We would like to thank the reviewer for his careful reading of the manuscript and for the comments that improved the quality of the manuscript. All comments have been addressed and a point-by-point answer is provided in the following (in blue after the corresponding comment).

Major comments:

1. The manuscript is lengthy at 28 pages, but nearly half are devoted to introduction and methods that use a lot of space to describe standard procedures in machine learning. For example, section 2.2.1 goes into great detail about OOB samples. A section in feature selection using random forest is necessary (Figure 1 is very good) but the background is very standard, and can be shortened and refer the reader to appropriate background references. Section 2.2.2, 2.2.3 include a lot of background in RNNs and CNNs that can also be shortened as it is standard procedure in machine learning and not specific to forecasting cyclone tracks. Same for 3.3, 3.4 in normalization and evaluation criteria which do not need to be in separate sections; overall, the sections before the results can be reorganized for conciseness and avoid replicating a lot of existing background literature on the topic.

Reply: Thank you for your valuable advice. We reorganized the structure of sections 2 and 3 in the revised manuscript, merging the random forests method and de-vortexing method into data preprocessing, and merging RNNs and CNNs into the model framework of section 3.2. The standard background part of the three methods presented in the manuscript (Random Forest/RNNs/CNNs), was deleted and replaced with references, besides, we deleted sections 3.3 and 3.4.

2. I would like to suggest authors to be more careful in the introduction in regards to the strengths and limitations of NWP, statistical models, and deep learning, avoiding potential biases towards methods that are not deep learning. For example, line 49-50 says that NWP models have “limitations in methods” requiring “numerous calculations”. Framing of these limitations is needed. Is computational performance of these numerous calculations unacceptable? The work presented in this manuscript is of course efficient, giving results in seconds. But how long do NWP models take? The authors then mention “accurate mathematical descriptions of physical atmospheric mechanisms”. NWP models aren’t always exact and can involve many approximations and parameterizations. The authors may be trying to convey that NWP models require description of physical processes versus machine learning methods that learn from data. But that is not a limitation, it would be a property of different approaches to modeling.

Reply: Thank you for your valuable advice. We agree with the reviewer that these statements may be potentially biased. Line 49-50 “there are limitations in methods relying on high-performance computers and requiring precise initial conditions.” was changed to“there are limitations in methods relying on high-performance computers and requiring precise initial conditions.”
However, due to the computer performances, mode resolutions, and sizes of the selected areas being different, it is difficult to determine the running time of numerical model prediction.

The authors also criticize statistical models, saying "manual feature selection is unable to produce accurate predictions". The inaccuracy would need to be characterized (and cited where appropriate) in order to reach this conclusion.

Reply: L60-61: We added the "CLP5 had the largest Mean absolute error(MAE) of all models for TCs occurring from the Eastern Pacific and North Atlantic (Boussioux et al., 2022). " before "manual feature selection is unable to produce accurate predictions." and added the "Li-Min et al. (2009) used the BP neural network to predict that the average distance error of the 6h movement track of six typhoons in 2005 improved by 36.9km, compared with CLIPER." after it.

reference

Boussioux, L., Zeng, C., Guenais, T., and Bertsimas, D.: Hurricane Forecasting: A Novel Multimodal Machine Learning Framework, Weather and Forecasting, https://arxiv.org/abs/2011.06125, 2022.

Giffard-Roisin, S., Yang, M., Charpiat, G., Kégl, B., and Monteleoni, C.: Fused Deep Learning for Hurricane Track Forecast from Reanalysis Data, Climate Informatics Workshop Proceedings 2018, Boulder, United States, 2018-09-19, https://hal.science/hal-01851001, 2018.

Li-min, S., Gang, F. U., Xiang-chun, C., and Jian, Z.: Application of BP neural network to forecasting typhoon tracks, Journal of Natural Disasters, 18, 104-111, 2009.

In the following paragraph about deep learning, authors give very specific accuracies (e.g., L83, L85, L91) and strengths of this method. For completeness and fair comparison, I suggest authors also give conventional methods similar statistics and strengths, and avoid vague, uncited description of limitations.

Reply: Lines 80-87： "Gao et al. (2018) used long short-term memory (LSTM) to predict typhoon tracks in the Northwest Pacific Ocean; the ratio of the cyclone training set and test set was set at 8:2, and the 24-h prediction error could reach 105 km. Alemany et al. (2018) proposed an RNN based on a grid system to predict hurricanes in the Atlantic, potentially improving the 6-h prediction accuracy with a root mean square error (RMSE) of 0.11 for the test set. Kim et al. (2018) performed a TC identification task based on ConvLSTM to train WRF-simulated data, and the results are significantly better than those of a convolutional neural network (CNN)." introduce RNNs (one of the deep learning methods) applied to TC track forecasts that occurs in different sea basins and different years, resulting in different training sets and test sets. Another deep learning method, CNNs, is also introduced in lines 89-91:" Giffard-Roisin et al. (2020) combined historical trajectory data with wind field reanalysis data as input to a CNN and predicted Atlantic hurricane tracks since 1979, with an average error of 32.9130 km for 24-h predictions." They are not comparable. These examples are to illustrate the feasibility of the application of RNNs in TC track forecasts. In addition, the author only gives 6h prediction results in the article. It is difficult to unify the standard.

3. Text in figures is at times hard to read because of the small font size. Please also make the fonts consistent, e.g., Arial. I suggest going through the futures to ensure consistency in presentation and that all figures are clear and readable.

Reply: Thank you for your valuable advice. After careful examination, we adjusted the size of the text in the figures and unified the fonts for Times New Roman.

There are also other statements in the introduction that require revision. I include them in the specific comments below.

Specific comments:

1. Abstract, L29: please include the average distance errors of the CMO forecast results as well for comparison.
Reply: "(27.57km and 59.09km)" has been added.

2.L82 uses 24-h prediction distance error for LSTM, then L85 uses 6-h RMSE, L91 uses 6-h distance error. If possible, please be more consistent in error metrics.
Reply: This question has been answered above in the third paragraph of the second major comment.

3. L86: authors say Kim et al. (2018) are "significantly better" than those of a CNN. How much?
Reply: We added "The average precision of the forecast was improved by 78.99%." after "and the results are significantly better than those of a convolutional neural network (CNN)."

4.L97: please define MLP, first time the acronym has appeared in the manuscript.
Reply: Suggestion adopted. MLP is modified to the full name ' Multi-Layer Perceptron'.

5.L101: "Previous studies have shown…". Which previous studies? Please provide references.
Reply: Perhaps what I want to express is the above studies. I have modified it in the manuscript.

6. L102: "Still, most of them have neglected to describe and analyze the meteorological factors that affect the movement of TCs, ignoring valuable features." Which studies? Did this neglect of meteorological factors significantly affect performance, compared to studies that have considered these factors? Please also give examples of these "valuable features".
Reply: Thank you for your valuable advice. We added "The 6-hour average distance error between predicted and real location by the fusion network (wind+track) is 32.9 km, while the network prediction results without adding wind variables are 35km (Giffard-Roisin et al., 2018). ", which indicates that the addition of meteorological field variables can effectively improve the prediction accuracy.

reference
Giffard-Roisin, S., Yang, M., Charpiat, G., Kégl, B., and Monteleoni, C.: Fused Deep Learning for Hurricane Track Forecast from Reanalysis Data, Climate Informatics Workshop Proceedings 2018, Boulder, United States, 2018-09-19, https://hal.science/hal-01851001, 2018.

7. L126-127: Do you mean that the Coriolis parameter is included in the predictors?

Reply: No, the Coriolis parameter of the typhoon is included in the input variable. In order to avoid ambiguous statements, L126-127: "In addition, the Coriolis parameter corresponding to the latitude of the past 24 h influences the geostrophic deflection force on the TCs." was changed to "The Coriolis parameters corresponding to the latitude of the TCs in the past 24 hours are also included."

8. L128-133 describes a TC bias to northwest; I am having trouble following the reason for this paragraph. Is this the reason for the geographically asymmetrical data selection in L142-147 (3)? If yes, then why is this not done for (1) & (2)?

Reply: There is a formatting error at the end of this paragraph that should be merged with the next paragraph. I deleted this sentence "Because they are influenced by the earth's rotation, TCs will be biased to the northwest (Kitade, 1981). " L128-136: Both observational and theoretical studies have shown that TC movement is closely related to large-scale airflow fields (Holland, 1983), and TC movement is mainly affected by the steering flow (Brand et al., 1981; Chan, 1984). Interactions among weather systems, the subtropical anticyclone, Westerlies, and the Tibetan High will also affect the movement of cyclones (George and Gray, 1976; Chan et al., 1980). The geopotential heights of 300 hpa, 500 hpa, and 700 hpa are selected as the locations for the high, middle, and low-level circulation data, respectively. In addition, the underlying surface conditions must be considered, and, in the case of a weak guidance environment, TCs tend to move toward warmer sea-surface temperatures (Sun et al., 2017; Katsube and Inatsu, 2016)." describes several meteorological factors affecting the TCs movement including the steering flow, sea surface temperature, and weather systems, corresponding to uv, sst and hgt described in the next content. First of all, I refer to a Chinese paper on the selection of the meteorological field variable division area size, and it is symmetrical in the zonal direction but asymmetrical in the meridian direction. The main reason is that TCs tend to move north, so they are mostly affected by the weather system in the north, especially the subtropical anticyclone.

9. L143, L145 "10 degree radius". Do you mean extended by a 10 degree distance in each direction, since a 21x21 grid is formed?

Reply: Here is my misstatement. The "radius of 10 degrees" should be changed to "centered the typhoon, extend 10 degrees outward in the zonal and meridian direction respectively, and form a square matrix with a 21x21 grid.

10. Figure 4: I suggest also adding the RMSE for the test set for the three recurrent neural networks inset in the figures for ease of comparison.

Reply: Suggestion adopted.

11. L414, Figure 9 legend: 2106 -> 2016.

Reply: The numbers in parentheses originally indicate the numbers of the TCs. After I review several works of literature, these numbers should be changed to years.

12. L416-417: add "using deep learning methods" at the end of the opening sentence.

Reply: Suggestion adopted.