

Incorporation of model accuracy in gravitational wave Bayesian inference

In the format provided by the
authors and unedited

Incorporation of model accuracy in gravitational-wave Bayesian inference - Supplementary Material

Charlie Hoy^{1†}, Sarp Akçay², Jake Mac Uilliam² and Jonathan E. Thompson^{3,4}

¹Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth, PO1 3FX, UK ²School of Mathematics & Statistics, University College Dublin, Dublin, D4, Ireland ³Theoretical Astrophysics Group, California Institute of Technology, Pasadena, CA, 91125, U.S.A ⁴Mathematical Sciences & STAG Research Centre, University of Southampton, Southampton, SO17 1BJ, UK. † e-mail: charlie.hoy@port.ac.uk

To demonstrate the improved accuracy of the NR informed approach compared to Evidence informed and Standard algorithms, we compared the posterior distributions obtained with each method against those obtained with NRSUR7DQ4; the leading generic spin numerical relativity surrogate model. We only considered the `SXS:BBH:0926` numerical relativity simulation since it was the main focus of this work.

In Supplementary Figure 1 we see that our NR informed approach infers a posterior that significantly overlaps with NRSUR7DQ4. For the inference of the binary masses we see that NR informed has support for slightly larger primary masses, but the secondary mass remains comparable with NRSUR7DQ4. Similarly, when considering the inferred component spins, we see that our NR informed approach infers a comparable marginalized one-dimensional posterior for the effective perpendicular spin, with slightly more support for larger effective parallel spin. Importantly, we see that the Evidence informed and Standard approaches fail to capture the true value. Specifically, the Evidence informed result has low overlap with the NRSUR7DQ4 result and the Standard result unnecessarily inflates the uncertainty for the effective perpendicular spin. Interestingly, while both our NR informed approach and NRSUR7DQ4 capture the true effective spins of the binary with their two-dimensional 90% credible intervals, the majority of posterior support is to lower effective perpendicular and larger effective parallel spins. In general, we found that our NR informed approach obtained the most statistically similar one-dimensional posterior probability distributions to the surrogate out of the methods considered in this work.

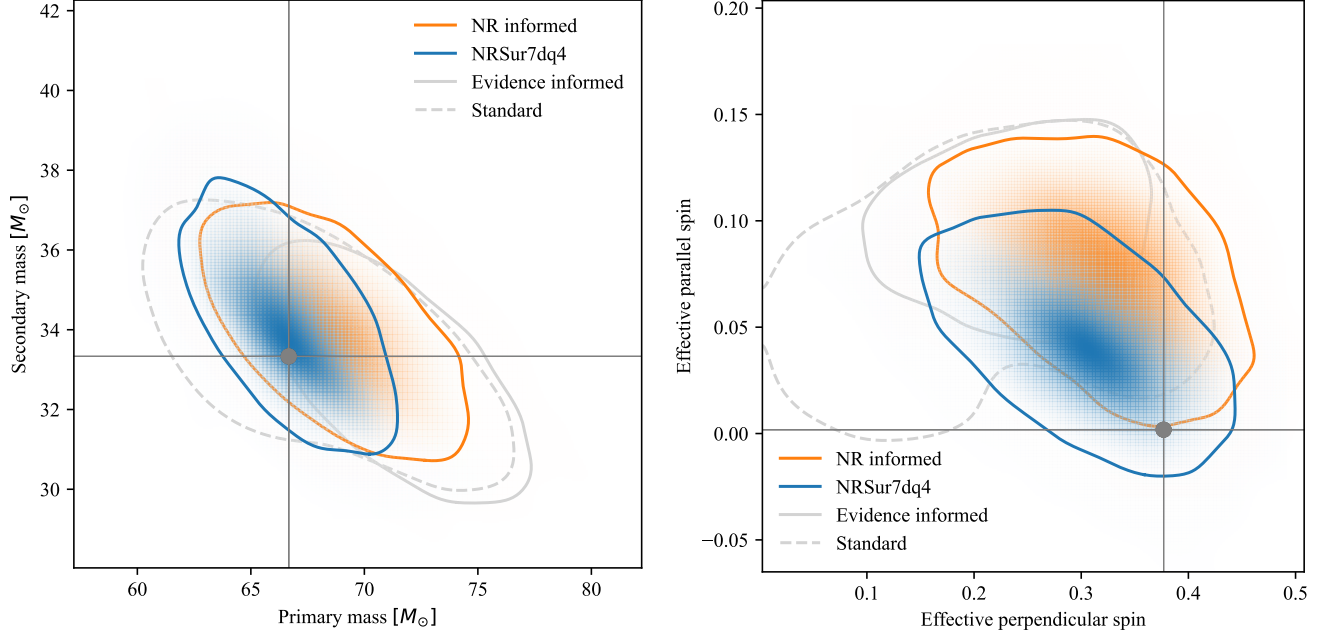
We additionally analysed the `SXS:BBH:0143`^{1,2} and `SXS:BBH:1156`^{1,2} numerical relativity simulations produced by the SXS collaboration with our NR informed approach, as well as the Evidence informed and Standard algorithms. For the case of `SXS:BBH:0143`, we found largely overlapping posteriors between methods, with most of the one-dimensional marginalized 90% confidence intervals containing the true value. For the case of `SXS:BBH:1156`, we found that our method outperforms the Standard and Evidence informed analyses despite being in the extrapolation regime of our technique. Here, we provide further details.

In Supplementary Figure 2 we show the results obtained with our method, NR informed, vs. the Evidence informed and Standard analyses for the parameter recovery of the `SXS:BBH:0143` numerical relativity simulation. We see that in general, all methods give largely overlapping posteriors. When inspecting the one-dimensional marginalized posterior distributions, provided in the associated data release, the biggest disagreement is seen in the inferred secondary

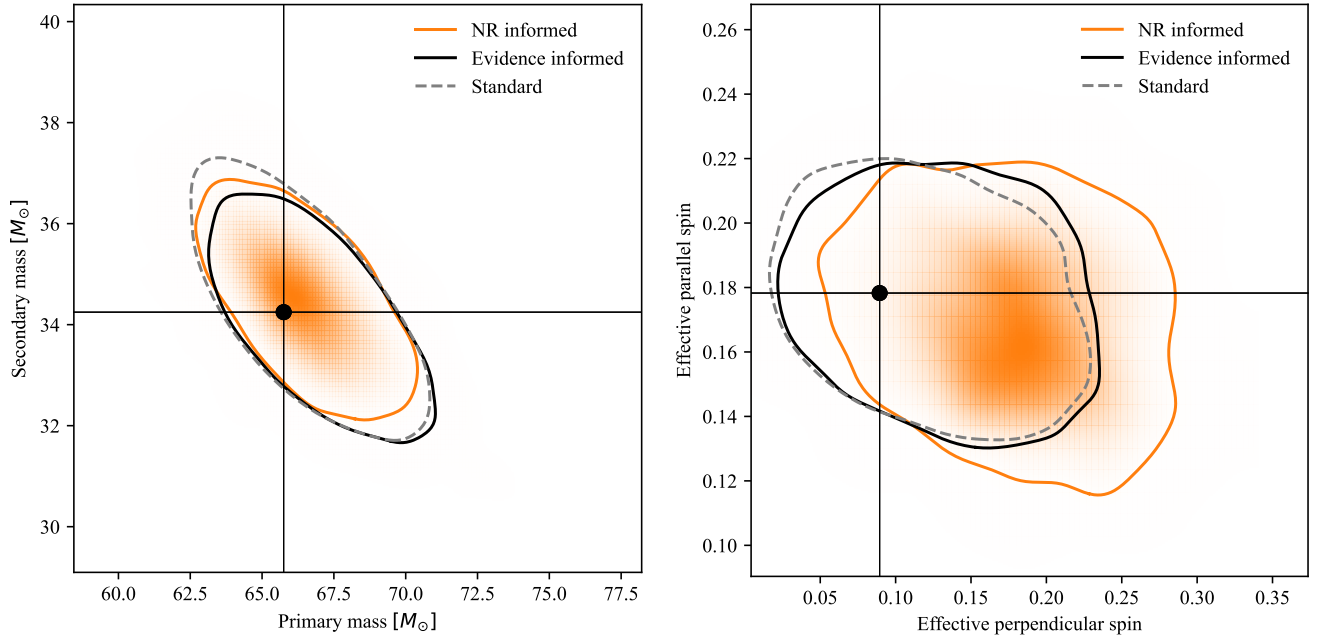
spin magnitude of the binary: our method prefers a rapidly spinning black hole, while the other methods are largely uninformative with a spin ~ 0.25 . The reason is because our method primarily chooses the SEOBNRv5PHM model, while the Evidence informed result is primarily using IMRPHENOMTPHM. In this region of the parameter space we find that SEOBNRv5PHM is the most accurate model, hence why it is favoured in our method: we find that SEOBNRv5PHM is $\sim 1.4\times$ more accurate than IMRPHENOMTPHM and $\sim 1.8\times$ more accurate than IMRPHENOMXPHM. Given that the individual spin components are difficult to measure at present-day detector sensitivities³, it is not surprising that the biggest difference between methods is seen for the secondary spin. Since the secondary spin contribution is also suppressed for asymmetric mass ratio binaries, it is possible that the mismatch interpolant is missing information, leading to poorly chosen models in this region of the parameter space. A mismatch interpolant that is a function of the individual spin components may solve this problem, but it would require a significantly larger mismatch dataset. We leave this to future work. It could also be independent of the interpolant: the interpolant could be accurately describing the mismatch, but SEOBNRv5PHM may not be accurate enough to avoid parameter biases, despite being the most accurate model of those considered in this region of the parameter space.

In Supplementary Figure 3 we compare results for the `SXS:BBH:1156` numerical relativity simulation. Although all methods infer the injected value within 90% credible interval, we see that our method outperforms the others, especially seen for the high total mass ($M = 100 M_{\odot}$) injection: only our method contains the injected spin within the 40% confidence interval. When inspecting the one-dimensional marginalized posterior distributions, provided in the associated data release, we consistently see a more accurate inference of the binary parameters with our method. As with `SXS:BBH:0143`, we see the worst performance for the inferred spin of the smaller black hole in the binary. Although the injected value is contained within the 90% confidence interval, the Evidence informed and Standard analyses capture the true value more accurately. For the low total mass ($M = 75 M_{\odot}$) injection, we observe a better performance by all methods. We argue that this is because the GW produced from a lower mass system spends more time within the sensitive region of the GW detectors, meaning that all GW models better capture the true source parameters of the binary, as long as the primary is not strongly precessing.

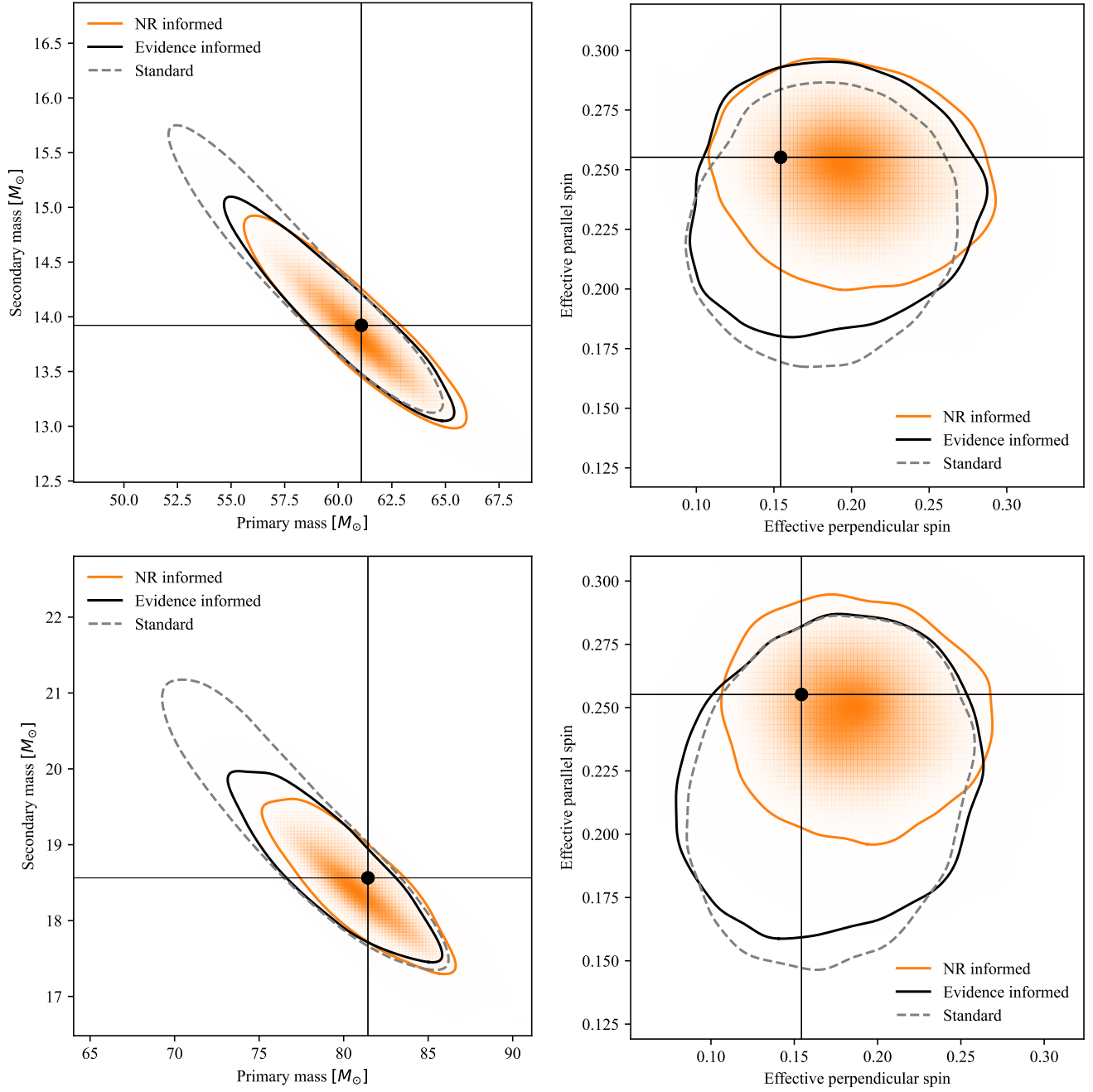
1. Blackman, J. *et al.* Numerical relativity waveform surrogate model for generically precessing binary black hole mergers. *Phys. Rev. D* **96**, 024058 (2017). 1705.07089.
2. Boyle, M. *et al.* The SXS Collaboration catalog of binary black hole simulations. *Class. Quant. Grav.* **36**, 195006 (2019). 1904.04831.
3. Pürrer, M., Hannam, M. & Ohme, F. Can we measure individual black-hole spins from gravitational-wave observations? *Phys. Rev. D* **93**, 084042 (2016). 1512.04955.



Supplementary Figure 1 | Two-dimensional posterior probabilities obtained in our analysis of the SXS:BBH:0926 numerical relativity simulation. The left panel shows the measurement of the primary and secondary mass of the binary, and the right panel shows the inferred effective parallel and perpendicular spin components (as defined in equations 11,12). The contours represent 90% credible intervals and the grey cross hairs indicate the true value. The Evidence-informed and Standard analyses are set to a lighter color than in other Figures to highlight the comparison between NR informed and NRSUR7DQ4.



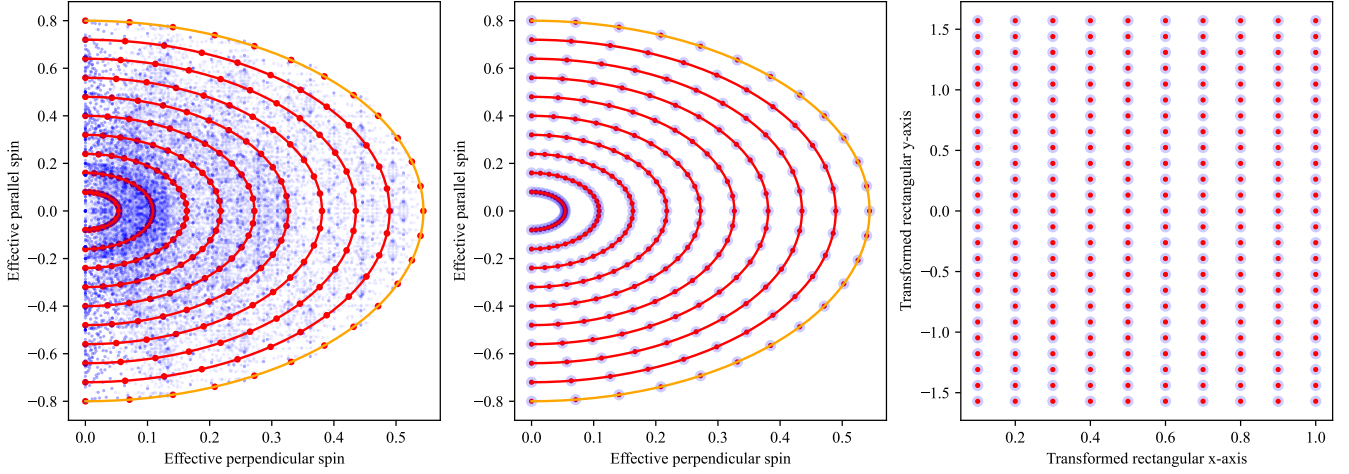
Supplementary Figure 2 | Two-dimensional posterior probabilities obtained in our analysis of the SXS:BBH:0143 numerical relativity simulation. The left panel shows the measurement of the primary and secondary mass of the binary, and the right panel shows the inferred effective parallel and perpendicular spin components (as defined in equations 11,12). The contours represent 90% credible intervals and the black cross hairs indicate the true value.



Supplementary Figure 3 | Supplementary Figure 3: Two-dimensional posterior probabilities obtained in our analysis of the SXS:BBH:1156 numerical relativity simulation. The left column shows the measurement of the primary and secondary mass of the binary, and the right column shows the inferred effective parallel and perpendicular spin components (as defined in equations (11,12)). The top row shows a low total mass injection of $M = 75 M_{\odot}$ and the bottom row shows a high total mass injection of $M = 100 M_{\odot}$. The contours represent 90% credible intervals and the black cross hairs indicate the true value.

Primary spin, χ_1/m_1	0.45	mbits
Secondary spin, χ_2/m_2	0.51	mbits
Mass ratio, $q = m_2/m_1$	0.89	mbits
Phase, ϕ	2.1	mbits
Polarization, Ψ	0.82	mbits
Right ascension α	0.99	mbits
Declination δ	1.3	mbits

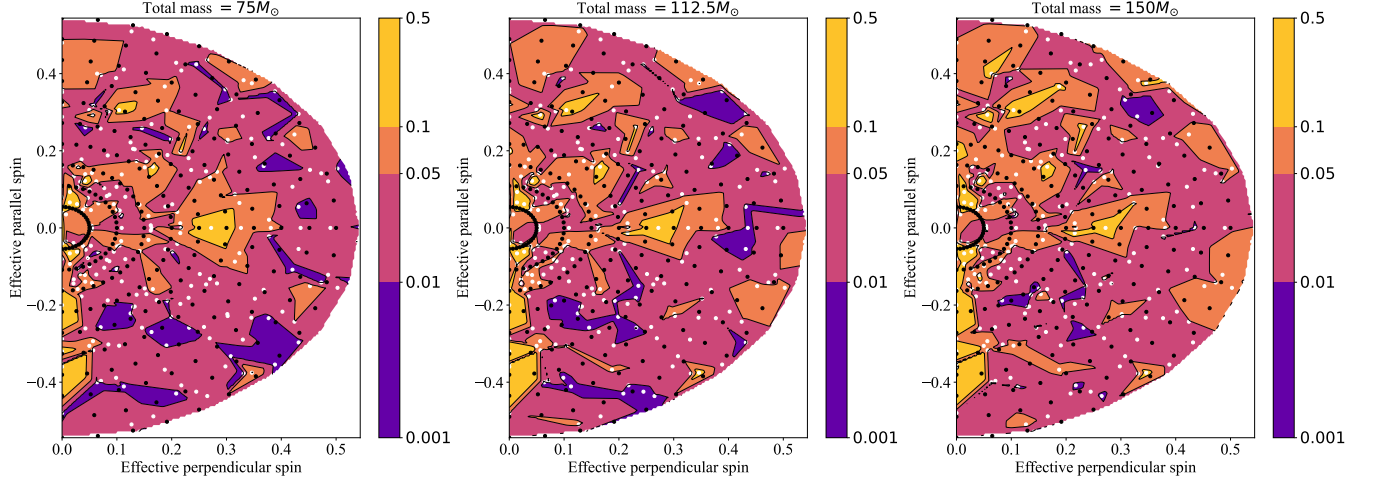
Supplementary Table 1 | Comparison of Jensen-Shannon divergences when using the true mismatch. Jensen-Shannon divergences between posteriors obtained when using true mismatch and the mismatch interpolant for an aligned-spin injection and recovery. Column one lists the parameters for which we display the Jensen-Shannon divergences of the posteriors in column two. These Jensen-Shannon divergences are reported in the base 2 logarithm and reported in millibits (mbits).



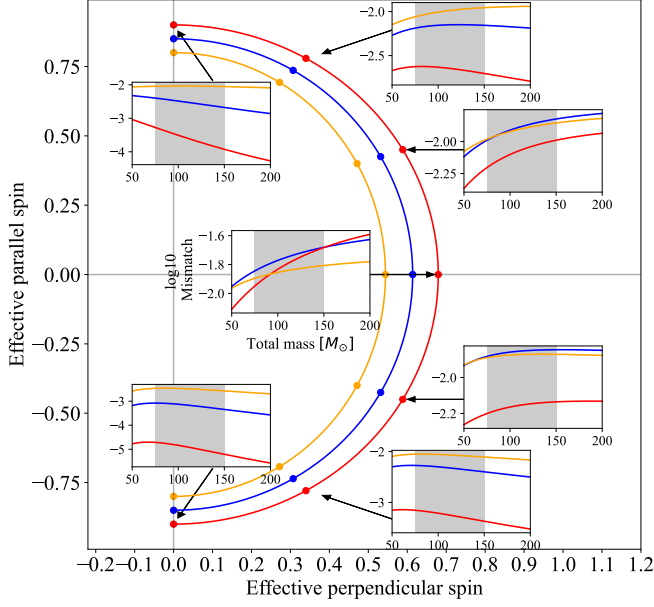
Supplementary Figure 4 | Construction of the two dimensional elliptical and rectangular grids. We show the construction of the two dimensional elliptical and rectangular grids for the $\{x, y\} = \{\text{Effective perpendicular spin}, \chi_{\perp}, \text{Effective spin}, \chi_{\text{eff}}\}$ pairing. The blue dots in the left panel represent the ~ 50000 points obtained from a regularly spaced grid in the intrinsic parameter space $\{q, \bar{S}_1, \bar{S}_2, \theta_1, \theta_2, \Delta\phi\}$. The orange curve is the outermost prolate ellipse whose axes are determined via equation (13). The red curves are concentric ellipses retaining the same aspect ratio as the orange one. These are parametrized by r, s in equations (16a, 16b). The red and orange dots situate the elements of the 10×25 elliptical grid where the azimuthal coordinate is sampled in steps of $\pi/24$ from $-\pi/2$ to $\pi/2$. Though this grid size is somewhat arbitrary, it is limited by computational resources such as the runtimes to generate the mismatches (see Sec. 4.5) and to construct the fits. The middle panel shows the same elliptical grid of the left panel overlaid, in faint blue, with the actual grid points that we determine by finding the roots of equations (17a, 17b). On average, each blue dot is offset by 10^{-12} from the nearest, exactly positioned red dot. Finally, in the right panel, we show the corresponding rectangular grid. The red dots are exact, hence simply given by the transformed rectangular coordinates $X = 0.1, 0.2, \dots, 1, Y = -\pi/2, -\pi/2 + \pi/24, \dots, \pi/2$ and the blue dots are mapped from the blue dots of the middle panel via equations (19a, 19b).

Model	y	$\{n_i, n_j, n_k\}$	# Params	$\Delta_{\text{rel}}^{(1), \text{ver}}$	$\Delta_{\text{rel}}^{(2), \text{ver}}$	$1 - \tilde{R}^2$	χ^2/DoF
SEOBNRv5PHM	χ_{eff}	$\{4, 3, 2\}$	220	0.0032	-0.013	0.0049	0.050
SEOBNRv5PHM	$\chi_{\parallel}$	$\{5, 3, 2\}$	264	0.0026	0.0041	0.0052	1.3
IMRPHENOMTPHM	χ_{eff}	$\{5, 3, 2\}$	264	0.0039	-0.012	0.012	0.99
IMRPHENOMTPHM	$\chi_{\parallel}$	$\{4, 3, 2\}$	220	0.0039	-0.0090	0.0077	0.076
IMRPHENOMXPHM-ST	χ_{eff}	$\{4, 3, 2\}$	220	0.0038	-0.019	0.010	1.2
IMRPHENOMXPHM-ST	$\chi_{\parallel}$	$\{4, 3, 2\}$	220	0.0032	-0.0021	0.0078	0.74

Supplementary Table 2 | Metrics for the mismatch fitting function. Relevant metrics for the fits to the $\log_{10}$ of the mismatches between NRSUR7DQ4 and the models listed in the first column. The fitting function is given in equation (20) with the transformation from the $\{X, Y\}$ coordinates to the elliptical $\{x, y\}$ done via equations (19a, 19b). Column two lists our choice for y representing whether we used χ_{eff} or $\chi_{\parallel}$ when constructing the elliptical fit training grid, *e.g.*, shown for χ_{eff} in Supplementary Figure 4. Column three lists the upper limits of the triple summation used in equation (20) which then determines the total number of fitting parameters c_{ijkl} displayed in column four. Columns five and six present the values for the two relative difference measures of equations (22, 23) applied to the verification sets. Finally, column seven lists $1 - \tilde{R}^2$, where $\tilde{R}^2$ is the reduced R square and column eight the chi square over the degrees of freedom. Overall, we see that the fits based on the $y = \chi_{\parallel}$ grid are slightly more faithful to the verification data and $\tilde{R}^2 \gtrsim 0.99$ for all fits.



Supplementary Figure 5 | Fit unfaithfulness over the verification set. We plot the absolute value of the relative difference, *i.e.*, Δ_{rel}^k [see equation (21)], between the data for the $\log_{10}$ of the NRSUR7DQ4-SEOBNRV5PHM mismatches and the fit (20) to it with equations (19a, 19b) substituted. The panels from left to right correspond to total masses of $75M_\odot$, $112.5M_\odot$, $150M_\odot$. The effective perpendicular, $\chi_\perp$, and effective parallel spins, $\chi_\parallel$, are defined in equations (11, 12), respectively. The white dots mark the verification set data points while the black dots mark the training set points. The color scale is logarithmic. Overall, we observe an absolute relative difference of less than 0.05 for most of the parameter space.



Supplementary Figure 6 | Extrapolation of the fits. We show the performance of the fit to the $\log_{10}$ of the NRSUR7DQ4-SEOBNRv5PHM mismatches extrapolated to $\{q, \bar{S}_{1,2}\} = \{1/5, 0.85\}$ (blue) and $\{1/6, 0.9\}$ (red) as well as the fit at the edge of its training region with $\{q, \bar{S}_{1,2}\} = \{1/4, 0.8\}$ (orange). The red, blue, orange dots accordingly mark the positions of seven cases along each corresponding ellipse with the above $\{q, \bar{S}_{1,2}\}$ values and the remaining intrinsic parameters chosen such that the dots trace each ellipse in angular steps of $\pi/6$. The smaller orange ellipse was shown previously in Supplementary Figure 4. Pointing to each $\Phi = \text{constant}$ dot is an inset showing the plot of the fit from $M = 50M_\odot$ to $200M_\odot$, but at each separate elliptical coordinate. The shaded gray region in each inset has been placed to remind the reader our training range of $M \in [75, 150]M_\odot$. From the insets, we see that the extrapolation does not appear to be pathological and the blue curves mostly lay between the red and orange ones as expected.