Dear Editor,

This is the referee report for the draft titled "Possible Causes of False General Relativity Violations in Gravitational Wave Observations" by A. Gupta et al.

In this paper, the authors conduct a thorough investigation into potential sources of false violations of General Relativity (GR) in gravitational wave (GW) data. They offer a comprehensive discussion of possible contributing factors, which they categorize into three main areas: noise artifacts, waveform systematics, and astrophysical effects. A detailed checklist for identifying strong candidates of GR violation is deferred to a forthcoming publication.

The manuscript is well-structured, clearly written, and accessible. Each potential cause is explained in sufficient detail, allowing even non-experts to grasp the key concepts. The conclusions appear robust and could be of significant interest to the broader community.

I am pleased to recommend this paper for publication in SciPost Physics Community Reports, with some minor comments attached below that the authors might consider addressing to further enhance the paper.

- In Sec. 2.1, it would be helpful if the authors could include examples of both instrumental and environmental sources of non-stationarity. Additionally, they mention that these effects should not impact searches for massive binary black holes but could influence neutron star analyses. I wonder whether future subsolar searches with experiments like the Einstein Telescope and Cosmic Explorer might also be affected by these effects.

- In Sec. 2.2, they comment that glitches exhibit characteristics similar to CBC signals with extreme mass ratios and large spins. Do they have a physical understanding of such similarity?

- In Sec. 2.3, the authors say that detection and parameter estimation of overlapping signals are not a significant concern, without however providing a motivation. They add few comments on line 252 and 253, but it would be useful to expand the above paragraph. Similarly, on line 262 the authors state that when results are combined on a population level, the biases tend to smoothen out. This sentence seem to be in contrast with the last three sentences (line 275-278), where they comment that population analyses may accumulate biases (see also first line on page 23). A clarification on this point is needed.

- In Sec. 2.5, they comment that parameters that suffer large biases are those related to the GW amplitude. The paper would benefit from further explanation of this property.

- In Sec. 3.1.1, the authors comment about the impact of standard astrophysical scenarios on eccentricity. The authors could mention as well the role of accretion and dynamical friction on eccentricity, see https://arxiv.org/pdf/2010.15151.

- In Sec. 3.1.2, they state that not including subdominant tidal effects, such as dynamical tides, can lead to biases in the estimation of tidal parameters. I was wondering if something similar could happen for nonlinear tides (mentioned in their Ref. [184]). Adding further details could make the paragraph more comprehensive. In the same paragraph, they also mention the role of universality relations (which hold to a certain accuracy) in GW data. Does this intrinsic (subpercent-) accuracy impact on the analyses?

- In Sec. 3.1.4, the authors talk about testing of no-hair theorems. It would be useful if they could stress the distinction between primary and secondary hair, to be more precise.

- In Sec. 3.2.1, they discuss about memory effects, see end of page 15, Table 1 (cross on O4) and Sec. 5.2, on a single event basis. The paper would benefit of further additions about similar considerations on a population-based analysis of memory effects and the so-called "stacking" procedure. Would similar conclusions hold on a population level?

- On lines 626-630, the authors propose that a probability distribution modelled through a multidimensional Gaussian mixture could account for uncertainty estimates due to sub-optimal fits, and could mitigate waveform systematics at high SNR. I was wondering if such approach could in principle loose any putative non-Gaussian information in the data. Moreover, the sentence on line 666 is a bit unclear, may they rewrite it?

- In Sec. 4.2, the authors discuss environmental effects around binary black holes, not considering clouds of ultralight bosons as environment, which are instead thoroughly discussed in Sec. 4.3.2 from line 914. I was wondering if the paper would benefit of moving that discussion to the above section, where environments are discussed. However, if the authors prefer to keep them there, I would at least suggest adding a sentence about them and a reference to the paragraph in Sec. 4.3.2 (even though I do not see why these clouds should be better described as mimickers and not as environments).

- In Sec. 4.3.2, can they clarify why black hole mimickers with no horizon generally have a much smaller GW absorption cross section than black holes? Furthermore, they could also mention tidal disruption as a changing feature.

- In the paragraph starting on line 951, I was wondering if mass gap events could fall within the category of events detected in regions of the parameter space disfavored by astrophysical population models (even though hierarchical mergers could falsify this possibility).

- On line 1092, the authors forgot to add g/cm^3 after $\rho > 10$.

- One possible suggestion for the authors is to highlight key sentences for each of the potential false violations, helping readers to quickly grasp the main takeaways.