Reply to the Referee report on DQ12326/Marsat

We would like to thank the anonymous referee for this report and apologize for the delay in implementing the corrections. We describe below the steps taken to address the referee's remarks, as well as other minor corrections.

- *For the supermassive BH signals, the inferences in this paper use a waveform model with no spins. (Described in sec III B). This very significant simplification; it greatly simplifies the modelling (e.g. the ROM only interpolates in 1D) and may obscure some parameter degeneracies. I ask the authors to consider emphasising this point more prominently in the introduction and/or abstract. The other key simplification (namely the accelerated likelihood calculation of a sparse grid of frequencies) is already clearly pointed out in the abstract.*

We emphasized more this point in the abstract and introduction. It should be noted that the inclusion of aligned spins, although important quantitatively (especially for merger-dominated signals), is not expected to change the qualitative picture of degeneracies in the set of extrinsic parameters (as we saw in further studies), due to the weak correlations between intrinsic and extrinsic parameters. Misaligned spins causing precession, on the other hand, could change qualitatively the inference and also affect extrinsic parameters, since they change the mode content of the waveform and its structure. To our knowledge, this remains to be explored. We added a few sentences in the introduction to reflect this.

Edit in the abstract: *We present the first simulations of Bayesian inference for the parameters of massive black hole systems including consistently the merger and ringdown of the signal, as well as higher harmonics, albeit neglecting spins.*

Edit in the introduction: *In the present study, we will consider a somewhat reduced parameter space, not including the effects of the spins of the two black holes. These could lead to more complicated waveforms, notably through precession effects, which we leave for future work. It should be noted, however, that aligned spin components, though quantitatively important, are not expected to change many of our conclusions due to the relative disconnect between the inference of intrinsic and extrinsic parameters.*

- *Eq. 20 describes an approximation to the transfer function (single link redshift observables) based on a separation of timescales. I would find*

useful if this approximation could be explained in more detail in the text. For example, what are the different timescales involved?

We expanded around the reference to [16], where this is discussed in extensive details, and clarified this point in the text.

Edit above eq. (20) : *How to take the Fourier transform of (**??**) was investigated in details in the perturbative formalism of [**?**]. The main timescales to compare are the radiation-reaction timescale $\propto 1/\sqrt{\bar{\omega}}$ of the binary and the LISA orbital timescale* 1yr. *At leading order in the separation of these timescales, for a signal that is chirping fast enough, we have simply*

- *When discussing degeneracies in the sky. Is it possible to compare the match between the injected signal and the best fitting signal from the other side of the sky? Currently in this is discussed in section VIB where the authors use log-likelihood values to quantify the importance of the second mode. I think using match values might be more familiar to many in the community.*

The reason we chose a likelihood measure here is that it is both easy to compute and easy to interpret in terms of probability (density). Because of the scaling of likelihoods with SNR^2, the meaning of a given mismatch figure depends on the SNR of the signal, and for LISA this can be in the thousands (vs. few tens for LIGO/Virgo). Mismatches are also usually computed by optimizing over time, phase and polarization, which we are not doing with our estimates. We did not compute non-optimized mismatches but added a discussion in the text to emphasize these points.

Edit in Sec. VI B: *We choose here to show a likelihood measure instead of the perhaps more familiar mismatch between signals. Mismatches are by definition normalized and therefore SNR-insensitive, while the log-likelihood scales with SNR^2. Since LISA could observe SNR values in the hundreds or thousands, intuition based on previous familiarity with mismatch measures used in the LIGO/Virgo context might be lost.*

- *In sec III E: "In this process we also impose the constrain...", constraint?.*

Thanks for noticing this, typo corrected.

Other changes

There was a typo in eqs. (67)-(68), with the real part being taken in the wrong place. The $\langle \ell m | \ell' m' \rangle$ mode overlaps are defined as complex quantities. This does not affect our results.

We added $m > 0$ in eq. (13).