On the Optimal Choice of Nucleosynthetic Yields, Initial Mass Function, and Number of SNe Ia for Chemical Evolution Modeling

Oliver Philcox1,2, Jan Rybizki2, and Thales A. Gutcke2
1 Institute of Astronomy, Madingley Road, Cambridge, CB3 0HA, UK; ohep2@cam.ac.uk
2 Max Planck Institute for Astronomy, Königstuhl 17, D-69117 Heidelberg, Germany

Received 2017 December 14; revised 2018 April 26; accepted 2018 May 14; published 2018 June 29

Abstract
To fully harvest the rich library of stellar elemental abundance data available, we require reliable models that facilitate our interpretation of them. Galactic chemical evolution (GCE) models are one such set, a key part of which are the selection of chemical yields from different nucleosynthetic enrichment channels, predominantly asymptotic giant branch stars, Type Ia supernovae (SNe Ia), and core-collapse supernovae (CC-SNe). Here we present a scoring system for yield tables based on their ability to reproduce protosolar abundances within a simple parameterization of the GCE modeling software Chempy, which marginalizes over galactic parameters describing simple stellar populations (SSPs) and interstellar medium physics. Two statistical scoring methods are presented, based on Bayesian evidence and leave-one-out cross-validation, and are applied to five CC-SN tables, for (a) all mutually available elements and (b) a subset of the nine most abundant elements. We find that the yields of Prantzos et al. (P18, including stellar rotation) and Chieffi & Limongi (C04) best reproduce protosolar abundances for the two cases, respectively. The inferred best-fit SSP parameters for case (b) are $\alpha_{IMF} = -2.45^{+0.15}_{-0.11}$ for the initial mass function high-mass slope and $N_{b} = 1.29^{+0.45}_{-0.31} \times 10^{-3} M_{\odot}^{-1}$ for the SN Ia normalization, which are broadly consistent across tested yield tables. Additionally, we demonstrate how Chempy can be used to dramatically improve elemental abundance predictions of hydrodynamical simulations by plugging tailored best-fit SSP parameters into a Milky Way analog from Gutcke & Springel. Our code, including a comprehensive tutorial, is freely available and can additionally provide SSP enrichment tables for any combination of parameters and yield tables.

Key words: Galaxy: evolution – methods: statistical – stars: fundamental parameters – stars: luminosity function, mass function – Sun: abundances

1. Introduction

In order to understand the chemical enrichment of galaxies and interpret the growing wealth of observations contributing to the elemental abundance space, we require accurate models that are able to trace the chemical content of the interstellar medium (ISM) as a function of cosmic time, taking into account chemical, physical and dynamical processes (Matteucci 2003). At the heart of almost every model including chemical evolution are stellar enrichment processes that are integrated over the stellar initial mass function (IMF) and used to predict the elemental feedback of a simple stellar population (SSP).

For the most important (in terms of bulk metal contribution) stellar enrichment channels, namely, core-collapse supernovae (CC-SNe), Type Ia supernovae (SNe Ia), and asymptotic giant branch (AGB) stars, there exists a vast menagerie of published yield tables with substantial differences, as shown for CC-SN and AGB yields in Romano et al. (2010, hereafter RK10) and Mollá et al. (2015, hereafter MC15). Even for SN Ia yields the discrepancies between authors have recently seen an increase due to the availability of 3D simulations (Seitenzahl et al. 2013), an effect that is likely to be emulated with the other nucleosynthetic channels (Müller 2016). Similarly, the most important SSP parameters, i.e., the high-mass slope of the IMF and the incidence of SNe Ia (Côté et al. 2016a; Rybizki et al. 2017, hereafter RJR17), have a substantial spread of literature values (Côté et al. 2016a, Table 7).

Despite these crucial uncertainties, recent cosmological hydrodynamical simulations have produced increasingly realistic abundance distributions (e.g., Few et al. 2012; Naiman et al. 2018). In order to improve on these results and include more elements in the analysis, parameter studies to determine the best combination of yield set and SSP parameters need to be carried out. Hydrodynamical simulations are computationally far too expensive to use for such searches (Marinacci et al. 2014); therefore, it is common to rely on geometrically simplified galactic chemical evolution (GCE) models assuming a well-mixed ISM.

Previous studies have investigated the effect of varying the SSP parameters (e.g., Romano et al. 2005; Vincenzo et al. 2015; Côté et al. 2016a), the employed yields (e.g., Gibson 2002; RK10), or both (e.g., Gibson 1997; Kobayashi et al. 2011b; MC15; Andrews et al. 2017) on the predicted abundance distributions, but usually only one parameter at a time is left free, or if not, only a single set of yields is considered. From such analysis no firm conclusion can be drawn on the optimal parameter and/or yield set since restricting to a single set of yields (and likewise to fixed parameter values) has been shown to produce significantly biased results (Gibson 1998; RJR17).

In the literature, the two main investigations into yield set comparison are RK10 and MC15. Both test different AGB star and CC-SN yield tables against an extensive set of observables using a fiducial GCE model with radial zones. Whereas RK10 analyze many different elemental abundances and discuss in depth observational uncertainties and the effects of nucleosynthetic modeling assumptions, they lack a measure for goodness of fit. MC15 improves on this by employing a $\chi^2$ statistic on
their binned observational constraint. The drawback of this is that outliers are penalized disproportionately, preventing them from including elements with less well established yields. Even though RK10 (MC15) test two (six) IMFs, they hold all other GCE parameters fixed during their yield set comparison. This severely underestimates both degeneracies with the yield tables, and uncertainties in the best yield table determination arising from GCE parameters (Côté et al. 2016a).

In this paper we present a flexible technique capable of generating scores for nucleosynthetic yields only relying on a single well-established observational constraint, the solar elemental abundances (Asplund et al. 2009). We present two different statistical metrics and apply them to five CC-SN tables using (a) all available elements and (b) the subset of the nine most abundant ones. The flexible GCE model *Chempy* (RJR17) is used as a framework and sped up by means of neural networks such that we are able to properly marginalize over nuisance chemical evolution parameters for the first time in such an analysis. To account for uncertainties in the model and yield tables, we include a variable error model whose shape parameter is also marginalized out. Thus, this technique can determine the best combination of yields and, as a side effect, provide best-fit SSP parameters for any given set of yields and elements. To demonstrate this functionality, we predict best-fit values for the IMFs high-mass slope and SN Ia normalization for a Milky-Way-like magnetohydrodynamical simulation (Gutcke & Springel 2017) using the AREPO code (Springel 2010), under the constraint of a fixed yield set. The predicted parameters clearly improve the resulting abundance distribution of the simulation, as can be seen in Figure 7.

Our analysis can be extended to other observations predicted by chemical evolution models (e.g., additional stellar elemental abundances, age–metallicity relations, halo gas metallicity or metallicity distribution functions). It can also be applied to score the yield tables of other chemical enrichment channels and is intended to guide modelers of nucleosynthetic yields who wish to test their own tables coming from different parameter studies (e.g., rotation vs. static, rapid vs. delayed detonation, mass cut; Meynet & Maeder 2002; Fryer et al. 2012; Pignatari et al. 2016) against observational constraints of GCE models. An additional feature of the code is the ability to produce SSP yield tables (net yield as a function of time and metallicity) for any given set of yields and SSP parameters similar to the module presented in Ritter et al. (2017a).

### Table 1
Free *Chempy* and Model Error Parameters Used in This Study, with Their Prior Values and Gaussian Widths

| Parameter | Description | $\theta_{\text{prior}} \pm \sigma_{\text{prior}}$ | Limits | Approximated Prior Based on: |
|-----------|-------------|---------------------------------|--------|----------------------------------|
| $\alpha_{\text{IMF}}$ | High-mass slope of the CH03 Table 1 IMF | $-2.3 \pm 0.3$ | $[-4, -1]$ | CH03, Table 1 |
| $\log_{10}(N_{\text{SN}})$ | Number of SNe Ia exploding per $M_\odot$ over 15 Gyr | $-2.75 \pm 0.3$ | $[-5, -1]$ | Maoz & Mannucci (2012, Table 1) |
| $\log_{10}(\text{SFE})$ | Star formation efficiency governing the infall and ISM | $-0.3 \pm 0.3$ | $[-3, 2]$ | Bigiel et al. (2008) |
| $\log_{10}(\text{SFR}_{\text{peak}})$ | SFR peak in Gyr (scale of $\gamma$-distribution with $k = 2$) | $0.55 \pm 0.1$ | $[-1, 1]$ | van Dokkum et al. (2013, Figure 4(b)) |
| $x_{\text{out}}$ | Fraction of stellar feedback outflowing to the corona | $0.5 \pm 0.1$ | $[0, 1]$ | RJR17, Table 1 |
| $\log_{10}(\beta)$ | Beta distribution shape parameter (Figure 1) | $1.0 \pm 0.5$ | $[0, \infty)$ | Initial studies at fixed $\beta$ |

A Python implementation is freely available online (Philcox & Rybizki 2018). This includes a tutorial on implementing new yield tables, changing the set of elements and free chemical evolution parameters, training the neural network, running the Markov Chain Monte Carlo (MCMC) analysis, and computing both scores. The authors are happy to assist with this process upon request.

We begin with a discussion of the *Chempy* model and its study-specific modifications in Section 2. Section 3 describes the observational constraints along with the choice of implemented yield tables and elements, and then the two statistical scores are presented in Section 4. Results for the different CC-SN yield tables and element subsets are explored in Section 5. In this section we also demonstrate the practicality of our model by using our parameter predictions in a galaxy formation simulation. We conclude with a summary in Section 6.

### 2. Specifics of the GCE Model

#### 2.1. Implementation of *Chempy*

A complete discussion of the *Chempy* model can be found in RJR17. For convenience we recapitulate the main characteristics in the following section. *Chempy* is a simple one-zone GCE model that computes the evolution of a localized region of a well-mixed ISM throughout cosmic time. It is an open-box model that self-consistently incorporates both primordial infall and a self-enriching gas halo via a fixed outflow fraction and infall constrained by the gas required to sustain the star formation. Using input parameters describing SSP and ISM physics (parameterizing the IMF, SN Ia rate, star formation rate (SFR, modeled as a gamma function), star formation efficiency (SFE), and the gas outflow fraction) together with hyperparameters (e.g., nucleosynthetic yields), *Chempy* can produce predictions, primarily of the elemental abundances of the ISM over time. By comparing mock observations derived from these to real data, we may construct a likelihood function, which can be sampled within a Bayesian framework via the MCMC routine emcee (Foreman-Mackey et al. 2013), thereby constraining the input parameters. Here, the key functionality of *Chempy* is that of a “black box,” producing a set of model elemental abundances at the time of solar birth for input parameter vector $\theta$.

https://github.com/oliverphilcox/ChempyScoring.
In this study, the Chempy free parameters \( \theta \) (Table 1) have been updated (see RJR17, Table 1), changing the IMF parameterization to that of Chabrier (2003, Table 1; hereafter CH03) with variable high-mass slope \( \alpha_{\text{IMF}} \) and considering the SFR peak parameter in logarithmic form. In addition, the formerly used corona mass factor, \( \log_{10}(f_{\text{corona}}) \), is fixed to its prior value of 0.30, since it was found to be degenerate with the outflow fraction \( x_{\text{out}} \). Also, the SN Ia delay time distribution parameter \( \log_{10}(\tau_{\text{Ia}}) \) is set to \(-0.80\), since this cannot be meaningfully constrained by protosolar abundances alone. A Gaussian prior is assumed for all five Chempy parameters with mean and standard deviation as given in Table 1.

The simulation time step was increased by a factor of five to 0.5 Gyr, since this was found to give sufficiently precise results in much reduced computation time. Furthermore, before running emcee, an initial parameter-space optimization is now performed via the scipy package (Jones et al. 2001), allowing for faster convergence. To prevent unrealistically low SFRs, we impose the constraint that the SFR at the time of the Sun’s birth should be at least 5% of the median value, i.e., 5% of the SFR if the SFR was constant.\(^4\)

### 2.2. Error Parameterization

In previous versions of Chempy no account has been made for model errors that derive from a number of sources, including modeling assumptions, yield table inaccuracies, and missing stellar enrichment channels. The effect of this is that a few badly reproduced elements (e.g., K in RJR17, Figure 14) strongly dominate the likelihood and therefore bias the posterior parameter distribution.

Here we introduce a model error distribution, \( f(\sigma_m) \), which is marginalized over, treating each element equally for simplicity. An associated shape parameter is used to favor large or small model errors and is integrated over. This clearly does not fully account for the different errors in each element, but it is able to make first-order corrections and mitigate biases arising from poorly predicted elements.

Each element in the observational data set \( \mathcal{O}_i \) (of size \( n\mathcal{O} \)) has abundance \( \mathcal{O}_{i,i} \) and error \( \sigma_{\text{obs},i} \), which can be compared to the predicted value, \( d_{i,i} \), and model error, \( \sigma_m \), using

\[
\mathcal{L}(\sigma_m) = \prod_{i=1}^{n\mathcal{O}} \frac{1}{(2\pi\sigma_i^2)^{0.5}} \exp \left( -\frac{(\mathcal{O}_{i,i} - d_{i,i})^2}{2\sigma_i^2} \right) \tag{1}
\]

(see RJR17, Equation (9)) for combined error \( \sigma_i^2 = \sigma_{\text{obs},i}^2 + \sigma_m^2 \).

A beta function was chosen for the error parameterization with probability density function (pdf)

\[
f(\sigma_m; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \sigma_m^{\alpha-1}(1 - \sigma_m)^{\beta-1}, \tag{2}
\]

for \( 0 < \sigma_m < 1 \), with shape parameters \( \alpha \) (set to unity) and \( \beta \), where \( \Gamma(z) \) is the gamma function.

By varying \( \beta \in [1, \infty) \), we can change the error tolerance, including the limiting cases of uniform and zero model error, as shown in Figure 1. The likelihood at fixed \( \beta \), \( \mathcal{L}(\beta) \), is then computed using

\[
\mathcal{L}(\beta) = \int_0^1 f(\sigma_m; 1, \beta) \mathcal{L}(\sigma_m) d\sigma_m, \tag{3}
\]

approximated via numerical integration over a grid of \( \sigma_m \) evenly spaced in [0, 1]. We allow the shape parameter to vary freely inside the aforementioned limits, with larger values of \( \beta \) indicating a preference of the model for smaller errors.

### 2.3. Neural Networks

Creation of a yield table score requires many computations of the posterior. In order to speed up the calculation, a neural network is implemented to predict the abundance output for a given set of input parameters, \( \theta \), practically acting as a fast interpolator. We make use of the PyTorch software,\(^5\) creating a simple network with five inputs (\( \theta \), excluding \( \beta \)) and \( n_{\theta} \) outputs (\( d_i \)).

The training set for the network is a grid of \( 10^5 \) points, with 10 values of each parameter drawn from an inverse Gaussian distribution spanning \( 2.8 \sigma_{\text{prior}} \) in parameter space (such that all points are inside the prior limits). Network hyperparameters were optimized using an independent data set to reduce bias, giving a 30-neuron network with one hidden tanh layer (LeCun et al. 1998; Karpathy 2017). An Adam optimizer (Kingma & Ba 2014) with learning rate 0.007 is used to train the network over 5000 epochs with the L1 loss function.

The network errors (discussed in Appendix A) are negligible compared to those of the observational data set, and the overall effect of the implementation is to reduce computation time from 600 to 7 ms for each posterior evaluation, although we note that the network must be retrained for each new yield set.

### 3. Data and Yield Table Selection

#### 3.1. Stellar Abundance Data

To assess the likelihood of model parameters, we must compare Chempy’s predictions with observed stellar abundances. Modern spectral libraries (e.g., APOGEE; Majewski et al. 2017)

---

\(^4\) This is merely to ensure that the SFR parameter does not peak at very early times, which would result in practically no SFR at the solar birth. The code would otherwise still provide ISM abundances in this case, which will generally be higher and therefore biased.

\(^5\) pytorch.org.
provide many such examples. Clearly, the model complexity increases with the number of data points, as ISM parameters must be allowed to vary between stars (see RIR17, Figure 16). This vastly inflates the number of free parameters, and we must also consider errors in the age of each star, adding further computational expense. Here we adopt the simplest possible prescription, using only protosolar abundances to assess how this can constrain the choice of yield table. 

[X/Fe] and [Fe/H] solar abundances are used from Asplund et al. (2009), as independently verified by analysis of meteorites and comets (Lawler et al. 1989; Lodders et al. 2009). We apply the heavy-element corrections of Turcotte & Wimmer-Schweingruber (2002) to convert these to protosolar abundances, adding 0.01 (0.04) dex to the [He/Fe] ([Fe/H]) abundance and increasing the uncertainties by 0.01 dex for each data point. These are compared to Chempy’s predictions at time \( t_p \sim \tau_\odot \) for current time \( t_p \) and solar age \( \tau_\odot = 4.5 \) Gyr (Dziembowski et al. 1999), assuming negligible age error. No further observational constraint is used in our analysis.

### 3.2. Yield Tables

Nucleosynthetic yields are implemented for three main processes: AGB feedback, CC-SNe, and SNe Ia, using a mass range of 0.5–8 (8–40) \( M_\odot \) for AGB (CC-SN) events (Smartt 2009). For higher stellar masses up to the IMF limit of 100 \( M_\odot \), a new process was implemented to simulate the effects of “failed SNe” (Adams et al. 2017), returning 75% of the initial stellar composition to the ISM via winds without enrichment, leaving the remainder to supposedly form a black hole. This increases the mass in remnants by \( \sim 6\% \) and decreases the alpha enhancement by \( \sim 35\% \). It also reduces extrapolation error, since the CC-SN yield tables we use do not calculate explosions for stars above 40 \( M_\odot \) (see Table 2).

To provide a diagnostic test of the statistical analysis, we compare five CC-SN yield tables, listed in Table 2, along with the default Karakas (2010) (AGB) and Seitenzahl et al. (2013) (SN Ia) yields from RIR17, Table 2. Here logarithmic interpolation is used to compute the yields for any given metallicity. Some element of bias will be introduced by imperfections in the yield sets that are held constant in the study, although these are fairly well understood compared to CC-SN yields, and their nucleosynthetic contribution is much smaller.

Only net yields are used here, such that the tables provide newly synthesized material and the remainder comes from the initial SSP composition. This requires that the C04 and W17 yields, originally presented in gross format, are modified to convert them into net yields utilizing the relevant initial abundances: Anders & Grevesse (1989) solar abundances and West & Heger (2013) abundances for the two models, respectively. In addition, we note that the N13 yields are based on those of Kobayashi et al. (2006) (excluding their hypernova yields), and the currently unpublished W17 yields are those used in Cô et al. (2016b). The R17 yields come in two detonation flavors: “rapid” and “delayed” (Fryer et al. 2012), of which we use the latter. The P18 yields are based on those of A. Chiefi & M. Limongi (2018, in preparation) and are averaged over three rotation speeds using a metallicity-dependent rotation distribution (see Prantzos et al. 2018, Figure 4). None of the other tested CC-SN yields include rotation in their calculation, and all but C04 include stellar winds.

### 3.3. Choice of Elements

The statistical methods below can be used to calculate scores for any combination of elements. The implemented CC-SN tables provide yields for all elements up to Ge, but we exclude Li, Be, and B, as these abundances are significantly altered by unmodeled nucleosynthetic pathways. This gives a total of 29 tracked elements resulting in 28 predictions (\( n_{\text{el}} = 28 \)), as H is only implicitly included via normalization in the abundances. No statistical weighting is applied to the other elements, though the Fe prediction affects each data point via the chosen [X/Fe] normalization. To demonstrate how the resulting scores depend on the chosen elements, we also apply our analysis to the main set of elements used in hydrodynamical simulations (e.g., Springel 2010), which are the nine most abundant by mass fraction (H, He, C, N, O, Ne, Mg, Si, Fe), giving \( n_{\text{el}} = 8 \). Adaptation to other sets of elements is straightforward and is described in the online tutorial.

### 4. Objective Scoring Models

Here we propose two complimentary methods to calculate an objective yield table score, discussing their implementation and limitations.

#### 4.1. Bayes Factor Approach

To compare yield sets \( \chi \) given observations \( \mathcal{O} \), we may simply calculate the ratio of their Bayesian evidences (e.g., Kass & Raftery 1995; Baier-Jones 2012):

\[
K_B = \frac{P(O | \chi_1)}{P(O | \chi_2)}.
\]

This is the classical Bayes factor, and if \( K_B \gg 1 \), we may conclude that \( \chi_1 \) is the better predictor of \( \mathcal{O} \). The evidence is found via integration:

\[
P(O | \chi_i) = \int P(O | \theta, \beta, \chi_i) d\theta d\beta
\]

\[
= \int L(O | \theta, \beta, \chi_i) P(\theta, \beta) d\theta d\beta,
\]

where \( \chi_i \) represent the choice of yield table.

---

6 For a comparison between models with a CC-SN mass limit of 40 \( M_\odot \) (including failed SNe) and 100 \( M_\odot \) (with no failed SNe), see Tables 4(c) and (d).

7 Retraining a test yield set for linear interpolation instead shows that this does not affect the analysis; comparative scores are consistent within their stated errors, and all posterior predictions are consistent within 0.2\%.

---

Table 2: CC-SN Yield Tables Compared in This Study and Their Mass and Metallicity Ranges

| Abbr. | Yield Table | Masses | Metallicities |
|-------|-------------|--------|---------------|
| C04   | Chiefi & Limongi (2004) | [13, 35] | [0, 0.02] |
| N13   | Nomoto et al. (2013) | [13, 40] | [0.001, 0.05] |
| W17   | C. West & A. Heger (2018, in preparation) | [13, 30] | [0, 0.3] |
| R17   | Ritter et al. (2017b) | [12, 25] | [0.0001, 0.02] |
| P18   | Prantzos et al. (2018) | [13, 120] | [0.0000134, 0.0134] |
for posterior \( P(O|\chi) \), likelihood \( L(O|\theta, \beta, \chi) \), and combined prior \( P(\theta, \beta) \) (Table 1), with the domain taken as the entire permitted parameter range. We marginalize over \( \beta \) to take into account all error model shapes, assuming a broad Gaussian prior with mean and width estimated from initial studies at various fixed values of \( \beta \).

The posterior is tightly peaked in parameter space about the median parameter values (shown in Figure 2 for the two limiting values of \( \beta \)) and can thus be fairly well fit using a multivariate Gaussian (with covariances drawn from the MCMC pdf output). The parameter-space integration, i.e., calculation of the Bayesian evidence, can then be performed efficiently via a Monte Carlo importance sampling routine (Press et al. 2002) implemented with the \texttt{scikit-monaco} Python package.\(^9\) Accuracy of the method has been checked using a more rudimentary (and slower) Monte Carlo method via the \texttt{mcint} package.\(^{10}\) A numerical precision of 0.5% was obtained using 50,000 posterior samples, which takes around 5 minutes on a modern eight-core machine.

As noted in Appendix A, the neural network has increasing error toward the edges of parameter space, itself truncated by the finite prior domains (Table 1). Computation of Equation (5) for increasingly broad parameter ranges shows that there are negligible contributions from regions outside \( 3\sigma_{\text{prior}} \), validating our approximations.

We can thus obtain the Bayesian evidence, henceforth denoted by \( S_B \), for any combination of yield sets, although this method retains significant dependence on the prior, \( P \), which is set using reasonable broad distributions (see Table 1).

### 4.2. Cross-validation Approach

An alternative scoring method for yield tables is considered employing the technique of leave-one-out cross-validation (LOO-CV; e.g., Vehtari & Lampinen 2002; Bailer-Jones 2012). Using \textit{Chempy} and \textit{emcee}, we can predict the posterior pdf for the six free parameters, using only \( n_{\text{el}} - 1 \) of the \( n_{\text{el}} \) available data points. Each set of posterior parameters in the pdf is then fed into \textit{Chempy} to produce a model elemental abundance pdf for the excluded element, which can be well approximated by a Gaussian.

Using the predicted Gaussian mean, \( \mu_i \), and standard deviation, \( \sigma_{\text{pred},i} \), we may construct the \( i \)th element likelihood:

\[
L_{i}^{\text{CV}} = \left(2\pi\sigma_i^2\right)^{-0.5} \exp \left(-\frac{(\mu_i - O_{i})^2}{2\sigma_i^2}\right),
\]

which compares the optimal element prediction with the observational abundances using the combined error \( \sigma_i^2 = \sigma_{\text{pred},i}^2 + \sigma_{\text{obs},i}^2 \). This is iterated over all elements, giving an overall score

\[
S_C^{\text{V}} = \left[\prod_{i=1}^{n_{\text{el}}} L_{i}^{\text{CV}}(\chi)\right]^{1/n_{\text{el}}},
\]

taking the geometric mean for normalization, and allowing comparison between scores with different \( n_{\text{el}} \). Here the shape parameter \( \beta \) affects the results indirectly through broadening of the MCMC pdf. As before, yield tables are compared using the relative score \( K_{\text{CV}} = S_C(\chi)/S_C(\chi_i) \), which can be computed in around half an hour on a modern eight-core machine for \( n_{\text{el}} = 28 \).

### 5. Results

Here we discuss the results of applying MCMC and the above metrics to the CC-SN tables defined in Section 3.2, with both sets of elements (Section 3.3). In addition, Section 5.3 shows how the code can be used to constrain the IMF slope and SN Ia normalization parameter for galaxy formation simulations, using AREPO as an example.

---

\(^9\) Bugnion 2013, \url{http://scikit-monso.com/atlasthatscikit.org/}

\(^{10}\) Snowsill 2011, \url{http://pypi.python.org/pypi/mcint/}
5.1. Using All Available Elements

5.1.1. Posterior Parameter Distributions

Table 3(a) shows the results of running an MCMC analysis on the models with each choice of CC-SN yields for all available elements. \( \log_{10}(\beta) \) is well constrained, with a peak value between 0.5 and 0.8, significantly different from both the prior and zero (the latter would indicate a flat error model). Larger \( \beta \) favors smaller model errors, so from this alone one could surmise that the P18 and C04 yields best represent the data, although this does not take into account the full complexities of parameter space incorporated in the Bayesian score.

Meaningful constraints can only be placed on the SSP parameters. The ISM parameters, treated as nuisance parameters in this study, closely reproduce the priors as can be seen in Table 4 in Appendix B. This is unlike the constraints found in RJR17 and results from the newly introduced model error that “flattens” the posterior pdf (see Figure 2, where the lower panel shows the behavior in the absence of an error model). The median of \( \alpha_{\text{IMF}} \) is mostly consistent with the fiducial value of \(-2.3\), except more tightly constrained. Differences to the results of RJR17, Table 4 arise from the addition of the “failed SN” routine, which gives less CC-SN enrichment, and also from the different IMF parameterization used in this study. The W17 \( \alpha_{\text{IMF}} \) median posterior is anomalously high, most likely due to the large fraction of black holes produced by their SN model.

The SN Ia normalization, \( N_{\text{Ia}} \), is consistent with \( 10^{-3}\, M_{\odot}^{-1} \) and a little lower than our prior, with C04 favoring a lower SN Ia rate than N13, matching the behavior seen in RJR17. This underlines the possibility for parameters to be fine-tuned (within the ranges compatible with other observational constraints) for a chosen yield set in order to best match protosolar abundances, as will be demonstrated in Section 5.3.

5.1.2. Scoring Metrics

Figure 2 shows a slice of the log posterior function in the \( \alpha_{\text{IMF}} - \log_{10}(N_{\text{Ia}}) \) plane for two values of \( \beta \) corresponding to flat (top panel) and approximately zero (bottom panel) error models. This uses the \( n_{\text{el}} = 28 \) data set and N13 yields for illustration. Both the scale bar and confidence intervals demonstrate the strongly peaked nature of the function, and it is clear that the vast majority of the integrated posterior comes from the central regions, as previously assumed. The posterior obtained for \( \beta = 1 \) is many orders of magnitude greater than that for \( \beta = 1000 \), since we only allow for errors in the former case, giving a broader posterior. In addition, the confidence intervals shrink considerably as \( \beta \) increases, since the optimal parameters become more tightly constrained.

Table 3(a) gives the results of the Bayes and LOO-CV scores of each CC-SN yield table. The uncertainty estimates on \( S_B \) coming from the errors in the Monte Carlo integration procedure are negligible and are thus not included. Other sources of uncertainty could, for example, be arising from Chempy modeling assumptions or the neural network implementation. For \( S_{CV} \), the uncertainty is given by the results’ standard deviation from running the analysis 10 times (the MCMC results will change and affect the posterior predictions).
For this set of elements, the Bayes scores ($S_B$) show clearly that P18, W17, and C04 yields better reproduce protosolar abundances, with an overall preference for P18 and a relative Bayes factor of greater than $10^3$ for either compared to N13 or R17 yields. The LOO-CV prescription ($S_{CV}$) is in accordance with this, having comparable scores for P18, W17, and C04, albeit with weaker improvements over N13 and R17, although still statistically significant. Direct comparison of the two scores is difficult owing to the different metrics applied, but it is clear that the preferences are stronger for the Bayes scores. We may thus conclude that the P18, and W17 models are the optimal sets in this case, with the caveat that the W17 yields predict an unreasonable top-heavy IMF (due to a large amount of black hole production). We thus take P18 yields to be the best predictor of protosolar abundances using this set of elements.\footnote{We attribute this predominantly to the inclusion of stellar rotation in the P18 models, since rerunning the analysis using P18 yields without rotation gives much lower Bayes and LOO-CV scores ($-1.52$ and $-1.09^{+0.02}_{-0.01}$, respectively).}

5.1.3. Posterior Element Predictions

A benefit of the LOO-CV approach is that we can visualize the predictions of each element, as derived from the MCMC parameter pdf with that element excluded. Figure 3 shows this for all available elements, comparing the five yield sets and protosolar observations. The median values and error bars are taken from the Gaussian fitting of the pdf for each element, as described in Section 4.2. For comparison we also plot the abundances from RK10, Figure 23 normalized to Asplund et al. (2009) solar abundances.

From the plot it is clear that there are still significant inadequacies in current yield tables, as many predicted abundances are far from their observational counterparts (missing nucleosynthetic channels could be an alternative explanation, as well as oversimplifications in the GCE model). Some elements are particularly poorly predicted by certain yield sets, for example, K and Cl by N13 and C04 (and to a lesser extent P18), Sc by N13 and W17, and Al and Cr by R17. This indicates that their physics is not well understood, as noted by Argast et al. (2002), Kobayashi et al. (2006, 2011a), and RJR17, among others, although Ritter et al. (2018) overcome the deficiency of the odd-Z elements by using interacting convective O and C shells in massive stars.

The RK10 abundances are shown for their optimal “Model 15,” using their fixed GCE model with yields from Karakas (2010) for AGB stars, Iwamoto et al. (1999) for SNe Ia, and Kobayashi et al. (2006) for CC-SNe (similar to our N13 yields, although they apply a hypernova fraction of unity for stars with $m > 20 M_\odot$ and extrapolate to $M_{CC-SN,max} = 100 M_\odot$), supplemented by He, C, N, and O from pre-SN yields of Meynet & Maeder (2002), Hirschi et al. (2005), Hirschi (2007), and Ekström et al. (2008). Their solar abundance pattern is very similar to our N13 results except for C, N, and O, which is as expected since the yields are almost identical except for the CNO elements where pre-SN yields are included, which consider rotation and produce significant N in metal-poor stars (see RK10, Figure 3).

The approach of RK10 is completely different from that presented here, as they change the yields and attempt to fit as many observational constraints as possible while keeping the other GCE parameters fixed (which are chosen and validated using other data). We highlight their work here to show that we both find the major limitation in reproducing the solar elemental abundance pattern to be the yield sets used and possibly missing nucleosynthetic enrichment channels.

Similarly, a comparison of N13 and R17 is depicted in Figure 2 of Côté et al. (2017a) for the abundances of O, Mg, Si, Ti, Mn, and Ni with respect to Fe (using $M_{CC-SN,max} = 30 M_\odot$). The two yield sets show the same relative difference in abundance pattern as is visible in our Figure 3, with R17 yields...
producing more Ni and Ti and less O and Mg compared to N13 in both studies. For all yield sets we see an underproduction of Sc and Ti. One possible reason might be that neutron star winds (not included in the yield tables) contribute significantly to those elements (Pruet et al. 2005). Young et al. (2006) argue that transitioning from 1D to 3D models will increase Ti and Ni yields. However, Ni is already overproduced by all CC-SN yield sets, though we also expect significant contributions (40%–70%) of Ni by SNe Ia (RJR17, Figure 13). In addition, we note a consistent overabundance of Si and S for all CC-SN tables. Fryer et al. (2018, Figure 8) clearly demonstrate that this might be remedied by decreasing the CC-SN explosion energies.

Despite the inclusion of a variable error model, both our statistical methods are significantly affected by poorly reproduced elements, and exclusion of certain elements, such as Sc, could thus drastically change and reorder the overall scores. This indicates the necessity of matching the elemental subset for which the score is calculated to that of the simulation.

### 5.2. Subset of Most Abundant Elements

The above analysis can similarly be performed for any subset of elements, for example, the nine most abundant, which are usually tracked in hydrodynamical simulations. These are the most well understood elements; thus, we would expect our CC-SN yield tables to perform better in this context.

#### 5.2.1. Posterior Parameter Distributions

Tables 3(b) and 4(b) give the relevant MCMC output for the five yield tables, and we see that in almost all cases the posteriors are far greater than those previously obtained, with much less distinction between yield sets. This is a direct result of the removal of the most poorly predicted elements from the analysis. Similarly, the median values of $\beta$ are greater for most yield sets, indicating a preference for smaller model errors, suggesting a better fit to the data. The $1\sigma$ confidence intervals of $\alpha_{\text{IMF}}$ and $\log_{10}(N_{\text{el}})$ for all yield tables can be seen to overlap with the respective intervals from the $n_{\text{el}}=28$ optimization (i.e., Table 3(a)).

For C04, N13, W17, and P18, fewer CC-SNe and more SNe Ia are favored in order to reproduce protosolar abundances (compared to the $n_{\text{el}}=28$ run, considering $\alpha_{\text{IMF}}$ and $\log_{10}(N_{\text{el}})$). This is most likely because formerly underproduced elements (e.g., Sc, K, or Cl as visible from Figure 3) are now excluded and illustrates the bias that can be introduced by element choice (in combination with wrongly predicted elements). In contrast, the R17 yields produce more CC-SNe and fewer SNe Ia when using the reduced $n_{\text{el}}=8$ subset. From the pattern of Figure 3, it is clear that R17 underproduces the light $\alpha$-elements and overproduces some heavier elements (e.g., P, Cl, Cr), leading to this opposite effect. We also note that R17 produces black holes for all stars with $m > 25 M_\odot$, giving less overall enrichment, resulting in the need for the IMF to become top-heavy (see Sukhbold et al. 2016, for an in-depth discussion of the black hole explosion “landscape”).

For P18 yields, we note that the maximum posterior is in fact reduced, indicating that the 29-element score is less affected by poorly reproduced elements. Comparison of Figures 3 and 4 shows that this is due to the significant bias from the severely underproduced Mg in the nine-element case. This is reduced for $n_{\text{el}}=28$ owing to an optimal parameter set giving greater Mg production and for the scoring metrics, the effect of the geometric averaging over elements.

If we omit the pathological cases of W17 and R17 (due to their high “failed SN” fraction), then $\alpha_{\text{IMF}}$ (log$_{10}(N_{\text{el}})$) is constrained to around $-2.35 (-2.9)$ by C04, N13, and P18. The IMF high-mass slope, inferred here for the Milky Way from chemical evolution constraints, is slightly steeper than the canonical value, in agreement with recent inferences using star counts in M31 (Weisz et al. 2015). Studies by both RK10 and MC15 favor a slightly bottom-heavy Kroupa et al. (1993) IMF with a high-mass slope of $-2.7$, which results in about 2.5 times less CC-SN explosions compared to the CH03 IMF with $\alpha_{\text{IMF}} = -2.3$, which we more or less recover (see MC15, Table 5). At the same time, we are utilizing a smaller CC-SN mass range with $M_{\text{CC-SN,max}} = 40 M_\odot$, which reduces the number of CC-SN by 27% (see Romano et al. 2005, Table 1) and excludes the enrichment from supermassive stars, which usually relies on extrapolation. RK10 does this up to $100 M_\odot$, whereas MC15 simply integrate up to the respective highest-mass grid point. In addition, the inferred IMF also depends on the yield sets and observational constraints applied, which differ for all three works.

As another example, Suzuki & Maeda (2018) explicitly address the issue of which mass ranges explode as CC-SNe and where explosions fail using GCE arguments with C04 yields together with solar O and Fe observations. They conclude that $M_{\text{CC-SN,max}} = 100 M_\odot$ best reproduces observations, but their analysis suffers from yield table extrapolation and fixed GCE parameters. When allowing for the latter (as here), those abundances are well recovered using C04 yields and $M_{\text{CC-SN,max}} = 40 M_\odot$ (see Figure 4). This illustrates that keeping GCE parameters fixed will inevitably bias the inference.

#### 5.2.2. Scoring Metrics

The Bayes scores are several orders of magnitude greater than for $n_{\text{el}}=28$ in most cases as a result of the exclusion of

---

12 We note that it is difficult to infer the IMF high-mass slope from star counts from within the Milky Way owing to incomplete knowledge of the most massive stars (e.g., Rybizki & Just 2015; Dib et al. 2017).
poorly reproduced elements. In accordance, $\beta$, $S_B$, and $S_{CV}$ are largest for C04 with this choice of elements, albeit with a value of $\log_{10}(S_{CV}) = 0.23$, which is still not compatible with the self-observation score (obtained by setting the model “data” equal to a Gaussian realization of the protosolar observations), which has $\log_{10}(S_{CV}) = 0.64_{-0.06}^{+0.04}$. In $S_B$ we note a relative preference of $\approx 10$ for C04 over N13 and W17, which are themselves favored over R17 and P18 by a significant factor.

The LOO-CV scores are less decisive, with relative scores of around 1.6 (4) between the C04 and N13 (R17) yields, for example, meaning that detailed conclusions are more difficult. Both metrics are in agreement, however, that the R17 and P18 yields are the worst predictor of protosolar abundances for the CC-SN yields and element set tested here. As for the maximum posteriors, we note that, in the case of P18, $S_B$ is lower for the nine-element subset, again as a consequence of the greater influence of the Mg-induced bias in this case. For the LOO-CV score, we report no significant change, due to the weaker dependence of this score on poorly reproduced elements.

The ranking varies for the two scores and also for the different elemental subsets, although overall C04 seems to score well in both, whereas R17 tends to come last. MC15 also prefers C04 CC-SN yields even though their elemental subset (C, N, O, Fe, and [$\alpha$/Fe]) is different from ours and other observational constraints like the metallicity distribution function (MDF) and star formation history are included. At the same time, they find that the AGB yields are less important (despite half their elements being strongly affected by AGB feedback) and that a Kroupa et al. (1993) IMF is preferred. They also face the problem of how to factor in results from half their elements being strongly affected by AGB feedback, and that a Kroupa et al. (1993) IMF is preferred. They also face the problem of how to factor in results from their $\chi^2$ metrics; hence, a second overall score is reported with the MDF being left out, resulting in C04 and N13 scoring best together with the Karakas (2010) AGB yields. RK10 also favors N13 CC-SN yields over the classical Woosley & Weaver (1995) values (C04 yields were not included in their study).

5.2.3. Posterior Element Predictions

Figure 4 plots the LOO-CV predictions for each element, showing far better agreement between observations and predictions than in Figure 3, as expected. The majority of the predictions are consistent with observations at a $2\sigma$ level, although there are still significant discrepancies for Mg using W17, R17, and P18 yields and a significant overproduction of Fe by R17.

The yield set predictions for Mg and Si are broadly self-consistent yet significantly different from the protosolar data, with Mg (Si) always underproduced (overproduced). This could be mitigated in CC-SN modeling by decreasing the explosion energy of the piston (Heger & Woosley 2010; Fryer et al. 2018, Figure 8), or it could be indicating errors in the solar abundance determination, though that seems less likely (Amarsi & Asplund 2017).

In Figure 5 we show predictions of our modeling in comparison to APOGEE red clump (RC) stars (Bovy et al. 2014). We select 4832 such objects from APOGEE DR14 (Abolfathi et al. 2018) subject to the constraints $5 \, \text{kpc} < R_{\text{gal}} < 9 \, \text{kpc}$, $|z| < 1$, and with the APOGEE flags ANDFLAG, ASPCAPFLAG, EXTRATARG being zero. For the Chempy model we show our C04 results with $n_d = 8$. The distribution is sampled using 250 randomly drawn evolutionary tracks from the MCMC posterior. For each track the respective SFR was taken into account, together with the RC age distribution. An observational error of 0.05 dex in [Fe/H] and 0.03 dex in [$\alpha$/Fe] was added to simulate real data. Further details on this procedure are given in Just & Rybizki (2016, Section 3). Similar to RJR17, Figure 11, we see that the MDF of the APOGEE data does not possess the long tail to lower metallicities that is seen in the Chempy mock observations; instead, they produce a greater number of supersolar-metallicity stars. In addition, the highest densities of both model and data are at the origin, i.e., solar values of [Fe/H] and [$\alpha$/Fe], as expected since the Chempy parameters are optimized to reproduce protosolar abundances. We caution that this compares a selection of RC stars with predictions of a Chempy model that was optimized to reproduce only eight protosolar abundances, and thus the spatial selection in the data does not represent our solar one-zone model (a proxy of which is impossible to find), nor do the individual elements fit equally well. For these reasons, the APOGEE data have not been used as constraining data for Chempy in this work.

5.3. Testing Inferred SSP Parameters in Hydrodynamical Simulations

It is not certain a priori that our predicted best-fit SSP parameters, obtained by marginalizing over Chempy ISM parameters, will be useful for a hydrodynamical simulation with a completely different ISM physics representation, since specific model assumptions can significantly bias the inferred parameters, as shown in Côté et al. (2017b). Even if the ISM model were similar, Chempy uses constant parameter values over the whole period of galactic evolution, some of which (e.g., the SFE) will vary over time and space in a hydrodynamical simulation.

5.3.1. Optimizing Parameters for a Specific Yield Set

To test this, we apply our method to predict two SSP parameters that we subsequently insert into a zoom-in simulation of a Milky-Way-like galaxy using the cosmological
magnetohydrodynamical code AREPO. The fiducial model is identical to the simulation of “Halo 16” (level 5) in Gutcke & Springel (2017, hereafter GS17), which uses initial conditions from the Auriga project (Grand et al. 2017) and has been shown to match a variety of observational constraints from the Milky Way. The yield set is identical to the one introduced in the IllustrisTNG simulations (Pillepich et al. 2018, Table 2), henceforth referred to as TNG yields. These are implemented into 

Chempy, with $M_{\text{CC-SN,max}}$ and the SN Ia delay time modified to match the simulation values of 100 $M_\odot$ and 40 Myr, respectively. This ensures that our IMF and SN Ia model (and yield set) are identical to those of GS17; hence, the SSP enrichment over time is the same for both codes if the same SSP parameters are used. The fiducial values in GS17 are $\log \theta(\alpha) = -2.89$ and $\alpha_{\text{IMF}} = -2.3$ (using a CH03 IMF).

Three 

Chempy analyses were performed using $n_{\alpha} = 8$:

1. Leaving all five 

Chempy parameters and $\beta$ free, using TNG yields with the appropriate $M_{\text{CC-SN,max}}$ and SN Ia delay time values ($\text{TNG}_{\text{IMF & SN Ia free}}$).
2. As in (1), but fixing $\alpha_{\text{IMF}}$ to $-2.3$ ($\text{TNG}_{\text{SN Ia free}}$).
3. As in (2), but also fixing $\log \theta(\alpha)$ to $-2.89$ ($\text{TNG}_{\text{fiducial}}$).

A graphical summary of the resulting chemical abundance tracks drawn from the posterior of this inference can be inspected in Figure 7. The resulting best-fit parameters and scores are summarized at the end of Table 3. As expected, the scores increase with every additional free parameter, and the effect of altering the IMF is found to be stronger than that of the SN Ia normalization. The Bayes score increases by almost $10^3$ when allowing both SSP parameters to vary freely compared to the fiducial case where they are fixed. We note that it also supersede the score found for our other yield sets even though the TNG yield set is very similar to our N13 yield set. The reason is most likely that the TNG yields have additional grid points at 9 and 12 $M_\odot$ from the Portinari et al. (1998) yields, which allows for a higher flexibility, as the N13 yields are extrapolated from 13 $M_\odot$ down to 8 $M_\odot$. In the case with only $N_{\text{Ia}}$ free, the number of SNe Ia per $M_\odot$ almost doubles, but when the IMF is also allowed to vary, $N_{\text{Ia}}$ becomes close to the fiducial value and $\alpha_{\text{IMF}}$ decreases to $-2.68$.

5.3.2. Comparing Hydrodynamical Simulation Predictions to Observational Data

Next, these results are tested by rerunning the hydrodynamical simulation after inserting the derived SSP parameters. In Figure 7, the predicted abundance distribution in the $[\alpha/Fe]$ versus [Fe/H] plane is shown for the hydrodynamical simulations with each set of SSP parameters, with the left panel showing the fiducial case (see GS17, Figure 6). The middle panel displays the simulation using the optimized parameter $\log \theta(\alpha) = -2.63$, while the right panel uses the parameters $\log \theta(\alpha) = -2.87$ and $\alpha_{\text{IMF}} = -2.68$, both left free during the 

Chempy analyses. Here $\alpha$ is defined as the mass-weighted average of O, Mg, and Si. The abundance distributions are compared to those of the APOGEE RC stars (Bovy et al. 2014; Abolfathi et al. 2018; Ness et al. 2018), plotted in black contours. Stars from the simulation are weighted according to the RC age distribution (Bovy et al. 2014, Equation (11)) and selected according to the APOGEE spatial selection of 5 kpc $< R_{\text{Gal}} < 9$ kpc and $|z| < 1$ kpc. These data were not used during the 

Chempy optimization and are depicted to illustrate the typical variation of Milky Way disk stars. Additionally, we show in blue (gray) the median from 250 abundance tracks randomly drawn from the posterior (prior) of the respective 

Chempy inference specified in Section 5.3.1. The dashed (dotted) lines enclose the 1$\sigma$ (2$\sigma$) contours of these tracks.

Whereas no Sun-like star is produced by the simulation with the fiducial SSP parameters (since the abundance distribution is far too $\alpha$-enhanced), there are stars with [Fe/H] = $[\alpha/Fe] = 0$ for the run using the $\log \theta(\alpha)$ prediction, similar to the findings in Naiman et al. (2018, Figure 4), where $N_{\text{Ia}}$ is increased by a factor of 3.5. However, $\alpha$-enhancement is only decreased by a small fraction, still being 0.2 dex overabundant with respect to the APOGEE stars. Considering the simulation using both optimized SSP parameters, the resulting abundance distribution of the Milky-Way-like hydrodynamical simulation matches the APOGEE data very closely, in both slope and $[\alpha/Fe]$ scatter of the distribution. Some discrepancy remains, with a few more stars in the metal-poor regime and a metal-rich tail not found in the observational data. The latter may arise from too efficient radial migration in the simulation.

Since no additional parameters were changed for these simulations, we note that many simulation predictions of the fiducial run are likely altered when modified SSP parameters are used. Most notably, the feedback parameters were not adjusted, and the new $N_{\text{Ia}}$ and $\alpha_{\text{IMF}}$ values possibly have a significant impact on the resulting properties of the galaxy. Seeing that our goal here is to show how 

Chempy can be used to optimize the $\alpha$-abundance predictions, such investigation is beyond the scope of this paper.

References

Appendix B, Table 4(d).
5.3.3. Why Can Our Best-fit Parameters Improve Hydrodynamical Simulations?

As can be seen from Figure 7, the Chempy median posterior (solid blue line) is close to the running median of the hydrodynamical simulation for tested SSP parameter sets. Notably, the variances of the two distributions are similar when both IMF parameters are allowed to vary (right panel), but the Chempy track is far narrower than that of the hydrodynamical simulation in the fiducial case with SSP parameters held constant, implying that variation of the ISM parameters causes only a minor spread in the evolutionary tracks. As noted previously, in the fiducial model we see that the posterior ISM parameters mostly reproduce the prior, and thus there is negligible difference between the Chempy prior and posterior tracks in this case. For the right panel, it is clear that the posterior distributions (and observational data) lie far from the prior evolutionary tracks, a consequence of the steeper IMF found to be preferred for this scenario. Despite the differing distribution widths, the clear correlation between the two simulations implies that our simple GCE model Chempy can produce a satisfactory reproduction of the evolutionary tracks of far more complex simulations. This may be rationalized by noting that, although the ISM physics of the hydrodynamical simulation has a completely different parameterization and includes more effects than our simple GCE model, the hyperparameters are still tuned to reproduce global ISM constraints, including the Kennicutt–Schmidt law (Springel & Hernquist 2003, Figure 3) and the cosmic star formation history (Vogelsberger et al. 2013, Figure 6), which were also incorporated into the Chempy ISM parameterization and priors. Because the SSP enrichment engines are the same for both models, the optimized SSP parameters coming from a Chempy MCMC inference are seen to be useful predictors for the chemical abundance tracks of hydrodynamical simulations. The increased ISM complexity simply adds to the variance of the abundance distribution, while the median track is conserved. This outcome could be biased because only a single galaxy was considered and it is known that the chemical enrichment of galaxies of a certain mass depends on their evolutionary environment (Vogelsberger et al. 2013, Figure 5; Blancato et al. 2017, Figure 11). However, we expect the "environmental" abundance prediction noise at that galaxy mass to be lower than the [α/Fe] discrepancy of 0.2 dex found here. At the same time, we are optimizing for eight abundances (as in Figure 6) but only consider the α-enhancement in the APOGEE data, so we could have optimized for fewer elements. Similarly, the previously used protosolar observational data could be supplemented with more data points (or the whole distribution) from APOGEE to more stringently optimize the Chempy parameters, allowing the implementation of constraints from sub- and supersolar metallicity. Even in this simple form, however, we note a marked improvement in the hydrodynamical simulation abundance predictions using our optimized SSP parameters.

6. Summary and Conclusions

In this paper we have constructed a scoring system for nucleosynthetic yield tables and tested it on five different CC-SN yields using a simple observational data set, the protosolar abundances. The software developed can be utilized to make informed decisions regarding yield table choice for simulations and to produce optimal Galactic and stellar parameters for any subset of traced elements.

Our simulation framework has been the flexible GCE modeling software Chempy (Rybizki et al. 2017), with the addition of an error model to account for simulation inadequacies and yield table errors. Two scoring metrics were considered, based on Bayesian evidences and LOO-CV, giving overall scores that, unlike other approaches, marginalize over the main GCE parameters: IMF slope, SN Ia normalization, SFR, SFE, and outflow fraction. In addition, the error parameter is left free, with lower model error implying better yield table predictions. We show that the resulting abundance distributions (in the [α/Fe]–[Fe/H] plane) of the Chempy simulation (with optimized parameters) are comparable to observational data, despite being constrained only by protosolar data.

The yield table ranking of both metrics depends on the choice of elements; thus, we have tested two different sets: (a) using all available elements, and (b) using the subset of the nine most abundant elements (as commonly tracked by hydrodynamical simulations). The two scores yield

![Figure 7. Distribution of stellar particles in the Halo 16 simulation at $z = 0$ in the [$\alpha$/Fe] vs. [Fe/H] plane for the fiducial simulation (left), the simulation using the optimized value of $N_{\text{fl}}$ (middle), and that with the optimal values of both $N_{\text{fl}}$ and $\sigma_{\text{IMF}}$ (right). The black contours denote the logarithmic density isocurves of APOGEE red clump stars (Bovy et al. 2014) from DR14 (Abolfathi et al. 2018), using the same selection as in Figure 5. Both the spatial and red clump age selections are also applied to the hydrodynamical simulation, with this figure following the same analysis as GS17, Figure 6. To this we add blue (gray) lines showing the median abundance tracks from the respective Chempy posterior (prior), with the dashed (dotted) lines enclosing the 1σ (2σ) deviations of all tracks.](image-url)
qualitatively similar results for most yield sets and favor Prantzos et al. (2018) (P18) and Chieffi & Limongi (2004) (C04) yields for cases (a) and (b), respectively. C04 and West & Heger (2017, in preparation) yields also score well for case (a), but the latter requires an unrealistically top-heavy IMF (due to their model’s inclusion of many failed SNe). For our two CC-SN yield tables, which do not include failed SNe, C04 and Nomoto et al. (2013) (N13), we infer their mutual encompassing 1σ range to be $-2.30 > a_{\text{IMF}} > -2.63$ and $-2.68 > \log_{10}(N_{\text{He}}) > -3.01$ when using the smaller (more reliable) elemental subset. CC-SN models that include rotation (the P18 yields) substantially improve the yield table predictions; poor results for the nine-element case can be attributed specifically to the severe underproduction of Mg in this model.

We find the Bayesian metric to be more discriminative and computationally cheaper, whereas the cross-validation has the advantage of being a useful tool for identifying poorly predicted elements. In case (a) we find a consistent overproduction (underproduction) of Si, S, and Ni (Sc and Ti) regardless of the tested yield table. For Sc and Ti, missing neutron star wind enrichment could potentially explain this (Pruet et al. 2005). In case (b) we find that all CC-SN yield tables overpredict Si and underpredict Mg, which could be remedied by reducing the explosion energy (Heger & Woosley 2010; Fryer et al. 2018, Figure 8). This also applies to S from case (a).

The utility of the code in predicting optimal simulation parameters is shown via application to a hydrodynamic zoom-in simulation (Gutcke & Springel 2017) using a fixed yield set (Table 2 Pillepich et al. 2018) both with and without the constraint of fixing the IMF. The resulting abundance distribution of this simulation, which is heuristically similar to the Chempy predictions, is significantly improved when both of our best-fit SSP parameters are used, as validated by comparison to an independent data set (see Figure 7). The improvement by only optimizing the number of SNe Ia is much less, and the IMF seems to have a stronger impact on the abundance prediction. The reason why a simple GCE optimization of SSP parameters can improve the abundance patterns produced by hydrodynamical simulations, despite strongly differing ISM physics implementations, is because both types of simulation try to reproduce average ISM relations (Springel & Hernquist 2003; Vogelsberger et al. 2013).

In addition, our analysis may be used to test the improvement when modeling extra nucleosynthetic channels, for example, neutron star mergers (Côté et al. 2017a) or super-AGB stars (Siess 2010; Doherty et al. 2015). Extending the observational constraints to include more stars or other GCE data (see MC15) is also straightforward.

The code has been made publicly available, including a comprehensive online tutorial and a method to produce SSP enrichment tables similar to Ritter et al. (2017a).
We would like to thank Kareem El-Badry, Coryn Bailer-Jones, Benoit Côté Chris Fryer, Brad Gibson, Robert Grand, David Hogg, Samuel Jones, Annalisa Pillepich, and Hans-Walter Rix for fruitful discussions regarding this work, as well as Alexander Heger, Jill Naiman, Nikos Prantzos, Christian Ritter, and Christopher West for providing us with yield tables and Donatella Romano for providing us with her solar abundance predictions. In addition, we would like to thank Walter Rix for fruitful discussions regarding this work, as well as Alexander Heger, Jill Naiman, Nikos Prantzos, Christian Ritter, and Christopher West for providing us with yield tables and Donatella Romano for providing us with her solar abundance predictions. In addition, we would like to thank Walter Rix for fruitful discussions regarding this work.

The Astrophysical Journal, 861:40 (15pp), 2018 July 1

Appendix A

Testing the Neural Network

To test the accuracy of the neural network, additional data sets were constructed with each parameter drawn from a uniform distribution of $3\sigma_{\text{prior}}$ width centered on the prior mean. Figure 8 shows a corner plot of the five-dimensional Chempy parameter space, $\theta$, with colored points representing the median L1 error between the neural network and Chempy at that location in the parameter-space slice. This is done for $n_{\text{el}} = 28$ and N13 yields, and histograms of the test data set are shown on the diagonal. Average neural network errors across the $3\sigma_{\text{prior}}$ range were 0.0097 ± 0.0081, and the figure shows that in the central regions, where the posterior peak concentrates, the error is far smaller. Crucially, these errors are insignificant compared to those of the minimum observational error of 0.02 dex. The network errors can be further reduced by training for a greater number of epochs, although this requires additional computation time.

Appendix B

Posterior Distributions and Scores for All Runs

Table 4 includes the results from Table 3, as well as the ISM parameters and the maximum posterior value for the Chempy runs previously described. In addition, we include TNG yield runs with $M_{\text{CC-SN,max}} = 40 M_\odot$ (here denoted TNG$_{40}$), which are more comparable to our five tested yield sets. The TNG$_{40}$ runs yield similar results to the previous $M_{\text{CC-SN,max}} = 100 M_\odot$ (TNG$_{100}$) runs, but with slightly poorer scores and a less bottom-heavy IMF. The Bayesian evidence for both TNG runs is highest when only the IMF is left free and $N_{\text{Ia}}$ is fixed to the fiducial value (which is also recovered when both SSP parameters are left free). This results in a higher score because the parameter space is collapsed along this axis at a favorable value.

For TNG$_{100}$, even the maximum achieved posterior is higher in the lower-dimensional parameter space, though one would expect that leaving an additional parameter free would result in a better fit. Because of the high dimensionality, the MCMC never samples the exact peak, which will be hit much more easily by a Markov chain that has to sample in a lower-dimensional space (under the condition that the missing parameter is set close to the optimal value).

The ISM parameters mainly reproduce the priors, albeit with a small correlation with the IMF high-mass slope. If it tends to be top-heavy, as in case (a), the SFE is slightly higher, the SFR peaks earlier, and the outflow fraction is lower, all of which slightly increase the ISM metallicity. The correlation points in the other direction when the IMF becomes slightly bottom-heavy, as in case (b).
### Table 4
Inferred *Chempy* and Model Error Parameters (16th, 50th, and 84th Percentiles of \( \theta_{\text{posterior}} \)) Obtained from an MCMC Analysis for Each Yield Set

| Yield Set | Scores | Error Parameter | SSP Parameters | ISM Parameters |
|-----------|--------|-----------------|---------------|---------------|
| \( \chi \) | \( \log_{10}(S_\theta) \) | \( \log_{10}(S_{UV}) \) | Posterior \( \log_{10}(P_{\text{max}}) \) Priors: | | |
| | \( \mu \) | \( \sigma \) | | \( \mu \) | \( \sigma \) | \( \mu \) | \( \sigma \) | \( \mu \) | \( \sigma \) |
| \( \alpha_{\text{IMF}} \) | \( -2.3 \pm 0.3 \) | | \( -2.75 \pm 0.3 \) | | | | |
| \( \log_{10}(N_{\text{HI}}) \) | | | | \( -0.3 \pm 0.3 \) | | \( 0.55 \pm 0.10 \) | | \( 0.5 \pm 0.1 \) |
| \( \log_{10}(S_{F E}) \) | | | | | | | |
| \( \log_{10}(S_{\text{FRE}}) \) | | | | | | | |
| \( x_{\text{out}} \) | | | | | | | |

(a) All 29 Available Elements

| \( \chi \) | \( \mu \) | \( \sigma \) | \( \mu \) | \( \sigma \) | \( \mu \) | \( \sigma \) | \( \mu \) | \( \sigma \) |
| C04 | -1.21 | 0.39 | 0.77 | 0.30 | -2.30 | 0.11 | -3.14 | 0.16 |
| N13 | -5.69 | -7.35 | 0.58 | 0.29 | -2.32 | 0.15 | -2.90 | 0.19 |
| W17 | -0.78 | -3.91 | 0.72 | 0.32 | -1.83 | 0.18 | -3.29 | 0.15 |
| R17 | -6.11 | -9.18 | 0.50 | 0.36 | -2.29 | 0.23 | -2.91 | 0.21 |
| P18 | 0.86 | -7.36 | 0.79 | -0.33 | -2.13 | 0.23 | -2.93 | 0.16 |

(b) Nine Most Abundant Elements

| \( \chi \) | \( \mu \) | \( \sigma \) | \( \mu \) | \( \sigma \) | \( \mu \) | \( \sigma \) | \( \mu \) | \( \sigma \) |
| C04 | 1.61 | 10.53 | 1.01 | 0.32 | -2.45 | 0.15 | -2.89 | 0.13 |
| N13 | 0.65 | 7.99 | 0.83 | 0.37 | -2.52 | 0.12 | -2.79 | 0.11 |
| W17 | 0.73 | 7.26 | 0.90 | 0.30 | -2.19 | 0.17 | -3.09 | 0.16 |
| R17 | -0.49 | 3.44 | 0.78 | 0.29 | -2.00 | 0.22 | -3.22 | 0.21 |
| P18 | -0.01 | 5.25 | 0.72 | 0.35 | -2.22 | 0.20 | -2.84 | 0.18 |

(c) Like (b) but for TNG Yields with \( M_{\text{CC-SN, max}} = 100 M_\odot \)

| \( \chi \) | \( \mu \) | \( \sigma \) | \( \mu \) | \( \sigma \) | \( \mu \) | \( \sigma \) | \( \mu \) | \( \sigma \) |
| TNG\_fiducial | -0.89 | -0.98 | 1.54 | 0.64 | -2.3 | fixed | -2.89 | fixed |
| TNG\_SN la free | -0.31 | -0.15 | 3.41 | 0.80 | -2.3 | fixed | -2.63 | fixed |
| TNG\_IMF free | 2.12 | 0.34 | 12.26 | 1.17 | -2.72 | 0.05 | -2.89 | fixed |
| TNG\_IMF & SN la free | 1.82 | 0.29 | 12.01 | 1.13 | -2.68 | 0.09 | -2.87 | -0.09 |

(d) Like (c) but with \( M_{\text{CC-SN, max}} = 40 M_\odot \)

| \( \chi \) | \( \mu \) | \( \sigma \) | \( \mu \) | \( \sigma \) | \( \mu \) | \( \sigma \) | \( \mu \) | \( \sigma \) |
| TNG\_fiducial | -0.40 | -0.69 | 2.81 | 0.75 | -2.3 | fixed | -2.89 | fixed |
| TNG\_SN la free | -0.45 | -0.20 | 4.00 | 0.76 | -2.3 | fixed | -2.69 | fixed |
| TNG\_IMF free | 1.69 | 0.20 | 11.18 | 1.02 | -2.64 | -0.06 | -2.89 | fixed |
| TNG\_IMF & SN la free | 1.19 | 0.18 | 11.41 | 1.07 | -2.64 | -0.08 | -2.90 | -0.11 |

Note. The parameter priors and the maximum posterior obtained are also given. Results are given for (a) the set of all available elements with \( n_d = 28 \); (b) the Arepo most abundant element subset (\( n_d = 8 \)); (c) the predicted parameters for the GS17 simulations, using Pillepich et al. (2018), Table 2) TNG yields with \( M_{\text{CC-SN, max}} = 100 M_\odot \) (the default of IllustrisTNG) for five combinations of optimized SSP parameters; and (d) the same as (c) but using \( M_{\text{CC-SN, max}} = 40 M_\odot \), matching this work.
