Machine Learning Detects Multiplicity of the First Stars in Stellar Archaeology Data

Tilman Hartwig[1,2,3] , Miho N. Ishigaki[3,4,5] , Chiaki Kobayashi[3,6] , Nozomu Tominaga[3,4,7] , and Ken'ichi Nomoto[3]

[1] Institute for Physics of Intelligence, School of Science, The University of Tokyo, Bunkyo, Tokyo 113-0033, Japan; miho.ishigaki@nao.ac.jp
[2] Department of Physics, School of Science, The University of Tokyo, Bunkyo, Tokyo 113-0033, Japan
[3] Kavli Institute for the Physics and Mathematics of the Universe (WPI), The University of Tokyo Institutes for Advanced Study, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan
[4] National Astronomical Observatory of Japan, Mitaka, Tokyo 181-8588, Japan
[5] Astronomical Institute, Tohoku University, 6-3, Aramaki, Aoba-ku, Sendai, Miyagi 980-8578, Japan
[6] Centre for Astrophysics Research, Department of Physics, Astronomy and Mathematics, University of Hertfordshire, Hatfield AL10 9AB, UK
[7] Department of Physics, Faculty of Science and Engineering, Konan University, 8-9-1 Okamoto, Kobe, Hyogo 658-8501, Japan
Received 2022 November 7; revised 2023 February 1; accepted 2023 February 6; published 2023 March 22

Abstract

In unveiling the nature of the first stars, the main astronomical clue is the elemental compositions of the second generation of stars, observed as extremely metal-poor (EMP) stars, in the Milky Way. However, no observational constraint was available on their multiplicity, which is crucial for understanding early phases of galaxy formation. We develop a new data-driven method to classify observed EMP stars into mono- or multi-enriched stars with support vector machines. We also use our own nucleosynthesis yields of core-collapse supernovae with mixing fallback that can explain many of the observed EMP stars. Our method predicts, for the first time, that $31.8\% \pm 2.3\%$ of 462 analyzed EMP stars are classified as mono-enriched. This means that the majority of EMP stars are likely multi-enriched, suggesting that the first stars were born in small clusters. Lower-metallicity stars are more likely to be enriched by a single supernova, most of which have high carbon enhancement. We also find that Fe, Mg. Ca, and C are the most informative elements for this classification. In addition, oxygen is very informative despite its low observability. Our data-driven method sheds a new light on solving the mystery of the first stars from the complex data set of Galactic archeology surveys.

Unified Astronomy Thesaurus concepts: Population III stars (1285); Population II stars (1284); Milky Way formation (1053); Support vector machine (1936)

Supporting material: machine-readable table

1. Introduction

Big Bang nucleosynthesis has produced hydrogen, helium, and trace amounts of lithium in the first minutes of the Universe. All heavier elements were synthesized and released by stars and their violent final fates, such as supernova (SN) explosions. The crucial transition from a primordial Universe to a Universe enriched with heavier elements (summarized as "metals" by astronomers) was initiated by the first stars, also termed Population III stars. These stars formed in pristine minihalos around redshift 6–30 (Bromm et al. 2002; Yoshida et al. 2003; Magg et al. 2016; Jaacks et al. 2019; Liu & Bromm 2020a; Skinner & Wise 2020; Kulkarni et al. 2021; Schauer et al. 2021; Hartwig et al. 2022). They ended the cosmic dark ages, they provided the first metals, they contributed to the reionization of the Universe, they might have provided the seeds for the first supermassive black holes (Woods et al. 2019), and they have set the scene for all subsequent galaxy formation (Taylor & Kobayashi 2014; Dayal & Ferrara 2018; Chen et al. 2020; Washinoue & Suzuki 2021). Despite their importance for cosmology and intensive studies in the last decades (Glover 2005; Greif 2015; Klessen 2019; Haemmerlé et al. 2020), only a little is known about the first stars.

Population III stars are believed to be more massive than present-day stars because primordial gas cools slower than metal-enriched gas and the resulting scale for fragmentation is therefore larger (Bromm & Loeb 2003). This theoretical consideration on the masses of Population III stars is supported by the lack of observations of surviving metal-free stars. If Population III stars were less massive than $\sim 0.8\,M_\odot$, their lifetime should be longer than the age of the Universe and we could observe metal-free stars in the Milky Way. Based on their nondetection, the lower-mass limit of the initial mass function (IMF) of the first stars is $\gtrsim 0.65\,M_\odot$ with a confidence level of 95% (Hartwig et al. 2015; Ishiyama et al. 2016; Magg et al. 2018, 2019; Rossi et al. 2021). Further observational constraints based on gravitational waves (Kinugawa et al. 2014; Hartwig et al. 2016; Liu & Bromm 2020b), high-z SN explosions (Hummel et al. 2012; Hartwig et al. 2018a; Regős et al. 2020; Rydberg et al. 2020), the chemical composition of quasar broad line regions (Yoshii et al. 2022), or direct detection (Schauer et al. 2020; Grisdale et al. 2021; Riaz et al. 2022) may become available in the future.

Numerical simulations of Population III star formation are an important addition to deepen our understanding of primordial star formation. All recent simulations that resolve the protostellar disk show fragmentation of the primordial gas and predict that the first stars should form in small clusters (Clark et al. 2011; Greif 2015; Liao & Turk 2019; Sharda et al. 2020; Sugimura et al. 2020; Wollenberg et al. 2020). Susa (2019) has shown that the number of fragments in a Population III–forming minihalo increases with time after the formation of the first protostar, and concluded that we should

expect 10–50 Population III protostars per minihalo. However, no numerical approach has simulated the formation process until radiative feedback halts accretion onto the protostars (Hosokawa et al. 2016) and the zero age main sequence (ZAMS) of Population III stars is reached. Therefore, we cannot make conclusive statements about the final masses and multiplicity of Population III stars, because the gas accretion, stellar mergers, and possible ejections out of the minihalo are not yet fully simulated. In conclusion, all recent simulations suggest that Population III stars could form in a small cluster, but we do not yet have an observational confirmation for this conjectural result of numerical simulations.

The multiplicity of the first stars is not only an academic curiosity, but has several direct implications. The Population III multiplicity defines the number of primordial stars per minihalo, and combined with the characteristic mass of Population III stars, this quantity allows us to constrain the star formation efficiency of primordial gas. The resulting cosmic number density of Population III stars also sets tighter constraints on their contribution to reionization, to the Lyman–Werner background, and to the number of available stellar mass seed black holes. Moreover, stellar multiplicity is a necessary condition to form stellar binaries. Population III stellar binaries (Stacy & Bromm 2013; Sugimura et al. 2020; Liu et al. 2021) might provide binary black holes that can be observed with gravitational wave detectors, or they might evolve into X-ray binaries that can provide X-rays at high redshift or an additional source for cosmic reionization.

Another independent approach to indirectly constrain the nature of the first stars is stellar archeology (e.g., Frebel & Norris 2015). The first SNe have released metals into the previously pristine interstellar medium (ISM). Second-generation stars can form out of this enriched gas, which still carries the specific chemical fingerprint of the first SNe. Some of these second-generation stars have survived until the present day and we can observe them as old, metal-poor stars in the Milky Way. If we can determine the chemical composition of such extremely metal-poor (EMP, [Fe/H][8] $\leqslant -3$) stars with high-resolution spectroscopy, we can infer the nature of the first SNe and therefore constrain the properties of the first stars (e.g., Umeda & Nomoto 2003; Tominaga et al. 2014; Placco et al. 2016; Fraser et al. 2017; Ishigaki et al. 2018; Choplin et al. 2019; Hansen et al. 2020; Placco et al. 2021; Skúladóttir et al. 2021; Hartwig et al. 2022).

Previously, a standard assumption of stellar archeology has been that EMP stars are mono-enriched, i.e., that one EMP star contains metals from only one enriching SN (Audouze & Silk 1995). Based on this assumption, the initial mass function of the first SNe was observationally estimated (Ishigaki et al. 2018). However, recent numerical simulations and semi-analytical models (Hartwig et al. 2018b, 2019) suggest that Population III stars form in small clusters and that there should be multiple SN explosions in one minihalo. Consequently, we expect that at least some of the observed EMP stars are multi-enriched. While it is in principle possible to assume multi-enrichment in traditional abundance fitting methods, it increases the number of free parameters to fit the chemical composition of an EMP star and therefore weakens the predictive power (Chan & Heger 2017; Salvadori et al. 2019).

In this study, we develop a new method with supervised machine learning (Section 2) to constrain the number of enriching Population III SNe, and we apply the method to 462 EMP stars from literature. Our method allows us to constrain the multiplicity of the first stars, for the first time (Section 3). Section 4 and Section 5 provide a discussion and our conclusions.

2. Methodology

This is the first approach to classify a representative set of EMP stars into mono- and multi-enriched. Therefore, there is no convenient ground truth for this problem, i.e., a catalog that labels observed EMP stars with the number of SNe that enriched the gas out of which they formed. In order to deconvolve the physical properties of the first stars from complex observational data, we therefore took the following steps: (i) we carefully selected a sample of EMP stars; (ii) we then created a realistic subset of synthetic SN yields, (iii) which we then used to generate representative mock observations to (iv) train and optimize our Support Vector Machine (SVM) classifier.

2.1. EMP Star Observations

In this study, we analyzed EMP stars with [Fe/H] $\leqslant -3$. We base this metallicity threshold on two main criteria: first, this is the conventional threshold to define EMP stars (Beers & Christlieb 2005). Second, the metallicity contribution from Population III SNe dominates the overall metal mass up to [Fe/H] ~ -3 (Hartwig et al. 2018b; Ishigaki et al. 2021), which implies that EMP stars are mainly enriched by the first stars. We selected 432 EMP stars from the SAGA database (Suda et al. 2008), added 26 stars (Cohen et al. 2013; Yong et al. 2013; Roederer et al. 2014) from the compilation by Ishigaki et al. (2018) that have not yet been listed in the SAGA database, and added four recent stars at the lowest metallicities (Aguado et al. 2018; Nordlander et al. 2019; Placco et al. 2021; Skúladóttir et al. 2021).

Our analysis is based on abundance ratios of metals with respect to each other and not with respect to hydrogen. There are two main reasons for this: First, we want to analyze the observed stars independently of their absolute metallicity. This enables us to extend the analysis in the future to stars with [Fe/H] > -3, where we also expect interesting, multi-enriched stars with contributions from Population III SNe (Salvadori et al. 2019; Ishigaki et al. 2021; Aguado et al. 2023). Second, we created mock observations to mimic the observed stars. Modeling the mixing of metals with pristine gas after Population III SNe is a stochastic process (Tarumi et al. 2020) that requires high-resolution hydrodynamical simulations (Ritter et al. 2015; Sluder et al. 2016) for every single star, which is beyond the scope of current computational capacities.

The data quality and availability of various elements differ among EMP stars, depending on, e.g., the instrument, signal-to-noise ratio, and covered spectral range. We created the mock observations to have the same observational properties as the sample of observed EMP stars that we analyzed. Moreover, we only included stars with a detected [Fe/H] value (and not upper limits) to investigate trends with metallicity. For all possible combinations of abundance ratios with elements between carbon and zinc, we calculate their observability as a fraction of the EMP star sample for which this abundance ratio is measured. For example, the abundance ratio [Ca/Fe] is

---

[8] Defined as [X/H] $= \log_{10}(N_X/N_H) - \log_{10}(N_{X,\odot}/N_{H,\odot})$ throughout the paper, where N_X is the abundance of metals, N_H is the abundance of hydrogen, and $N_{X,\odot}$ and $N_{H,\odot}$ are the solar abundances of these (Asplund et al. 2009).

available for 452/462 EMP stars in our sample, and [Ca/Mn] is still available for ∼63% of EMP stars. It should be noted that upper detection limits were counted as nondetections because they are not useful for our supervised classification. We include the following 13 elements, for reasons we will explain below: C, O, Na, Mg, Al, Si, Ca, Cr, Mn, Fe, Co, Ni, Zn.

These observed abundances reflect the current composition of the star in the stellar photosphere. However, the surface carbon abundance of a star can be reduced due to CN processing in the upper red giant branch. Hence, the observed carbon abundance can be smaller than the natal carbon abundance, which is relevant for this study. To correct for this effect, we apply the carbon corrections based on those of Placco et al. (2014), which take into account the metallicity and surface gravity.[9] While this correction can affect the classification of individual EMP stars, it does not affect the mean fraction of multi-enriched EMP stars in our sample. Throughout the paper, we will use the corrected carbon abundances, and only in Appendix B do we show results without the carbon corrections.

2.2. Theoretical Supernova Models

We used of a set of Population III SN yields (Ishigaki et al. 2018), which are based on pre-SN and explosive nucleosynthesis calculations (Umeda et al. 2000; Umeda & Nomoto 2005; Tominaga et al. 2007). Both the mass cut and mixing are treated as free parameters under the framework of the mixing-fallback model (Umeda & Nomoto 2002, 2005; Tominaga et al. 2007). The mass cut is constrained empirically to include cases where the explosion is not spherically symmetric, in which case a single value of the mass cut may not be defined well (Umeda & Nomoto 2002; Tominaga 2009; Ezzeddine et al. 2019). These yields have been proven to be successful in reproducing the chemical composition of individual EMP stars, including carbon-enhanced metal-poor (CEMP) stars (Tominaga et al. 2007; Ishigaki et al. 2018), as well as to explain the more complex chemical evolution of the Milky Way (Kobayashi et al. 2006) and Damped Lyα (DLA) Systems (Kobayashi et al. 2011b).

To ensure that we only took into account yields of realistic Population III SNe, i.e., SNe that occur in nature, we determined how successful a particular SN yield model is at reproducing observed abundance patterns of EMP stars. We did this by calculating the reduced χ^2 between each model and the selection of EMP stars (Ishigaki et al. 2018) (differences between the $[X/H]_i$ values), taking into account element-specific observational uncertainties. For every yield model, we noted the mean χ^2 that it can provide, averaged over all observations.

The goal was to identify a threshold so that all yield models with χ^2 below this value can be considered realistic SN yields. To find such an an optimal, data-driven threshold, we compared the distributions of the 78 considered abundance ratios between the selected subsets (sets of SN yield models with χ^2 below a certain threshold) and our set of 462 observed EMP stars. We found that the range $\chi^2 < 15.1$ provides the best agreement between the abundance ratios of observed EMP stars and Population III SN yields (see Section 4.4 for more details). This threshold on χ^2 selects 512 different SN yield models.

We also added 27 yield models that were identified as best fits to individual EMP stars (Ishigaki et al. 2018) and that were

not yet included. While the first subset accounts for yields that are a good fit, on average, to many EMP stars, the second subset guarantees that more exotic yields that are a good fit to only one or few EMP stars are included. These additional 27 yield models include abundances with low [Mg/Fe] and high [C/Fe].

Although we select a theoretical yield set under the assumption of mono-enrichment to determine the best fits, the yield set is representative of both mono- and multi-enrichment. This is because multi-enriched SN yields usually form a centrally concentrated subset in the abundance space of mono-enriched yields. With this approach, we can reject the null hypothesis that all EMP stars are mono-enriched.

Our subset of theoretical yields comprises 539 SN yield models, including progenitor stars of masses (13, 15, 25, 40, and 100) $M_\odot$ and explosion energies of (0.5, 1, 10, 30, and 60) × 10^{51} erg (Ishigaki et al. 2018). Most of the selected SN models are normal CCSNe with progenitor masses in the range 13–25 $M_\odot$ or hypernovae with 25–40 $M_\odot$, and 24% of our training set comprises faint SNe with ejected masses of ^{56}Ni $\lesssim 10^{-3} M_\odot$, which decays to Fe.

2.3. Mock Observations

In this section, we explain the main formalism to discriminate enrichment by one versus multiple SNe. To create mock observations of mono-enriched stars, we directly took the masses of yields from the selected 539 Population III models.

We created mock observations of multi-enriched stars by adding the yields of multiple randomly selected SN yield models together. Here, we implicitly assume that the metals from enriching SNe mix homogeneously. While this assumption may not be valid under all conditions (Ritter et al. 2015; Sluder et al. 2016), it avoids the inclusion of unconstrained mixing fractions. We did not take into account the effects of inhomogeneous mixing between different elements (Chiaki & Tominaga 2020) or the energy-limited hydrogen dilution mass (Magg et al. 2020).

Multi-enrichment could result from any combination of ⩾2 SNe. In a pilot study, we attempted to discriminate between different levels of multi-enrichment from two, three, four, or five SNe. However, the discrimination of N-fold enrichment is very degenerate, and it is already challenging to discriminate enrichment from one and two SNe. Therefore, we only distinguished between mono- and multi-enrichment in this study, and we did not aim to quantify the level of multi-enrichment. First, this makes our prediction more robust because we mitigate the degeneracy for various levels of multi-enrichment. Second, we are mainly interested in discriminating mono- and multi-enrichment, and the exact level of multi-enrichment is only of secondary importance. Therefore, we created multi-enriched mock observations by combining two-fold, three-fold, four-fold, and five-fold enrichment into one set with the same size as the set of mono-enriched mock observations. In addition, we verified that the exact number and combination of SNe that we include in the multi-enriched set does not affect the final classification of mono- or multi-enrichment.

We augmented the amount of multi-enriched mock observations by a factor 16 to include additional possible combinations of multi-enrichment. With four different levels of multi-enrichment (enrichment from two to five SNe) and a training set size of 50%, this provided us with approximately 17,000

---

multi-enriched mock observations for training, and we generate the same amount of mono-enriched mock observations. In all our sets for training, cross-validation, and blind testing, the fraction of mono- and multi-enriched samples is always 50% each. In other words, our prior assumption is that mono- and multi-enrichment are equally likely (see Section 4.3 for more details).

Once we calculated the abundance ratios of our mock observations, we applied the observational masks to mimic the observability: for each of our 462 EMP stars, we determined the observability masks, i.e., an array that contains the information if a certain abundance ratio is observable. We then applied these masks to our validation and test data. This guaranteed that we used the same information content when analyzing the mock observations, as it is available when we eventually evaluate the actual EMP stars.

In this approach, we began with 13 elements, namely C, O, Na, Mg, Al, Si, Ca, Cr, Mn, Fe, Co, Ni, and Zn, which enables the construction of 78 independent abundance ratios. We wanted to keep the number of elements small, since the complexity of the model scales with the number of elements and more complex models are prone to overfitting. Therefore, we only used elements for which we have a sufficient number of observations for EMP stars. We also excluded Sc and Ti because our theoretical models cannot correctly resolve the production of these elements (Tominaga 2009; Ishigaki et al. 2018; Kobayashi et al. 2020).

While N can also provide constraints on Population III SNe, theoretical uncertainties of N yields are large due to the potential presence of stellar rotation (Meynet & Maeder 2002; Hirschi 2007; Choplin 2019). Since the number of N measurements is also small, we do not use N in this work.

Finally, to account for theoretical uncertainties, we added random scatter to the mock observations. We introduce a matrix of error bars for specific combinations of abundance ratios because some of the elements share the same physical or observational reasons for uncertainties (see Appendix A). It should also be noted that we include observational uncertainties at classification stage with bootstrap sampling (see below).

2.4. Ground Truth Sample

Figure 1 presents the distribution of mock observations enriched by one (orange circles) or multiple (blue triangles) SNe. In this example, we used [C/Mg] and [Ca/Fe] as the two dimensions, because we show below that these are the most informative elements for this classification purpose. In general, mono-enriched stars span a larger area in this abundance space, whereas multi-enriched stars are more clustered toward the center. This makes sense because the abundance ratio of a multi-enriched star is the weighted mean of the individual abundance ratios of the individual SNe that contributed to the enrichment. In other words, by combining two SN yields, the resulting abundance ratios will never be more extreme than the abundance ratios of any of the individual contributing SNe. This figure also illustrates a degeneracy where the abundance ranges are largely overlapping if only a small number of elements are used. As a consequence of this degeneracy, mono-enriched stars can be classified more reliably because there are regions of the parameter space that can only be reached by the yields of single SNe. Hence, we mostly focus on the probabilities and fractions for mono-enrichment, as they are more reliable.
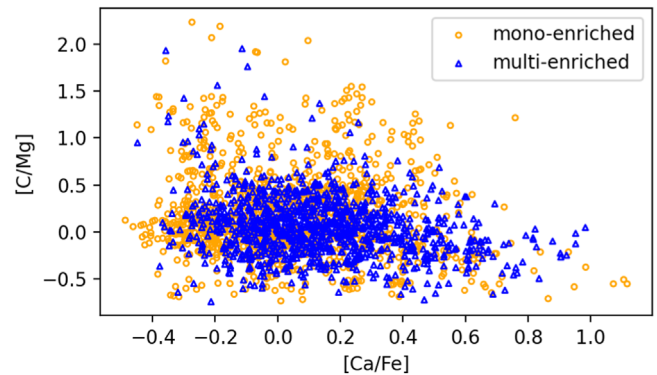
Figure 1. Ground truth training data for EMP stars enriched by one (orange circles) or multiple (blue triangles) SNe. Multi-enriched EMP stars are more centrally concentrated because their yields are a weighted average of individual SNe. Exceptions to this trend can result from theoretical uncertainties that are added as scatter.

2.5. Supervised Classification with Support Vector Machines

The basis of this approach are SVMs, a supervised machine-learning technique that iteratively finds a hyperplane in the feature space that optimally discriminates two classes (Cortes & Vapnik 1995). SVMs have been applied successfully to various astrophysical tasks (Wadadekar 2005; Huertas-Company et al. 2008; Małek et al. 2013; Marton et al. 2016). If the data are not linearly separable (like in our case), the SVM attempts to find an optimum by minimizing the number of misclassifications and their distance to the decision boundary.

We show below that four dimensions are optimal for the SVMs in this problem. Hence, the training data are points $x_i \in \mathbb{R}^4$ with their associated classes $y_i \in \{-1, 1\}$. The learning goal for the linear SVM is to find optimal values for w, b, and ζ that minimize

$$\frac{1}{2}w^T w + C \sum_i \zeta_i \qquad (1)$$

under the constraint

$$y_i(w^T x_i + b) \geqslant 1 - \zeta_i, \qquad (2)$$

with $\zeta_i > 0$ and the index i running over all training examples (Chang & Lin 2011). The second equation guarantees that every point is on the correct side of the decision boundary, or at most ζ_i away from their correct margin boundary. The first equation aims at maximizing the margin between the two classes by minimizing $||w||$. The second term minimizes the allowed tolerance ζ_i with the regularization parameter, for which we find $C = 1$ to be an optimal choice. The final classification of the SVM for a new observation x_i is obtained by evaluating the sign of $w^T x_i + b$.

Instead of using the points x_i directly in the linear SVM, one can also augment the data or replace the x_i with a kernel function $\phi(x_i)$. This kernel trick might enable the transformation of training data that are not linearly separable in 4D into linearly separable data in higher dimensions. We found that radial basis functions with a kernel width of $\gamma = 1$ are an ideal choice for our specific classification problem.

For the training, we assumed that all abundance ratios are observable. For the cross-validation and blind test data, we

accounted for the fact that certain abundances and combinations of abundance ratios are not always observable in all EMP stars.

2.6. Ensemble Learning

A single supervised classifier is susceptible to overfitting. Consequently, a single SVM has a higher generalization error and might perform poorly on new, unseen data. To overcome this problem, we employed ensemble learning and combined multiple SVMs into one final prediction. One additional advantage of ensemble learning for our approach is that we can fully exploit the available abundance ratios.

The optimal dimension for the SVM is a balance between two factors: a higher dimension provides better classification accuracy because it is based on more information and the SVM has more flexibility to identify an optimal decision boundary. On the other hand, each dimension introduces one new abundance ratio, which might not be observable for EMP stars. If only one element is not available, this SVM cannot be used for such an observation. Thus, we find that 4D is a good compromise.

We combined multiple SVMs to avoid overfitting and mitigate missing data. The strategy was that we trained several independent SVMs and let them vote on the final classification. In addition, we performed bootstrap sampling to mimic observational uncertainties. The final classification for one star is an average of up to $N_B \times N_{SVM}$ individual classifications, where $N_B = 49$ is the number of bootstrap samples and $N_{SVM} \leqslant 10$ is the number of SVMs used in this paper.

We first resample the observed abundance ratios 49 times, based on the uncertainty matrix (see Appendix A). For the number of resamplings, we have chosen an odd number in order to break ties. The results do not change with a larger number of resamplings. For each sample, we obtain a classification result ($p_{mono,i} \in [0, 1]$, i.e., 0 for mono-enriched and 1 for multi-enriched in our case). For an individual EMP star, we obtain its mean value and uncertainty over the bootstrap samples in the form $p_{mono} = \bar{x} \pm \sigma_x$. For the entire sample of EMP stars, we first calculate the mean fraction of mono-enriched EMP stars as $f_{mono,i}$ for one bootstrap sample. Then, we calculate the mean value and uncertainty of the mono-enriched fraction for the entire sample, based on the N_B bootstrap samples. Throughout the paper, we use this latter approach (resampling the ensemble of EMP stars) as our fiducial model and report the uncertainty of individual EMP stars wherever appropriate.

If the abundance ratios for some SVMs are not available, they cannot provide a prediction, and we take the average only from SVMs that can be used for that specific observation. Therefore, it is essential to select an optimal configuration of SVMs that fully exploits the available abundance ratios. With 78 different abundance ratios, there are over one million possible combinations to choose the four dimensions of an SVM. To find the optimal combination, we proceeded in the following manner: we created all possible quadruples of abundance ratios and sorted them by observability. Then, we selected the top 10 most observable abundance ratio quadruples and augmented them with 6 additional quadruples so that each of the 13 considered elements is included at least once in a quadruple. Next, we trained 16 SVMs with these most promising quadruples of abundance ratios as dimensions. This provided us with 16 different 4D SVMs that are trained on different abundance ratios. Then, we used the cross-validation set to test how all possible combinations of these 16 SVMs would perform.

We selected our optimal model based on the validation accuracy and several other constraints: we aimed for a maximum fraction of (mock) observations to be classified. Hence, we prefer combinations of abundance ratios with high observability. Moreover, we required a symmetric confusion matrix and a large dynamical range of possible predictions. Our optimal fiducial model is based on 10 different 4D SVMs that use 24 different abundance ratios. The two elements Mn and Ni are not included in the final set of 10 SVMs. This selection is data-driven, based on the input yields and our optimization goals.

We also considered other classification algorithms, such as deep neural networks, decision trees, and a regressive approach. Since we created our own mock observations for the training, our problem is not limited by data. In such a case, most classification algorithms show a similar marginal performance because, with sufficient training data and a consequently dense training set, every classification problem becomes a nearest-neighbor search. We verified this in the initial phase of the project, and similar behavior was shown by independent studies on astrophysical and generic data (Bazell & Aha 2001; Lan et al. 2020). Moreover, SVMs produce inspectable models and therefore provide a certain intuition regarding how the decision boundary is derived. This is very important for the physical interpretation of the results and therefore provides an additional validation.
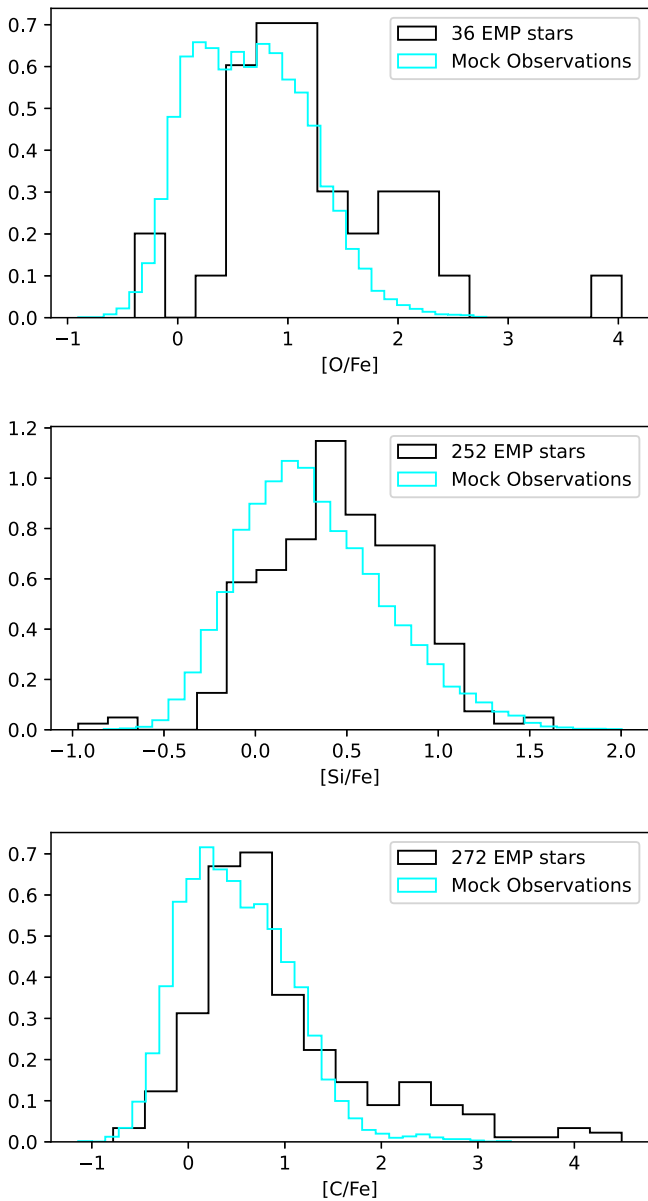
2.7. Are the Mock Data Realistic?

The first important question is if our training data are representative of the actual data (Acquaviva et al. 2020). For example, if we train a supervised classifier to discriminate cats and dogs, but then use it to distinguish elephants and penguins, the final classification will not be accurate, even if the test accuracy on a set of cats and dogs is good.

In Figure 2, we compare the abundance distributions from the mock observations and the EMP stars for three representative abundance ratios. In an appendix, we show the remaining 21 histograms for all abundance ratios used. By construction, our subset of Population III SN yields minimizes the mean KS distance between mock observations and real EMP stars. Phrased differently, it is the best subset that we could select for this purpose. Here, we check if the best is good enough.

For most of the abundance ratios, the distributions of mock and actual observations are similar. However, there are discrepancies in the distributions for a few abundance ratios, which may introduce a bias in our conclusions. There is no independent method to confirm if our mock observations are realistic. However, three factors support our approach to constructing the mock observations.

First: We verified that our final conclusion is not affected when we modify the subset of Population III SNe that we use as training set (see Section 4.4). In summary, we increased and decreased the number of included Population III SN yields by a factor of two, which did not significantly change the final results. Moreover, we shifted and scaled the values of the 10 most informative abundance ratios so that the means and standard deviations of their distributions match exactly with those of the distribution of the EMP stars. This did also not significantly affect the final classification. Therefore, our
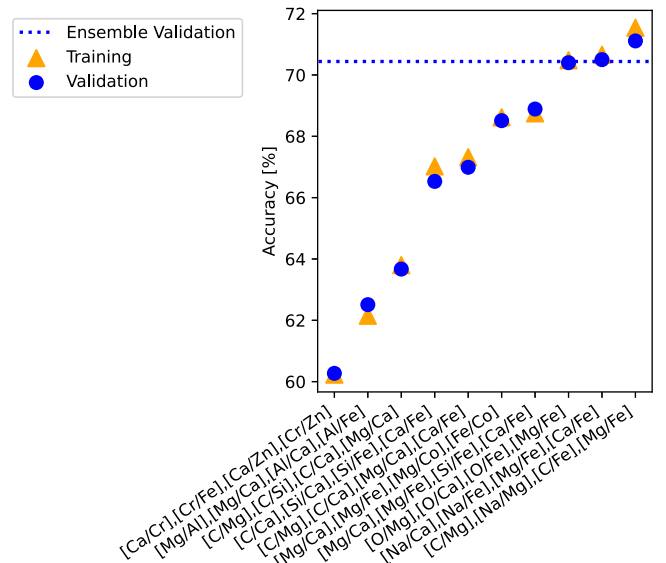
Figure 3. Prediction accuracy on the training set (triangles) and validation set (dots) for the individual SVMs. The dashed line shows the validation accuracy of the ensemble, i.e., the combined prediction of the 10 SVMs. Individual SVMs can achieve a higher accuracy than the ensemble method. However, such SVMs need specific elements (O, Na), which are not observable for all stars. The ensemble validation accuracy is valid for all classifications, irrespective of the number of observed elements. This high ensemble validation accuracy demonstrates the advantage of an ensemble learning approach. The order of the 10 SVMs from left to right is sorted by accuracy as a visual aid.

Figure 2. Comparison of the abundance distribution for the three most informative abundance ratios. While the distributions look qualitatively similar, the distribution of [O/Fe] is shifted by >0.5 dex, which might introduce a bias in the final classification.

training data might not be perfect, but we are confident that the final conclusion of this study is not affected by the exact selection of training data.

Second: The final decision of the SVMs is based on a majority vote. If the combined classifications of the independent SVMs is within 1σ of the decision boundary, we can consider the sample as not being classified. This makes the decision process more robust toward new, unseen observations. Thus, even if the ranges of mock and actual observations are not precisely overlapping, our classification process is sufficiently robust to handle outliers.

Third: If all these histograms for all abundance ratios matched exactly, we would not need this study. It would imply that we understand the enrichment and formation history of EMP stars sufficiently well. Unfortunately, this is not the case. The mock observations are constructed based on our current understanding of EMP enrichment. Therefore, the mismatch

between actual observations and mock observations highlights that our current understanding is not sufficient. Instead, based on our current state-of-the-art assumptions regarding Population III SNe and EMP stars, we construct our mock observations and thereby contribute to iteratively improving our understanding of the physics involved.

2.8. Accuracy of SVMs

We measured the prediction accuracy on the training set and on a second validation set that was not used in the training. We defined the accuracy as the fraction of correctly classified mock observations. The accuracy for the individual SVMs and for the combined prediction is presented in Figure 3. The validation accuracy provides an independent estimate on unseen data. The validation accuracy of the ensemble learning model is higher than most individual accuracies. This demonstrates the advantage of a weighted vote between the 10 SVMs compared to the prediction of a single SVM.

To further understand the predictions, errors, and associated biases, we illustrate the confusion matrix in Figure 4. Overall, the confusion matrix is symmetric: the misclassifications are almost equally distributed with ~13.8% and ~15.6%, and the correctly classified cases are also almost equal with ~34.2% and ~36.0%. Our prediction pipeline predicts mono-enrichment for ~48.0% and multi-enrichment for ~51.6% of the classified cases.

To further quantify the Bayes error and degeneracy of this problem, we repeated the training of our model, excluding theoretical or observational uncertainty. Under these optimal conditions, we found 79% of mock observations are correctly classified (compared to 70% with realistic uncertainties). Phrased differently, the intrinsic degeneracy of this classification task is a major challenge, and improved observational data

Figure 4. Confusion matrix of our ensemble classifier. The confusion matrix is balanced and symmetric.

can increase the accuracy, but will not be able to solve this underlying degeneracy completely.

3. Results

In this paper, we analyze whether EMP stars are likely to be enriched by a single SN or by multiple SNe, in order to explain the observed elemental abundances. For this purpose, we calculated nucleosynthesis yields of over 13,000 SNe covering all possible parameters for the first stars. Then, we trained an ensemble of SVMs on mock observations to classify high-resolution spectroscopic data of EMP stars ($[Fe/H] \leqslant -3$), using detailed chemical compositions from carbon to zinc assuming an equal contribution of mono- and multi-enrichment in our training sample. The possibility of mono-enrichment, which is determined by 10 SVMs and 49 bootstrap samplings, is shown in Figure 5(a). From the original 462 EMP stars, we exclude 35 stars because they have only three elements observed, which is not enough for a reliable classification. Moreover, three EMP stars have a predicted $p_{mono} = 0.5$ and can therefore not be assigned to any category. From the remaining 424 EMP stars, we find the average fraction of mono-enriched stars to be $31.8\% \pm 2.3\%$, where the standard deviation reflects observational uncertainties.

As the number of stars sharply decrease toward lower metallicities, both numbers of mono- and multi-enriched stars in our sample decrease at lower metallicities in panel (b). Panel (c) indicates a clear metallicity dependence of the ratio between them; stars at the lowest metallicities are likely to be mono-enriched, which has been assumed in numerous previous studies (Umeda & Nomoto 2003; Placco et al. 2015; Ishigaki et al. 2018) but has never been tested. At higher metallicities, stars tend to be multi-enriched. Our observation-based confirmation of this trend is remarkable because we do not use $[Fe/H]$ values to train our SVMs but instead use solely relative abundance ratios of various metals, excluding hydrogen. What is surprising is that the mono-enriched fraction is not 100% at $[Fe/H] \sim -4.5$, which means that some second-generation stars were already enriched by multiple SN explosions.

3.1. CEMP Stars

For EMP stars, the most notable feature is their carbon enhancement; a large fraction of EMP stars show a large carbon enhancement relative to iron (Beers & Christlieb 2005; Placco et al. 2014; Zepeda et al. 2022), and multiple populations have been identified in the diagram of $[C/H]$–$[Fe/H]$ (Bonifacio et al. 2015; Yoon et al. 2016). In Figure 6, we depict the carbon versus iron abundance of EMP stars, color-coded by the probability for mono-enrichment. We find a positive correlation

between p_{mono} and $[C/Fe]$. EMP stars with high $[C/Fe]$ are less likely to be multi-enriched. Specifically, 75 of 125 carbon-enhanced metal-poor stars (CEMP; $[C/Fe] > 0.7$; Aoki et al. 2007; Arentsen et al. 2022) are mono-enriched, and all 49 stars with $[C/Fe] > 1.5$ are mono-enriched. In the terms used in previous works (Rossi et al. 2005; Spite et al. 2013; Bonifacio et al. 2015; Yoon et al. 2016), all Group III stars are mono-enriched. The origin of this bi- or multimodality can be explained as follows: Multi-enrichment tends to average yields and makes them more centrally concentrated in the abundance space. Faint SNe, which are known to be important in the early Universe (Umeda & Nomoto 2003; Kobayashi et al. 2011b), produce only small amounts of iron due to their larger black hole than for normal SNe; mixing their yields with normal CCSNe attenuates the initially high $[C/Fe]$ from the faint SN to a smaller value. Therefore, it becomes more difficult for CEMP stars to form after multiple SNe have exploded in one minihalo (Jeon et al. 2021). On the other hand, for C-normal stars, once they are enriched by a normal SN, it becomes impossible to eliminate the possibility of additional enrichment from faint SNe in our analysis; hence our estimated number of SNe is a lower limit.

One could speculate if mono-enrichment and carbon enhancement are synonymous or one is a subset of the other. To allow readers their own conclusion based on the preferred threshold of carbon enhancement, we provide the quantitative classification data in the carbon-enhanced regime in Table 1.

The fraction of mono-enriched stars increases with $[C/Fe]$. The last column shows the fraction of all EMP stars for which carbon enhancement is a consequence of being mono-enriched *and* mono-enrichment is a consequence of being carbon enhanced. That is, the missing stars to 100% are those that are either mono-enriched but not carbon enhanced, or that are carbon enhanced but not mono-enriched. This fraction is highest around $[C/Fe] \sim 1.1$. It declines at higher $[C/Fe]$ because there are too many mono-enriched stars, which are not classified as carbon enhanced anymore due to the higher threshold. Phrased differently, if we want to define a physics-informed threshold for CEMP stars based on the ability to discriminate mono- from multi-enriched EMP stars, the best threshold would be around $[C/Fe] \sim 1.1$.

The classification of EMP stars and their distribution on the $[C/H]$–$[Fe/H]$ plane is affected by the carbon corrections. Therefore, we also provide a version of this figure without the carbon corrections in Appendix B.

3.2. Most Metal-poor Stars

In Table 2, we show the classification of the most iron-poor stars in our sample. It will be interesting to model their exact formation scenarios based on the number of enriching SNe in future works. However, one has to be cautious with the direct interpretation of the provided face values. About 70% of samples in the blind test set have been classified correctly (80% if we only take into account EMP stars for which the predicted p_{mono} is more than one standard deviation away from the decision boundary). Therefore, while the average fraction of mono-enriched stars is reliable, individual values for the number of enriching SNe should not be overinterpreted.
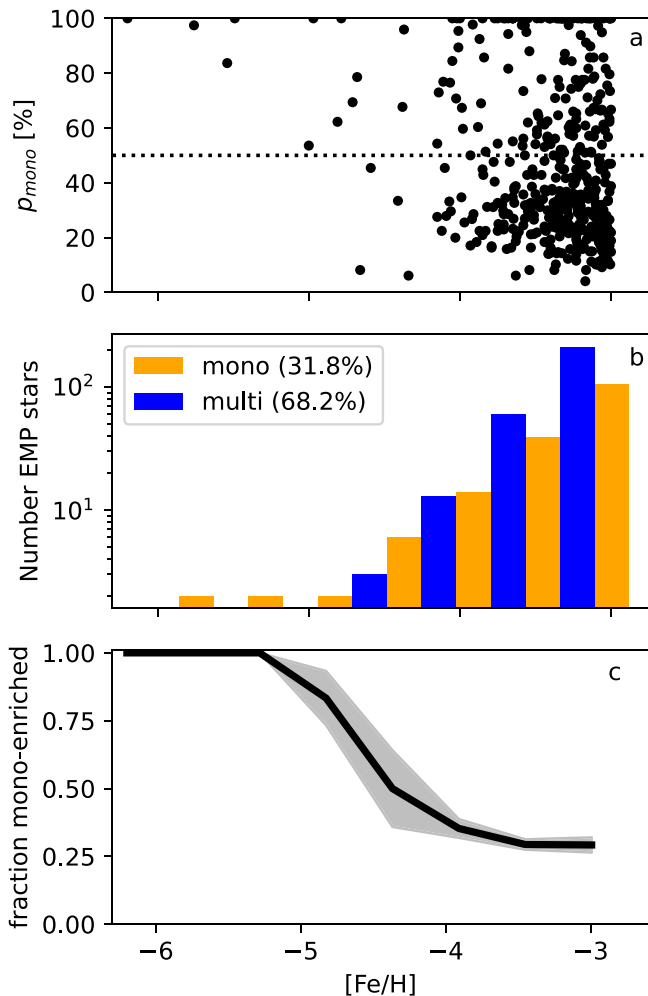
Figure 5. All panels show the results as a function of metallicity. Panel (a) shows the probability of mono-enrichment, p_{mono}, for individual EMP stars. Panel (b) shows a histogram with the absolute number of classified mono-enriched (orange) and multi-enriched (blue) EMP stars in each bin (bins are slightly offset for clarity). Panel (c) shows the fraction of mono-enriched EMP stars, and the gray contour illustrates scatter due to the bootstrap resampling of observational uncertainties.

3.3. Most Informative Elements

Finally, to understand what elements are most informative in the decision process, we calculate the permutation feature importance (Breiman 2001) of the classification pipeline. The feature importance for element X is defined as the decrease in cross-validation accuracy if we randomly shuffle the values of all abundances that include element X in the cross-validation data. This score indicates how much the model depends on a specific element. In our case, the maximum cross-validation accuracy is 70%, and the accuracy for random guessing is 50%. So possible values for the feature importance are in the range 0%–20%. We show the feature importance for the 11 used elements in Figure 7, where the most informative elements are Fe, Mg, Ca, and C. These are the elements that are most valuable to discriminate mono- from multi-enriched EMP stars in our analysis and are usually included in observational surveys.

Two main effects dominate the feature importance of an abundance ratio: first, if an abundance ratio spans a large range in [X/Y], finding a meaningful decision boundary is easier. The training data contain scatter to mimic theoretical

uncertainty. If the range over which abundance ratios in the training set are distributed is of the same order as the scatter, it is difficult to draw a reasonable decision boundary through this data. However, if the abundance ratios span a range that is significantly larger than the uncertainties, the SVM has more flexibility to identify an informative decision boundary to divide the data. The second effect is the observability. If an element is available for most EMP stars, the SVMs likely rely on it for the classification. To differentiate the feature importance from the observability, we divided the feature importance of an element by its respective observability. Although the order of most informative elements barely changes, we identify oxygen as very informative, relative to its low observability. Therefore, future observations of oxygen will be useful to distinguish between mono- and multi-enriched EMP stars.

4. Discussion

Our findings strongly indicate that a significant number of Population III–forming minihalos experience multiple SNe prior to EMP star formation, which suggests that most first stars formed in small clusters that contained multiple massive stars (Peebles & Dicke 1968) rather than as isolated massive stars (Doroshkevich et al. 1967). However, it is not easy to estimate the exact multiplicity of the first stars, because our result was obtained under the prior assumption that mono- and multi-enrichment are equally likely. Also, based on EMP star observations, the nature of the first stars can be investigated only for those exploded as SNe (Ishigaki et al. 2018).

Our finding of the need for multiplicity is consistent with recent hydrodynamical simulations of Population III star formation, which show fragmentation of the primordial gas cloud and predict that the first stars could form in small clusters (Clark et al. 2011; Hirano & Bromm 2017), resulting in multiple Population III SNe per minihalo. The number of fragments in a minihalo increases with time after the formation of the first protostar, and the number of Population III protostars per minihalo is expected to be 10–50 (Susa 2019). However, no numerical approach has simulated the formation process until the main-sequence stage of Population III stars, and hence it has not been possible to draw conclusions regarding the final masses and multiplicity of Population III stars.

Our result that Population III stars form in clusters is also supported by observations in the present-day Universe. At solar metallicity, we see that the binary frequency increases with stellar mass (Lada 2006; Duchêne & Kraus 2013; Janson et al. 2013) and that most massive stars form in binaries or higher-order systems systems (Zinnecker & Yorke 2007; Lee et al. 2020). Moreover, the close binary fraction seems to be anticorrelated with metallicity (Moe et al. 2019). Because we also expect metal-free stars to be massive, we can therefore expect that also they form in binaries, which requires more than one Population III star per minihalo.

4.1. Understanding the Decision Process

We use the abundance ratios [C/Mg] and [Ca/Fe] to illustrate the final classification in Figure 8, because these dimensions provide a high permutation importance. The general trend is that mono-enriched stars are located in the outskirts of the sampled region and multi-enriched stars are
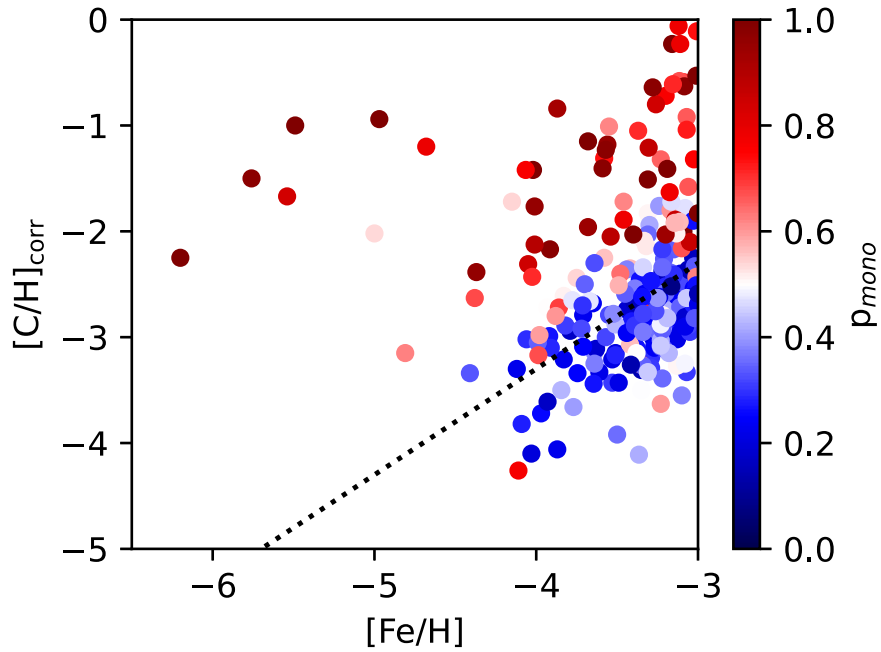
Figure 6. Carbon vs. iron abundance of EMP stars. The color bar shows the probability for mono-enrichment. The dashed line at [C/Fe] = 0.7 should guide the eye to highlight the range of CEMP stars. There is a trend that most CEMP stars are mono-enriched.

Table 1
Classification Results in the Carbon-enhanced Regime as a Function of the [C/Fe] Threshold

[C/Fe]$_{corr}$	N_{mono}	N_{multi}	CEMP ⇔ mono-enriched
>0.7	75	50	74.1%
>0.8	73	35	77.6%
>0.9	74	24	80.0%
>1.0	70	16	80.9%
>1.1	66	9	81.6%
>1.2	61	6	81.1%
>1.3	56	5	80.2%
>1.4	52	2	80.0%
>1.5	49	0	79.9%

Notes. The last column shows the fraction of all EMP stars for which mono-enrichment and carbon enhancement (based on the variable threshold in the first column) are synonymous. At [C/Fe] > 1.5, all CEMP stars are mono-enriched.

Table 2
List of the 10 Most Iron-poor Stars in Our Sample

Name	[Fe/H]	[C/Fe]$_{corr}$	N_{avail}	p_{mono} [%]
SMSS J160540.18−144323.1	−6.20	3.95	15	100
HE 1327−2326	−5.76	4.26	15	97^{+3}_{-8}
HE 0107−5240	−5.54	3.87	10	84 ± 16
SDSS J081554.26+472947.5	−5.49	4.49	15	100
SDSS J131326.89−001941.4	−5.00	2.98	21	54 ± 9
SDSS J092912.33+023817.0	−4.97	4.03	6	100
HE 0557−4840	−4.81	1.66	28	62 ± 14
SDSS J174259.67+253135.8	−4.79	3.62	3	100
SDSS J102915.14+172927.9	−4.71	⋯	10	69^{+31}_{-46}
HE 0233−0343	−4.68	3.48	10	79 ± 9

Notes. The columns show the name, iron abundance, carbon-to-iron ratio, number of available abundance ratios, and p_{mono} as result of our supervised classification. The full table is available online.
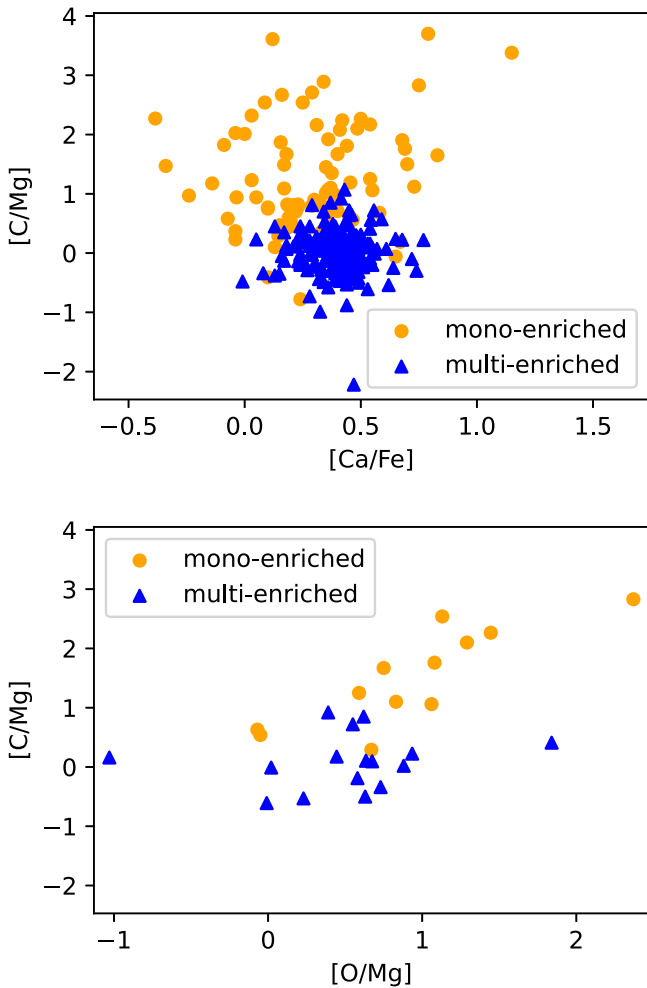
(This table is available in its entirety in machine-readable form.)
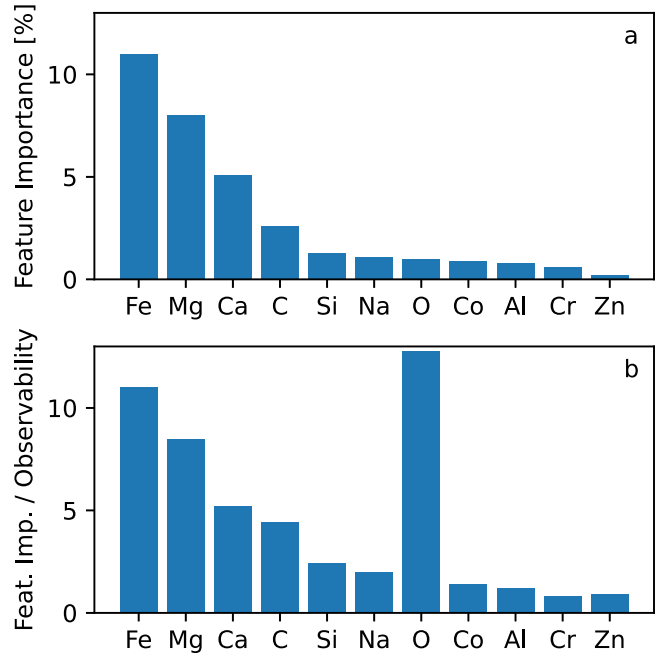
Figure 7. Panel (a): permutation feature importance for used elements. Panel (b): the feature importance divided by observability demonstrates that oxygen is informative despite its low observability.

more centrally concentrated. This trend is expected because multi-enrichment results in a weighted average of abundance ratios. Therefore, only mono-enriched stars can be found at extreme abundance ratios.

This figure also illustrates that two abundance ratios are not sufficient to confidently classify EMP stars. In these 2D projections, there are degenerate regions in which we find both mono- and multi-enriched EMP stars, e.g., around [Ca/Fe] ∼0.4 and [C/Mg] ∼0.5. However, these illustrations enable us to identify general trends, e.g., all EMP stars at [Ca/Fe] <0 are

Figure 9. Predicted fraction of multi-enrichment as a function of multi-enrichment in the validation data. All these results were obtained with our fiducial model that was trained on a data set that contains equal amounts of mono- and multi-enriched stars. The black dotted line shows the diagonal, to guide the eye.
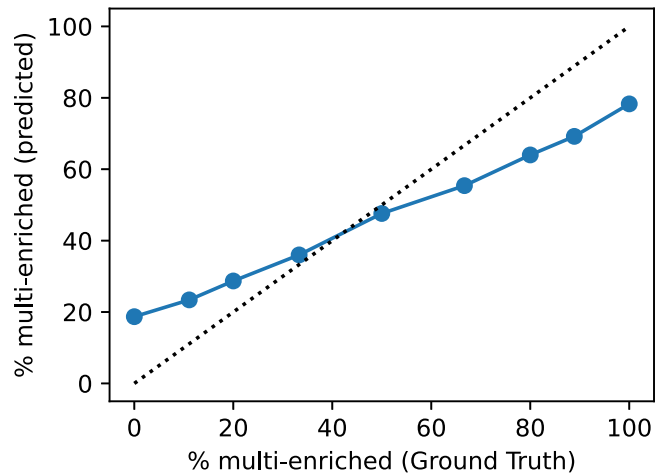

Figure 8. Illustration of the final classification in different 2D projections. The top panel uses the four most informative elements, and the right panel uses oxygen, which is more challenging to observe for EMP stars. These are 2D projections of a higher-dimensional ensemble learning classification. Therefore, the decision boundary does not appear as a smooth line in these representations.

mono-enriched and most EMP stars around [Ca/Fe] ~ 0.4 can be multi-enriched. Moreover, we recover the trend found earlier (Hartwig et al. 2018b) that EMP stars with [C/Mg] $\gtrsim 1$ are mostly mono-enriched.

4.2. Comparison to Previous Works

Kobayashi et al. (2011b) used the elemental abundance patterns of only a few DLAs to find that faint SNe seem to be the main enrichment source rather than pair-instability SNe. Welsh et al. (2019) have analyzed the chemical composition of the 11 most metal-poor DLA systems known at redshift $z < 5$. They use a stochastic model to infer the number of SNe that have contributed to the chemical enrichment of these systems. In contrast to EMP stars, DLAs provide a more direct way to study the chemical composition of gas in the early Universe (Zou et al. 2020). Welsh et al. (2019) find that these near-pristine gas clouds are enriched by $\lesssim 72$ SNe from massive stars. While the redshift of these DLAs ($2.6 \leqslant z \leqslant 5.0$) may be too low and their metallicity ($-3.5 \leqslant$ [Fe/H] $\leqslant -2.0$) may be too high to favor enrichment by only Population III SNe, their analysis shows that metal-poor gas at high redshift is enriched by multiple SNe. In a similar analysis, Welsh et al. (2021)

analyze the stochastic enrichment of metal-poor stars in the MW halo with metallicities of [Fe/H] $\leqslant -2.5$. This metallicity range might include enrichment from Pop II SNe (Ji et al. 2015; Ishigaki et al. 2021), to which their model is also sensitive. They find that these stars are enriched by 5^{+13}_{-3} SNe, which supports the notion that early star formation occurs in clusters. However, because of the metallicity range, their results do not allow a clear conclusion about the Population III multiplicity. In our study, we therefore focus on EMP stars with metallicities of [Fe/H] $\leqslant -3.0$, as we motivated above.

Compared to previous studies, our method and results are new in several regards. Previous attempts at classifying mono- and multi-enriched EMP stars used only few abundance ratios (Hartwig et al. 2018b, 2019; Welsh et al. 2021). In contrast, our new method is data-driven and maximizes the information gain from all observed abundances. Previous studies used a small, biased subset of metal-poor stars or included stars at [Fe/H] > -3 (Placco et al. 2018; Hansen et al. 2020; Rasmussen et al. 2020; Purandardas & Goswami 2021). Therefore, these studies are not representative of enrichment by Population III SNe. In summary, our method is the first data-driven analysis of a representative sample of EMP stars for which the enrichment was dominated by Population III SNe.

4.3. Prior Dependence

To develop our fiducial model, we have to assume an initial distribution of mono- and multi-enriched stars. Supervised classification algorithms are most robust when trained on balanced data sets, and thus we assume an equal distribution, i.e., 50% each, for our training set. This could affect p_{multi}, and we estimate the dependence as follows. First, we use our fiducial classification pipeline for stars that are at least one standard deviation away from the decision boundary, and apply it to validation data with different fractions of mono- and multi-enriched mock observations. We then check which fraction of validation samples was classified as multi-enriched as a function of the multi-enriched fraction in the validation data. The results can be seen in Figure 9.

Our fiducial model, which was trained under the assumption of 50% multi-enrichment, can predict a wide range of multi-enriched fractions from 20%–80%, once confronted with the
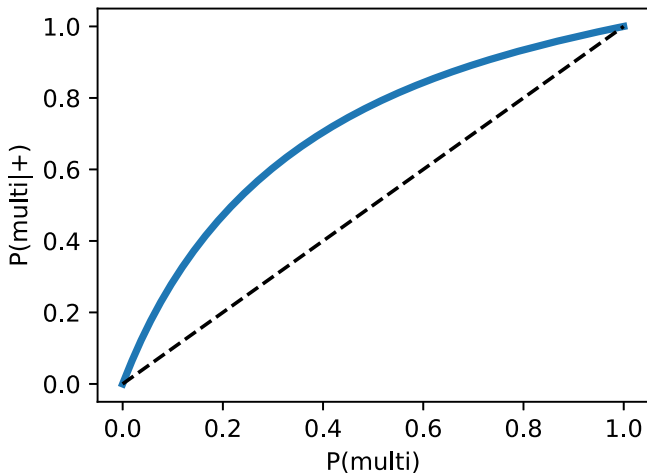
Figure 10. Probability that an EMP star that is classified as multi-enriched is actually multi-enriched, as a function of the assumed multi-enriched fraction. The dashed diagonal line is intended to guide the eye and emphasize the convex shape of the blue line.

data. Our model can also consistently reproduce the tendency, showing a linear trend in this figure. However, up to 20% of multi-enriched stars are misclassified. In an extreme case, even if all validation data come from one class, about 20% of samples can be misclassified. This result is related to the overall accuracy of our model of ∼80%. If all EMP stars in nature were mono-enriched, our model would still predict a multi-enriched fraction of 20%. Instead, we find a fraction of multi-enriched EMP stars of about 70%, which indicates that most EMP stars are multi-enriched.

Let us now calculate how reliably we can classify an EMP star to be multi-enriched as a function of the prior assumption. The unknown fraction of multi-enriched EMP stars is $P(\text{multi})$. We denote the probability that an EMP star classified as multi-enriched $(+)$ is actually multi-enriched as $P(\text{multi}|+)$. Moreover, the probability that we classify a multi-enriched star as multi-enriched is $P(+|\text{multi})$. The probability that we classify any star as multi-enriched is $P(+)$. Using Bayes' theorem, we can calculate the reliability of our multi-enriched predictions as

$$P(\text{multi}|+) = \frac{P(+|\text{multi})P(\text{multi})}{P(+)}$$
$$= \frac{P(+|\text{multi})P(\text{multi})}{P(+|\text{multi})P(\text{multi}) + P(+|\text{mono})P(\text{mono})}. \quad (3)$$

All components of the right side of this equation are known from the confusion matrix (see Figure 4), except for $P(\text{multi})$ and $P(\text{mono}) = 1 - P(\text{multi})$. Following Bottrell et al. (2022), we plot this function in Figure 10 as a function of the unknown $P(\text{multi})$. It is convex and always above the diagonal. If we were to relax our prior assumption of $P(\text{multi}) = 0.5$ and allow $0.16 \leqslant P(\text{multi}) \leqslant 0.84$ (central 64%), then the possible range of $P(\text{multi}|+)$ would be 0.41–0.95. This means that the probability for a star that is classified as multi-enriched to be actually multi-enriched is >41%, even if we allow variations in the prior assumption.

While our supervised machine-learning model depends on the prior assumption $P(\text{multi})$, this analysis can help to understand its quantitative influence and to correct the classification results in light of better prior assumptions in the

future. For example, if a better estimate for a prior of this classification problem is available in the future, one can use these calculation to update our results.

If we start with the fair prior assumption of $P(\text{multi}) = 0.5$ for the ensemble and assume that, for each individual star, the prior probability for multi-enrichment is flat between 0%–100%, then the marginalized distribution of $P(\text{multi}|+)$ corresponds to the posterior distribution for the probability that an EMP star is multi-enriched, given that we classify it as multi-enriched. This posterior is skewed toward multi-enrichment, which supports the conclusion that the majority of EMP stars are multi-enriched.

4.4. Variations of the Input Yields

We generate our training and test data based on theoretical Population III SN yields. There is no independent method to confirm whether the distribution of SN yields is realistic. To verify if our assumptions and training data are reasonable, we confirm that the final results of our study are sufficiently robust with respect to the exact choice of Population III SN yields.

The first test is unphysical, but provides intuition with regard to how the prediction might change with different distributions of input yields. As we saw previously, the distributions of abundances are not identical between the observed EMP stars and the mock data that we use for training. In an attempt to equalize these distributions, we perform two transformations to minimize this discrepancy. First, we shift the distributions of mock observations so that their mean value is identical to the mean value of EMP stars. Second, in addition to the shift, we also scale the mock observations so that they have the same standard deviation as the observed abundance ratios of EMP stars. As mentioned above, the shift and scaling are not physical and should only demonstrate how robust the model is with with respect to changes in the input data.

For the second test, we increase and decrease the number of Population III yields that we include to generate the mock data. In the fiducial model, we select theoretical yields that have $\chi^2 < 15.1$, where χ^2 quantifies how well these yields fit individual EMP stars (Ishigaki et al. 2018). We ran one more restrictive case with yields that fulfill $\chi^2 < 13$ and one less restrictive case with $\chi^2 < 17$. These two cases roughly halved and doubled the number of included Population III yields.

The results of these tests can be seen in Figure 11. We compare the predicted p_{mono} for all EMP stars between the fiducial model and the two variations. In most cases, the predictions for an EMP star from the fiducial model and from one of the variations are very similar, i.e., most points are close to the diagonal. Most importantly, there are no catastrophic failures with classifications far from the diagonal.

4.5. Enrichment with Neutron-capture Elements

The SVMs were trained to classify EMP stars based on how many Population III CCSNe have enriched the gas out of which they formed. The classification is based on elemental abundances up to Zn. However, there are other channels for chemical enrichment that we did not consider explicitly and that may produce elements heavier than Zn. As two representative abundance ratios, we analyze [Ba/Fe] and [Eu/Fe] as typical tracers for the s-process and r-process, respectively. While our model does not use these abundance ratios for the classification, this information is available for
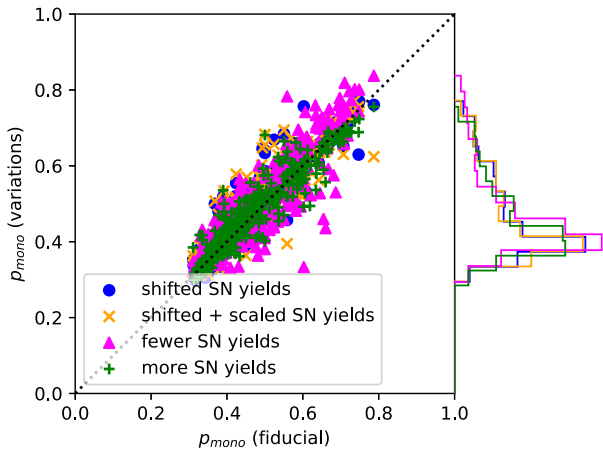
Figure 11. Predicted p_{mono} for all EMP stars. The vertical axis shows the value of four variations, and the horizontal axis shows the value in the fiducial model. Most points are close to the diagonal line, which illustrates that the final prediction for EMP stars is robust with respect to the exact selection of input yields. The vertical histograms on the right show the marginal distributions of the four variations.

Figure 12. p_{mono} as a function of [Ba/Fe] (cyan squares), [Eu/Fe] (orange circles), and [Ba/Eu] (magenta crosses).

some observed EMP stars, which allows us to compare our predictions to the abundances of neutron-capture elements.

If these alternative enrichment channels, which we do not explicitly account for, do not provide significant amounts of elements lighter than Zn, our approach is robust (but see Yong et al. (2021) for the co-production), because then we account for all enrichment channels (specifically Population III CCSNe) that should dominate the chemical composition of EMP stars. However, we need to be careful regarding binary mass transfer from a companion star (Suda et al. 2004; Arentsen et al. 2019). This mass transfer can add s-process elements (such as Ba) but also carbon to the EMP stars of interest. For an EMP star that was enriched in carbon via binary mass transfer, our training set, which includes carbon only from Population III CCSNe, would not not be representative anymore.

One could try to exclude EMP stars from the analysis if they are highly enriched in neutron-capture elements or are in close binaries. However, there is no physically motivated boundary to define such a cut, and it is certainly not one threshold value of, e.g., [Ba/Fe], below which only Population III SNe contribute to the chemical enrichment of EMP stars. Moreover, excluding s- or r-process-enriched EMP stars from this analysis could bias the result if their enrichment is dominated by Population III SNe. Therefore, we decided to keep EMP stars in our sample even if they are enriched in neutron-capture elements.

In Figure 12, we show p_{mono} as a function of [Ba/Fe], [Eu/Fe], and [Ba/Eu]. Most stars for which [Eu/Fe] is available (orange points) are multi-enriched, but there is one CEMP-r star ([Eu/Fe] $\gtrsim 1$; Beers & Christlieb 2005), which could be enriched by a single enrichment source such as magneto-rotational hypernovae (Yong et al. 2021).

A larger variation of p_{mono} is seen for the stars for which [Ba/Fe] is available (cyan squares); out of 12 s-process-enriched EMP stars with [Ba/Fe] > 1, only one is multi-enriched. For the stars with both Ba and Eu abundance measurements (magenta crosses in Figure 12), three stars show [Ba/Eu] > 0.5, which suggests the s-process origin of neutron-capture elements in these stars (e.g., Arlandini et al. 1999). All s-process-enriched EMP stars with [Ba/Fe] > 0.5 and $p_{mono} > 0.5$ are also carbon enhanced with [C/Fe] > 0.7, i.e.,
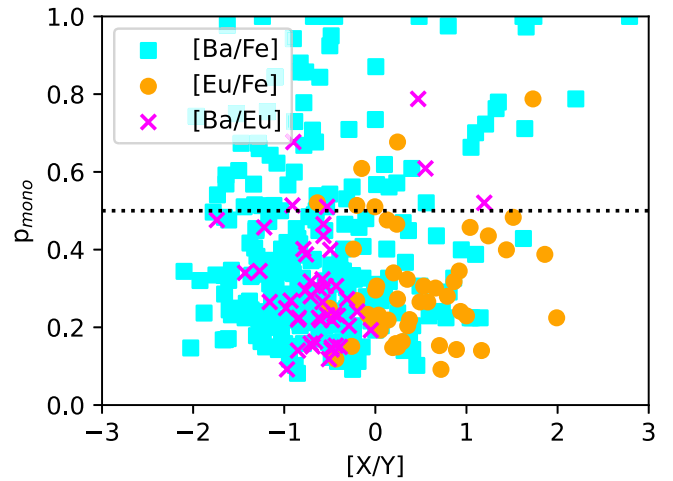
CEMP-s stars. Based on our results, these stars are likely to be enriched by one Population III CCSN. However, they might also have received heavy elements via binary mass transfer, which might attenuate their nomenclature as strictly mono-enriched.

Figure 12 also shows that the majority of stars with available [Ba/Eu] are compatible with the r-process origin of neutron-capture elements in these stars (e.g., Arlandini et al. 1999). The origin of r-process elements is debated, with various proposed enrichment channels (Metzger et al. 2008; Tanaka & Li 2014; Haynes & Kobayashi 2019; Ji et al. 2019; Brauer et al. 2021; Matsuno et al. 2021; Tarumi et al. 2021). In our data, we do not see any trend of p_{multi} with [Eu/Fe], which implies that our classification is agnostic with respect to the Eu abundance, which could mean that the dominant channel for the production of Eu does not produce significant amounts of elements between C and Zn.

Barium and carbon might have a similar origin, and many s-process-enriched stars are also C-enriched. If other enrichment channels (such as binary mass transfer) contribute to the enrichment, our SVM, which was only trained on CCSNe, might not be able to classify such unfamiliar abundances correctly. Such unfamiliar yields are instead classified as mono-enriched. Moreover, most Ba-enhanced stars are also CEMP stars, which makes them more likely to be mono-enriched (see Figure 6). Therefore, s-enriched stars might appear as mono-enriched.

In summary, EMP stars that are enriched in neutron-capture elements might bias our classification. However, of the 12 stars with [Ba/Fe] > 1.0, only one is classified as multi-enriched. So if we were to exclude these stars from the classification, the fraction of multi-enriched stars would increase, which strengthens our final conclusion that most EMP stars are multi-enriched.

5. Conclusions

We have used supervised machine learning trained on a set of nucleosynthesis yields from Ishigaki et al. (2018) to classify a representative set of 462 EMP stars from the literature according to the number of SNe that have enriched the gas out of which they formed. Under the prior assumption that mono- and multi-enrichment are equally likely, we find that

31.8% ± 2.3% of EMP stars are classified as mono-enriched. Our study is the first attempt to constrain the number of enriching SNe for EMP stars. Throughout the training, validation, and blind test process, we followed best practices for supervised machine learning and verified that the distribution of our mock observations are robust.

Our model develops a physical intuition without being explicitly trained to do so, e.g., discerning the dependence of multiplicity on metallicity (Figure 5) and carbon enhancement (Figure 6). Specifically, we find the the fraction of mono-enrichment increases from about 30% at [Fe/H] ∼ −3 to 100% at [Fe/H] ≲ −5. Moreover, our model offers a physical explanation of the origin of the carbon abundance bimodality in EMP stars (Bonifacio et al. 2015; Yoon et al. 2016); we find that most CEMP stars at [Fe/H] ⩽ −3 are mono-enriched, which is consistent with the theory that these form out of gas that was enriched by a faint SN, and all EMP stars with [C/Fe] > 1.5 are mono-enriched.

It may be possible to derive the exact number of SNe with a larger number of stars and elements (with error bars) in ongoing and future spectroscopic surveys, which is the most informative observational approach to unveil the episode of the first star formation in the early Universe. We have also identified Fe, Mg, Ca, C, and O as very informative elements for this classification. Future observations of such elements and smaller uncertainties of the measured abundances will improve our predictions in the future. By training our classifier on independent SN yields from other models (e.g., Heger & Woosley 2010; Limongi & Chieffi 2012), we can also test if our results are independent of the assumed yields.

Software: The source code was written in Python, is based on Scikit Learn, and is available online.[10] Furthermore, we used Python (Van Rossum & Drake 2009), numpy (Harris et al. 2020), scipy (Virtanen et al. 2020), matplotlib (Hunter 2007), and scikit-learn (Pedregosa et al. 2011).

Appendix A
Observational and Theoretical Uncertainties

Observational errors that we can obtain from observational papers often do not include the major error source, which is the NLTE effect (Andrievsky et al. 2007; Lind et al. 2011; Zhao et al. 2016; Mashonkina et al. 2017; Nordlander & Lind 2017), and the prediction from stellar atmosphere modeling is confirmed by a galactic chemical evolution (GCE) model for

Na and Al (also for K; not confirmed for Mn and Cu). The GCE model also predicts a strong NLTE effect for Cr I observations (Kobayashi et al. 2006; Sneden et al. 2016). In addition, the 3D effect should also be included, in particular for C, N, and O abundances (Amarsi et al. 2019), and possibly for Mg (Bergemann et al. 2017) and Mn (Bergemann et al. 2019). These effects depend on the masses, metallicities, and evolutionary stages of the observed stars. The corrections for different elements may be correlated, but we do not have a good understanding of the correlation of the corrections among various elements. Therefore, we assume a single matrix for all of our sample (mostly taken from the SAGA database), including all of these possible effects. The errors of abundance ratios relative to Fe are set based on previous results (Kobayashi et al. 2020), and we calculate the errors of the other combination of elements as the squared sum:

$$\sigma_{\rm obs}([X/Y]) = \sqrt{\sigma_{\rm obs}([X/Fe])^2 + \sigma_{\rm obs}([Y/Fe])^2}. \quad (A1)$$

For theoretical models, a few factors should be taken into account (see Section 3.6 of Kobayashi et al. (2020) for detailed discussion). The main error sources can be summarized as (1) nuclear reactions including neutrino processes, (2) stellar rotation and any mixing during hydrostatic burning, (3) mixing during SN explosion, and (4) fallback. Apart from the effect (1), these effects are not independent, and thus the errors should not be treated as the squared sum. As a result, there are multiple elements that are similarly affected mainly by one effect, and in that case, the errors of these elemental abundance ratios are small. Therefore, we provide a matrix of theoretical errors of all used combinations of elemental abundances in Table 3. These values are estimated by comparing our stellar evolution calculations with/without mixing, and our 1D and 2D nucleosynthesis calculations with different mixing and fallback (Umeda et al. 2000; Kobayashi et al. 2006; Tominaga et al. 2007; Tominaga 2009; Kobayashi et al. 2011a; Nomoto et al. 2013; Kobayashi et al. 2020).

The largest error can be seen for Na and Al, due to effect (2). Among α elements (O, Mg, Si, and Ca), the error of [(O, Ca)/Fe] is set to be smaller than that of [(Mg, Si)/Fe]; this is suggested by the GCE model, and the reason is likely to be effect (1). Iron peak elements (including Ti) can be affected by all of these effects, and the impact of each effect can be evaluated at each nucleosynthetic region (a layer in 1D) inside the SN ejecta. Namely, Cr, Mn are mainly produced in the incomplete Si-burning region, while Fe, Ni, Co, and Zn are produced in the complete Si-burning region (Kobayashi et al. 2006), which results in the smaller errors for Cr/Mn, Ni/Fe, and Co/Zn. The errors for Co and Zn are larger due to the effects (3 and 4, possibly 2 and 1 as well) of aspherical explosions, which are included in our 1D mixing-fallback model but may not be fully understood (C. Kobayashi & N. Tominaga 2023, in preparation). These dependencies are confirmed by comparison of the GCE model to the NLTE abundances of high-resolution observations (Kobayashi et al. 2020). We obtain the matrix in Table 3, taking into account all of these nonlinear effects in nuclear astrophysics, and this matrix can in principle be used for other nucleosynthesis yield sets.

Table 3
Observational (Top Right Triangle, [X(row)/Y(column)]) and Theoretical (Bottom Left Triangle, [X(column)/Y(row)]) Uncertainties

	C	O	Na	Mg	Al	Si	Ca	Cr	Mn	Fe	Co	Ni	Zn
C	⋯	±0.28	$^{+0.28}_{-0.56}$	$^{+0.36}_{-0.28}$	$^{+0.54}_{-0.28}$	±0.28	±0.28	$^{+0.45}_{-0.28}$	$^{+0.36}_{-0.28}$	±0.20	±0.28	±0.28	±0.28
O	±0.10	⋯	$^{+0.28}_{-0.54}$	$^{+0.36}_{-0.28}$	$^{+0.54}_{-0.28}$	±0.28	±0.28	$^{+0.45}_{-0.28}$	$^{+0.36}_{-0.28}$	±0.20	±0.28	±0.28	±0.28
Na	$^{+0.10}_{-0.50}$	$^{+0.10}_{-0.50}$	⋯	$^{+0.58}_{-0.28}$	$^{+0.71}_{-0.28}$	$^{+0.54}_{-0.28}$	$^{+0.54}_{-0.28}$	$^{+0.64}_{-0.28}$	$^{+0.58}_{-0.28}$	$^{+0.50}_{-0.20}$	$^{+0.54}_{-0.28}$	$^{+0.54}_{-0.28}$	$^{+0.54}_{-0.28}$
Mg	±0.10	±0.10	$^{+0.50}_{-0.10}$	⋯	$^{+0.54}_{-0.36}$	$^{+0.28}_{-0.36}$	$^{+0.28}_{-0.36}$	$^{+0.45}_{-0.36}$	±0.36	$^{+0.20}_{-0.30}$	$^{+0.28}_{-0.36}$	$^{+0.28}_{-0.36}$	$^{+0.28}_{-0.36}$
Al	$^{+0.10}_{-0.50}$	$^{+0.10}_{-0.50}$	±0.20	$^{+0.10}_{-0.50}$	⋯	$^{+0.28}_{-0.54}$	$^{+0.28}_{-0.54}$	$^{+0.45}_{-0.54}$	$^{+0.36}_{-0.54}$	$^{+0.20}_{-0.50}$	$^{+0.28}_{-0.54}$	$^{+0.28}_{-0.54}$	$^{+0.28}_{-0.54}$
Si	$^{+0.10}_{-0.20}$	$^{+0.10}_{-0.20}$	$^{+0.50}_{-0.20}$	$^{+0.10}_{-0.20}$	$^{+0.50}_{-0.20}$	⋯	±0.28	$^{+0.45}_{-0.28}$	$^{+0.36}_{-0.28}$	±0.20	±0.28	±0.28	±0.28
Ca	$^{+0.10}_{-0.20}$	$^{+0.10}_{-0.20}$	$^{+0.50}_{-0.20}$	$^{+0.10}_{-0.20}$	$^{+0.50}_{-0.20}$	±0.20	⋯	$^{+0.45}_{-0.28}$	$^{+0.36}_{-0.28}$	±0.20	±0.28	±0.28	±0.28
Cr	$^{+0.10}_{-0.15}$	$^{+0.10}_{-0.15}$	$^{+0.50}_{-0.15}$	$^{+0.10}_{-0.15}$	$^{+0.50}_{-0.15}$	$^{+0.20}_{-0.15}$	$^{+0.20}_{-0.15}$	⋯	$^{+0.36}_{-0.45}$	$^{+0.20}_{-0.40}$	$^{+0.28}_{-0.45}$	$^{+0.28}_{-0.45}$	$^{+0.28}_{-0.45}$
Mn	$^{+0.10}_{-0.20}$	$^{+0.10}_{-0.20}$	$^{+0.50}_{-0.20}$	$^{+0.10}_{-0.20}$	$^{+0.50}_{-0.20}$	±0.20	±0.20	±0.10	⋯	$^{+0.20}_{-0.30}$	$^{+0.28}_{-0.36}$	$^{+0.28}_{-0.36}$	$^{+0.28}_{-0.36}$
Fe	±0.10	±0.10	$^{+0.50}_{-0.10}$	±0.10	$^{+0.50}_{-0.10}$	$^{+0.20}_{-0.10}$	$^{+0.20}_{-0.10}$	$^{+0.15}_{-0.10}$	$^{+0.20}_{-0.10}$	⋯	±0.20	±0.20	±0.20
Co	$^{+0.10}_{-0.30}$	$^{+0.10}_{-0.30}$	$^{+0.50}_{-0.30}$	$^{+0.10}_{-0.30}$	$^{+0.50}_{-0.30}$	$^{+0.20}_{-0.30}$	$^{+0.20}_{-0.30}$	$^{+0.15}_{-0.30}$	$^{+0.20}_{-0.30}$	$^{+0.05}_{-0.25}$	⋯	±0.28	±0.28
Ni	$^{+0.10}_{-0.15}$	$^{+0.10}_{-0.15}$	$^{+0.50}_{-0.15}$	$^{+0.10}_{-0.15}$	$^{+0.50}_{-0.15}$	$^{+0.20}_{-0.15}$	$^{+0.20}_{-0.15}$	±0.15	$^{+0.20}_{-0.15}$	$^{+0.05}_{-0.10}$	$^{+0.25}_{-0.10}$	⋯	±0.28
Zn	$^{+0.10}_{-0.30}$	$^{+0.10}_{-0.30}$	$^{+0.50}_{-0.30}$	$^{+0.10}_{-0.30}$	$^{+0.50}_{-0.30}$	$^{+0.20}_{-0.30}$	$^{+0.20}_{-0.30}$	$^{+0.15}_{-0.30}$	$^{+0.20}_{-0.30}$	$^{+0.05}_{-0.25}$	±0.20	$^{+0.10}_{-0.25}$	⋯

Notes. For example, the theoretical uncertainty of [C/O] is 0.10 dex and the observational uncertainty of [C/O] is 0.28 dex. The observational uncertainties are based on iron. For some abundance ratios, we provide the asymmetric errors. These values should be subtracted for correcting observed/modeled values, i.e., $[Na/Fe]_{obs}-^{+0.5}_{-0.2}$ and $[Fe/Na]_{model}-^{+0.5}_{-0.1}$.

Appendix B
Carbon Corrections

While the carbon corrections do not affect the mean ensemble classification, they do slightly affect the distribution in the [C/H] versus [Fe/H] plane. To allow readers a transparent comparison, we also present the results without carbon correction in Figure 13. In this representation, the fiducial CEMP boundary at [C/Fe] = 0.7 (dotted line) seems to better discriminate between mono- and multi-enrichment.
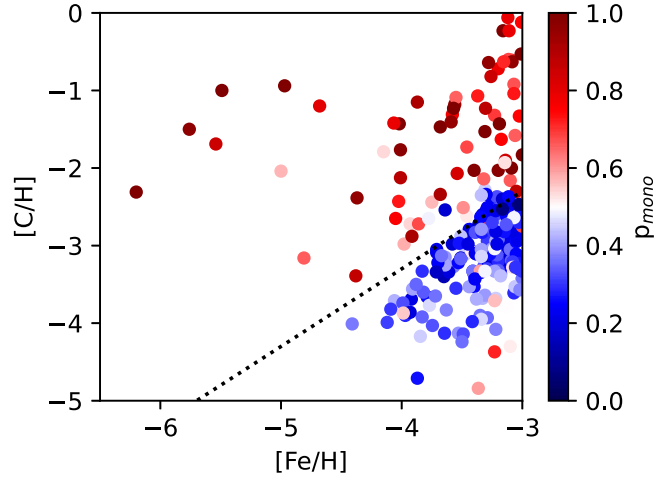

Figure 13. Same as Figure 6, but without carbon corrections. This means that we did not use the carbon corrections for the classification, nor for the vertical axis.

Appendix C
Decision Maps

To make our results available for the community, we made the source code public.[11] For a direct, qualitative comparison to our model, we provide additional classification maps in Figure 14. These maps show 2D projections of the final classification (same structure as Figure 8).
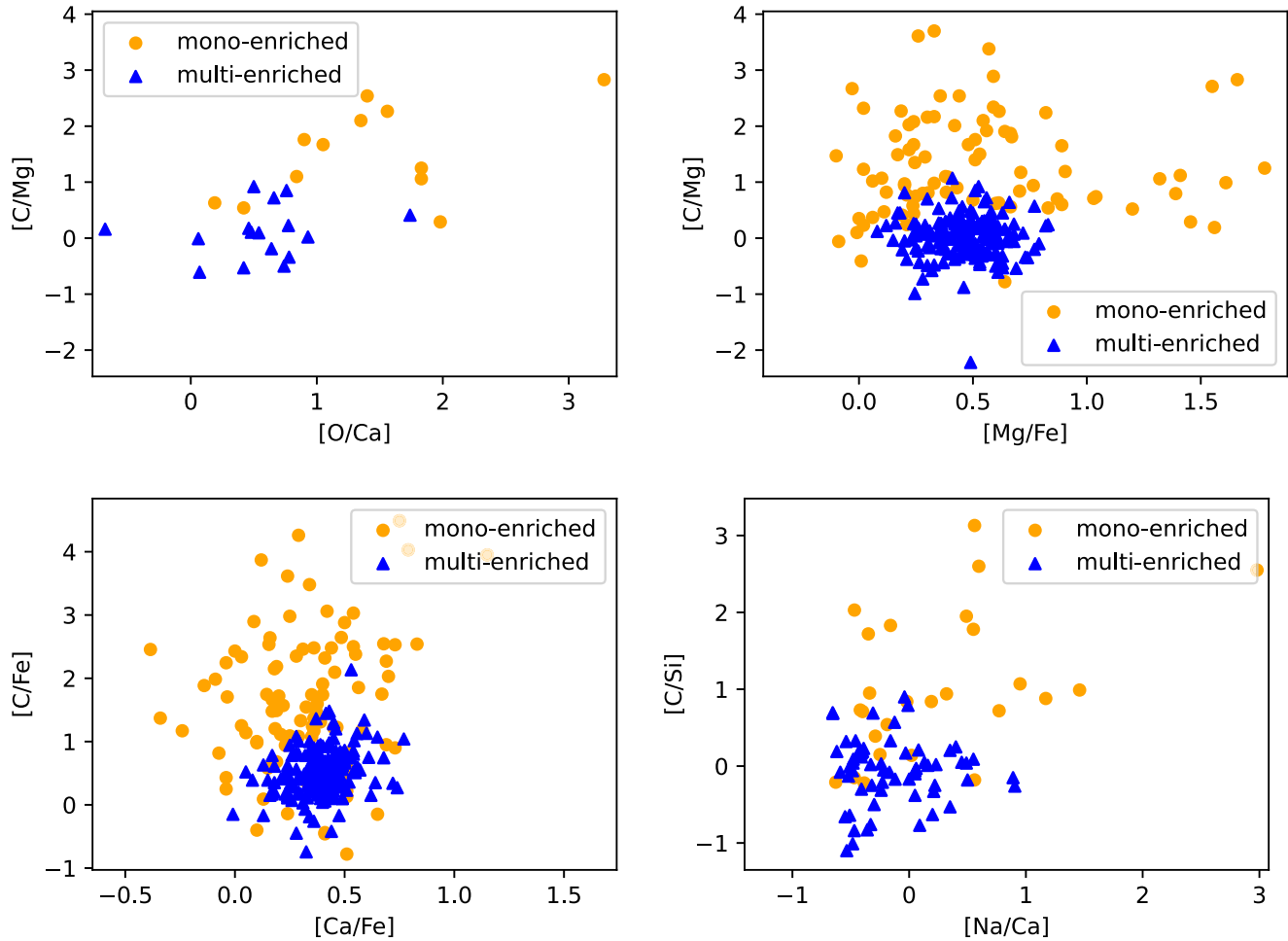
Figure 14. Two-dimensional classification maps for all EMP stars for which these abundances are available.

Appendix D
Mock Histograms

In addition to the three distributions in Figure 2, we provide the remaining histograms of abundances that are used in the classification in Figure 15.
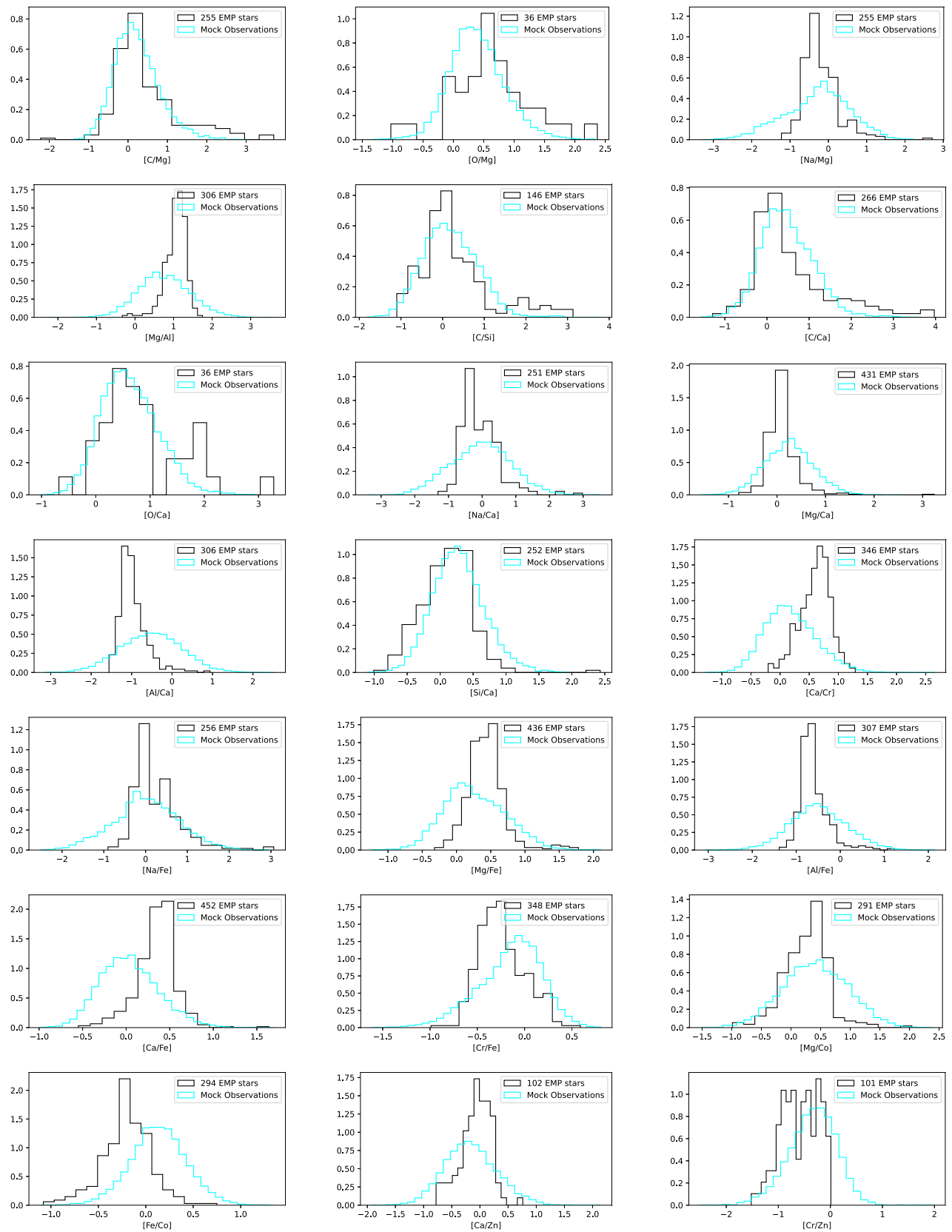

Figure 15. Same as Figure 2, but for the remaining 21 abundance ratios that were used.

ORCID iDs

Tilman Hartwig https://orcid.org/0000-0001-6742-8843
Miho N. Ishigaki https://orcid.org/0000-0003-4656-0241
Chiaki Kobayashi https://orcid.org/0000-0002-4343-0487
Nozomu Tominaga https://orcid.org/0000-0001-8537-3153
Ken'ichi Nomoto https://orcid.org/0000-0001-9553-0685

References

Acquaviva, V., Lovell, C., & Ishida, E. 2020, arXiv:2012.00066
Aguado, D. S., Allende Prieto, C., González Hernández, J. I., & Rebolo, R. 2018, ApJL, 854, L34
Aguado, D. S., Salvadori, S., Skúladóttir, A., et al. 2023, MNRAS, 520, 866,
Amarsi, A. M., Nissen, P. E., Asplund, M., Lind, K., & Barklem, P. S. 2019, A&A, 622, L4
Andrievsky, S. M., Spite, M., Korotin, S. A., et al. 2007, A&A, 464, 1081
Aoki, W., Beers, T. C., Christlieb, N., et al. 2007, ApJ, 655, 492
Arentsen, A., Placco, V. M., Lee, Y. S., et al. 2022, MNRAS, 515, 4082
Arentsen, A., Starkenburg, E., Shetrone, M. D., et al. 2019, A&A, 621, A108
Arlandini, C., Käppeler, F., Wisshak, K., et al. 1999, ApJ, 525, 886
Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, ARA&A, 47, 481
Audouze, J., & Silk, J. 1995, ApJL, 451, L49
Bazell, D., & Aha, D. W. 2001, ApJ, 548, 219
Beers, T. C., & Christlieb, N. 2005, ARA&A, 43, 531
Bergemann, M., Collet, R., Schönrich, R., et al. 2017, ApJ, 847, 16
Bergemann, M., Gallagher, A. J., Eitner, P., et al. 2019, A&A, 631, A80
Bonifacio, P., Caffau, E., Spite, M., et al. 2015, A&A, 579, A28
Bottrell, C., Hani, M. H., Teimoorinia, H., Patton, D. R., & Ellison, S. L. 2022, MNRAS, 511, 100
Brauer, K., Ji, A. P., Drout, M. R., & Frebel, A. 2021, ApJ, 915, 81
Breiman, L. 2001, Mach. Learn., 45, 5
Bromm, V., Coppi, P. S., & Larson, R. B. 2002, ApJ, 564, 23
Bromm, V., & Loeb, A. 2003, Natur, 425, 812
Chan, C., & Heger, A. 2017, in 14th Int. Symp. on Nuclei in the Cosmos (NIC2016), ed. S. Kubono et al. (Tokyo: Physical Society of Japan), doi:10.7566/JPSCP.14.020209
Chang, C.-C., & Lin, C.-J. 2011, ACM Trans. Intell. Syst. Technol., 2, 1
Chen, L.-H., Chen, K.-J., Tsai, S.-h., & Whalen, D. 2020, arXiv:2010.02212
Chiaki, G., & Tominaga, N. 2020, MNRAS, 498, 2676
Choplin, A. 2019, arXiv:1901.10708
Choplin, A., Tominaga, N., & Ishigaki, M. N. 2019, A&A, 632, A62
Clark, P. C., Glover, S. C. O., Smith, R. J., et al. 2011, Sci, 331, 1040
Cohen, J. G., Christlieb, N., Thompson, I., et al. 2013, ApJ, 778, 56
Cortes, C., & Vapnik, V. 1995, Mach. Learn., 20, 273
Dayal, P., & Ferrara, A. 2018, PhR, 780, 1
Doroshkevich, A. G., Zel'dovich, Y. B., & Novikov, I. D. 1967, SvA, 11, 233
Duchêne, G., & Kraus, A. 2013, ARA&A, 51, 269
Ezzeddine, R., Frebel, A., Roederer, I. U., et al. 2019, ApJ, 876, 97
Fraser, M., Casey, A. R., Gilmore, G., Heger, A., & Chan, C. 2017, MNRAS, 468, 418
Frebel, A., & Norris, J. E. 2015, ARA&A, 53, 631
Glover, S. 2005, SSRv, 117, 445
Greif, T. H. 2015, ComAC, 2, 3
Grisdale, K., Thatte, N., Devriendt, J., et al. 2021, MNRAS, 501, 5517
Haemmerlé, L., Mayer, L., Klessen, R. S., et al. 2020, SSRv, 216, 48
Hansen, C. J., Koch, A., Mashonkina, L., et al. 2020, A&A, 643, A49
Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Natur, 585, 357
Hartwig, T., Bromm, V., Klessen, R. S., & Glover, S. C. O. 2015, MNRAS, 447, 3892
Hartwig, T., Bromm, V., & Loeb, A. 2018a, MNRAS, 479, 2202
Hartwig, T., Ishigaki, M. N., Klessen, R. S., & Yoshida, N. 2019, MNRAS, 482, 1204
Hartwig, T., Magg, M., Chen, L.-H., et al. 2022, ApJ, 936, 45
Hartwig, T., Volonteri, M., Bromm, V., et al. 2016, MNRAS, 460, L74
Hartwig, T., Yoshida, N., Magg, M., et al. 2018b, MNRAS, 478, 1795
Haynes, C. J., & Kobayashi, C. 2019, MNRAS, 483, 5123
Heger, A., & Woosley, S. E. 2010, ApJ, 724, 341
Hirano, S., & Bromm, V. 2017, MNRAS, 470, 898
Hirschi, R. 2007, A&A, 461, 571
Hosokawa, T., Hirano, S., Kuiper, R., et al. 2016, ApJ, 824, 119
Huertas-Company, M., Rouan, D., Tasca, L., Soucail, G., & Le Fèvre, O. 2008, A&A, 478, 971
Hummel, J. A., Pawlik, A. H., Milosavljević, M., & Bromm, V. 2012, ApJ, 755, 72

Hunter, J. D. 2007, CSE, 9, 90
Ishigaki, M. N., Hartwig, T., Tarumi, Y., et al. 2021, MNRAS, 506, 5410
Ishigaki, M. N., Tominaga, N., Kobayashi, C., & Nomoto, K. 2018, ApJ, 857, 46
Ishiyama, T., Sudo, K., Yokoi, S., et al. 2016, ApJ, 826, 9
Jaacks, J., Finkelstein, S. L., & Bromm, V. 2019, MNRAS, 488, 2202
Janson, M., Lafrenière, D., Jayawardhana, R., et al. 2013, ApJ, 773, 170
Jeon, M., Bromm, V., Besla, G., Yoon, J., & Choi, Y. 2021, MNRAS, 502, 1
Ji, A. P., Drout, M. R., & Hansen, T. T. 2019, ApJ, 882, 40
Ji, A. P., Frebel, A., & Bromm, V. 2015, MNRAS, 454, 659
Kinugawa, T., Inayoshi, K., Hotokezaka, K., Nakauchi, D., & Nakamura, T. 2014, MNRAS, 442, 2963
Klessen, R. 2019, in Formation of the First Black Holes, ed. M. Latif & D. Schleicher (Singapore: World Scientific), 67
Kobayashi, C., Karakas, A. I., & Lugaro, M. 2020, ApJ, 900, 179
Kobayashi, C., Karakas, A. I., & Umeda, H. 2011a, MNRAS, 414, 3231
Kobayashi, C., Tominaga, N., & Nomoto, K. 2011b, ApJL, 730, L14
Kobayashi, C., Umeda, H., Nomoto, K., Tominaga, N., & Ohkubo, T. 2006, ApJ, 653, 1145
Kulkarni, M., Visbal, E., & Bryan, G. L. 2021, ApJ, 917, 40
Lada, C. J. 2006, ApJL, 640, L63
Lan, T., Hu, H., Jiang, C., Yang, G., & Zhao, Z. 2020, AdSpR, 65, 2052
Lee, Y.-N., Offner, S. S. R., Hennebelle, P., et al. 2020, SSRv, 216, 70
Liao, W.-T., & Turk, M. 2019, arXiv:1911.00610
Limongi, M., & Chieffi, A. 2012, ApJS, 199, 38
Lind, K., Asplund, M., Barklem, P. S., & Belyaev, A. K. 2011, A&A, 528, A103
Liu, B., & Bromm, V. 2020a, MNRAS, 497, 2839
Liu, B., & Bromm, V. 2020b, MNRAS, 495, 2475
Liu, B., Meynet, G., & Bromm, V. 2021, MNRAS, 501, 643
Magg, M., Hartwig, T., Agarwal, B., et al. 2018, MNRAS, 473, 5308
Magg, M., Hartwig, T., Glover, S. C. O., Klessen, R. S., & Whalen, D. J. 2016, MNRAS, 462, 3591
Magg, M., Klessen, R. S., Glover, S. C. O., & Li, H. 2019, MNRAS, 487, 486
Magg, M., Nordlander, T., Glover, S. C. O., et al. 2020, MNRAS, 498, 3703
Małek, K., Solarz, A., Pollo, A., et al. 2013, A&A, 557, A16
Marton, G., Tóth, L. V., Paladini, R., et al. 2016, MNRAS, 458, 3479
Mashonkina, L., Jablonka, P., Sitnova, T., Pakhomov, Y., & North, P. 2017, A&A, 608, A89
Matsuno, T., Hirai, Y., Tarumi, Y., et al. 2021, A&A, 650, A110
Metzger, B. D., Thompson, T. A., & Quataert, E. 2008, ApJ, 676, 1130
Meynet, G., & Maeder, A. 2002, A&A, 390, 561
Moe, M., Kratter, K. M., & Badenes, C. 2019, ApJ, 875, 61
Nomoto, K., Kobayashi, C., & Tominaga, N. 2013, ARA&A, 51, 457
Nordlander, T., Bessell, M. S., Da Costa, G. S., et al. 2019, MNRAS, 488, L109
Nordlander, T., & Lind, K. 2017, A&A, 607, A75
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, J. Mach. Learn. Res., 12, 2825
Peebles, P. J. E., & Dicke, R. H. 1968, ApJ, 154, 891
Placco, V. M., Beers, T. C., Santucci, R. M., et al. 2018, AJ, 155, 256
Placco, V. M., Frebel, A., Beers, T. C., et al. 2016, ApJ, 833, 21
Placco, V. M., Frebel, A., Beers, T. C., & Stancliffe, R. J. 2014, ApJ, 797, 21
Placco, V. M., Frebel, A., Lee, Y. S., et al. 2015, ApJ, 809, 136
Placco, V. M., Roederer, I. U., Lee, Y. S., et al. 2021, ApJL, 912, L32
Purandardas, M., & Goswami, A. 2021, ApJ, 912, 74
Rasmussen, K. C., Zepeda, J., Beers, T. C., et al. 2020, ApJ, 905, 20
Regős, E., Vinkó, J., & Ziegler, B. L. 2020, ApJ, 894, 94
Riaz, S., Hartwig, T., & Latif, M. A. 2022, ApJL, 937, L6
Ritter, J. S., Sluder, A., Safranek-Shrader, C., Milosavljević, M., & Bromm, V. 2015, MNRAS, 451, 1190
Roederer, I. U., Preston, G. W., Thompson, I. B., et al. 2014, AJ, 147, 136
Rossi, M., Salvadori, S., & Skúladóttir, Á 2021, MNRAS, 503, 6026
Rossi, S., Beers, T. C., Sneden, C., et al. 2005, AJ, 130, 2804
Rydberg, C.-E., Whalen, D. J., Maturi, M., et al. 2020, MNRAS, 491, 2447
Skúladóttir, Á, Salvadori, S., Amarsi, A. M., et al. 2021, ApJL, 915, L30
Salvadori, S., Bonifacio, P., Caffau, E., et al. 2019, MNRAS, 487, 4261
Schauer, A. T. P., Drory, N., & Bromm, V. 2020, ApJ, 904, 145
Schauer, A. T. P., Glover, S. C. O., Klessen, R. S., & Clark, P. 2021, MNRAS, 507, 1775
Sharda, P., Federrath, C., & Krumholz, M. R. 2020, MNRAS, 497, 336
Skinner, D., & Wise, J. H. 2020, MNRAS, 492, 4386
Sluder, A., Ritter, J. S., Safranek-Shrader, C., Milosavljević, M., & Bromm, V. 2016, MNRAS, 456, 1410
Sneden, C., Cowan, J. J., Kobayashi, C., et al. 2016, ApJ, 817, 53
Spite, M., Caffau, E., Bonifacio, P., et al. 2013, A&A, 552, A107

Stacy, A., & Bromm, V. 2013, MNRAS, 433, 1094
Suda, T., Aikawa, M., Machida, M. N., Fujimoto, M. Y., & Iben, I. J. 2004, ApJ, 611, 476
Suda, T., Katsuta, Y., Yamada, S., et al. 2008, PASJ, 60, 1159
Sugimura, K., Matsumoto, T., Hosokawa, T., Hirano, S., & Omukai, K. 2020, ApJL, 892, L14
Susa, H. 2019, ApJ, 877, 99
Tanaka, T. L., & Li, M. 2014, MNRAS, 439, 1092
Tarumi, Y., Hartwig, T., & Magg, M. 2020, ApJ, 897, 58
Tarumi, Y., Hotokezaka, K., & Beniamini, P. 2021, ApJL, 913, L30
Taylor, P., & Kobayashi, C. 2014, MNRAS, 442, 2751
Tominaga, N. 2009, ApJ, 690, 526
Tominaga, N., Iwamoto, N., & Nomoto, K. 2014, ApJ, 785, 98
Tominaga, N., Umeda, H., & Nomoto, K. 2007, ApJ, 660, 516
Umeda, H., & Nomoto, K. 2002, ApJ, 565, 385
Umeda, H., & Nomoto, K. 2003, Natur, 422, 871
Umeda, H., & Nomoto, K. 2005, ApJ, 619, 427
Umeda, H., Nomoto, K., & Nakamura, T. 2000, in The First Stars: Proc: MPA/ESO Workshop, ed. A. Weiss, T. G. Abel, & V. Hill (Berlin: Springer), 150

Van Rossum, G., & Drake, F. L. 2009, Python 3 Reference Manual (Scotts Valley, CA: CreateSpace)
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, NatMe, 17, 261
Wadadekar, Y. 2005, PASP, 117, 79
Washinoue, H., & Suzuki, T. K. 2021, MNRAS, 506, 1284
Welsh, L., Cooke, R., & Fumagalli, M. 2019, MNRAS, 487, 3363
Welsh, L., Cooke, R., & Fumagalli, M. 2021, MNRAS, 500, 5214
Wollenberg, K. M. J., Glover, S. C. O., Clark, P. C., & Klessen, R. S. 2020, MNRAS, 494, 1871
Woods, T. E., Agarwal, B., Bromm, V., et al. 2019, PASA, 36, e027
Yong, D., Kobayashi, C., Da Costa, G. S., et al. 2021, Natur, 595, 223
Yong, D., Norris, J. E., Bessell, M. S., et al. 2013, ApJ, 762, 27
Yoon, J., Beers, T. C., Placco, V. M., et al. 2016, ApJ, 833, 20
Yoshida, N., Abel, T., Hernquist, L., & Sugiyama, N. 2003, ApJ, 592, 645
Yoshii, Y., Sameshima, H., Tsujimoto, T., et al. 2022, AJ, 937, 61
Zepeda, J., Beers, T. C., Placco, V. M., et al. 2022, arXiv:2209.12224
Zhao, G., Mashonkina, L., Yan, H. L., et al. 2016, ApJ, 833, 225
Zinnecker, H., & Yorke, H. W. 2007, ARA&A, 45, 481
Zou, S., Petitjean, P., Noterdaeme, P., et al. 2020, ApJ, 901, 105