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

The authors spend a lot of space explaining the technical details of the standard machine learning and deep learning models (random forests, RNN, CNN). These can be shortened, or refer the readers to the detailed background references. In addition, the equations for the GRU and LSTM cells are hard to follow, thus they can be complemented with diagrams showing the flow of data in these cells.

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. In addition, the relevant diagrams of LSTM and GRU are added to the revised manuscript so that readers can better understand it.

In the model framework section, I find it hard to understand the network architecture that the authors used in this work.
I think it would be benefit to include a table detailing the network architecture.
In addition, in figure 3, I think the description of the figure could be revised to include more details such as: CNN kernel size, what the solid white arrows mean, what the dashed red arrow means, etc.
The authors do not mention the architecture of the RNN, LSTM, GRU that they used in this work. I think it would improve the clarity if they were included here.

Reply: Thank you for your valuable advice. We added an explanation of the arrows in Figure 3. Among them, the gray filling arrow represents the TimeDistributed layer that is applied to a series of tensors in the processing of the time dimension. The black solid arrow means the multidimensional tensor into a one-dimensional vector. The dashed black arrow represents the fully connected layer in the network framework, and the dashed red arrow means the merging of multiple vectors into one vector. The figure is revised as follows:

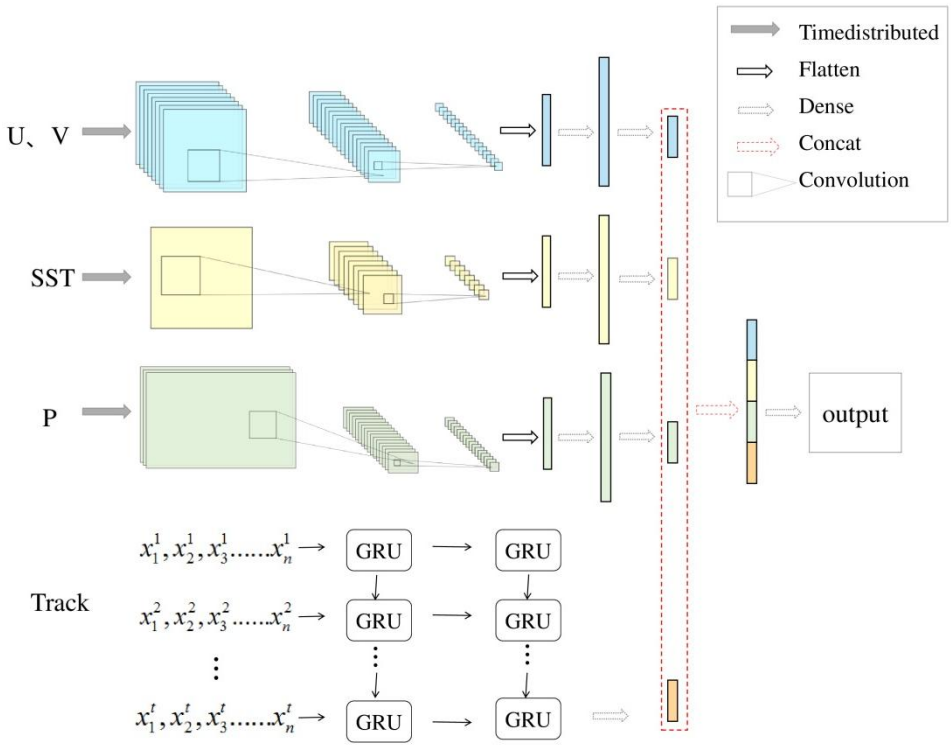

Figure 3: The model framework and network structure of GRU_CNN.

In addition, In order to show the network framework more clearly and facilitate readers reading and understanding, we added a table to list the input and output size of each layer in the network framework, including convolution kernel size, stride, and channel number.

The 2D-CNN-based encoder architecture of the pressure field branch

Table 1 Each layer architecture of the GRU_CNN

Layers	Kernel Size	Stride	Channel	Input Size	Output Size
Conv_uv	7×7	2	8	21×21	8×8
MaxPool_uv	4×4	4	16	8×8	2×2
Flatten _uv	-	-	16	2×2	64
Dense_uv_1	-	-	-	64	128
Dense_uv_2	-	-	-	128	32
Conv_sst	7×7	2	1	21×21	8×8
MaxPool_sst	4×4	4	8	8×8	2×2
Flatten_sst	-	-	8	2×2	32
Dense_sst_1	-	-	-	32	128
Dense_sst_2	-	-	-	128	32
Conv_p	14×25	4	3	46×81	9×15
MaxPool_p	5x11	4	16	9×15	2×2
Flatten _p	-	-	16	2×2	64
Dense_p_1	-	-	-	64	128
Dense_p_2	-	-	-	128	32
GRU_1	-	-	-	8×11	8×128

GRU_2	-	-	-	8×128	128
Dense_GRU	-	-	-	128	32
Concat_layer	-	-	-	-	128

Line 323-327 introduces some detailed descriptions of the three recurrent neural network frameworks. Firstly, RNN, LSTM, and GRU are all recurrent neural networks with similar structures and the parameters of the three networks are the same. Secondly, their architectures are actually included in the overall framework, which is a part of our proposed model GRU_CNN, so it is not highlighted. In the case of only inputting trajectory features, these three networks are used to compare which result is better, and then put it into our fusion model.

In the discussion of table 3 (L384-L392), the authors claim that the influence of SST and geopotential height gradually increases at long-term forecasts. Can the authors provide more explanation of why this is the case?

Reply: This conclusion is based on the statistically average value in table 3, which can not represent each tropical cyclone and can be regarded as the results of the whole. It is shown that as the forecast time increases, the proportion of the steering flow to the predicted value gradually weakens, so the sea surface temperature and geopotential height increase accordingly. However, we have not found relevant literature to explain this phenomenon. Meteorologically, sea surface temperature will drive the TCs to the warm sea surface, so it will not affect the movement of the TCs in a short time. The geopotential height represents the weather system at high and low altitudes, and it will affect the movement of TCs for a long time. When there is a subtropical anticyclone staying in the north of the TC, it will cause the cyclone to stagnate or move slowly, which involves the analysis of weather patterns in meteorology, It is very interesting, but beyond the scope of this manuscript.

Since the authors compare the performance of GRU_CNN with other methods: FAXAI, MITAG, and IN-FA in figures 7-9, I think it would be more convincing if the authors can also provide detailed comparison between these models like in the table 3.

Reply: Here, FAXAI, MITAG, and IN-FA are all the TCs' names, not the model method. We selected three of the 54 typhoons in the test sets and analyzed them. Figure 7-9 shows their actual and predicted paths. They all have the common characteristics of track turning. The other 51 TCs' forecast paths are shown in the supplementary document.

Specific comments:

L79: missing a space between a reference and the word “applied”

Reply: Suggestion adopted.

Figure 4: this figure could instead show the difference between the predicted longitudes/latitudes with the observed longitudes/latitudes to improve clarity and readability.

All figures’ texts and labels can be a bit bigger to improve readability.

Reply: Suggestion adopted.

L397: what are these methods: FAXAI, MITAG, IN-FA? Can you provide a short description and references for these methods?

Reply: This question has been answered above in the fourth General comment.