Unbinning global LHC analyses - Referee comments

Summary of paper

The manuscript shows how the application of SBI can be extended to global SMEFT analyses. A “derivative learning” approach is used to learn likelihood ratios over a high-dimensional observable space. These model how the kinematic observables for a particular process vary as a function of the SMEFT Wilson coefficients. The likelihood ratios are then used for inference, and compared to a traditional histogram-based method. The authors demonstrate a clear improvement in sensitivity, not only in constraining individual parameters, but also in disentangling the effects of multiple SMEFT operators. The learned likelihood ratios are then combined in a fit to multiple LHC processes, demonstrating how the advantages of SBI persist for global SMEFT analyses. The techniques described above are applied to the WW, WZ, WH and ZH processes.

Overall assessment

The paper is interesting as it shows for the first time (at least to my knowledge) the use of SBI in a global analysis context. The analysis appears technically sound in all aspects, and is presented in a clear way for the most part. That said, I think the paper would greatly benefit from a number of clarifications. Therefore I recommend that this paper is accepted after minor revisions, subject to the points raised below.

While the idea of combining likelihoods to improve sensitivity is a-priori known, it is still interesting to see this applied to SBI for LHC measurements. There are a few instances in the paper e.g. in section 4.2 where the authors use phrases like “*It remains to clarify* whether this advantage persists when combining…”, which is perhaps overstating the significance of the result.

It should be noted that the analysis in this paper is a much-simplified version of what would be required for an analysis based on real LHC data. Systematic uncertainties are not included. These are necessary to account for any mis-modeling in the simulation, and would complicate the procedure substantially, especially when combining correlated systematic effects across the different processes. The discussion of including systematics is minimal (end of section 3.2), and left to other references. In addition, the theory parameters are only considered for the signal process in each selection, while in reality they can also affect the background contributions. In the paper, it is not clear how the procedure can be extended to include this. Finally, LHC analyses typically use data-driven approaches to estimate the backgrounds and it remains to be seen how this can be absorbed into the SBI method. I would advise the authors to make these simplifications clearer, in both the introduction and conclusions sections of this paper. It would also help to expand the methodology section with a few paragraphs explaining how one may incorporate these limitations.

The use of language in the manuscript is generally good, but there are a few instances which can be improved. The comments below do not focus on the language/grammar unless it impacts the understanding. I recommend a careful language pass before publication.

Minor revisions/questions

General

- The use of the word “global” is a bit of a stretch as only WW, WZ, WH and ZH. I would consider rephrasing as “towards a global”, where possible.
- The naming “Simulation-based inference” does not distinguish from traditional (histogram) methods. Almost all analyses at the LHC are simulation-based inference. I propose that you change the phrasing throughout to e.g. “Neural simulation-based inference”.
- Throughout the paper, please switch “Higgs” -> “Higgs boson” e.g. “the Higgs discovery” should be “the Higgs boson discovery”.

Abstract

- Add “at the LHC” at the end of the first sentence to give context to the “processes” which you refer to in the second sentence. And then remove “high-profile”, it makes it sound like the processes you study are not high-profile (also to remove “high-profile” in introduction).

1.Introduction

- “Statistically optimal” is too strong a statement. While the Neyman-Pearson lemma tells us that the likelihood ratio is the most powerful hypothesis test, this is only true for simple tests (H0 vs H1) with no free parameters. I would just remove “statistically optimal” from the sentence.
- Add references to the most recent CMS and ATLAS global SMEFT analyses [https://arxiv.org/abs/2504.02958, https://cds.cern.ch/record/2816369/] . The statement that these global analyses are currently based on total rate or binned differential rate measurements is not strictly true, see ttX inclusion in Section 3.6 of the CMS paper.
- “and add fermionic contact terms”. What do you mean by “add” in this sentence?

2. Methodology

- Please describe what θ_0 is after equation 4.
- “The squared difference is averaged over combined parton-level and reconstruction-level”. Would “paired” instead of “combined” be a more appropriate word here?
- Please describe what the terms in Equation 6 are. For someone unfamiliar with SMEFT, this needs a careful description, particularly an explanation of how the Wilson coefficients become the set of theory parameters (theta) which we are trying to infer will help with clarity.

- Equation 8: i would advise adding the ‘ (prime) notation to identify that these terms are derivatives i.e. R’_i and R’’_ij. This would also help in Equation 17 etc.
- It is not clear to me how the neural network estimators listed in Equation 10 are trained. Is the idea that you use the loss function given in equation 5, where the target becomes the expressions in Equation 9. These estimators are independent of theta value, so what theta value do you use to generate the training data (I assume it is theta_0)? Can you provide a few sentences explaining these details, perhaps in section 3.3. It may also benefit to add an “approximately equal to” after equation 11, showing the sum of the neural network estimators to be absolutely clear on how we approximate the likelihood ratio.
- It is worth explicitly stating that the background components are independent of the theory parameters, e.g. after Equation 12.
- In section 2.4, there are a number of terms which are undefined e.g. n, Nexp. Also the replacement of the sum to the mean is specific to applying this method to Monte-Carlo simulation. What about if you apply this to real data at the LHC i.e. the observed limits? I would add a sentence explaining this.

3. Processes, operators, and training setup

- Please provide motivation as to why you have neglected flavor, CP-violating and higher-dimension operators (i.e. insensitive/mass scale suppression).
- The theory parameters (WCs of operators listed in Table 1) would also affect the backgrounds e.g. ZZ for WZ production. It is not clear to me how to extend the procedure discussed in Section 2.2, when you have multiple processes (signals or backgrounds) in the same analysis selection which are simultaneously dependent on the theory parameters. This also applies to contamination of the other signal processes in a given region e.g. WW landing in WZ region. Could you provide a paragraph at the end of Section 2.2 explaining how one may approach this problem? This would be necessary for a true global SBI analysis.
- What do you mean by the sentence “assuming that the trained signal-background classifier will favor the Higgs signal” ? Is this referring to the classifier in Equation 15? My interpretation is that this, omega(x), is a correction factor to weight the contributions in the likelihood ratio according to the signal-to-background ratio at point x. The use of the word “favor” is confusing to me here. Is this part of the sentence necessary? Also reword “measured with the measurements” at the end of the paragraph.
- Why use SMEFT predictions (NLO) at a higher-order than the generator (LO)? For consistency would it not make sense to use SMEFT@LO?
- The paragraph at the end of Section 3.2 simplifies the problem of systematic uncertainties too much. There are many systematic uncertainties (both theory and experimental) which affect the shape of the kinematic distributions. These can often be amongst the leading impacts in analyses involving the processes considered here. Then, the second comment “SBI extracts more kinematic information” does not align with the previous comment about how total rate is the dominant factor. I do agree that given the SBI approach learns more information about the kinematics, we can use that information

to better constrain/decorrelate the systematics from the theory parameters. However, I think the wording/clarity can be improved.

- Can you clarify, do you use a single MLP to regress all Ri and Rij? As in these are different output nodes of the network, and the loss is the sum of the mean squared difference over the nodes? How does this align with the later statement that R_cPhiD and R_cPhiD^2 are learned separately? Does this mean you have separate MLPs for these terms. A bit of extra detail in these paragraphs will help the reader understand the setup.
- Why use GELU for the likelihood ratio term estimators, but leaky-ReLU for the classifier? It would be worth adding some reasons for your choices of architecture/activation functions etc.
- I like the use of coverage to validate the learned likelihood. I'd suggest this deserves more attention than just an Appendix? The bias would also be a property that would be interesting to investigate i.e. what is the mean of the pull distribution over repeated samples: [fit - true/uncertainty] , as a function of true. Have you checked the bias for the SBI approach?

4. Results

- "For each type of limit, two contours…" -> "For each result, solid and dashed contours indicate the one-sigma and two-sigma regions, respectively."
- Figure 2,4,6,8,11 - Add luminosity and labeling of 1sigma and 2sigma intervals in the legend. Perhaps use thick and thin lines instead of bold and faded.
- "Is able to constraint each" -> "is able to constrain each"
- On the sentence: "It remains to clarify whether this advantage persists when combining all four processes". I think this is a given i.e. known a-priori that combining likelihoods will only enhance the sensitivity, and therefore overstates the significance of the result. I'd recommend changing to something like: "The following results show how these advantages persist when combining all four processes".
- Equation 29: It is worth stressing that this is a huge simplification, compared to what would have to be done in a real combined analysis at the LHC.
- Figure 10 caption is incorrect, you do not compare to the parton-level limits.

5. Conclusions

- Add a paragraph explaining the simplifications/caveats involved, and what would be necessary to extend this procedure to real analyses at the LHC e.g. inclusion of systematic uncertainties.
- Add a paragraph explaining how one does not need to use SBI for all processes i.e. it would be possible to combine a histogram-style analysis with an SBI analysis, assuming the two are statistically independent. This will likely be what the first combined SBI result looks like, where SBI is used in a few of the channels.

Appendix B

- "An improved extraction of the likelihood…". Add a comment on how one may improve the extraction? E.g. more complicated architecture? Better training samples?