Manuscript number: NHESS_2020_74
Full Title: A New View on Risk of Typhoon Occurrence in the Western North Pacific

General comments:
This manuscript describes a new and efficient method to produce TC event set in the Pacific basin. The TC events are detected from an ensemble data archive TIGGE, using an objective impact-oriented windstorm identification algorithm WiTRACK. This dataset contributes to existing synthetic datasets (mostly statistical basin-wide methods) as in this dataset the TCs are detected in GCMs, so that the complex physical processes of TCs are captured and hence TCs are physically realistic. More extreme TCs are found in the dataset and these data will help overcome the shortage in observational record. Overall, I think the results will be a nice contribution to the field of TC risk assessment. However, I have some basic questions about the methodology, the utility of certain results and I recommend major revisions prior to publication. Also, I would recommend careful editing of the manuscript. There are many terminologies in the manuscript and please make sure it is easy for readers to follow.

Major comments:
1) L191: The detection rates of historical TCs are reported here. However, will the detection algorithm produce more TCs? What fraction of TCs that the detection algorithm produce is real historical TCs?

2) The authors mention that one benefit of this dataset is "The TPEPS event set includes events which are unlikely but physically possible. This provides an important and unique advantage for typhoon risk assessment." Combined with Fig. 2, TC tracks in the detected dataset is very different with observations, and TPEPS tracks appear in locations with no historical tracks. If there is no historical track in some regions, are they supposed to be no storms or there can be storms but no storm has appeared in historical records due to the low probability? This needs to be explained.

3) The sensitivity and performance of four ensemble data archive are not well described. For example, in some dataset, the storms are much weaker than historical storms. And some models have biases in simulating extratropical cyclone transition. More explanations and descriptions of the data archive needs to be added. Also, how these biases would have an impact on the detection algorithm?

4) The authors have compared the TIGGE PEPS TCs with JRA-55 in terms of track density, landfall frequency, etc. How about other characteristics? For example, landfall intensity along coastline?

5) Fig. 7 shows the difference between TIGGE PEPS event set and observation. In the text, you have mentioned possible reasons for these differences. Is there possible way to reduce these differences, for example in the detection algorithm, to also remove low-impact storms? Also, you mentioned the ESSI, is there a way to quantify this index?

Minor comments:

L41-42: more recent papers should be added. Such as the following two recent models:

- Lee, C.-Y., M. K. Tippett, A. H. Sobel, and S. J. Camargo, 2018: An environmentally forced tropical cyclone hazard model. Journal of Advances in Modeling Earth Systems, 10 (1), 223–241.
- Jing, R., and N. Lin, 2020: An environment-dependent probabilistic tropical cyclone model. Journal of Advances in Modeling Earth Systems, 12 (3), e2019MS001 975.

L45: I didn't understand the sentence 'the typhoon event set might not be physically consistent'. What is 'physically' consistent?

L79: "The domain of this study covers the Western North Pacific (WNP), east and south-east Asia spanning from 85 E to 195E and 15 S to 75 N." Why data around equator is also used? There is no TCs forming around equator.

L102-104: Is there a reason why an old version of IBTrACS is used?

L152: "the accuracy of the LRC is about 90%" What is the fraction of TC (or positive samples?) Does there exist issue of imbalanced data?

L197: "Percentage of total TC windstorms as PEPS TCs can be treated as a proxy to quantify the forecast skill of the model." In Table 5, NCEP is almost twice of that in JMA, what does this percentage mean?

L203: do you mean Fig 3? Also, more explanations should be added in the text. I can't understand this figure.

L203: In Fig. 2, all TPEPS are much more similar with each other, comparing with JRA-55. How to explain this?

Fig4: The tracks in black are very easily messed up with the map. Probably change the color of coastline.

Fig5: The y-axis is not clear to me, please add more explanation.

Fig8: The colored dots for single center are too light to see. If this figure is to show distribution, I would recommend not using same color bar for single model and for TIGGE total.

Fig9: It's hard to see the distributions are in good agreement, probably can change to annual frequency instead of total number of landfall events. Also, the correlation coefficients could be used to show the landfall frequency in all TIGGE dataset is positively correlated with JRA-55.

Fig12: I can see your points in showing the grey dashed lines. But the lower bound curves can not show the trend properly. I would recommend add 75% or 80% confidence interval to show that the trends are same, but TIGGE PEPS event set has much narrower bounds.