MAPPING POPULATION SYNTHESIS EVENT RATES ON MODEL PARAMETERS. II.
CONVERGENCE AND ACCURACY OF MULTIDIMENSIONAL FITS

R. O’SHAUGHNESSY and V. KALOGERA
Department of Physics and Astronomy, Northwestern University, Evanston, IL 60208; oshaughn@northwestern.edu

AND

K. BELCZYNSKI
Department of Astronomy, New Mexico State University, Las Cruces, NM 88003

Received 2006 September 16; accepted 2007 June 7

ABSTRACT

Binary population synthesis calculations and associated predictions, especially event rates, are known to depend on a significant number of input model parameters with different degrees of sensitivity. At the same time, for systems with relatively low formation rates, such simulations are heavily computationally demanding, and therefore the needed explorations of the high-dimensional parameter space require major, often prohibitive, computational resources. In the present study, to better understand several key event rates involving binary evolution and binaries with two compact objects in Milky Way–like galaxies and to provide ways of reducing the computational costs of complete parameter space explorations, (1) we perform a methodical parameter study of the StarTrack population synthesis code, and (2) we develop a formalism and methodology for the derivation of multidimensional fits for event rates. We significantly generalize our earlier study, and we focus on ways of thoroughly assessing the accuracy of the fits. We anticipate that the efficient tools developed here can be applied in lieu of large-scale population calculations and will facilitate the exploration of the dependence of rate predictions on a wide range of binary evolution parameters. Such explorations can then allow one to derive the constraints on these parameters, given empirical rate constraints and accounting for fitting errors. Here we describe in detail the principles and practice behind constructing these fits, estimating their accuracy, and comparing them with observations in a manner that accounts for their errors.

Subject headings: binaries: close — black hole physics — stars: evolution — stars: neutron

1. INTRODUCTION

Models for binary stellar evolution and population syntheses are necessary to provide quantitative theoretical predictions for the relative likelihood of assorted events involving the evolution of binary stars. The resulting predictions are particularly critical when no empirical estimates exist for topics of immediate astrophysical interest, such as mergers of double compact objects (DCOs) through the emission of gravitational waves. The most practical and widely applied binary population synthesis codes available in the community, such as the StarTrack code, described in Belczynski et al. (2002) and significantly updated in Belczynski et al. (2007a), the BSE code, described in Hurley et al. (2002), the SeBa code, described in Portegies Zwart & Verbunt (1996), and the StarFaster code, described in Fryer et al. (1998), rely on a large set of fairly simple parameterized rules to characterize many complex and often ill-understood physical processes. Unfortunately, current binary population synthesis codes are greatly and often forbiddingly computationally demanding (depending on their level of sophistication); even with substantial simplifications, exploring the entire parameter space relevant to the simulations is beyond present-day computational capability.

However, observational information can provide us with constraints that help us improve our understanding of massive binary evolution. For example, pulsar searches continue to discover and refine observations of isolated pulsars and new binary pulsar systems; e.g., see Lorimer (2005) for a review. Specifically, the samples of binary pulsars with neutron star and relatively massive white dwarf companions have been used for a statistical derivation of empirical rate estimates for their formation (most recently, see Kim et al. 2003, 2004, Kalogera et al. 2005, and references therein).

In addition, many ground-based gravitational wave detectors now operating at or near design sensitivity (i.e., LIGO, GEO, TAMA) are designed to detect the late stages of DCO inspiral and merger. On the basis of only early-stage data, these instruments have already provided conservative upper limits to certain DCO merger rates (see, e.g., Abbott et al. 2005a, 2005b). With LIGO now very close to design sensitivity, a year of LIGO data could definitively exclude (and even possibly confirm) the most optimistic theoretical predictions for black hole–black hole (BH-BH) merger rates (see, e.g., O’Shaughnessy et al. [2005a] for a range of BH-BH merger rates arising from binary evolution in Milky Way–like galaxies). Thus, gravitational wave–based upper limits (and, eventually, detections) will shortly provide constraints on theoretical models of DCO formation.

Faced with the availability of empirical rate constraints, and yet at first unable to quantitatively impose them on population synthesis predictions because of the extremely limited exploration of the multidimensional parameter space, O’Shaughnessy et al. (2005a) realized that any single unambiguous population synthesis prediction could be sampled loosely and then fitted over the most sensitive population synthesis parameters. In the same study, we also presented a technique to accelerate the synthesis code used (StarTrack) to study a single target subpopulation, which we called “partitioning”: we used experience gained from prior simulations to reject binary parameters highly unlikely to produce the current event of interest.

O’Shaughnessy et al. (2005b) first applied these early fits to compare StarTrack-produced population synthesis predictions...
and the observed formation rate for neutron star–neutron star (NS-NS) binaries. Although only a small fraction (2%) of the StarTrack models appeared to be consistent with the constraints, conceptual challenges with seven-dimensional visualization prevented O’Shaughnessy et al. (2005a) from clearly describing the constraint-satisfying region. A forthcoming paper, O’Shaughnessy et al. (2006), will significantly extend that earlier preliminary analysis, adding more observational constraints, as well as providing a clearer investigation of the constraint-satisfying models.

In this paper, we extend and generalize the analysis of O’Shaughnessy et al. (2005a), and we present a thorough discussion of our much updated and vastly larger population synthesis archive (§ 2), and particularly of the fitting methods that we employ to extract predictions and assess their uncertainties and quality (§ 3). Because several important and yet not immediately obvious pitfalls must be identified and avoided when constructing, testing, and applying fits, in this paper we provide a thorough and pedagogical discussion of our fitting method and philosophy. For example, we explain how systematic fitting errors that were ignored in O’Shaughnessy et al. (2005b) can limit our ability to constrain the merger rates in the Milky Way. In a companion paper (O’Shaughnessy et al. 2006) we explore the astrophysical applications of these fits, emphasizing their use in deriving empirical constraints on population syntheses and rate predictions.

Although we develop the fitting formulation specifically for the results of the StarTrack code, the methods described in this paper can be applied to any family of population synthesis simulations. For this reason, we survey the physics underlying the StarTrack model only in our companion paper, O’Shaughnessy et al. (2006), in which these fits are used to cover the parameter space and apply empirical rate constraints from supernovae and binary pulsars.

2. POPULATION SYNTHESIS ARCHIVES

Population synthesis simulations can be extremely computationally demanding; even though StarTrack can fully evolve roughly $10^5$ binaries of interest $^2$ per CPU hour with modern processors, because some DCOs form very infrequently (e.g., black holes, which occur roughly once every $\sim 10^{-4}$ binaries evolved), a representative sample of stellar systems often contains $10^4$–$10^5$ binaries and requires hundreds of CPU hours to complete. In addition, since population synthesis rate predictions depend delicately on model parameters, the computation time needed to build up a sufficiently representative collection of stellar systems, one in which some event of interest occurs many times, varies considerably depending on astrophysical assumptions. Given the prohibitive computational demands of a brute force approach, we took advantage of several simplifications originally developed in O’Shaughnessy et al. (2005a) to assemble our archive of roughly 3000 population synthesis simulations, upon which our fits are predicated. In this section we briefly describe how these archives were generated and how we identify and extract event rates for several processes of interest.

2.1. StarTrack Population Synthesis Code

We estimate formation and merger rates for several classes of double compact objects using the StarTrack code, first developed by Belczynski et al. (2002) and recently significantly updated and tested as described in detail in Belczynski et al. (2007a). Like other population synthesis codes, StarTrack evolves some number $N$ of binaries from their birth (drawn randomly from specified birth distributions) to the present, tracking the stellar and binary parameters. Because binaries without any high-mass components cannot produce black holes or neutron stars, we configured StarTrack to simulate only those binaries whose heaviest initial component mass $m_1$ was greater than 4 $M_\odot$. In effect, the small number $N$ of binaries simulated mimic the results of a much larger simulation of size $N_{\text{eff}}$ in which $m_1$ can take on any value from the hydrogen-burning limit (0.08 $M_\odot$) to the maximum initial stellar mass that we allow (150 $M_\odot$):

$$N_{\text{eff}} = \frac{\int_{150 M_\odot}^{150 M_\odot} N \, dm(\phi)}{\tilde{N}} \approx 131N,$$

based on a broken Kroupa initial mass function $\phi(m)$ that is $\propto m^{-2.3}$ if $m \in [0.08, 0.5] M_\odot$, $\propto m^{-2.2}$ if $m \in [0.5, 1] M_\odot$, and $\propto m^{-2.7}$ if $m > 1 M_\odot$. In turn, a simulation of $N_{\text{eff}}$ stellar systems can be scaled up to represent the $N_s$ stellar systems in the Milky Way. The number of stellar systems $N_s$ in the Milky Way (and thus the normalization of population synthesis simulations) can be estimated in many ways, depending on the observational inputs used; for the purposes of this paper, we estimate $N_s$ from the average number of binary systems formed through steady star formation over $T = 10$ Gyr of star formation at $\dot{M} \simeq 3.5 M_\odot$ yr$^{-1}$ (Rana 1991; Lacey & Fall 1985):

$$N_s = \frac{MT}{\langle m_1 + m_2 \rangle} \frac{\dot{M}}{\langle m_1(1 + f_b(m_2/m_1)) \rangle} \simeq 7.6 \times 10^{10} \frac{T}{10 \text{ Gyr}} \frac{1}{1 + f_b}\frac{\langle m_2 \rangle}{\langle m_1 \rangle},$$

where $f_b$ is the initial binary fraction (defined as the fraction of the $N_s$ stellar systems that begin their life as binaries) and $\langle m_2 \rangle/\langle m_1 \rangle$ is the ratio of the average masses of the companion and the primary, respectively. (By construction, the binary fraction $f_b$ and the average mass ratio $\langle m_2/m_1 \rangle$ are between 0 and 1.) The proportionality constant $s = N_s/N_{\text{eff}}$ that is needed to scale the results of our simulations up to the universe is therefore

$$s = 5.9 \times 10^8 N^{-1} \frac{T}{10 \text{ Gyr}} \frac{1}{1 + f_b}\frac{\langle m_2 \rangle}{\langle m_1 \rangle}.$$

Note that parameters of the population synthesis model such as $f_b$ and $\langle m_2 \rangle/\langle m_1 \rangle$ influence the result at best by O(50%). These scaling relations allow us to estimate the merger rate implied by simulations ($\hat{R}$) via a surrogate ($\hat{R}$) based on the number $n$ of merger events seen in the simulation:

$$\hat{R} \equiv sn/T = 7.6 \times 10^{-3} \frac{n}{N} \frac{\dot{M}}{\langle m_1 + m_2 \rangle} \simeq (0.059 \text{ yr}^{-1}) \frac{n}{N} \frac{1}{1 + f_b}\frac{\langle m_2 \rangle}{\langle m_1 \rangle}.$$

(Unless otherwise noted, “tilde” quantities refer to estimates for the corresponding “normal” quantity for one individual simulation, based on only information from that one simulation.)

Given unlimited computational resources, this simulation could be repeated many times to measure the merger rate $\hat{R}$ implied by this model to any accuracy desired, or, equivalently, to measure the average number $\mu$ of merger events expected from this simulation:

$$\mu \equiv RT/s.$$
In practice, each simulation is performed once; therefore, the simulated number of merger events \( (n) \) is only statistically correlated to the average number of events expected \( \langle \mu \rangle \), with relative probabilities \( p(n|\mu) \) of any one simulation producing \( n \) events that are given by the Poisson distribution

\[
p(n|\mu) = \frac{\mu^n}{n!} e^{-\mu}.
\]

On average, our estimates of \( \bar{R} = sn/T \), and equivalently of the number of events \( \bar{\mu} = n \), will agree with the true properties of the underlying simulation: averaging over many trials, \( \langle \bar{R} \rangle = R \) and \( \langle \bar{\mu} \rangle = \mu \). However, an estimate based on any one specific simulation should differ from the mean by a characteristic relative amount

\[
\sqrt{\langle (\log \bar{R} - \log R)^2 \rangle} = \sqrt{\langle (\log \bar{\mu} - \log \mu)^2 \rangle} \\
\simeq \left[ \sqrt{\bar{\mu} \ln(10)} \right]^{-1} \simeq \left[ \sqrt{n \ln(10)} \right]^{-1}.
\]

2.2. Parameters Varied in Archives

Our extensive experience in modeling binary compact objects with StarTrack clearly indicates that there are seven parameters that strongly influence compact object merger rates (see, e.g., Belczynski et al. 2002): the supernova kick distribution (three parameters, \( \sigma_1, \sigma_2, \) and \( s \)), describing a superposition of two independent Maxwellians, the massive stellar wind strength \( w \) (one parameter), the common envelope energy transfer efficiency \( \alpha \) (one parameter), the fraction of mass accreted by the accretor in phases of nonconservative mass transfer \( f_a \) (one parameter), and the binary mass ratio distribution, as described by a negative power-law index \( r \) (one parameter).

We allow the dimensionless parameters \( \alpha, \lambda, f_a, w, \) and \( s \) to run from 0 to 1; the dimensionless parameter \( r \) can be between 0 and 3; and finally, we vary the dispersion of either component of a bi-modal Maxwellian \( (\sigma_1 \) and \( \sigma_2 \)) from 0 to 1000 km s\(^{-1}\). In addition, motivated by recent observations suggesting pulsars with masses near 2 \( M_\odot \) (see, e.g., Nice et al. 2005; Splaver et al. 2002), we assume the maximum neutron star mass to be 2.5 \( M_\odot \).³

To improve our statistics, we combine results from several different databases of simulations. The most extensive database samples \( \sigma_1 \in [200, 1200] \) and \( \sigma_2 \in [0, 200] \) very densely and was developed by O’Shaughnessy et al. (2005a) and O’Shaughnessy et al. (2005b). A second archive, significantly less dense due to computational resource limitations, allows both dispersions to run uniformly from 0 to 1000 km s\(^{-1}\). (Additional archives include, for example, a set chosen to better sample kick parameters that best correspond to observations of pulsar proper motions; Arzoumanian et al. 1999; Hobbs et al. 2005.) Consequently, as shown in Figure 1, our archived results do not uniformly sample the kick-related parameters through this range. This irregular sampling has two effects. First, having irregular sampling of the model parameters effectively corresponds to imposing nonflat priors on these parameters and hence biasing the resulting distribution function of merger rates that come directly from the database of runs; see Appendix A and, in particular, Figure 6. Therefore it is even more important to develop the fits, which then allow us to perform uniform sampling of the parameter space, and hence the derivation merger rate distributions assuming flat priors, as is our intention. Second, certain kick combinations are relatively undersampled, which likely plays a role in the relatively poor global convergence of fits for physical parameters, as is described in § 4. Nonetheless, our sampling rather thoroughly explores the most physically likely regimes suggested by Hobbs et al. (2005) and Arzoumanian et al. (1999).

2.3. Event Identification

Most events of physical interest are uniquely and fairly unambiguously identifiable within the code. Type II and Ib/c supernova (SN) events are distinguished by the presence or absence of a hydrogen-rich envelope at the supernova event. We also record DCOs that merge, so we can unambiguously determine the number of BH-BH, BH-NS, NS-NS, and white dwarf–neutron star (WD-NS) merger events that occur in a simulation. [These event classes are hereafter denoted by BHBB(m), BHNS(m), NSNS(m), and WDNS(m), respectively, for brevity.] Since the code also tracks binary eccentricity, for example, we can identify those WD-NS binaries that end their evolution with a nonzero eccentricity [denoted by WDNS(e)].

When constructing our archived population synthesis results, we do not record detailed information about the nature and amount of any mass transfer onto the firstborn NS. We therefore cannot determine the degree to which pulsars are recycled. However, we do record whether any mass transfer occurs. Thus, for the purposes of identifying a class of potentially recycled ("visible") wide NS-NS binaries [denoted by NSNS(vw)], we assume that any system undergoing non-CE mass transfer recycles its NS primary.

2.4. Practical Archive Generation and Resolution, with Partitions and Heterogeneous Targets

The accuracy to which each population synthesis event rate prediction is known, \( \sim 1/\sqrt{n} \), is uniquely set by the number of events \( n \) seen within that simulation. For this reason, O’Shaughnessy et al. (2005a) (1) designed their population synthesis runs to continue until a fixed number of events were seen and (2) used results only from such targeted simulations, where a minimum number of events

³ Such a high neutron star mass converts many merging binaries that we would otherwise interpret as BH-NS or even BH-BH binaries into merging NS-NS binaries, driving down the BH-NS and BH-BH rates significantly from the distributions shown in O’Shaughnessy et al. (2005b), as also described in § 4.
was guaranteed. In contrast, in this study, given limited computational resources and a wide range of targets for which predictions are needed, we extract all possible information from each simulation; whenever possible, we make an estimate of each event rate of interest.

However, it is important to note that most of our simulations employ some degree of accelerating simplification, which can bias estimates. To give the most extreme example: the most accurate estimates for the BH-BH merger rate come from population synthesis runs that evolve only a subset of possible progenitor binaries by using partitions. This subset has been shown to include the progenitors of the vast majority of BH-BH binaries, but very few progenitors of NS-NS binaries and other less massive DCOs (for more information, see O’Shaughnessy et al. [2005a]; we continue to employ the same partition they devised). Similarly, the vast majority of simulations used to study NS-NS event rates (1) use a similar partition to reduce contamination from white dwarf (WD) binaries and (2) terminate the evolution of any binary immediately after any WD forms. These strong biases make data immediately after any WD forms. These strong biases make data

3. FITTING: PRINCIPLES AND TESTS

Given the prohibitive computational demands of the direct population synthesis simulations described in § 2, we use fits based on archived results of population synthesis runs as a surrogate for repeated detailed simulations. Confidence in our results is therefore tied intimately to confidence in the quality of these fits. However, even low-order fits in seven dimensions involve many parameters: to fit any nontrivial function, we must fit roughly a handful of data points per parameter. To build confidence, we must show that the fit order chosen adequately describes the data without overfitting. More delicately, since these fits are used in the derivation of empirical constraints on the population synthesis models in our companion paper (O’Shaughnessy et al. 2006), we must also be able to show that the key end product, the “constraint-satisfying model region,” does not depend sensitively on the fit details or on random accidents in the data (i.e., any different Monte Carlo realization of our simulations should yield the same result). Finally, we note that our companion paper (O’Shaughnessy et al. 2006) will impose four independent constraints. In order to have more than 90% confidence that all models that satisfy all constraints are indeed inside the intersection of the four constraint-satisfying regions, we must at a minimum show that each individual constraint-satisfying prediction contains more than 0.90^{1/4} \approx 0.974 of the models that truly satisfy that constraint.

3.1. One-dimensional Model

Although we intend to build confidence explicitly in our ability to make predictions on the basis of seven-dimensional fits, some of the required principles, notation, and tools are best illustrated through a one-dimensional example. Thus, in this section we explore an arbitrarily chosen but known “population synthesis” model. Depending on one real parameter x with 0 \leq x \leq 1, this model on average will produce \mu(x) merger events out of N = 10^5 binaries, with \mu(x) chosen arbitrarily as

$$\mu(x) \equiv 10^{M(x)} = 5 + 20x^2.$$  (8)

Specifically, in this model we assume the following: that all stars are born in binaries (f_b = 1); that the smaller companion star has a mass randomly distributed from the hydrogen-burning limit to the primary’s mass, corresponding to \langle m_2/m_1 \rangle = 0.5; and that that star formation extends over an interval of T = 10 Gyr. Under these specific circumstances, as discussed in § 2 (eqs. [3] and [5]), a model that produces \mu(x) events on average out of 10^5 binaries corresponds to a merger rate of

$$R(x) = s \mu(x)/T \quad \approx 4 \times 10^{-7} \mu(x) \text{ yr}^{-1}. \quad (9)$$

An individual simulation of this model will produce some number \bar{n} of events, with \bar{n} statistically related to \mu(x) via the Poisson distribution (eq. [6]).

To estimate the merger rate as a function of x, we perform several simulations (N_e = 11) using different parameter values x = x_\alpha = (\alpha - 1)/10 for index values \alpha = 1–11, obtaining results n_\alpha and thus estimates \hat{R}_\alpha = n_\alpha/T, as shown in Figure 2. To better fit the merger rate over the many orders of magnitude that appear in practice, we choose our fit R(x) = 10^{M(x)} s/T to minimize the logarithmic difference between it and the simulations, weighted by the relative statistical uncertainties of each simulation, |n/\langle n \rangle - 1|^{-1/2}, from equation (7). (Here and throughout, we use “hat” symbols to denote our fit quantities; unlike “tilde” quantities, which estimate properties of an individual simulation, the “hatted” quantities generally depend on the results of all simulations simultaneously.) Specifically, to generate a qth-order polynomial fit the logarithm of the (Gaussian) maximum likelihood estimator for R(x), the best single estimate possible,4 Figure 2 shows the results of this minimization for the linear (q = 1) and quadratic (q = 2) fits:

$$\log \hat{R}_1 = \log(s/T) + 0.49 + 0.93x, \quad (12)$$

$$\log \hat{R}_2 = \log(s/T) + 0.35 + 1.49x - 0.547x^2. \quad (13)$$

4 We have also applied a maximum likelihood estimate based on Poisson rather than (approximate) Gaussian errors. For the problems explored here, for which many merger events are usually available, we see no significant difference between the results of this more statistically appropriate method and the conventional and pedagogically far simpler Gaussian maximum likelihood method.
Both performing simulations and fitting introduce errors, which we can quantify by the mean square deviation between the exact merger rate \( R(x) \) and its fit \( \hat{R}(x) \):

\[
I_q = \left\{ \int dx \left[ \log[\hat{R}(x)/R(x)] \right]^2 \right\}^{1/2}.
\]

For example, the value of \( I_1 \) is 0.075. Roughly speaking, \( I_q \) estimates the rms uncertainty in the fit: \( \hat{R}_q \) should lie within a factor of \( \sim 10^{\pm I_q/2} \) of \( R(x) \). For example, \( \hat{R}_1 \) should be accurate to within a factor of roughly 10 \( \pm 0.075 \approx 1.2 \), as is confirmed by Figure 2.

Similarly, the values of \( x \) that correspond to a given merger rate should be uncertain to roughly \( O(\sqrt{\partial R/\partial x}) \approx 0.1 \).

As noted previously, an extremely high level of confidence is needed in any prediction of a constraint-satisfying region. Let us use \( V(G, f) \) to denote the set of points that \( f \) maps to an interval \( G = \{g_{\text{min}}, g_{\text{max}}\} \), such that \( V(G, f) = \{x | f(x) \in G\} \); and \( C \) is an observation of that we demand our simulation to reproduce; then the fraction of \( V(C, \log R) \), the population synthesis models that truly satisfy the constraint, that are inside \( V(C, \log \hat{R}) \), the set of population synthesis models that are naively predicted to satisfy the constraint, based on the fit and the original constraint, is given by \( r_s(C, \log R | C, \log R) \), where \( r_s \) is generally defined by

\[
r_s(A, \log R | B, \log \hat{R}) = \frac{|V(A \log R) \cap V(B \log \hat{R})|}{|V(A \log R)|}
\]

for arbitrary intervals \( A \) and \( B \); here \( |V| \) is the volume of \( V \). In the example shown in Figure 2, the fraction \( r_s \) of the truly constraint-satisfying area that is predicted to satisfy the constraints is only 78%; if this constraint had to be combined with four other constraints of similar quality, fewer than \( (78\%)^4 \approx 37\% \) of all constraint-satisfying points would lie inside the intersection of all four naive predictions.

To increase the fraction of constraint-satisfying points included in a prediction—in the notation of this paper, to find a volume \( V \) that contains the “naive” prediction \( V(C, \log \hat{R}_q) \subset V \) but contains more of the truly constraint-satisfying points \( V(C, \log R) \)—the simplest option is simply to increase the size of the “constraint” interval. A larger constraint interval \( C^* \) containing the original would automatically include more points in its predicted region [now \( V(C^*, \log \hat{R}) \)]. While this prediction loses some of its reliability (i.e., we cannot ensure that only constraint-satisfying points lie in this region), we increase its robustness: by a good choice of \( C^* \), we can almost guarantee that all constraint-satisfying points are included. Specifically, if we change the interval \( C = \{c_{\text{min}}, c_{\text{max}}\} \) to the broader interval \( C^* = \{c_{\text{min}} - I, c_{\text{max}} + I\} \)—that is, if we increase the constraint region by the size of the characteristic error in the fit—then any point \( x \) that truly satisfies the constraints [in our notation, \( x \in V(C, \log R) \)] almost certainly lies within the larger “predicted” volume \( V(C^*, \log \hat{R}) \). Formally, assuming that the error at any point \( |\delta \log R| \) is likely less than \( I \), we conclude that

\[
\log \hat{R}(x) = \delta \log R(x) + \log R(x) > \delta \log R(x) + c_{\text{min}} > -I + c_{\text{min}},
\]

and similarly that \( \log \hat{R}(x) \leq I + c_{\text{max}} \). In the case shown in Figure 2, this error-widened prediction \( V(C^*(I)), \log \hat{R}_I \) includes both a higher fraction of constraint-satisfying points \( r_s \approx 92\% \) and a higher fraction of points that are not constraint-satisfying \((20\%)\).

All these calculations, however, depend on the fit order: higher fit orders \( q \) often provide more degrees of freedom with which to reproduce the true function \( \log R(x) \) and therefore estimate \( V(C, \log R) \). For example, if we assume arbitrarily small (constant) errors and arbitrarily dense sampling \( x_i \), the best possible \( q \)th-order polynomial least-squares fit is the orthogonal projection \( P_q \log R \) of \( \log R \) onto the space of \( q \)th-order polynomials, given in one dimension by

\[
(P_q \log R(x) = \sum_{l=0}^{q} \phi_l(x) \int_0^1 dy \phi_0(y) \log R(y)
\]

where \( \phi_l(x) \) is any set of orthonormal basis polynomials of degree \( \leq l \) [e.g., \( \phi_k(x) = (2l + 1)^{1/2} P_l(2x - 1) \), where \( P_l(x) \) are Legendre polynomials]. Therefore, the minimum possible error of a \( q \)th-order polynomial fit is the magnitude \( |P_{q} \log R| \) of the difference \( P_{q} \log R \equiv \log R - P_{q} \log R \) between \( \log R \) and this best fit, where \( \|f\| = \int_0^1 dx |f(x)|^2 \). As the fit order increases, this “truncation error”

\[
\epsilon_{r, q} \equiv \|P_{q} \log R\|
\]
decreases dramatically, as shown in Figure 3. However, the total error $I_q$ in the fit reflects a balance between (1) decreasing the truncation error with more degrees of freedom and (2) increasing the sampling error as fewer data points are available per degree of freedom. For this reason, an optimal fit order exists, which provides the smallest error in $\tilde{R}(x)$, indicated by a minimum in $I_q$. This optimal fit order will provide the most reliable estimate of $\tilde{V}(C, \log R)$. In principle, this one-dimensional example contains all the core concepts that we will apply to fitting seven-dimensional population synthesis results: the sampling uncertainty in each sample $n_a$ of simulated merger events; the fit and its inaccuracies; the need to increase the size of the constraint interval and thus the predicted constraint-satisfying region to include a larger fraction of constraint-satisfying events; and a specific procedure for doing so, based on the characteristic errors of the fit. In practice, however, we must substantially flesh out this outline in two ways: (1) we must add a reliable way for estimating the fit error and selecting the optimal fit order without using $I$, since its definition in equation (14) uses the fit itself and requires exact, a priori knowledge of $\tilde{R}(x)$ (the function being fitted) that is not available in practice; and (2) we must extend the concepts developed here to seven dimensions, given the dependence of our population simulations.

### 3.2. Seven-dimensional Analog

By virtue of existing in seven dimensions, the full population synthesis fitting problem is qualitatively different than the one-dimensional problem, even with the same number of points and the same characteristic error. In the one-dimensional case, the many densely packed and relatively accurate samples of $R$ permit very accurate fits to $\log R$; barring rapid variation in $\log R$ with $x$, one-dimensional fits can easily have accuracies that considerably exceed the limiting accuracy of any individual measurement. On the other hand, in seven dimensions the same number of data points are much more loosely spaced (geometrically, the characteristic spacing is proportional to $N^{-1/2}$) and must constrain many more degrees of freedom at the same polynomial order, yielding much less accurate fits. For this reason, rather than attempting to rescale the number and uncertainties of our population synthesis simulations to generate a tractable one-dimensional analog, we instead demonstrate convergence using a seven-dimensional model that best captures the relevant features of our population synthesis data.

Explicitly, in this seven-dimensional toy model, (1) the logarithm of the simulation size $\log N$ is Gaussian-distributed around $-4.5$ with an order-of-magnitude standard deviation, omitting simulations smaller than $10^3$ and larger than $10^7$; (2) each set of simulation parameters $x_a$ is chosen by uniformly selecting seven random numbers $0 \leq x_a^{\text{low}} \leq 1$; (3) the number of mergers $n_a$ observed in any particular simulation parameterized by the seven parameters $x_a$ is drawn from a Poisson distribution with mean $\mu(x_a) = \rho(x_a)N$, where the mean number of mergers per binary $\rho(x)$ is defined by

$$\rho(x) = \mu(x)/N = 10^{-4x} \cdot x^{-7.5}$$

and thus (4) implies a merger rate (again assuming that $f_x = 1$ and $\langle m_2 \rangle/\langle m_1 \rangle = 0.5$) of

$$\tilde{R}(x) = \rho(x)(0.04 \text{ yr}^{-1}).$$

More specifically still, to compare with the array of $O(2000)$ simulations of merging NS-NS binaries, we generate a “model archive” of $N_s = 2000$ randomly realized analogs, each defined by their parameter choices $x_a$, the specific simulation sizes $N_a$, the observed merger numbers $n_a$, and the merger rates $\tilde{R}_a = (0.04 \text{ yr}^{-1})(n_a/N_a)$. This seven-dimensional model qualitatively reproduces the distribution of merging double neutron star [NSNS(m)] simulations in $(n, N)$ and $R$ (Fig. 4). Quantitatively, this toy model is on average as accurate as our population synthesis simulations: the rms relative sampling error $\sigma_f$ (the “expected” standard deviation in the data, given the limiting error produced by sampling at each point),

$$\sigma^2_f = 1/N_s \sum_{n=1}^{N_s} \left[ n_a (\log 10)^2 \right]^{-1},$$

of our population synthesis simulations, $\sigma_f = 0.25$ (see Table 2), agrees with the corresponding value, $\sigma_f = 0.24$ (see Table 1), for this seven-dimensional analog. Furthermore, in both cases this average accuracy $\sigma_f$ is still slightly smaller than the range in $\log R$, as characterized by $\sigma_{\text{DD}}$ (the “data-data” standard deviation, measuring the range of the distribution):

$$\sigma^2_{\text{DD}} = 1/N_s \sum_{n=1}^{N_s} \left[ (\log \tilde{R}_a)^2 - \langle \log \tilde{R} \rangle^2 \right]$$

($\sigma_{\text{DD}} = 0.73$ for real simulations; $\sigma_{\text{DD}} = 0.42$ for our toy model). Finally, as can be seen by comparing $N_s$, the number of population synthesis simulations, to $N_{\text{sim}}$, the number of simulations with at least one event (Table 2), a significant fraction $O(10\%–50\%)$ of population synthesis simulations are totally unresolved; our seven-dimensional toy model has a similar fraction ($20\%$) of unresolved simulations.

### 3.3. Error Estimates and Convergence

Using the concrete and highly realistic seven-dimensional model problem described in § 3.2, we can not only conclusively demonstrate that fairly accurate fits are possible for population...
syntheses, but we can also find a reliable estimate of the rms error in the fit, an estimate that can be reliably applied to surround all constraint-satisfying points inside a predicted volume. Precisely as in the one-dimensional case, given a fixed archive of seven-dimensional simulations, we can estimate log $R$ with log $\hat{R}_q$, the unique polynomial of order $\leq q$ that minimizes $\chi^2_q$, using any integer $q$ (Table 1). With more degrees of freedom, higher order fits can much more easily reproduce the data, so their mean square difference from the data $\sigma_{DF}$,

$$\sigma_{DF}^2 \equiv \frac{1}{N} \sum_{a} [\log \hat{R}(x_a) - \log \hat{R}_a]^2,$$

(a “data-fit” comparison) decreases monotonically; see the fifth column of Table 1. But as these high-order polynomials increasingly fit every sampling-induced fluctuation in the data, the high-order fits increasingly deviate from the exact solution: the exact rms error $I_q$ is also shown in Table 1 (calculated by comparing the fit to the known seven-dimensional model $R(x)$ from eqs. (17) and (18)), showing that $I_q$ has a minimum near an optimal fit order. Near this optimal fit order, the characteristic fit error $I_q$ is substantially less than the characteristic range of the data ($\sigma_{DD}$); in fact, even less than the average statistical error of the sample, $\sigma_E$. In other words, although this fit is accurate to only 20%–30% at any point (based on $10^b - 1$ for $q = 1$–4), this accuracy is sufficient to apply constraints, given the many orders of magnitude in the range spanned by $R(x)$.

But while accurate fits do exist, the tools presented henceforth provide few ways to identify them and, particularly, to estimate their error without resorting to knowledge of the function being fitted. For example, as we increase the available number of degrees of freedom, the rms difference $\sigma_{DF}$ between the fit and the data decreases monotonically to zero. Even at or near the optimal fit order, in our experience, both $\sigma_{DF}$ and $\sigma_E$, the expected sampling-induced rms difference between data and predictions (eq. [19]), correlate only weakly with the true error in the fit ($I_q$). A good fit can be identified\(^6\) using the weighted squared difference $\chi^2_q$, which does have a minimum value for a good fit.\(^7\) However, to understand and estimate the fit-induced error, we must compare fits of successive orders with

$$J_q \equiv \left\{ \int d^7 x \left[ \log(\hat{R}_q/\hat{R}_{q-1}) \right]^2 \right\}^{1/2}.$$

This comparison between fits of different orders arises naturally from understanding how the error in the fit $I_q$ reflects a balance between (1) an increasing ability to fit log $R$ exactly with more basis polynomials and ever smaller truncation error, decreasing the magnitude $\| P_{\perp q} \log R \|$ for $P_{\perp q}$, and (2) a decreasing

---

**TABLE 1**

**BEHAVIOR OF SAMPLE FIT AS ORDER INCREASES**

| $q$ | $\chi^2$ | $J_q$ | $I_q$ | $\sigma_{DF}$ | $\sigma_{DD}$ | $\sigma_E$ |
|-----|----------|-------|-------|---------------|---------------|------------|
| 0   | 60.7     | 0.0   | 0.462 | 0.421         | 0.421         | 0.241      |
| 1   | 4.16     | 0.373 | 0.138 | 0.228         | 0.421         | 0.241      |
| 2   | 1.3      | 0.104 | 0.0813| 0.203         | 0.421         | 0.241      |
| 3   | 1.61     | 0.0475| 0.0929| 0.198         | 0.421         | 0.241      |
| 4   | 2.4      | 0.105 | 0.143 | 0.182         | 0.421         | 0.241      |

\(^6\) The best way to identify a good fit, given sampled data, is with a blind test, where data points not involved in the fitting process are compared against the fit. We have explored using blind test points and calculating $\chi^2_q$, the analogous weighted squared difference between the fit and these test points. In our test cases, $J_q$ provided just as much discriminating power between fits as a blind test.

\(^7\) In fact, $\chi^2_q$ also allows us to identify when the fit is as good as the sampling statistics allow, when $\chi^2_q \simeq 1$. 

---

**FIG. 4.** — Comparison between our toy model (left) and simulations of merging NS-NS binaries (right), demonstrating that our toy models are statistically similar to the output of real population synthesis calculations. Top: Scatter plot of the simulation size $N$ vs. the number of merger events $n$ seen. Bottom: Distribution of merger rates seen in the simulation. In the right panels, no corrections are made for the highly irregular patterns in which $N$ and $n$ were chosen, which introduces unavoidable biases in the distribution of log $R$. 

---
statistical accuracy with fewer independent degrees of freedom \(N_q - N_{\leq q}\) available to constrain the \(N_{\leq q}\) independent coefficients in the expansion. These two terms contribute independently to the error at any fit order, allowing us to estimate \(I_q\) by an incoherent superposition of truncation error and statistical error without explicitly calculating the best-fit \(\log \hat{R}\) itself:

\[
I_q \simeq \left\| P_{\perp q} \log R \right\|^2 + \sigma_E^2 \frac{N_{\leq q}}{N_q - N_{\leq q}},
\]

\[
N_{\leq q} = \frac{(d + q)!}{d! q!}.
\]

Here \(d = 7\) is the dimension in which our functions are defined and \(P_{\perp q}\) projects functions perpendicular to the space of basis polynomials of order less than or equal to \(q\) (cf. eq. [16]): for example, if \(\log R\) is a polynomial, \(P_{\perp q}\) selects those coefficients of \(\log R\) of order less than \(q\). In this and other cases, this expression very accurately reproduces \(I_q\); see, for example, Table 1.

When sampling errors are relatively small (i.e., \(n_s \gg 1\), so \(\sigma_E \sim 0\)), each successive approximation order \(\log \hat{R}_q \simeq P_{\perp q} \log R\) is found by orthogonal projection of \(\log R\); therefore, for low fit orders, the truncation error is just the magnitude of the next highest order correction to be applied, \(J_q\):

\[
\left\| P_{\perp q} \log R \right\| \simeq \left\| \log \hat{R}_{q+1} - \log \hat{R}_q \right\| = J_{q+1}.
\]

In other words, for low \(q\) we expect \(I_q \simeq J_{q+1}\), as is confirmed in Table 1.

From our experience with this and other model problems, the true rms fit error \(I_q\) is quite generally close to the change in the fit to the next order \(J_{q+1}\) and from the previous order \(I_q\). In particular, the difference between the current and next lowest order fit \(I_q\) has a minimum that is similar in magnitude and location to the minimum of \(I_q\); see, for example, Table 1. We therefore use the order at which \(I_q\) is smallest to select the optimal fit order and its value as the characteristic rms error in the fit.

### 3.4. Volume Estimation

This error estimate provides the key tool needed to reliably and algorithmically surround almost all of the points with merger rates \(R\) that are consistent with the constraints, using information only about the fit, to the high level of accuracy needed in order for us to trust the results of multiple constraints. For example, observations of Galactic binary pulsars suggest that double neutron star binaries merge in the Milky Way at a rate between \(c_{\text{min}} = \log(3 \times 10^{-5}\ \text{yr}^{-1})\) and \(c_{\text{max}} = \log(23 \times 10^{-5}\ \text{yr}^{-1})\). This constraint range turns out to lie on the high-end tail of what our model simulations produce (Kim et al. 2003; O'Shaughnessy et al. 2006).

If our seven-dimensional model exactly described double neutron star merger rates, then only a small fraction of model parameters \(x \in V(C, \log R)\) (5% by volume of the unit seven-dimensional cube, corresponding to \(|x| < 1.04\)) reproduce these observations. However, even though these merger rates correspond to the largest possible and therefore best-sampled predictions from population synthesis, the fit remains sufficiently inaccurate that only \(1 - r_s \simeq 18\%\) of all constraint-satisfying points would not be included in a prediction \(V(C, \log R)\) on the basis of the fit alone. But on the basis of Table 1, we can estimate the characteristic error in our fit by \(J \simeq 0.05\). Therefore, when we compensate for this uncertainty with a wider constraint interval \(C^* \equiv C^*(J)\), we much more accurately bound the set of constraint-satisfying points: \(1 - r_s(C^*, \log R) \approx 5\%\).

Strictly, the fraction of constraint-satisfying points that will be encompassed with a larger constraint interval \(C\) depends strongly on the width and placement of the original interval \(C\); intuitively, an extremely narrow constraint requires an extremely high-precision reconstruction of \(R(x)\), which is in turn possible only at high merger rates \(R\), where many events should be seen in the simulation. However, the case shown here, with a very tight constraint (with only 5% by volume of parameters satisfying the constraint), is substantially more difficult to satisfy than most individual observational constraints that can be applied in practice. Comparing our experience with model problems to the weak constraints available, we are confident that fewer than 5% of constraint-satisfying systems will be omitted from \(V(C, \log R)\) when these volumes are constructed on the basis of real population synthesis simulations and corresponding observations. (The results of our broader array of model problems are available online in O’Shaughnessy et al. [2007].)

### 4. Fitting: Population Synthesis Predictions

Almost exactly the same fitting techniques presented above are applied to our population synthesis data: we perform a weighted least-squares fit of polynomial-like basis functions over our seven-dimensional space for each of the event rates of interest [BBHBM(h), BHNS(m), NSNS(m), NSNS(vv), WDNS(e), WDNS(m), SN Ib/c, and SN II, as introduced in § 2.3]. For each fit, we evaluate the fit quality of several different polynomial orders to our data. To minimize the possibility of using more parameters than are allowed by the number of our data points, we choose as the “best” fit the order that minimizes the relative difference between fit orders \(I_q\) (see below). However, we know that the merger rate \(\hat{R}(x)\) must satisfy certain symmetries, based on the manner in which we represent the kick velocity dispersion as a sum of two Maxwellians (i.e., we can switch labels associated with the two distributions); therefore, we limit the manner in which those parameters associated with kicks can enter into the distribution (see Appendix B).

Table 2 summarizes the properties of the least-squares fits that we applied to our archived population synthesis results. The first column provides a brief label for the fit, as described in greater detail in the text. The next two columns summarize the amount of available information contained in our population synthesis archive: \(N_q\) is the number of population synthesis models with unbiased data (i.e., where all plausible progenitors for the target event have been included), whereas \(N_{\geq q}\) is a smaller number of models with unbiased data containing one or more events (i.e., for which an estimate of the rate, rather than merely an upper bound, is possible). The next block of five columns describes properties and diagnostics of a weighted least-squares fit applied to our data. The first two columns, those for \(N_{\leq q}\) and \(q\), merely indicate the number of degrees of freedom and the polynomial order involved in our fit \((N_{\leq q}\) differs from the \(N_{\leq q}\) introduced earlier due to the symmetry requirement described above). In all the cases shown here, the optimal polynomial order produces far fewer degrees of freedom than the population synthesis simulations (i.e., \(N_{\leq q} \ll N_{\geq q}\)).

The next three columns provide critical diagnostics of our fit: \(\chi^2, \sigma_{\text{DF}},\) and \(J_q\) (see eqs. [11], [19], and [22]). All three columns roughly measure our “goodness of fit”; based on our experience with model problems and our understanding of the values these quantities should take for fits dominated by statistical and truncation errors, all three universally indicate that truncation error dominates; that is, that our low-order polynomial basis functions
are not sufficiently general models to match the exact merger rate $R(x)$ implied by our population synthesis simulations. For example, the limiting value of $J_q$ is significantly above the level of sampling error $\sigma_E/N_s/N_{\bar{S}}$ suggested by the second term in equation (23); that is, above the level of error expected when averaging $N_s/N_{\bar{S}}$ simulation samples per degree of freedom, each with characteristic error $\sim \sigma_E$. This high level of error suggests that the fit accuracy is limited by truncation error—the inability to fit the exact form of $R(x)$ with our basis polynomials—rather than sampling uncertainties at each point.\(^8\) In addition, as illustrated in our discussion of seven-dimensional model problems, the latter two columns provide the logarithmic uncertainty in our fits; for example, fits with $J \geq 0.33$ are known to within a factor of 2 with 1 $\sigma$ confidence. These columns should be contrasted with the next two, which provide the average sampling-induced uncertainty $\sigma_E$ in the data (eq. [19]) and the characteristic range of merger rates $\sigma_{DD}$ seen in the simulation (eq. [20]). Our fits remain useful so long as their uncertainties are much less than the range of log $R$ (i.e., so long as $J \leq \sigma_{DD}$). As seen in Table 2, our best fit satisfies this requirement ($J \leq \sigma_{DD}$) for each population of interest.

The last column provides the fraction $\eta$ of simulations with no events when more than 10 would be expected on the basis of the fit. Based on an average uncertainty of a characteristic factor of 2 in the fit (our best fits have $J \simeq 0.33$, corresponding to a merger rate known to within a factor of $2 \simeq 10^2$), we would estimate that the fraction of simulations $\eta$ that produce no merger events when more than 10 should be seen on the basis of our uncertain estimate of the event rate should be comparable to or smaller than the fraction of simulations that should produce $10/2 = 5$ events but in fact produce zero, namely, $\eta \leq \rho(0, 10/2) \simeq 0.7\%$ (eq. [6]). However, regions with the lowest merger rates will be undersampled and therefore have characteristic uncertainties marginally larger (e.g., up to a factor of 3–4), leading to values for $\eta$ of a few to 10%.

4.1. Results

Supernovae.—Given our choice of the stellar mass interval that we probed in our simulation (i.e., $m_1 > 4 M_\odot$), supernovae occur extremely frequently, providing us with superb statistics at low cost. However, our limited set of basis functions can only with difficulty reproduce the observed variation in supernova rates: even though SN rates for models in our archive are at times known to 1% (i.e., involving $10^4$ or more sampled events), or 0.004 in the logarithm, our optimal fit differs significantly from the data, by $\sigma \simeq O(0.04)$ in the logarithm.

Nonetheless, as is discussed in our forthcoming paper (O’Shaughnessy et al. 2006), the supernova rate remains a striking success of the StarTrack population synthesis code and our normalization conventions (e.g., $M \simeq 3.5 M_\odot$ yr$^{-1}$). No matter what combination of population synthesis parameters we choose, the predicted SN rates lie well within the observational constraints found by Cappellaro et al. (1999).

WD-NS binaries.—As with supernovae, white dwarf–neutron star binaries occur fairly frequently, allowing us to accumulate fairly good statistics over a broad range of population synthesis parameters. In addition, based on the distribution of $N_s$ versus $n$ (i.e., as in Fig. 5), our sample shows no signs of systematic incompleteness; we appear to have covered the full range. Although our polynomial fits continue to introduce systematic errors, the resulting fit behaves well throughout the range.

BH-BH binaries.—Double black hole binaries, in contrast, occur extremely infrequently, especially with an assumed maximum NS mass of $2.5 M_\odot$, instead of $2.0 M_\odot$, as assumed in O’Shaughnessy et al. (2005a, 2005b). (The inefficient formation of coalescing BH-BH binaries is discussed and explained in more detail by Belczynski et al. [2006b].) Nonetheless, by using special-purpose partitions, O’Shaughnessy et al. (2005a) accumulated a large ($\simeq 500$ simulations) sample with good ($n \simeq 10$) statistics on a large subset of parameter space; although these simulations assumed a maximum NS mass of $2 M_\odot$, we post facto changed the maximum NS mass to $2.5 M_\odot$. By adjoining the results of general-purpose simulations not assured of good statistics, this sample has since been enlarged by a factor of $\simeq 5$, with emphasis on the same subset of parameter space presented in O’Shaughnessy et al. (2005a) (see Fig. 1). Finally, although the distribution of $n$ versus $N_s$ (Fig. 5) suggests that the lowest BH-BH merger rates may not be very well resolved, we have no reason to suspect that we have any significant underresolved region: simulations larger than $10^6$ binaries consistently produce several merger events ($n \gg 0$).

Nonetheless, the BHBHM(m) fit is comparatively poor: even though the average simulation with binary black holes produces enough to fairly accurately determine the rate (i.e., $\sigma_E \simeq 0.25$), our best fit is barely more accurate than approximating the average BH-BH merger rate seen in simulations (i.e., compare the characteristic fit error $J_q \simeq 0.67$ with the range of BH-BH merger rates seen in simulations, $\sigma_{DD} \simeq 0.78$).

NS-NS and BH-NS binaries.— Despite (or perhaps because of) a concerted effort to accumulate good statistics targeted specifically to these classes, fits for NS-NS and BH-NS event rates are statistically implausible as measured by $\chi^2$. Judging from Figure 5, the NSNS(m) and BHNS(m) are comparatively well

---

**TABLE 2**

| Type       | $N_s$ | $N_{\bar{S}}$ | $\bar{N}_{\bar{S}}$ | $q$ | $\chi^2$ | $\sigma_{DD}$ | $J_q$ | $\sigma_E$ | $\sigma_{DD}$ | $\eta(10)$ |
|------------|-------|---------------|----------------------|----|---------|---------------|------|------------|---------------|----------|
| BHBHM(m)   | 2930  | 1201          | 141                  | 3  | 15.5    | 0.77          | 0.67 | 0.25       | 0.78          | 12       |
| BHNS(m)    | 2533  | 1334          | 141                  | 3  | 11.1    | 0.53          | 0.40 | 0.23       | 0.71          | 2        |
| NSNS(m)    | 2803  | 2382          | 141                  | 3  | 18.4    | 0.37          | 0.22 | 0.14       | 0.63          | 5        |
| NSNS(vw)   | 1325  | 1087          | 141                  | 3  | 16.7    | 0.39          | 0.37 | 0.13       | 0.76          | 8        |
| WDNS(c)    | 1770  | 1564          | 141                  | 3  | 12.5    | 0.34          | 0.20 | 0.14       | 0.57          | 10       |
| WDNS(m)    | 1770  | 1658          | 141                  | 3  | 16.4    | 0.34          | 0.19 | 0.13       | 0.45          | 11       |
| SN Ib/c    | 1482  | 1482          | 141                  | 3  | 7.8     | 0.07          | 0.06 | 0.02       | 0.11          | N/A      |
| SN II.......| 1482  | 1482          | 141                  | 3  | 6.0     | 0.04          | 0.02 | 0.03       | 0.17          | N/A      |

---

\(^8\) In addition, when fitting supernova rates as a function of population synthesis parameters for single stars, where only one of our parameters (wind strength) enters, we find that the rate has a moderately complex functional form that requires high-degree polynomials to fit. We expect similarly complex behavior in the multidimensional case and interpret the large $\chi^2$ correspondingly.
resolved: longer simulations fairly consistently have a lower chance of producing \( n = 0 \) results. For this reason, we strongly suspect that some feature of these underlying rate functions is poorly described by our basis functions; we intend to more thoroughly test this hypothesis (with better statistics) by comparing these fits to non-parametric estimates in a future paper.

5. CONCLUSIONS

To develop a more comprehensive understanding of population synthesis predictions and to allow those theoretical predictions to be systematically compared with observations of the end products of high-mass single and binary stellar evolution, we have fitted eight predictions from the StarTrack code over seven of its input parameters. These fits are available on request to the first author. In a companion paper, O’Shaughnessy et al. (2006), we apply these fits along with estimates of their systematic errors to discover robust constraints on the seven parameters that enter into population synthesis. In addition, we have demonstrated that in analogous model problems, the constraint-satisfying region defined by using these fits can, under appropriate conditions, very accurately trace the underlying constraint-satisfying region. Finally, we have presented a thorough diagnostic formalism, including a large list of diagnostic quantities and tests (\( I, J_q, \sigma_E, \sigma_{DD}, \chi^2, \eta, r_+ \), etc.), that can be applied to studies of fits to large archives of any (individual) population synthesis code.

We thank the NCSA for providing us with the resources used to perform many of the computations in this text. This work is partially supported by NSF Gravitational Physics grant PHYS-0353111, a David and Lucile Packard Foundation Fellowship in Science and Engineering, and a Cottrell Scholar Award from the Research Corporation to V. K. K. B. acknowledges support from the Polish Science Foundation (KBN) grant 1P03D02228. V. K. is grateful for the hospitality and partial financial support of the KITP, UCSB for the period of 2006 June and July.

APPENDIX A

SAMPLING, PRIOR DISTRIBUTIONS, AND THE LIKELIHOOD OF RESULTS

In Figure 6 we compare the distribution of merger rates derived directly from our current database of simulations (dashed lines) to the distribution of merger rates derived from the corresponding fit (solid lines) obtained using the methodology described here. Unlike Figure 3 of O’Shaughnessy et al. (2005b), in which the simulations being fitted had parameters drawn randomly from the entire range of parameters allowed, in this study the simulations were not randomly distributed through the entire parameter range; see, for example, Figure 1. In effect, the rate distributions drawn from the simulations and the fits assume significantly different priors for population synthesis model parameters, the former favoring lower kicks. Therefore, although the distributions appear different in shape, they reflect the same physical process, just with different priors regarding which model parameters are likely. In particular, the fitted rate results represent the unbiased distributions of rates with uniform coverage of the seven-dimensional model parameter space. Furthermore, these distributions fully account for uncertainties in the fits, as described in § 4.

To phrase the same statement more abstractly, our prior expectations about the relative likelihood of population synthesis parameters (e.g., kicks) influence our expectations for the relative likelihood of different merger rates. If at any point our knowledge of binary properties, evolution, and compact object kicks improves to a degree such that we can confidently move away from flat priors into favoring certain more specific priors for the model parameters, then the shape of the probability distributions derived from fits will change accordingly, expressing the influence of the adopted priors.
APPENDIX B

POLYNOMIAL BASIS FUNCTIONS

To parameterize supernova kick distributions, StarTrack employs three parameters, $\sigma_1$, $\sigma_2$, and $s$, that represent the superposition of two Maxwellian kick distributions with probabilities $s$ and $1 - s$. The physical predictions associated with $(\sigma_1, \sigma_2, s)$ are therefore identical to those of $(\sigma_2, \sigma_1, 1 - s)$. To improve the physical significance of our fit, we have chosen to employ basis polynomials that enforce this requirement to all orders.

Specifically, rather than allowing for homogeneous basis polynomials in these parameters, we use the following, for arbitrary $p$ and $q$:  

$$\sigma_1^p s^q + \sigma_2^q (1 - s)^q$$  

(B1)  

$$2\sigma_1^p \sigma_2^p s (1 - s) + \sigma_1^{2p} s^2 + \sigma_2^{2p} (1 - s)^2.$$  

(B2)

Because $s$ must enter in a heterogeneous manner to preserve our desired symmetry, these basis polynomials are of a fixed order in all kick parameters. For the purposes of order counting when constructing fits, the first polynomial is denoted to be of order $p + q$, and the second is of order $p$.

REFERENCES

Abbott, B., et al. 2005a, Phys. Rev. D, 72, 082001  
———. 2005b, Phys. Rev. D, 72, 082002

Arzoumanian, Z., Cordes, J. M., & Wasserman, I. 1999, ApJ, 520, 696

Belczynski, K., Kalogera, V., & Bulik, T. 2002, ApJ, 572, 407

Belczynski, K., Kalogera, V., Rasio, F., Taam, R., Zezas, A., Bulik, T., Maccarone, T. J., & Ivanova, N. 2007a, ApJS, in press (astro-ph/0511811)

Belczynski, K., Taam, R. E., Kalogera, V., Rasio, F. A., & Bulik, T. 2007b, ApJ, 662, 504

Cappellaro, E., Evans, R., & Turatto, M. 1999, A&A, 351, 459

Fryer, C., Burrows, A., & Benz, W. 1998, ApJ, 496, 333

Hobbs, G., Lorimer, D. R., Lyne, A. G., & Kramer, M. 2005, MNRAS, 360, 974

Hurley, J. R., Tout, C. A., & Pols, O. R. 2002, MNRAS, 329, 897

Kalogera, V., Kim, C., Lorimer, D. R., Ihm, M., & Belczynski, K. 2005, in ASP Conf. Ser. 328, Binary Radio Pulsars, ed. F. A. Rasio & I. H. Stairs (San Francisco: ASP), 261

Kim, C., Kalogera, V., & Lorimer, D. R. 2003, ApJ, 584, 985

Kim, C., Kalogera, V., Lorimer, D. R., & White, T. 2004, ApJ, 616, 1109

Lacey, C. G., & Fall, S. M. 1985, ApJ, 290, 154

Lorimer, D. R. 2005, Living Rev. Relativ., 8, 7

Nice, D. J., Splaver, E. M., Stairs, I. H., Löhrner, O., Jessner, A., Kramer, M., & Cordes, J. M. 2005, ApJ, 634, 1242

O'Shaughnessy, R., Kalogera, V., & Belczynski, K. 2005a, ApJ, 620, 385

———. 2007, preprint (astro-ph/0609465v1)

O'Shaughnessy, R., Kim, C., Fragos, T., Kalogera, V., & Belczynski, K. 2005b, ApJ, 633, 1076

O'Shaughnessy, R., Kim, C., Kalogera, V., & Belczynski, K. 2006, ApJ, submitted (astro-ph/0610076)

Portegies Zwart, S. F., & Verbunt, F. 1996, A&A, 309, 179

Rana, N. C. 1991, ARA&A, 29, 129

Splaver, E. M., Nice, D. J., Arzoumanian, Z., Camilo, F., Lyne, A. G., & Stairs, I. H. 2002, ApJ, 581, 509