research thesis is considerably more optimistic about successful SN explosions at higher M_{CO} than

other recipes that are widely used, such as F12, the anticipated result is that the BBH merger rates will be lower than those predicted with those. The consequences for gravitational wave astronomy will need to be further explored.

Gravitational waves (Ch. 4.): An optimistic preliminary result obtained in Sect. 4.3. is that a more general gravitational wave model, i.e. one that depends on a greater number of BNS source parameters, is achievable with a manageable number of numerical relativity simulations while being accurate and efficient enough for real-time searches of signals at advanced LVK sensitivity. The next step is the construction of a gravitational wave model which predicts the BNS post-merger emission at a variable mass ratio q as an additional degree of freedom. This means that the regression problem introduced in Sect. 4.2.2. needs to be extended as

$$h_+(t) = \text{KNN}(t - t_0; M_{\text{tot}}, q) e^{i\phi_0}, \quad (5.1)$$

$$h_\times(t) = h_+(t - t_0) e^{i\frac{\pi}{2}}. \quad (5.2)$$

Following the methodology elaborated in Sect. 4., the minimal parameter space sampling density for reliable reconstruction of source signals at colored noise conditions will now need to be determined over the extended (M_{tot}, q) parameter space, at a fixed EOS. For a more general conclusion, it is of interest to repeat this study for an exemplary soft EOS – such as APR4 – and a stiff EOS.

While this extension of the parameter space is straightforward, the more challenging task is the parametrization of the post-merger emission at a variable EOS. The resulting gravitational wave model would predict

$$h_+(t) = \text{KNN}(t - t_0; M_{\text{tot}}, q, \text{EOS}) e^{i\phi_0} \quad (5.3)$$

as a function of the mass parameters M_{tot} and q , and at variable EOS assumptions. The optimal method for parametrizing the dependence of the GW emission on different EOSs using the same predictive framework has yet to be found. One possible approach, applicable to cold EOSs, is to use the polytropic EOS

$$P(\epsilon) \propto \epsilon^\gamma \quad (5.4)$$

where P is the inner pressure, ϵ the energy density and γ the polytropic index. For a fixed EOS, the polytropic model could be used to optimize the choice of γ over different ranges in energy density. For example, three different indices γ_1, γ_2 and γ_3 could be the modeling choices over the ranges $(\epsilon_a^{(1)}, \epsilon_b^{(1)})$, $(\epsilon_a^{(2)}, \epsilon_b^{(2)})$ and $(\epsilon_a^{(3)}, \epsilon_b^{(3)})$, respectively. Here, $\epsilon_a^{(i)}$ and $\epsilon_b^{(i)}$ define the upper and lower limits of the density interval, over which the polytropic index $i = 1, 2, 3$ is the appropriate choice. Different nuclear EOS

$$\text{EOS} \leftarrow (\gamma_1, \dots, \gamma_n) \quad (5.5)$$

are then characterized by different values in $\gamma_1, \dots, \gamma_n$, where n is the number of spectral indices used for the piecewise polytrope modeling.

A different, more general approach to parametrizing the EOS is to use deep learning based techniques for encoding different EOS in a latent space. This approach has already been tested in [Whittaker et al. \(2022\)](#), but requires further exploration. The idea is to use an encoder neural network architecture to find a low-dimensional EOS representation using latent variables. Each EOS, is then encoded by a set of variables $\alpha_1, \dots, \alpha_k$,

$$\text{EOS} \leftarrow \alpha_1, \dots, \alpha_k, \quad (5.6)$$

that span a latent space of dimension k , vectorized by $\vec{\alpha}$.

The construction of a general GW model that takes the EOS dependence of the BNS post-merger emission into account is necessary to address the long-term scientific goal: to help facilitate a future detection of a BNS post-merger GW signal, with interferometry at aLIGO sensitivity or next-generation detectors, that can be used to derive constraints on the nuclear EOS.

Acknowledgments

There are several people who contributed to accomplishment of this thesis, whom I owe my deep gratitude. First of all, dear Fritz (Röpke) and Fabian (Schneider), thank you for the opportunity to carry out this research under your guidance, for the knowledge you shared and the constructive feedback. There is a lot I learned from you, Fritz and Fabian, and there still is much more to learn. You remain role models for me in several regards.

Next, what would PhD life have been without the great colleagues in the PSO group? Be it the fruitful science-related discussions, computer-science related advice, or just the many laughs we had during our relaxing lunchtime breaks or morning coffees: Giovanni, Javier, Robert, Marco, Alex, Vijay, Mike, the former members such as Hans, Chris and Theo, and the recent newcomers - thank you for having been such an amazing team! We have truly grown together over the years.

Dear Sara, thank you for your loving support, encouragement and patience with me.

Dear parents, I am very grateful for your continuous care, advice and curiosity that helped me all the way along. Dear sister Mascha, our exchange on art & science related topics, and my admiration of your paintings have always been source of inspiration and creative thought on some sub-conscious level for me. My sweet niece Pauline and nephew Gregory, you radiate so much energy, and we have such truly fun times when we see one another.

Christian Fendt, I am thankful for having had the possibility to be part of the IMPRS graduate program. It has shaped my background knowledge in astrophysics by so much when I look back. IMPRS graduate school fellows from the 16th generation – it has been great to exchange with you, share ideas and learn from one another.

List of publications

Papers used in this thesis:

- [1] **K. Maltsev**, F. R. N. Schneider, F. K. Röpke, A. I. Jordan, G. A. Qadir, W. E. Kerzendorf, K. Riedmiller and P. van der Smagt; *Scalable stellar evolution forecasting. Deep learning emulation versus hierarchical nearest-neighbor interpolation*, 2024, <https://ui.adsabs.harvard.edu/abs/2024A&A...681A..86M>; A&A 681 A86.
- [2] **K. Maltsev**, F. R. N. Schneider, I. Mandel, A. Heger, B. Müller, F. K. Röpke and E. Laplace; *Explosibility criteria for the neutrino-driven supernova mechanism*; to-be-submitted to A&A.
- [3] T. Soultanis, **K. Maltsev**, A. Bauswein, K. Chatziioannou, F. K. Röpke and N. Stergioulas; *Gravitational-wave model for neutron star merger remnants with supervised learning*; arXiv:2405.09513; <https://ui.adsabs.harvard.edu/abs/2024arXiv240509513S>; submitted to Phys. Rev. D.

Papers not used in this thesis:

- [1] **K. Maltsev**; *Thermodynamics of Classical Schwarzschild Black Holes*; <https://doi:10.1134/S1063772921100218>; Astron. Rep. 65 10, 2021.
- [2] **K. Maltsev**; *Stellar gravitational collapse, singularity formation and theory breakdown*; https://doi:10.1142/9789811269776_0298; World Sci., ISBN 9789811269776, 2023.
- [3] K. Riedmiller, P. Reiser, E. Bobkova, **K. Maltsev**, K., G. Gryn'ova, P. Friederich and F. Gräter; *Substituting density functional theory in reaction barrier calculations for hydrogen atom transfer in proteins*; <https://doi:10.1039/D3SC03922F>, Chem. Sci. vol. 15, 2024.

Bibliography

- Aasi, J. et al. (2015). Advanced LIGO. *Class. Quant. Grav.*, 32:074001.
- Abbott, B. P., Abbott, R., Abbott, T. D., Abraham, S., Acernese, F., Ackley, K., Adams, C., Adhikari, R. X., Adya, V. B., Affeldt, C., et al. (2020). GW190425: Observation of a Compact Binary Coalescence with Total Mass $\sim 3.4 M_{\odot}$. *Astrophys. J. Lett.*, 892(1):L3.
- Abbott, B. P., Abbott, R., Abbott, T. D., Acernese, F., Ackley, K., Adams, C., Adams, T., Addesso, P., Adhikari, R. X., Adya, V. B., et al. (2017a). GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral. *Phys. Rev. Lett.*, 119(16):161101.
- Abbott, B. P., Abbott, R., Abbott, T. D., Acernese, F., Ackley, K., Adams, C., Adams, T., Addesso, P., Adhikari, R. X., Adya, V. B., et al. (2017b). Search for Post-merger Gravitational Waves from the Remnant of the Binary Neutron Star Merger GW170817. *Astrophys. J. Lett.*, 851(1):L16.
- Abbott, B. P. et al. (2017). Exploring the Sensitivity of Next Generation Gravitational Wave Detectors. *Class. Quant. Grav.*, 34(4):044001.
- Abbott, B. P. et al. (2018a). GW170817: Measurements of neutron star radii and equation of state. *Phys. Rev. Lett.*, 121(16):161101.
- Abbott, B. P. et al. (2018b). Prospects for Observing and Localizing Gravitational-Wave Transients with Advanced LIGO, Advanced Virgo and KAGRA. *Living Rev. Rel.*, 21(1):3.
- Abbott, B. P. and et al. (2019). *Detectability of Gravitational Waves from Binary Compact Object Mergers*. Springer.
- Abbott, B. P. et al. (2019). Properties of the binary neutron star merger GW170817. *Phys. Rev. X*, 9(1):011001.
- Abbott, B. P. et al. (2020). Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA. *Living Rev. Rel.*, 23(1):3.
- Acernese, F. et al. (2015). Advanced Virgo: a second-generation interferometric gravitational wave detector. *Class. Quant. Grav.*, 32(2):024001.
- Achille, A., Paolini, G., Mbeng, G., and Soatto, S. (2021). The information complexity of learning tasks, their structure and their distance. *Information and Inference: A Journal of the IMA*, 10(1):51–72.

- Ackley, K. et al. (2020). Neutron Star Extreme Matter Observatory: A kilohertz-band gravitational-wave detector in the global network. *Publ. Astron. Soc. Austral.*, 37:e047.
- Adams, S. M., Kochanek, C. S., Gerke, J. R., and Stanek, K. Z. (2017). The search for failed supernovae with the Large Binocular Telescope: constraints from 7 yr of data. *Mon. Not. Roy. Astron. Soc.*, 469(2):1445–1455.
- Agrawal, P., Hurley, J., Stevenson, S., Szécsi, D., and Flynn, C. (2020). The fates of massive stars: exploring uncertainties in stellar evolution with METISSE. *Mon. Not. Roy. Astron. Soc.*, 497(4):4549–4564.
- Aguilera-Dena, D. R., Müller, B., Antoniadis, J., Langer, N., Dessart, L., Vigna-Gómez, A., and Yoon, S.-C. (2023). Stripped-envelope stars in different metallicity environments. II. Type I supernovae and compact remnants. *Astron. Astrophys.*, 671:A134.
- Akmal, A., Pandharipande, V., and Ravenhall, D. (1998). The Equation of state of nucleon matter and neutron star structure. *Phys. Rev. C*, 58:1804–1828.
- Aldering, G., Humphreys, R. M., and Richmond, M. (1994). SN 1993J: The Optical Properties of its Progenitor. *Astron. J.*, 107:662.
- Arcavi, I., Hosseinzadeh, G., Brown, P. J., Smartt, S. J., Valenti, S., Tartaglia, L., Piro, A. L., Sanchez, J. L., Nicholls, B., Monard, B. L. A. G., Howell, D. A., McCully, C., Sand, D. J., Tonry, J., Denneau, L., Stalder, B., Heinze, A., Rest, A., Smith, K. W., and Bishop, D. (2017). Constraints on the Progenitor of SN 2016gkg from Its Shock-cooling Light Curve. *Astrophys. J. Lett.*, 837(1):L2.
- Ashton, G., Hübner, M., Lasky, P. D., Talbot, C., Ackley, K., Biscoveanu, S., Chu, Q., Divakarla, A., Easter, P. J., Goncharov, B., Vivanco, F. H., Harms, J., Lower, M. E., Meadors, G. D., Melchor, D., Payne, E., Pitkin, M. D., Powell, J., Sarin, N., Smith, R. J. E., and Thrane, E. (2019). Bilby: A user-friendly bayesian inference library for gravitational-wave astronomy. *The Astrophysical Journal Supplement Series*, 241(2):27.
- Asplund, M., Grevesse, N., Sauval, A. J., and Scott, P. (2009). The Chemical Composition of the Sun. *Ann. Rev. Astron. Astrophys.*, 47(1):481–522.
- Ba, J. L., Kiros, J. R., and Hinton, G. E. (2016). Layer normalization.
- Baiotti, L. (2019). Gravitational waves from neutron star mergers and their relation to the nuclear equation of state. *Progress in Particle and Nuclear Physics*, 109:103714.
- Barrett, J. W., Mandel, I., Neijssel, C. J., Stevenson, S., and Vigna-Gómez, A. (2017). Exploring the Parameter Space of Compact Binary Population Synthesis. In Brescia, M., Djorgovski, S. G., Feigelson, E. D., Longo, G., and Cavuoti, S., editors, *Astroinformatics*, volume 325, pages 46–50.

- Bauswein, A., Baumgarte, T. W., and Janka, H.-T. (2013). Prompt merger collapse and the maximum mass of neutron stars. *Phys. Rev. Lett.*, 111:131101.
- Bauswein, A., Blacker, S., Lioutas, G., Soutanis, T., Vijayan, V., and Stergioulas, N. (2021). Systematics of prompt black-hole formation in neutron star mergers. *Phys. Rev. D*, 103:123004.
- Bauswein, A. and Janka, H.-T. (2012). Measuring neutron-star properties via gravitational waves from neutron-star mergers. *Phys. Rev. Lett.*, 108:011101.
- Bauswein, A., Janka, H.-T., Hebeler, K., and Schwenk, A. (2012). Equation-of-state dependence of the gravitational-wave signal from the ring-down phase of neutron-star mergers. *Phys. Rev. D*, 86:063001.
- Bauswein, A., Janka, H.-T., and Oechslin, R. (2010a). Testing approximations of thermal effects in neutron star merger simulations. *Phys. Rev. D*, 82:084043.
- Bauswein, A., Just, O., Janka, H.-T., and Stergioulas, N. (2017). Neutron-star Radius Constraints from GW170817 and Future Detections. *Astrophys. J. Lett.*, 850(2):L34.
- Bauswein, A., Oechslin, R., and Janka, H.-T. (2010b). Discriminating strange star mergers from neutron star mergers by gravitational-wave measurements. *Phys. Rev. D*, 81:024012.
- Bauswein, A. and Stergioulas, N. (2015). Unified picture of the post-merger dynamics and gravitational wave emission in neutron star mergers. *Phys. Rev. D*, 91:124056.
- Bauswein, A. and Stergioulas, N. (2019). Spectral classification of gravitational-wave emission and equation of state constraints in binary neutron star mergers. *Journal of Physics G Nuclear Physics*, 46(11):113002.
- Bauswein, A., Stergioulas, N., and Janka, H.-T. (2016). Exploring properties of high-density matter through remnants of neutron-star mergers. *European Physical Journal A*, 52:56.
- Bazot, M., Bourguignon, S., and Christensen-Dalsgaard, J. (2012). A Bayesian approach to the modelling of α Cen A. *Mon. Not. Roy. Astron. Soc.*, 427(3):1847–1866.
- Beasor, E. R., Hosseinzadeh, G., Smith, N., Davies, B., Jencson, J. E., Pearson, J., and Sand, D. J. (2024). JWST Reveals a Luminous Infrared Source at the Position of the Failed Supernova Candidate N6946-BH1. *Astrophys. J.*, 964(2):171.
- Bellinger, E. P., Angelou, G. C., Hekker, S., Basu, S., Ball, W. H., and Guggenberger, E. (2016). Fundamental Parameters of Main-Sequence Stars in an Instant with Machine Learning. *Astrophys. J.*, 830(1):31.
- Bernuzzi, S. (2020). Neutron star merger remnants. *General Relativity and Gravitation*, 52(11):108.

- Bernuzzi, S., Dietrich, T., and Nagar, A. (2015). Modeling the complete gravitational wave spectrum of neutron star mergers. *Phys. Rev. Lett.*, 115:091101.
- Bischi, B., Binder, M., Lang, M., Pielok, T., Richter, J., Coors, S., Thomas, J., Ullmann, T., Becker, M., Boulesteix, A.-L., Deng, D., and Lindauer, M. (2023). Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. *WIREs Data Mining and Knowledge Discovery*, 13(2):e1484.
- Boccioli, L. and Fragione, G. (2024). Remnant masses from 1 D + core-collapse supernovae simulations: Bimodal neutron star mass distribution and black holes in the low-mass gap. *Phys. Rev. D.*, 110(2):023007.
- Boian, I. and Groh, J. H. (2018). Catching a star before explosion: the luminous blue variable progenitor of SN 2015bh. *Astron. Astrophys.*, 617:A115.
- Bose, S., Chakravarti, K., Rezzolla, L., Sathyaprakash, B. S., and Takami, K. (2018). Neutron-star radius from a population of binary neutron star mergers. *Phys. Rev. Lett.*, 120:031102.
- Branch, M. A., Coleman, T. F., and Li, Y. (1999). A subspace, interior, and conjugate gradient method for large-scale bound-constrained minimization problems. *SIAM Journal on Scientific Computing*, 21(1):1–23.
- Brandes, L., Weise, W., and Kaiser, N. (2023). Inference of the sound speed and related properties of neutron stars. *Phys. Rev. D*, 107:014011.
- Breschi, M., Bernuzzi, S., Chakravarti, K., Camilletti, A., Prakash, A., and Perego, A. (2022). KiloHertz Gravitational Waves From Binary Neutron Star Mergers: Numerical-relativity Informed Postmerger Model. *arXiv e-prints*, page arXiv:2205.09112.
- Breschi, M., Bernuzzi, S., Zappa, F., Agathos, M., Perego, A., Radice, D., and Nagar, A. (2019). KiloHertz gravitational waves from binary neutron star remnants: Time-domain model and constraints on extreme matter. *Phys. Rev. D*, 100:104029.
- Broekgaarden, F. S., Berger, E., Stevenson, S., Justham, S., Mandel, I., Chruślińska, M., van Son, L. A. C., Wagg, T., Vigna-Gómez, A., de Mink, S. E., Chattopadhyay, D., and Neijssel, C. J. (2022). Impact of massive binary star and cosmic evolution on gravitational wave observations - II. Double compact object rates and properties. *Mon. Not. Roy. Astron. Soc.*, 516(4):5737–5761.
- Brott, I., Evans, C. J., Hunter, I., de Koter, A., Langer, N., Dufton, P. L., Cantiello, M., Trundle, C., Lennon, D. J., de Mink, S. E., Yoon, S. C., and Anders, P. (2011). Rotating massive main-sequence stars. II. Simulating a population of LMC early B-type stars as a test of rotational mixing. *Astron. Astrophys.*, 530:A116.
- Brown, G. E., Heger, A., Langer, N., Lee, C. H., Wellstein, S., and Bethe, H. A. (2001). Formation of high mass X-ray black hole binaries. *New Astronomy*, 6(7):457–470.

- Burrows, A. and Goshy, J. (1993). A Theory of Supernova Explosions. *Astrophys. J. Lett.*, 416:L75.
- Burrows, A. and Lattimer, J. M. (1987). Neutrinos from SN 1987A. *Astrophys. J. Lett.*, 318:L63.
- Burrows, A. and Vartanyan, D. (2021). Core-collapse supernova explosion theory. *Nature*, 589(7840):29–39.
- Byrd, R. H., Schnabel, R. B., and Shultz, G. A. (1988). Approximate solution of the trust region problem by minimization over two-dimensional subspaces. *Mathematical Programming*, 40(1):247–263.
- Byrne, C. M., Stanway, E. R., Eldridge, J. J., McSwiney, L., and Townsend, O. T. (2022). The dependence of theoretical synthetic spectra on α -enhancement in young, binary stellar populations. *Mon. Not. Roy. Astron. Soc.*, 512(4):5329–5338.
- Capano, C. D., Tews, I., Brown, S. M., Margalit, B., De, S., Kumar, S., Brown, D. A., Krishnan, B., and Reddy, S. (2020). Stringent constraints on neutron-star radii from multimessenger observations and nuclear theory. *Nature Astronomy*, 4:625–632.
- Carroll, S. M. (1997). Lecture notes on general relativity.
- Chatziioannou, K. (2020). Neutron-star tidal deformability and equation-of-state constraints. *General Relativity and Gravitation*, 52(11):109.
- Chatziioannou, K., Clark, J. A., Bauswein, A., Millhouse, M., Littenberg, T. B., and Cornish, N. (2017). Inferring the post-merger gravitational wave emission from binary neutron star coalescences. *Phys. Rev. D*, 96:124035.
- Chatziioannou, K. and Han, S. (2020). Studying strong phase transitions in neutron stars with gravitational waves. *Phys. Rev. D*, 101(4):044019.
- Chatziioannou, K., Yagi, K., Klein, A., Cornish, N., and Yunes, N. (2015). Probing the internal composition of neutron stars with gravitational waves. *Phys. Rev. D*, 92(10):104008.
- Chieffi, A. and Limongi, M. (2020). The Presupernova Core Mass-Radius Relation of Massive Stars: Understanding Its Formation and Evolution. *Astrophys. J.*, 890(1):43.
- Choi, J., Dotter, A., Conroy, C., Cantiello, M., Paxton, B., and Johnson, B. D. (2016). Mesa Isochrones and Stellar Tracks (MIST). I. Solar-scaled Models. *Astrophys. J.*, 823(2):102.
- Church, R. P., Tout, C. A., and Hurley, J. R. (2009). N-body simulations with live stellar evolution. *Publications of the Astronomical Society of Australia*, 26(1):92–102.

- Clark, J. A., Bauswein, A., Stergioulas, N., and Shoemaker, D. (2016). Observing gravitational waves from the post-merger phase of binary neutron star coalescence. *Classical and Quantum Gravity*, 33(8):085003.
- Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20:273–297.
- Couch, S. M., Warren, M. L., and O’Connor, E. P. (2020). Simulating Turbulence-aided Neutrino-driven Core-collapse Supernova Explosions in One Dimension. *Astrophys. J.*, 890(2):127.
- Creevey, O. L., Sordo, R., Pailler, F., Frémat, Y., Heiter, U., Thévenin, F., Andrae, R., Fouesneau, M., Lobel, A., Bailer-Jones, C. A. L., Garabato, D., Bellas-Velidis, I., Brugaletta, E., Lorca, A., Ordenovic, C., Palicio, P. A., Sarro, L. M., Delchambre, L., Drimmel, R., Rybizki, J., Torralba Elipe, G., Korn, A. J., Recio-Blanco, A., Schultheis, M. S., De Angeli, F., Montegriffo, P., Abreu Aramburu, A., Accart, S., Álvarez, M. A., Bakker, J., Brouillet, N., Burlacu, A., Carballo, R., Casamiquela, L., Chiavassa, A., Contursi, G., Cooper, W. J., Dafonte, C., Dapergolas, A., de Laverny, P., Dharmawardena, T. E., Edvardsson, B., Le Fustec, Y., García-Lario, P., García-Torres, M., Gomez, A., González-Santamaría, I., Hatzidimitriou, D., Jean-Antoine Piccolo, A., Kontiza, M., Kordopatis, G., Lanzafame, A. C., Lebreton, Y., Licata, E. L., Lindstrøm, H. E. P., Livanou, E., Magdaleno Romeo, A., Manteiga, M., Marocco, F., Marshall, D. J., Mary, N., Nicolas, C., Pallas-Quintela, L., Panem, C., Pichon, B., Poggio, E., Rielet, F., Robin, C., Santoveña, R., Silvelo, A., Slezak, I., Smart, R. L., Soubiran, C., Süveges, M., Ulla, A., Utrilla, E., Valenari, A., Zhao, H., Zorec, J., Barrado, D., Bijaoui, A., Bouret, J.-C., Blomme, R., Brott, I., Cassisi, S., Kochukhov, O., Martayan, C., Shulyak, D., and Silvester, J. (2023). Gaia data release 3 - astrophysical parameters inference system (apsis). i. methods and content overview. *A&A*, 674:A26.
- Criswell, A. W., Miller, J., Woldemariam, N., Souldanis, T., Bauswein, A., Chatziioannou, K., Coughlin, M. W., Jones, G., and Mandic, V. (2023). Hierarchical bayesian method for constraining the neutron star equation of state with an ensemble of binary neutron star postmerger remnants. *Phys. Rev. D*, 107:043021.
- Curry, H. B. (1944). The method of steepest descent for non-linear minimization problems. *Quarterly of Applied Mathematics*, 2:258–261.
- Cutler, C. and Flanagan, É. E. (1994). Gravitational waves from merging compact binaries: How accurately can one extract the binary’s parameters from the inspiral waveform? *Phys. Rev. D.*, 49(6):2658–2697.
- da Silva Schneider, A., O’Connor, E., Granqvist, E., Betranhandy, A., and Couch, S. M. (2020). Equation of State and Progenitor Dependence of Stellar-mass Black Hole Formation. *Astrophys. J.*, 894(1):4.
- Davies, B. and Beasor, E. R. (2018). The initial masses of the red supergiant progenitors to Type II supernovae. *Mon. Not. Roy. Astron. Soc.*, 474(2):2116–2128.

- Davies, B., Crowther, P. A., and Beasor, E. R. (2018). The luminosities of cool supergiants in the Magellanic Clouds, and the Humphreys-Davidson limit revisited. *Mon. Not. Roy. Astron. Soc.*, 478(3):3138–3148.
- Del Pozzo, W., Li, T. G. F., Agathos, M., Van Den Broeck, C., and Vitale, S. (2013). Demonstrating the Feasibility of Probing the Neutron-Star Equation of State with Second-Generation Gravitational-Wave Detectors. *Phys. Rev. Lett.*, 111(7):071101.
- Dietrich, T., Coughlin, M. W., Pang, P. T. H., Bulla, M., Heinzl, J., Issa, L., Tews, I., and Antier, S. (2020). Multimessenger constraints on the neutron-star equation of state and the Hubble constant. *Science*, 370(6523):1450–1453.
- Dietrich, T., Hinderer, T., and Samajdar, A. (2021a). Interpreting binary neutron star mergers: describing the binary neutron star dynamics, modelling gravitational waveforms, and analyzing detections. *General Relativity and Gravitation*, 53(3):27.
- Dietrich, T., Hinderer, T., and Samajdar, A. (2021b). Interpreting binary neutron star mergers: describing the binary neutron star dynamics, modelling gravitational waveforms, and analyzing detections. *General Relativity and Gravitation*, 53(3):27.
- Dietrich, T., Ujevic, M., Tichy, W., Bernuzzi, S., and Brüggmann, B. (2017). Gravitational waves and mass ejecta from binary neutron star mergers: Effect of the mass ratio. *Phys. Rev. D.*, 95(2):024029.
- Dingle, K., Camargo, C. Q., and Louis, A. A. (2018). Input–output maps are strongly biased towards simple outputs. *Nature Communications*, 9(1).
- Dotter, A. (2016). MESA Isochrones and Stellar Tracks (MIST) 0: Methods for the Construction of Stellar Isochrones. *Astrophys. J. Supp. Ser.*, 222(1):8.
- Dudi, R., Dietrich, T., Rashti, A., Brüggmann, B., Steinhoff, J., and Tichy, W. (2022). High-accuracy simulations of highly spinning binary neutron star systems. *Phys. Rev. D.*, 105(6):064050.
- Easter, P. J., Ghonge, S., Lasky, P. D., Casey, A. R., Clark, J. A., Hernandez Vivanco, F., and Chatziioannou, K. (2020). Detection and parameter estimation of binary neutron star merger remnants. *Phys. Rev. D*, 102:043011.
- Easter, P. J., Lasky, P. D., Casey, A. R., Rezzolla, L., and Takami, K. (2019). Computing fast and reliable gravitational waveforms of binary neutron star merger remnants. *Phys. Rev. D*, 100:043005.
- Eldan, R. and Shamir, O. (2016). The power of depth for feedforward neural networks. In Feldman, V., Rakhlin, A., and Shamir, O., editors, *29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pages 907–940, Columbia University, New York, New York, USA. PMLR.
- Elias-Rosa, N., Van Dyk, S. D., Li, W., Silverman, J. M., Foley, R. J., Ganeshalingam, M., Mauerhan, J. C., Kankare, E., Jha, S., Filippenko, A. V., Beckman, J. E., Berger, E., Cuillandre, J.-C., and Smith, N. (2011). The Massive Progenitor of the Possible Type II-Linear Supernova 2009hd in Messier 66. *Astrophys. J.*, 742(1):6.

- Ertl, T., Janka, H. T., Woosley, S. E., Sukhbold, T., and Ugliano, M. (2016). A Two-parameter Criterion for Classifying the Explodability of Massive Stars by the Neutrino-driven Mechanism. *Astrophys. J.*, 818(2):124.
- Ertl, T., Woosley, S. E., Sukhbold, T., and Janka, H. T. (2020). The Explosion of Helium Stars Evolved with Mass Loss. *Astrophys. J.*, 890(1):51.
- Fang, Q. and Maeda, K. (2023). Inferring the Progenitor Mass-Kinetic Energy Relation of Stripped-envelope Core-collapse Supernovae from Nebular Spectroscopy. *Astrophys. J.*, 949(2):93.
- Fang, Q., Maeda, K., Kuncarayakti, H., Tanaka, M., Kawabata, K. S., Hattori, T., Aoki, K., Moriya, T. J., and Yamanaka, M. (2022). Statistical Properties of the Nebular Spectra of 103 Stripped-envelope Core-collapse Supernovae. *Astrophys. J.*, 928(2):151.
- Fix, E. and Hodges, J. L. (1989). Discriminatory analysis - nonparametric discrimination: Consistency properties. *International Statistical Review*, 57:238.
- Foucart, F., Haas, R., Duez, M. D., O'Connor, E., Ott, C. D., Roberts, L., Kidder, L. E., Lippuner, J., Pfeiffer, H. P., and Scheel, M. A. (2016). Low mass binary neutron star mergers: Gravitational waves and neutrino emission. *Phys. Rev. D*, 93:044019.
- Fragos, T., Andrews, J. J., Bavera, S. S., Berry, C. P. L., Coughlin, S., Dotter, A., Giri, P., Kalogera, V., Katsaggelos, A., Kovelakas, K., Lalvani, S., Misra, D., Srivastava, P. M., Qin, Y., Rocha, K. A., Román-Garza, J., Serra, J. G., Stahle, P., Sun, M., Teng, X., Trajceviski, G., Tran, N. H., Xing, Z., Zapartas, E., and Zevin, M. (2023). POSYDON: A General-purpose Population Synthesis Code with Detailed Binary-evolution Simulations. *Astrophys. J. Supp. Ser.*, 264(2):45.
- Fraser, M., Takáts, K., Pastorello, A., Smartt, S. J., Mattila, S., Botticella, M. T., Valenti, S., Ergon, M., Sollerman, J., Arcavi, I., Benetti, S., Bufano, F., Crockett, R. M., Danziger, I. J., Gal-Yam, A., Maund, J. R., Taubenberger, S., and Turatto, M. (2010). On the Progenitor and Early Evolution of the Type II Supernova 2009kr. *Astrophys. J. Lett.*, 714(2):L280–L284.
- Friedman, J. L. and Stergioulas, N. (2020). Astrophysical implications of neutron star inspiral and coalescence. *International Journal of Modern Physics D*, 29(11):2041015.
- Fryer, C. L., Belczynski, K., Wiktorowicz, G., Dominik, M., Kalogera, V., and Holz, D. E. (2012). Compact Remnant Mass Function: Dependence on the Explosion Mechanism and Metallicity. *Astrophys. J.*, 749(1):91.
- Gal-Yam, A., Leonard, D. C., Fox, D. B., Cenko, S. B., Soderberg, A. M., Moon, D.-S., Sand, D. J., Program, C. C. C., Li, W., Filippenko, A. V., Aldering, G., and Copin, Y. (2007). On the progenitor of sn 2005gl and the nature of type iin supernovae. *The Astrophysical Journal*, 656(1):372.

- Ganapathy, D., McCuller, L., Rollins, J. G., Hall, E. D., Barsotti, L., and Evans, M. (2021). Tuning Advanced LIGO to kilohertz signals from neutron-star collisions. *Phys. Rev. D*, 103(2):022002.
- Gerke, J. R., Kochanek, C. S., and Stanek, K. Z. (2015). The search for failed supernovae with the Large Binocular Telescope: first candidates. *Mon. Not. Roy. Astron. Soc.*, 450(3):3289–3305.
- Gilkis, A. and Arcavi, I. (2022). How much hydrogen is in Type Ib and IIb supernova progenitors? *Monthly Notices of the Royal Astronomical Society*, 511(1):691–712.
- Glorot, X. and Bengio, Y. (2010). Understanding the difficulty of training deep feed-forward neural networks. In Teh, Y. W. and Titterton, M., editors, *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, volume 9 of *Proceedings of Machine Learning Research*, pages 249–256, Chia Laguna Resort, Sardinia, Italy. PMLR.
- Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106:746 – 762.
- Goodfellow, I., Bengio, Y., Courville, A., and Bach, F. (2017). *Deep Learning*. MIT Press.
- Goswami, S., Silva, L., Bressan, A., Grisoni, V., Costa, G., Marigo, P., Granato, G. L., Lapi, A., and Spera, M. (2022). Impact of very massive stars on the chemical evolution of extremely metal-poor galaxies. *Astronomy amp; Astrophysics*, 663:A1.
- Gramacy, R. B. (2020). *Surrogates: Gaussian Process Modeling, Design and Optimization for the Applied Sciences*. Chapman Hall/CRC, Boca Raton, Florida. <http://bobby.gramacy.com/surrogates/>.
- Hahnloser, R. H. R., Sarpeshkar, R., Mahowald, M. A., Douglas, R. J., and Seung, H. S. (2000). Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature*, 405(6789):947–951.
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825):357–362.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.
- Heger, A., Fryer, C. L., Woosley, S. E., Langer, N., and Hartmann, D. H. (2003). How massive single stars end their life. *The Astrophysical Journal*, 591(1):288.
- Heger, A., Müller, B., and Mandel, I. (2023). Black Holes as the End State of Stellar Evolution: Theory and Simulations. In Haiman, Z., editor, *The Encyclopedia of Cosmology. Set 2: Frontiers in Cosmology. Volume 3: Black Holes*, pages 61–111.

- Heger, A. and Woosley, S. E. (2010). Nucleosynthesis and Evolution of Massive Metal-free Stars. *Astrophys. J.*, 724(1):341–373.
- Hernandez Vivanco, F., Smith, R., Thrane, E., Lasky, P. D., Talbot, C., and Raymond, V. (2019). Measuring the neutron star equation of state with gravitational waves: The first forty binary neutron star merger observations. *Phys. Rev. D*, 100(10):103009.
- Higson, E., Handley, W., Hobson, M., and Lasenby, A. (2019). Dynamic nested sampling: an improved algorithm for parameter estimation and evidence calculation. *Statistics and Computing*, 29(5):891–913.
- Ho, T. K. (1995). Random decision forests. In *Proceedings of 3rd International Conference on Document Analysis and Recognition*, volume 1, pages 278–282 vol.1.
- Hopkins, P. F., Kereš, D., Faucher-Giguère, C.-A., and Murray, N. (2018). Stellar feedback in galaxies and the origin of galaxy-scale winds. *Monthly Notices of the Royal Astronomical Society*, 481(3):4133–4188.
- Horiuchi, S., Nakamura, K., Takiwaki, T., Kotake, K., and Tanaka, M. (2014). The red supergiant and supernova rate problems: implications for core-collapse supernova physics. *Mon. Not. Roy. Astron. Soc.*, 445:L99–L103.
- Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366.
- Hotokezaka, K., Kiuchi, K., Kyutoku, K., Muramushi, T., Sekiguchi, Y.-i., Shibata, M., and Taniguchi, K. (2013). Remnant massive neutron stars of binary neutron star mergers: Evolution process and gravitational waveform. *Phys. Rev. D*, 88:044026.
- Hotokezaka, K., Kyutoku, K., Okawa, H., Shibata, M., and Kiuchi, K. (2011). Binary neutron star mergers: Dependence on the nuclear equation of state. *Phys. Rev. D*, 83:124008.
- Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101.
- Hurley, J. R., Pols, O. R., and Tout, C. A. (2000). Comprehensive analytic formulae for stellar evolution as a function of mass and metallicity. *Mon. Not. Roy. Astron. Soc.*, 315(3):543–569.
- Hurley, J. R., Tout, C. A., and Pols, O. R. (2002). Evolution of binary stars and the effect of tides on binary populations. *Mon. Not. Roy. Astron. Soc.*, 329(4):897–928.
- Huth, S., Pang, P. T. H., Tews, I., Dietrich, T., Le Fèvre, A., Schwenk, A., Trautmann, W., Agarwal, K., Bulla, M., Coughlin, M. W., and Van Den Broeck, C. (2022). Constraining neutron-star matter with microscopic and macroscopic collisions. *Nature*, 606(7913):276–280.

- Iorio, G., Mapelli, M., Costa, G., Spera, M., Escobar, G. J., Sgalletta, C., Trani, A. A., Korb, E., Santoliquido, F., Dall’Amico, M., Gaspari, N., and Bressan, A. (2023). Compact object mergers: exploring uncertainties from stellar and binary evolution with SEVN. *Mon. Not. Roy. Astron. Soc.*, 524(1):426–470.
- Isenberg, J. and Nester, J. (1980). Canonical Gravity. In *General Relativity and Gravitation. Vol. 1. One hundred years after the birth of Albert Einstein. Edited by A. Held. New York*, volume 1, page 23.
- Ivakhnenko, A. G. and Lapa, V. (1967). *Cybernetics and forecasting techniques*, volume 8. American Elsevier Publishing Company.
- Izzard, R. G. and Jermyn, A. S. (2023). Circumbinary discs for stellar population models. *Mon. Not. Roy. Astron. Soc.*, 521(1):35–50.
- Jacot, A., Gabriel, F., and Hongler, C. (2018). Neural tangent kernel: Convergence and generalization in neural networks. *CoRR*, abs/1806.07572.
- James, G., Witten, D., Hastie, T., Tibshirani, R., and Taylor, J. (2023). *An Introduction to Statistical Learning with Applications in Python*. Springer Texts in Statistics. Springer, Cham.
- Janka, H., Langanke, K., Marek, A., MARTINEZPINEDO, G., and MULLER, B. (2007). Theory of core-collapse supernovae. *Physics Reports*, 442(1–6):38–74.
- Janka, H. T. and Keil, W. (1998). Perspectives of Core-Collapse Supernovae beyond SN 1987A. In Labhardt, L., Binggeli, B., and Buser, R., editors, *Supernovae and cosmology*, page 7.
- Janka, H. T., Müller, B., Kitaura, F. S., and Buras, R. (2008). Dynamics of shock propagation and nucleosynthesis conditions in O-Ne-Mg core supernovae. *Astron. Astrophys.*, 485(1):199–208.
- Jørgensen, B. R. and Lindegren, L. (2005). Determination of stellar ages from isochrones: Bayesian estimation versus isochrone fitting. *Astron. Astrophys.*, 436(1):127–143.
- Justham, S., Podsiadlowski, P., and Vink, J. S. (2014). Luminous Blue Variables and Superluminous Supernovae from Binary Mergers. *Astrophys. J.*, 796(2):121.
- Kamlah, A. W. H., Leveque, A., Spurzem, R., Arca Sedda, M., Askar, A., Banerjee, S., Berczik, P., Giersz, M., Hurley, J., Belloni, D., Kühmichel, L., and Wang, L. (2022). Preparing the next gravitational million-body simulations: evolution of single and binary stars in NBODY6++GPU, MOCCA, and MCLUSTER. *Mon. Not. Roy. Astron. Soc.*, 511(3):4060–4089.
- Kankare, E., Kotak, R., Pastorello, A., Fraser, M., Mattila, S., Smartt, S. J., Bruce, A., Chambers, K. C., Elias-Rosa, N., Flewelling, H., Fremling, C., Harmanen, J., Huber, M., Jerkstrand, A., Kangas, T., Kuncarayakti, H., Magee, M., Magnier, E., Polshaw, J., Smith, K. W., Sollerman, J., and Tomasella, L. (2015). On the triple peaks of snhunt248 in ngc 5806. *A&A*, 581:L4.

- Kasen, D. and Bildsten, L. (2010). Supernova Light Curves Powered by Young Magnetars. *Astrophys. J.*, 717(1):245–249.
- Kaufman, C. G., Schervish, M. J., and Nychka, D. W. (2008). Covariance tapering for likelihood-based estimation in large spatial data sets. *Journal of the American Statistical Association*, 103(484):1545–1555.
- Kennedy, M. C. and O’Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63.
- Kerzendorf, W. E. and Sim, S. A. (2014). A spectral synthesis code for rapid modelling of supernovae. *Monthly Notices of the Royal Astronomical Society*, 440(1):387–404.
- Kerzendorf, W. E., Vogl, C., Buchner, J., Contardo, G., Williamson, M., and van der Smagt, P. (2021). Dalek: A Deep Learning Emulator for TARDIS. *Astrophys. J. Lett.*, 910(2):L23.
- Kilpatrick, C. D., Drout, M. R., Auchettl, K., Dimitriadis, G., Foley, R. J., Jones, D. O., DeMarchi, L., French, K. D., Gall, C., Hjorth, J., Jacobson-Galán, W. V., Margutti, R., Piro, A. L., Ramirez-Ruiz, E., Rest, A., and Rojas-Bravo, C. (2021). A cool and inflated progenitor candidate for the Type Ib supernova 2019yvr at 2.6 yr before explosion. *Mon. Not. Roy. Astron. Soc.*, 504(2):2073–2093.
- Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization.
- Kippenhahn, R. (1990). *Stellar Structure and Evolution*. Springer, Berlin.
- Kitaura, F. S., Janka, H. T., and Hillebrandt, W. (2006). Explosions of O-Ne-Mg cores, the Crab supernova, and subluminous type II-P supernovae. *Astron. Astrophys.*, 450(1):345–350.
- Kolmogorov, A. N. (1963). On tables of random numbers. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, 25(4):369–376.
- Koposov, S., Speagle, J., Barbary, K., Ashton, G., Bennett, E., Buchner, J., Scheffler, C., Cook, B., Talbot, C., Guillochon, J., Cubillos, P., Ramos, A. A., Johnson, B., Lang, D., Ilya, Dartiailh, M., Nitz, A., McCluskey, A., and Archibald, A. (2023). joshspeagle/dynesty: v2.1.3.
- Kruckow, M. U., Tauris, T. M., Langer, N., Kramer, M., and Izzard, R. G. (2018). Progenitors of gravitational wave mergers: binary evolution with the stellar grid-based code COMBINE. *Mon. Not. Roy. Astron. Soc.*, 481(2):1908–1949.
- Ksoll, V. F., Ardizzone, L., Klessen, R., Koethe, U., Sabbi, E., Robberto, M., Gouliermis, D., Rother, C., Zeidler, P., and Gennaro, M. (2020). Stellar parameter determination from photometry using invertible neural networks. *Mon. Not. Roy. Astron. Soc.*, 499(4):5447–5485.

- Lackey, B. D. and Wade, L. (2015). Reconstructing the neutron-star equation of state with gravitational-wave detectors from a realistic population of inspiralling binary neutron stars. *Phys. Rev. D*, 91(4):043002.
- Landry, P., Essick, R., and Chatziioannou, K. (2020). Nonparametric constraints on neutron star matter with existing and upcoming gravitational wave and pulsar observations. *Phys. Rev. D*, 101(12):123007.
- Laplace, E., Justham, S., Renzo, M., Götberg, Y., Farmer, R., Vartanyan, D., and de Mink, S. E. (2021). Different to the core: The pre-supernova structures of massive single and binary-stripped stars. *Astron. Astrophys.*, 656:A58.
- Laplace, E., Schneider, F. R. N., and Podsiadlowski, P. (2024). It’s written in the massive stars: The role of stellar physics in the formation of black holes. *arXiv e-prints*, page arXiv:2409.02058.
- Lee, J., Xiao, L., Schoenholz, S., Bahri, Y., Novak, R., Sohl-Dickstein, J., and Pennington, J. (2019). Wide neural networks of any depth evolve as linear models under gradient descent. In Wallach, H., Larochelle, H., Beygelzimer, A., dAlché-Buc, F., Fox, E., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.
- Legred, I., Chatziioannou, K., Essick, R., Han, S., and Landry, P. (2021). Impact of the PSR J 0740+6620 radius constraint on the properties of high-density matter. *Phys. Rev. D.*, 104(6):063003.
- Li, T., Davies, G. R., Lyttle, A. J., Ball, W. H., Carboneau, L. M., and García, R. A. (2022). Modelling stars with Gaussian Process Regression: augmenting stellar model grid. *Mon. Not. Roy. Astron. Soc.*, 511(4):5597–5610.
- Liebling, S. L., Palenzuela, C., and Lehner, L. (2021). Effects of high density phase transitions on neutron star dynamics. *Classical and Quantum Gravity*, 38(11):115007.
- Limongi, M. and Chieffi, A. (2018). Presupernova Evolution and Explosive Nucleosynthesis of Rotating Massive Stars in the Metallicity Range $-3 \leq [\text{Fe}/\text{H}] \leq 0$. *Astrophys. J. Supp. Ser.*, 237(1):13.
- Liu, H., Ong, Y.-S., Shen, X., and Cai, J. (2020). When gaussian process meets big data: A review of scalable gps. *IEEE Transactions on Neural Networks and Learning Systems*, 31(11):4405–4423.
- Lyttle, A. J., Davies, G. R., Li, T., Carboneau, L. M., Leung, H.-H., Westwood, H., Chaplin, W. J., Hall, O. J., Huber, D., Nielsen, M. B., Basu, S., and García, R. A. (2021). Hierarchically modelling Kepler dwarfs and subgiants to improve inference of stellar properties with asteroseismology. *Mon. Not. Roy. Astron. Soc.*, 505(2):2427–2446.
- Maggiore, M. (2000). Gravitational wave experiments and early universe cosmology. *Physics Reports*, 331(6):283–367.

- Maggiore, M. (2007). *Gravitational Waves: Volume 1: Theory and Experiments*. Oxford University Press.
- Maggiore, M. et al. (2020). Science Case for the Einstein Telescope. *JCAP*, 03:050.
- Maione, F., De Pietri, R., Feo, A., and Löffler, F. (2017). Spectral analysis of gravitational waves from binary neutron star merger remnants. *Phys. Rev. D*, 96:063011.
- Maltsev, K., Schneider, F. R. N., Mandel, I., Heger, A., Müller, B., Röpke, F. K., and Laplace, E. (2024a). Explodability criteria for the neutrino-driven supernova mechanism (to-be-submitted to *a&a*). In *EAS2024*, page 2047.
- Maltsev, K., Schneider, F. R. N., Röpke, F. K., Jordan, A. I., Qadir, G. A., Kerzendorf, W. E., Riedmiller, K., and van der Smagt, P. (2024b). Scalable stellar evolution forecasting. Deep learning emulation versus hierarchical nearest-neighbor interpolation. *Astron. Astrophys.*, 681:A86.
- Mandel, I. and Müller, B. (2020). Simple recipes for compact remnant masses and natal kicks. *Mon. Not. Roy. Astron. Soc.*, 499(3):3214–3221.
- Mapelli, M., Dall’Amico, M., Bouffanais, Y., Giacobbo, N., Arca Sedda, M., Artale, M. C., Ballone, A., Di Carlo, U. N., Iorio, G., Santoliquido, F., and Torniamenti, S. (2021). Hierarchical black hole mergers in young, globular and nuclear star clusters: the effect of metallicity, spin and cluster properties. *Monthly Notices of the Royal Astronomical Society*, 505(1):339–358.
- Mapelli, M., Spera, M., Montanari, E., Limongi, M., Chieffi, A., Giacobbo, N., Bressan, A., and Bouffanais, Y. (2020). Impact of the Rotation and Compactness of Progenitors on the Mass of Black Holes. *Astrophys. J.*, 888(2):76.
- Margalit, B. and Metzger, B. D. (2017). Constraining the maximum mass of neutron stars from multi-messenger observations of gw170817. *The Astrophysical Journal Letters*, 850(2):L19.
- Martynov, D., Miao, H., Yang, H., Vivanco, F. H., Thrane, E., Smith, R., Lasky, P., East, W. E., Adhikari, R., Bauswein, A., Brooks, A., Chen, Y., Corbitt, T., Freise, A., Grote, H., Levin, Y., Zhao, C., and Vecchio, A. (2019). Exploring the sensitivity of gravitational wave detectors to neutron star physics. *Phys. Rev. D.*, 99(10):102004.
- Maund, J. R., Fraser, M., Smartt, S. J., Botticella, M. T., Barbarino, C., Childress, M., Gal-Yam, A., Inserra, C., Pignata, G., Reichart, D., Schmidt, B., Sollerman, J., Taddia, F., Tomasella, L., Valenti, S., and Yaron, O. (2013). Supernova 2012ec: identification of the progenitor and early monitoring with PESSTO. *Mon. Not. Roy. Astron. Soc.*, 431:L102–L106.
- McKay, M. D., Beckman, R. J., and Conover, W. J. (1979a). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245.

- McKay, M. D., Beckman, R. J., and Conover, W. J. (1979b). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245.
- Messenger, C., Takami, K., Gossan, S., Rezzolla, L., and Sathyaprakash, B. S. (2014). Source redshifts from gravitational-wave observations of binary neutron star mergers. *Phys. Rev. X*, 4:041004.
- Meyer, R. and Christensen, N. (2016). Gravitational Waves: A Statistical Autopsy of a Black Hole Merger. *Significance*, 13(2):20–25.
- Miravet-Tenés, M., Castillo, F. L., De Pietri, R., Cerdá-Durán, P., and Font, J. A. (2023). Prospects for the inference of inertial modes from hypermassive neutron stars with future gravitational-wave detectors. *Phys. Rev. D.*, 107(10):103053.
- Mockus, J. (1982). The bayesian approach to global optimization. In Drenick, R. F. and Kozin, F., editors, *System Modeling and Optimization*, pages 473–481, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Morozova, V., Piro, A. L., Renzo, M., Ott, C. D., Clausen, D., Couch, S. M., Ellis, J., and Roberts, L. F. (2015). Light Curves of Core-collapse Supernovae with Substantial Mass Loss Using the New Open-source SuperNova Explosion Code (SNEC). *Astrophys. J.*, 814(1):63.
- Müller, B. (2019). A critical assessment of turbulence models for 1D core-collapse supernova simulations. *Mon. Not. Roy. Astron. Soc.*, 487(4):5304–5323.
- Müller, B. (2020). Hydrodynamics of core-collapse supernovae and their progenitors. *Living Reviews in Computational Astrophysics*, 6(1):3.
- Müller, B., Heger, A., Liptai, D., and Cameron, J. B. (2016). A simple approach to the supernova progenitor-explosion connection. *Mon. Not. Roy. Astron. Soc.*, 460(1):742–764.
- Müller, B. and Janka, H. T. (2015). Non-radial instabilities and progenitor asphericities in core-collapse supernovae. *Mon. Not. Roy. Astron. Soc.*, 448(3):2141–2174.
- Nakamura, K., Takiwaki, T., Kuroda, T., and Kotake, K. (2015). Systematic features of axisymmetric neutrino-driven core-collapse supernova models in multiple progenitors. , 67(6):107.
- Ni, W., Tan, S. K., Ng, W. J., and Brown, S. D. (2012). Moving-window gpr for nonlinear dynamic system modeling with dual updating and dual preprocessing. *Industrial & engineering chemistry research*, 51(18):6416–6428.
- Nichani, E., Radhakrishnan, A., and Uhler, C. (2020). Increasing depth leads to u-shaped test risk in over-parameterized convolutional networks.

- Nicholl, M., Margalit, B., Schmidt, P., Smith, G. P., Ridley, E. J., and Nuttall, J. (2021). Tight multimessenger constraints on the neutron star equation of state from GW170817 and a forward model for kilonova light-curve synthesis. *Mon. Not. Roy. Astron. Soc.*, 505(2):3016–3032.
- O’Brien, J. T., Kerzendorf, W. E., Fullard, A., Williamson, M., Pakmor, R., Buchner, J., Hachinger, S., Vogl, C., Gillanders, J. H., Flörs, A., and van der Smagt, P. (2021). Probabilistic Reconstruction of Type Ia Supernova SN 2002bo. *Astrophys. J. Lett.*, 916(2):L14.
- O’Connor, E. and Ott, C. D. (2010). A new open-source code for spherically symmetric stellar collapse to neutron stars and black holes. *Classical and Quantum Gravity*, 27(11):114103.
- O’Connor, E. and Ott, C. D. (2011). Black Hole Formation in Failing Core-Collapse Supernovae. *Astrophys. J.*, 730(2):70.
- O’Connor, E. and Ott, C. D. (2013). The Progenitor Dependence of the Pre-explosion Neutrino Emission in Core-collapse Supernovae. *Astrophys. J.*, 762(2):126.
- Oechslin, R. and Janka, H.-T. (2007). Gravitational waves from relativistic neutron-star mergers with microphysical equations of state. *Phys. Rev. Lett.*, 99:121102.
- Oechslin, R., Rosswog, S., and Thielemann, F.-K. (2002). Conformally flat smoothed particle hydrodynamics application to neutron star mergers. *Phys. Rev. D*, 65:103005.
- Oechslin, R., Janka, H.-T., and Marek, A. (2007). Relativistic neutron star merger simulations with non-zero temperature equations of state* - i. variation of binary parameters and equation of state. *A&A*, 467(2):395–409.
- Page, M. A., Goryachev, M., Miao, H., Chen, Y., Ma, Y., Mason, D., Rossi, M., Blair, C. D., Ju, L., Blair, D. G., Schliesser, A., Tobar, M. E., and Zhao, C. (2021). Gravitational wave detectors with broadband high frequency sensitivity. *Communications Physics*, 4(1):27.
- Pang, P. T. H., Tews, I., Coughlin, M. W., Bulla, M., Van Den Broeck, C., and Dietrich, T. (2021). Nuclear-Physics Multi-Messenger Astrophysics Constraints on the Neutron-Star Equation of State: Adding NICER’s PSR J0740+6620 Measurement. *arXiv e-prints*, page arXiv:2105.08688.
- Papish, O. and Soker, N. (2011). Exploding core collapse supernovae with jittering jets. *Mon. Not. Roy. Astron. Soc.*, 416(3):1697–1702.
- Paschalidis, V. and Stergioulas, N. (2017). Rotating stars in relativity. *Living Reviews in Relativity*, 20(1):7.
- Patton, R. A. and Sukhbold, T. (2020). Towards a realistic explosion landscape for binary population synthesis. *Mon. Not. Roy. Astron. Soc.*, 499(2):2803–2816.

- Patton, R. A., Sukhbold, T., and Eldridge, J. J. (2022). Comparing compact object distributions from mass- and presupernova core structure-based prescriptions. *Mon. Not. Roy. Astron. Soc.*, 511(1):903–913.
- Paunzen, E. and Netopil, M. (2006). On the current status of open-cluster parameters. *Monthly Notices of the Royal Astronomical Society*, 371(4):1641–1647.
- Paxton, B., Bildsten, L., Dotter, A., Herwig, F., Lesaffre, P., and Timmes, F. (2011). Modules for Experiments in Stellar Astrophysics (MESA). *Astrophys. J. Supp. Ser.*, 192(1):3.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Pejcha, O. and Thompson, T. A. (2015). The Landscape of the Neutrino Mechanism of Core-collapse Supernovae: Neutron Star and Black Hole Mass Functions, Explosion Energies, and Nickel Yields. *Astrophys. J.*, 801(2):90.
- Pellegrino, C., Howell, D. A., Terreran, G., Arcavi, I., Bostroem, K. A., Brown, P. J., Burke, J., Dong, Y., Gilkis, A., Hiramatsu, D., Hosseinzadeh, G., McCully, C., Modjaz, M., Newsome, M., Gonzalez, E. P., Pritchard, T. A., Sand, D. J., Valenti, S., and Williamson, M. (2022). The Diverse Properties of Type Icn Supernovae Point to Multiple Progenitor Channels. *Astrophys. J.*, 938(1):73.
- Pesios, D., Koutalios, I., Kugiumtzis, D., and Stergioulas, N. (2024). Predicting binary neutron star postmerger spectra using artificial neural networks.
- Pols, O. R., Schröder, K.-P., Hurley, J. R., Tout, C. A., and Eggleton, P. P. (1998). Stellar evolution models for $Z = 0.0001$ to 0.03 . *Mon. Not. Roy. Astron. Soc.*, 298(2):525–536.
- Prialnik, D. (1996). *An Introduction to the Theory of Stellar Structure and Evolution*. Cambridge University Press.
- Puecher, A., Dietrich, T., Tsang, K. W., Kalaghatgi, C., Roy, S., Setyawati, Y., and Van Den Broeck, C. (2023). Unraveling information about supranuclear-dense matter from the complete binary neutron star coalescence process using future gravitational-wave detector networks. *Phys. Rev. D.*, 107(12):124009.
- Raaijmakers, G., Greif, S. K., Hebel, K., Hinderer, T., Nisanke, S., Schwenk, A., Riley, T. E., Watts, A. L., Lattimer, J. M., and Ho, W. C. G. (2021). Constraints on the Dense Matter Equation of State and Neutron Star Properties from NICER’s Mass-Radius Estimate of PSR J0740+6620 and Multimessenger Observations. *Astrophys. J. Lett.*, 918(2):L29.
- Radice, D., Perego, A., Zappa, F., and Bernuzzi, S. (2018). Gw170817: Joint constraint on the neutron star equation of state from multimessenger observations. *The Astrophysical Journal Letters*, 852(2):L29.

- Rasmussen, C. E. (2004). *Gaussian Processes in Machine Learning*, pages 63–71. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Reitze, D., Adhikari, R. X., Ballmer, S., Barish, B., Barsotti, L., Billingsley, G., Brown, D. A., Chen, Y., Coyne, D., Eisenstein, R., Evans, M., Fritschel, P., Hall, E. D., Lazzarini, A., Lovelace, G., Read, J., Sathyaprakash, B. S., Shoemaker, D., Smith, J., Torrie, C., Vitale, S., Weiss, R., Wipf, C., and Zucker, M. (2019). Cosmic Explorer: The U.S. Contribution to Gravitational-Wave Astronomy beyond LIGO. In *Bulletin of the American Astronomical Society*, volume 51, page 35.
- Rezzolla, L., Most, E. R., and Weih, L. R. (2018). Using gravitational-wave observations and quasi-universal relations to constrain the maximum mass of neutron stars. *The Astrophysical Journal Letters*, 852(2):L25.
- Riley, J., Agrawal, P., Barrett, J. W., Boyett, K. N. K., Broekgaarden, F. S., Chattopadhyay, D., Gaebel, S. M., Gittins, F., Hirai, R., Howitt, G., Justham, S., Khandelwal, L., Kummer, F., Lau, M. Y. M., Mandel, I., de Mink, S. E., Neijssel, C., Riley, T., van Son, L., Stevenson, S., Vigna-Gómez, A., Vinciguerra, S., Wagg, T., and Willcox, R. (2022). Rapid stellar and binary population synthesis with COMPAS. *The Astrophysical Journal Supplement Series*, 258(2):34.
- Rocha, K. A., Andrews, J. J., Berry, C. P. L., Doctor, Z., Katsaggelos, A. K., Pérez, J. G. S., Marchant, P., Kalogera, V., Coughlin, S., Bavera, S. S., Dotter, A., Fragos, T., Kovelakas, K., Misra, D., Xing, Z., and Zapartas, E. (2022). Active learning for computationally efficient distribution of binary evolution simulations. *The Astrophysical Journal*, 938(1):64.
- Romero-Shaw, I. M., Talbot, C., Biscoveanu, S., D’Emilio, V., Ashton, G., Berry, C. P. L., Coughlin, S., Galaudage, S., Hoy, C., Hübner, M., Phukon, K. S., Pitkin, M., Rizzo, M., Sarin, N., Smith, R., Stevenson, S., Vajpeyi, A., Arène, M., Athar, K., Banagiri, S., Bose, N., Carney, M., Chatziioannou, K., Clark, J. A., Colleoni, M., Cotesta, R., Edelman, B., Estellés, H., García-Quirós, C., Ghosh, A., Green, R., Haster, C. J., Husa, S., Keitel, D., Kim, A. X., Hernandez-Vivanco, F., Magaña Hernandez, I., Karathanasis, C., Lasky, P. D., De Lillo, N., Lower, M. E., Macleod, D., Mateu-Lucena, M., Miller, A., Millhouse, M., Morisaki, S., Oh, S. H., Ossokine, S., Payne, E., Powell, J., Pratten, G., Pürrer, M., Ramos-Buades, A., Raymond, V., Thrane, E., Veitch, J., Williams, D., Williams, M. J., and Xiao, L. (2020). Bayesian inference for compact binary coalescences with BILBY: validation and application to the first LIGO-Virgo gravitational-wave transient catalogue. *Mon. Not. Roy. Astron. Soc.*, 499(3):3295–3319.
- Rosswog, S. (2015). SPH Methods in the Modelling of Compact Objects. *Living Reviews in Computational Astrophysics*, 1(1):1.
- Ruiz, M., Shapiro, S. L., and Tsokaros, A. (2018). Gw170817, general relativistic magnetohydrodynamic simulations, and the neutron star maximum mass. *Phys. Rev. D*, 97:021501.

- Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1985). Learning internal representations by error propagation. Technical report, California Univ San Diego La Jolla Inst for Cognitive Science.
- Ryabchikova, T., Piskunov, N., Pakhomov, Y., Tsymbal, V., Titarenko, A., Sitnova, T., Alexeeva, S., Fossati, L., and Mashonkina, L. (2016). Accuracy of atmospheric parameters of FGK dwarfs determined by spectrum fitting. *Mon. Not. Roy. Astron. Soc.*, 456(2):1221–1234.
- Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). Design and Analysis of Computer Experiments. *Statistical Science*, 4(4):409 – 423.
- Sana, H., de Mink, S. E., de Koter, A., Langer, N., Evans, C. J., Gieles, M., Gosset, E., Izzard, R. G., Le Bouquin, J. B., and Schneider, F. R. N. (2012). Binary Interaction Dominates the Evolution of Massive Stars. *Science*, 337(6093):444.
- Santoliquido, F., Mapelli, M., Giacobbo, N., Bouffanais, Y., and Artale, M. C. (2021). The cosmic merger rate density of compact objects: impact of star formation, metallicity, initial mass function, and binary evolution. *Monthly Notices of the Royal Astronomical Society*, 502(4):4877–4889.
- Sarin, N. and Lasky, P. D. (2021a). Multimessenger astronomy with a kHz-band gravitational-wave observatory. *arXiv e-prints*, page arXiv:2110.10892.
- Sarin, N. and Lasky, P. D. (2021b). The evolution of binary neutron star post-merger remnants: a review. *General Relativity and Gravitation*, 53(6):59.
- Sasli, A., Karnesis, N., and Stergioulas, N. (2023). Exploring the Potential for Detecting Rotational Instabilities in Binary Neutron Star Merger Remnants with Gravitational Wave Detectors. *arXiv e-prints*, page arXiv:2311.10626.
- Schaback, R. and Wendland, H. (2006). Kernel techniques: From machine learning to meshless methods. *Acta Numerica*, 15:543–639.
- Schneider, F. R. N., Langer, N., de Koter, A., Brott, I., Izzard, R. G., and Lau, H. H. B. (2014). Bonnsai: a Bayesian tool for comparing stars with stellar evolution models. *Astron. Astrophys.*, 570:A66.
- Schneider, F. R. N., Podsiadlowski, P., and Laplace, E. (2023). Bimodal Black Hole Mass Distribution and Chirp Masses of Binary Black Hole Mergers. *Astrophys. J. Lett.*, 950(2):L9.
- Schneider, F. R. N., Podsiadlowski, P., and Laplace, E. (2024). Pre-supernova evolution and final fate of stellar mergers and accretors of binary mass transfer. *Astron. Astrophys.*, 686:A45.
- Schneider, F. R. N., Podsiadlowski, P., and Müller, B. (2021). Pre-supernova evolution, compact-object masses, and explosion properties of stripped binary stars. *Astron. Astrophys.*, 645:A5.

- Schneider, F. R. N., Ramírez-Agudelo, O. H., Tramper, F., Bestenlehner, J. M., Castro, N., Sana, H., Evans, C. J., Sabín-Sanjulián, C., Simón-Díaz, S., Langer, N., Fossati, L., Gräfener, G., Crowther, P. A., de Mink, S. E., de Koter, A., Gieles, M., Herrero, A., Izzard, R. G., Kalari, V., Klessen, R. S., Lennon, D. J., Mahy, L., Maíz Apellániz, J., Markova, N., van Loon, J. T., Vink, J. S., and Walborn, N. R. (2018a). The VLT-FLAMES Tarantula Survey. XXIX. Massive star formation in the local 30 Doradus starburst. *Astron. Astrophys.*, 618:A73.
- Schneider, F. R. N., Sana, H., Evans, C. J., Bestenlehner, J. M., Castro, N., Fossati, L., Gräfener, G., Langer, N., Ramírez-Agudelo, O. H., Sabín-Sanjulián, C., Simón-Díaz, S., Tramper, F., Crowther, P. A., de Koter, A., de Mink, S. E., Dufton, P. L., Garcia, M., Gieles, M., Hénault-Brunet, V., Herrero, A., Izzard, R. G., Kalari, V., Lennon, D. J., Maíz Apellániz, J., Markova, N., Najarro, F., Podsiadlowski, P., Puls, J., Taylor, W. D., van Loon, J. T., Vink, J. S., and Norman, C. (2018b). An excess of massive stars in the local 30 Doradus starburst. *Science*, 359(6371):69–71.
- Scutt, O. J., Murphy, S. J., Nielsen, M. B., Davies, G. R., Bedding, T. R., and Lyttle, A. J. (2023). Asteroseismology of δ Scuti stars: emulating model grids using a neural network. *Mon. Not. Roy. Astron. Soc.*, 525(4):5235–5244.
- Settles, B. (2009). Active learning literature survey. Computer Sciences Technical Report 1648, University of Wisconsin–Madison.
- Shibata, M. (2005). Constraining nuclear equations of state using gravitational waves from hypermassive neutron stars. *Phys. Rev. Lett.*, 94:201101.
- Shibata, M., Fujibayashi, S., Hotokezaka, K., Kiuchi, K., Kyutoku, K., Sekiguchi, Y., and Tanaka, M. (2017). Modeling gw170817 based on numerical relativity and its implications. *Phys. Rev. D*, 96:123012.
- Shibata, M., Taniguchi, K., and Uryū, K. b. o. (2005). Merger of binary neutron stars with realistic equations of state in full general relativity. *Phys. Rev. D*, 71:084021.
- Skilling, J. (2004). Nested Sampling. In Fischer, R., Preuss, R., and Toussaint, U. V., editors, *Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, volume 735 of *American Institute of Physics Conference Series*, pages 395–405.
- Skilling, J. (2006). Nested sampling for general Bayesian computation. *Bayesian Analysis*, 1(4):833 – 859.
- Smartt, S. J. (2009). Progenitors of Core-Collapse Supernovae. *Ann. Rev. Astron. Astrophys.*, 47(1):63–106.
- Smartt, S. J. (2015). Observational constraints on the progenitors of core-collapse supernovae: The case for missing high-mass stars. *Publications of the Astronomical Society of Australia*, 32.

- Smith, N., Li, W., Miller, A. A., Silverman, J. M., Filippenko, A. V., Cuillandre, J.-C., Cooper, M. C., Matheson, T., and Van Dyk, S. D. (2011). A Massive Progenitor of the Luminous Type II_n Supernova 2010jl. *Astrophys. J.*, 732(2):63.
- Soultanis, T., Bauswein, A., and Stergioulas, N. (2022). Analytic models of the spectral properties of gravitational waves from neutron star merger remnants. *Phys. Rev. D*, 105:043020.
- Soultanis, T., Maltsev, K., Bauswein, A., Chatziioannou, K., Roepke, F. K., and Stergioulas, N. (2024). Gravitational-wave model for neutron star merger remnants with supervised learning. *arXiv e-prints*, page arXiv:2405.09513.
- Speagle, J. S. (2020). dynesty: a dynamic nested sampling package for estimating Bayesian posteriors and evidences. *Monthly Notices of the Royal Astronomical Society*, 493(3):3132–3158.
- Springel, V., Pakmor, R., Pillepich, A., Weinberger, R., Nelson, D., Hernquist, L., Vogelsberger, M., Genel, S., Torrey, P., Marinacci, F., and Naiman, J. (2018). First results from the IllustrisTNG simulations: matter and galaxy clustering. *Mon. Not. Roy. Astron. Soc.*, 475(1):676–698.
- Srivastava, V., Davis, D., Kuns, K., Landry, P., Ballmer, S., Evans, M., Hall, E. D., Read, J., and Sathyaprakash, B. S. (2022). Science-driven tunable design of cosmic explorer detectors. *The Astrophysical Journal*, 931(1):22.
- Steiner, A. W., Hempel, M., and Fischer, T. (2013). Core-collapse Supernova Equations of State Based on Neutron Star Observations. *Astrophys. J.*, 774(1):17.
- Stergioulas, N., Bauswein, A., Zagkouris, K., and Janka, H.-T. (2011). Gravitational waves and non-axisymmetric oscillation modes in mergers of compact object binaries. *Monthly Notices of the Royal Astronomical Society*, 418(1):427–436.
- Stone, M. (2018). Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):111–133.
- Sukhbold, T. and Woosley, S. E. (2014). The Compactness of Presupernova Stellar Cores. *Astrophys. J.*, 783(1):10.
- Sukhbold, T., Woosley, S. E., and Heger, A. (2018). A High-resolution Study of Presupernova Core Structure. *Astrophys. J.*, 860(2):93.
- Taddia, F., Sollerman, J., Fremling, C., Barbarino, C., Karamehmetoglu, E., Arcavi, I., Cenko, S. B., Filippenko, A. V., Gal-Yam, A., Hiramatsu, D., Hosseinzadeh, G., Howell, D. A., Kulkarni, S. R., Laher, R., Lunnan, R., Masci, F., Nugent, P. E., Nyholm, A., Perley, D. A., Quimby, R., and Silverman, J. M. (2019). Analysis of broad-lined Type Ic supernovae from the (intermediate) Palomar Transient Factory. *Astron. Astrophys.*, 621:A71.
- Taggart, R. J. (2022). Point forecasting and forecast evaluation with generalized huber loss. *Electronic Journal of Statistics*, 16(1).

- Takahashi, K., Takiwaki, T., and Yoshida, T. (2023). Monotonicity of the Cores of Massive Stars. *Astrophys. J.*, 945(1):19.
- Takami, K., Rezzolla, L., and Baiotti, L. (2015). Spectral properties of the post-merger gravitational-wave signal from binary neutron stars. *Phys. Rev. D*, 91:064001.
- Tanikawa, A., Yoshida, T., Kinugawa, T., Takahashi, K., and Umeda, H. (2020). Fitting formulae for evolution tracks of massive stars under extreme metal-poor environments for population synthesis calculations and star cluster simulations. *Mon. Not. Roy. Astron. Soc.*, 495(4):4170–4191.
- Taylor, S. R. and Gerosa, D. (2018). Mining gravitational-wave catalogs to understand binary stellar evolution: A new hierarchical Bayesian framework. *Phys. Rev. D*, 98(8):083017.
- Temaj, D., Schneider, F. R. N., Laplace, E., Wei, D., and Podsiadlowski, P. (2024). Convective-core overshooting and the final fate of massive stars. *Astron. Astrophys.*, 682:A123.
- Timmes, F. X., Woosley, S. E., and Weaver, T. A. (1996). The Neutron Star and Black Hole Initial Mass Function. *Astrophys. J.*, 457:834.
- Tofallis, C. (2015). A better measure of relative prediction accuracy for model selection and model estimation. *Journal of the Operational Research Society*, 66(8):1352–1362.
- Tootle, S. D., Papenfort, L. J., Most, E. R., and Rezzolla, L. (2021). Quasi-universal behavior of the threshold mass in unequal-mass, spinning binary neutron star mergers. *The Astrophysical Journal Letters*, 922(1):L19.
- Torres-Rivas, A., Chatziioannou, K., Bauswein, A., and Clark, J. A. (2019). Observing the post-merger signal of gw170817-like events with improved gravitational-wave detectors. *Phys. Rev. D*, 99:044014.
- Tringali, M. C., Puecher, A., Lazzaro, C., Ciolfi, R., Drago, M., Giacomazzo, B., Vedovato, G., and Prodi, G. A. (2023). Morphology-independent characterization method of postmerger gravitational wave emission from binary neutron star coalescences. *Classical and Quantum Gravity*, 40(22):225008.
- Tsang, K. W., Dietrich, T., and Van Den Broeck, C. (2019). Modeling the postmerger gravitational wave signal and extracting binary properties from future binary neutron star detections. *Phys. Rev. D*, 100:044047.
- Ugliano, M., Janka, H.-T., Marek, A., and Arcones, A. (2012). Progenitor-explosion Connection and Remnant Birth Masses for Neutrino-driven Supernovae of Iron-core Progenitors. *Astrophys. J.*, 757(1):69.

- Valenti, S., Taubenberger, S., Pastorello, A., Aramyan, L., Botticella, M. T., Fraser, M., Benetti, S., Smartt, S. J., Cappellaro, E., Elias-Rosa, N., Ergon, M., Magill, L., Magnier, E., Kotak, R., Price, P. A., Sollerman, J., Tomasella, L., Turatto, M., and Wright, D. E. (2012). A Spectroscopically Normal Type Ic Supernova from a Very Massive Progenitor. *Astrophys. J. Lett.*, 749(2):L28.
- Van Tooren, C. and Haas, T. (1993). A site investigation strategy using moving window kriging and automated semivariogram modelling. In *Contaminated Soil'93: Fourth International KfK/TNO Conference on Contaminated Soil 3–7 May 1993, Berlin, Germany*, pages 609–622. Springer.
- Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, pages 5–42.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272.
- Weaver, T. A., Zimmerman, G. B., and Woosley, S. E. (1978). Presupernova evolution of massive stars. *Astrophys. J.*, 225:1021–1029.
- Wellstein, S. and Langer, N. (1999). Implications of massive close binaries for black hole formation and supernovae. *Astron. Astrophys.*, 350:148–162.
- Whittaker, T., East, W. E., Green, S. R., Lehner, L., and Yang, H. (2022). Using machine learning to parametrize postmerger signals from binary neutron stars. *arXiv e-prints*, page arXiv:2201.06461.
- Wijngaarden, M., Chatziioannou, K., Bauswein, A., Clark, J. A., and Cornish, N. J. (2022). Probing neutron stars with the full premerger and postmerger gravitational wave signal from binary coalescences. *Phys. Rev. D*, 105:104019.
- Wilson, J. R., Mathews, G. J., and Marronetti, P. (1996). Relativistic numerical model for close neutron-star binaries. *Phys. Rev. D*, 54:1317–1331.
- Woosley, S. E. and Bloom, J. S. (2006). The Supernova Gamma-Ray Burst Connection. *Ann. Rev. Astron. Astroph.*, 44(1):507–556.
- Woosley, S. E., Heger, A., and Weaver, T. A. (2002). The evolution and explosion of massive stars. *Reviews of Modern Physics*, 74(4):1015–1071.
- Yang, H., Paschalidis, V., Yagi, K., Lehner, L., Pretorius, F., and Yunes, N. (2018). Gravitational wave spectroscopy of binary neutron star merger remnants with mode stacking. *Phys. Rev. D*, 97:024049.
- Yoon, S.-C., Dessart, L., and Clochiatti, A. (2017). Type Ib and IIb Supernova Progenitors in Interacting Binary Systems. *Astrophys. J.*, 840(1):10.
- Zhugge, X., Centrella, J. M., and McMillan, S. L. W. (1996). Gravitational radiation from the coalescence of binary neutron stars: Effects due to the equation of state, spin, and mass ratio. *Phys. Rev. D.*, 54(12):7261–7277.