Supplementary Material of "VQLTI: Long-Term Tropical Cyclone Intensity Forecasting with Physical Constraints"

Anonymous submission

The Details of Experiments

In the actual experiments, the VQLTI forecasts are made at 6-hour intervals, but for convenience of presentation, we only show the 24-hour, 48-hour, 72-hour, 96-hour, and 120-hour forecasts in the tables. The TC dataset itself consists of TC cases with different time lengths, which means that different input and output window lengths will significantly affect the number of test samples. Taking the input of 24 hours and the forecast of 120 hours as an example, the window size is 144 hours, which means that TC data with a life cycle of less than 144 hours will be excluded from the test set, reducing the number of test samples. To ensure as many test samples as possible, we only calculate the error of the final forecast value in the forecast window, that is, the 24-hour forecast error selects the last forecast error in the 24-hour forecast window, and the 120-hour forecast error selects the last forecast error in the 120-hour forecast window. Therefore, each time the forecast window length is changed, a new test needs to be performed on the test set. Finally, the test sample sizes for the 24-hour, 48-hour, 72-hour, 96-hour, and 120-hour forecasts in 2022 are 1141, 897, 700, 545, and 416, respectively.

Additional Experiments

Comparison of Forecasting Skill

The experiments in this paper have verified the superiority of VQLTI in both long-term experiments and real-time forecasting experiments. In this section, the experiment will use ECMWF-IFS as the baseline model to compare the forecasting skills of each model. The calculation method of forecasting skill is the same as mentioned in the main text. The forecast skill is often expressed as:

$$sf(\%) = 100 \times \frac{eb - ef}{eb} \tag{1}$$

where eb represents the error of the baseline model, while ef denotes the error of the model currently under evaluation. A positive value indicates that the forecasting performance is better than the baseline model, while a negative value indicates that the forecasting performance is worse than the baseline model.

As shown in Figure 1, the comparison of forecasting skills for long-term forecasts globally in 2019-2020, the VQLTI forecasting skill is all positive, with MSLP forecasting skill being $23.27\%-40.02\%$ and MSW forecasting skill being $35.65\%-42.51\%$ from 24h to 120h, far exceeding other models. As shown in Figure 2, the comparison of forecasting skills for real-time forecasts globally in 2022, the VQLTI (real-time) forecasting skill is also positive, with MSLP forecasting skill being $2.93\%-18.63\%$ and MSW forecasting skill being $34.07\%-50.27\%$ from 24h to 120h, still outperforming ECMWF-IFS and far exceeding other deep learning-based models.

Qualitative Analysis

The experiments in the previous sections are all quantitative analyses. This section will qualitatively analyze the VQLTI's 2022 real-time forecasting results, as shown in Figures 3 and 4. The figures display the forecast distributions for 6-hour, 24-hour, 72-hour, and 120-hour lead times. The closer the forecast results are to the blue diagonal line, the closer they are to the true results. The 6-hour forecast results are mostly concentrated along the diagonal line, and as the forecast lead time increases, the results become more dispersed. We use the least squares method to fit the forecast data, as shown by the red lines in the figures. The 6-hour red line closely aligns with the diagonal line, indicating smaller forecast errors. The red lines for the 24-hour, 72-hour, and 120-hour forecasts have similar angles with the diagonal line, suggesting that their overall errors are comparable, which also indicates that the long-term forecast error growth of VQLTI(real-time) is relatively slow.

The figures also use color depth to represent the distribution of the forecasts, and the red pentagram marks the position of maximum density. The maximum density is concentrated in the region of weaker TC intensity (larger MSLP values and smaller MSW values), while the forecasts are more dispersed in the region of stronger TC intensity. Based on previous experiments (Huang et al. 2022; Wang, Li, and Zheng 2024), the stronger the TC intensity, the more challenging the forecasting, which may be due to the lack of training samples for stronger TC intensity events, and future work can be improved in this regard.

In addition, we also present an example to intuitively demonstrate the superiority of VQLTI (real-time). As shown in Figure 5 and 6, we select the 120-hour forecast of hurricane FRANK in the 2022 real-time forecast for compar-

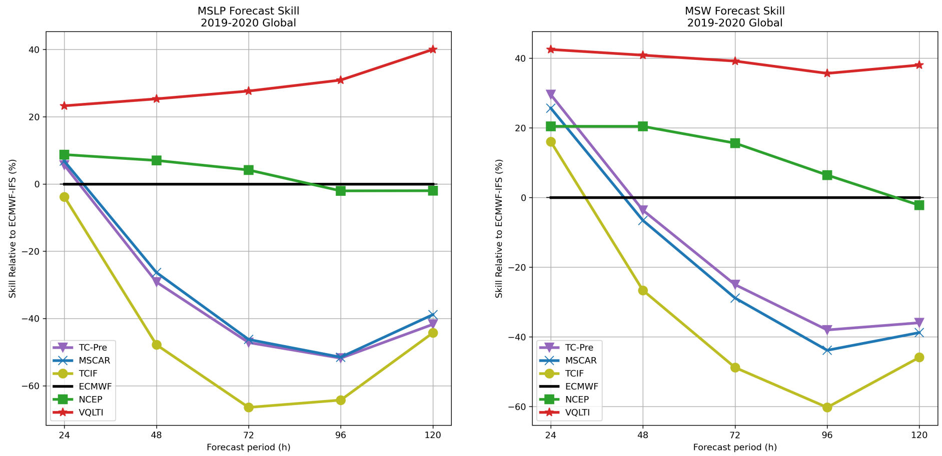

Figure 1: The comparison of forecasting skills for long-term forecasts globally in 2019-2020. The baseline model is ECMWF-IFS.

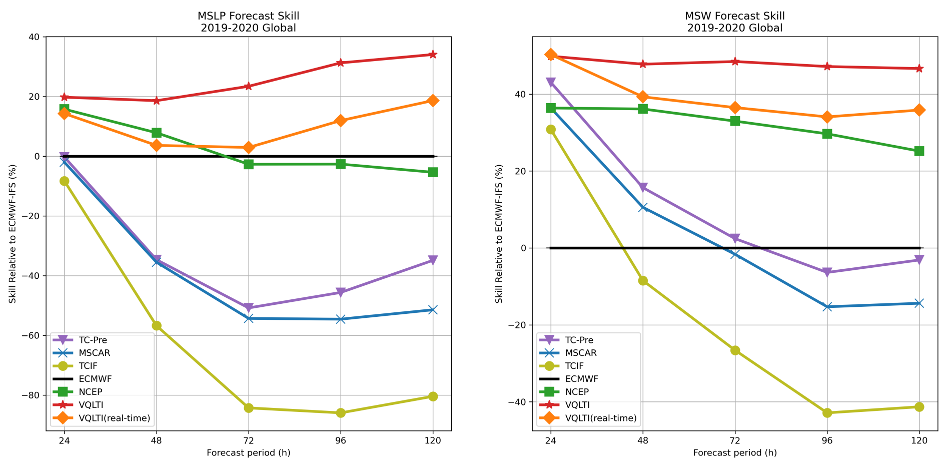

Figure 2: The comparison of forecasting skills for real-time forecasts globally in 2022. The baseline model is ECMWF-IFS.

ison. The blue curve represents the ground truth, and the red line represents the 120-hour forecast for that time. It is clearly observed that the forecasts of TC-Pre, MSCAR, and TCIF fail to capture the changing trend of FRANK, while VQLTI (real-time) exhibits strong long-term forecasting performance, with its forecast curve closely matching the ground truth curve.

References

Huang, C.; Bai, C.; Chan, S.; and Zhang, J. 2022. MMSTN: A Multi-Modal Spatial-Temporal Network for Tropical Cyclone Short-Term Prediction. *Geophysical Research Letters*, 49(4): e2021GL096898.

Wang, C.; Li, X.; and Zheng, G. 2024. Tropical cyclone intensity forecasting using model knowledge guided deep learning model. *Environmental Research Letters*, 19(2): 024006.

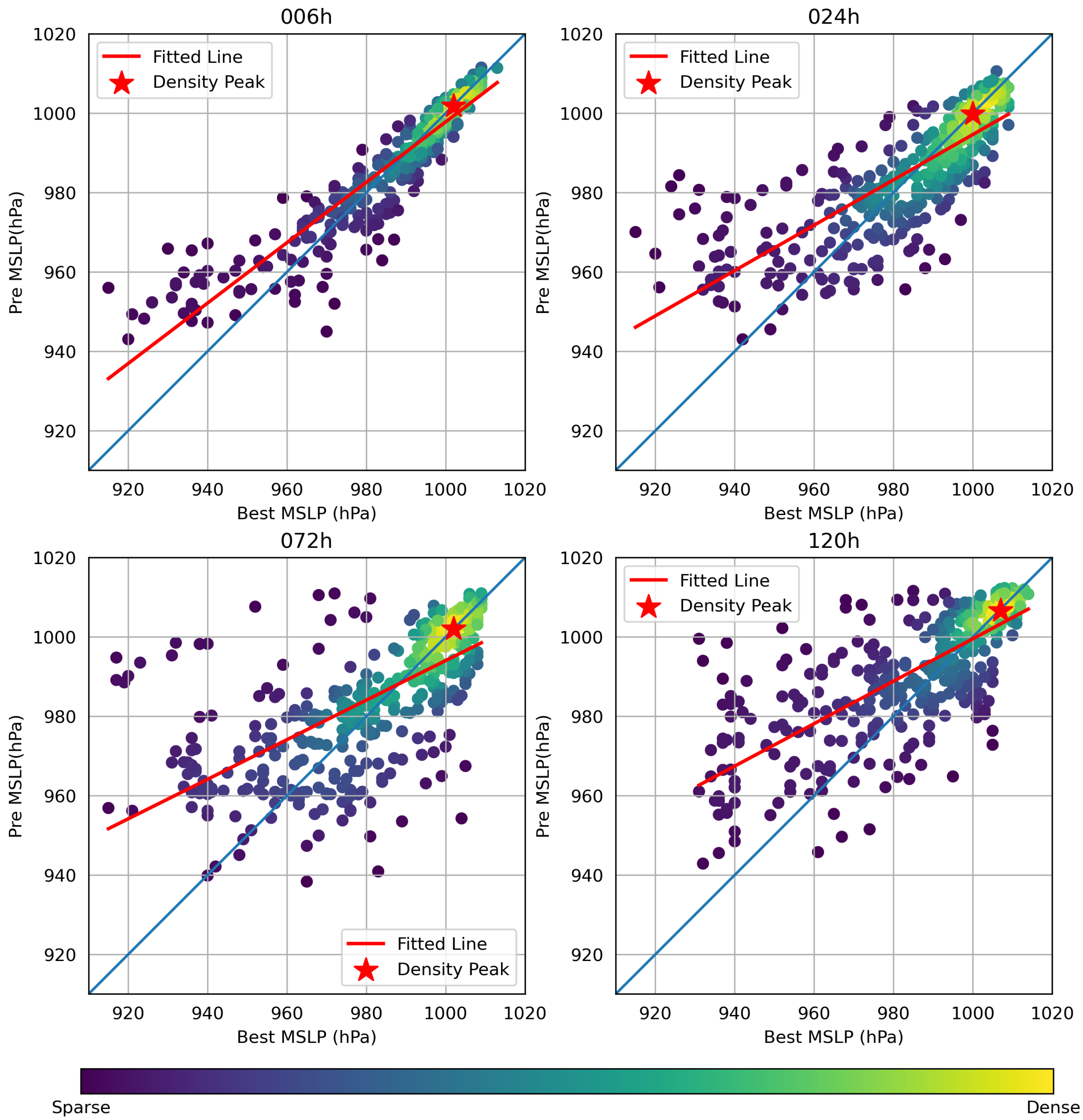

Figure 3: VQLTI's MSLP forecast distribution in 2022 real-time forecasts. The figures show the MSLP forecast distributions of VQLTI at 6-hour, 24-hour, 72-hour, and 120-hour lead times for 2022 real-time forecasts. The blue line represents the diagonal line, the red line is the least squares fit of the forecast data, and the red pentagram marks the location of maximum forecast density.

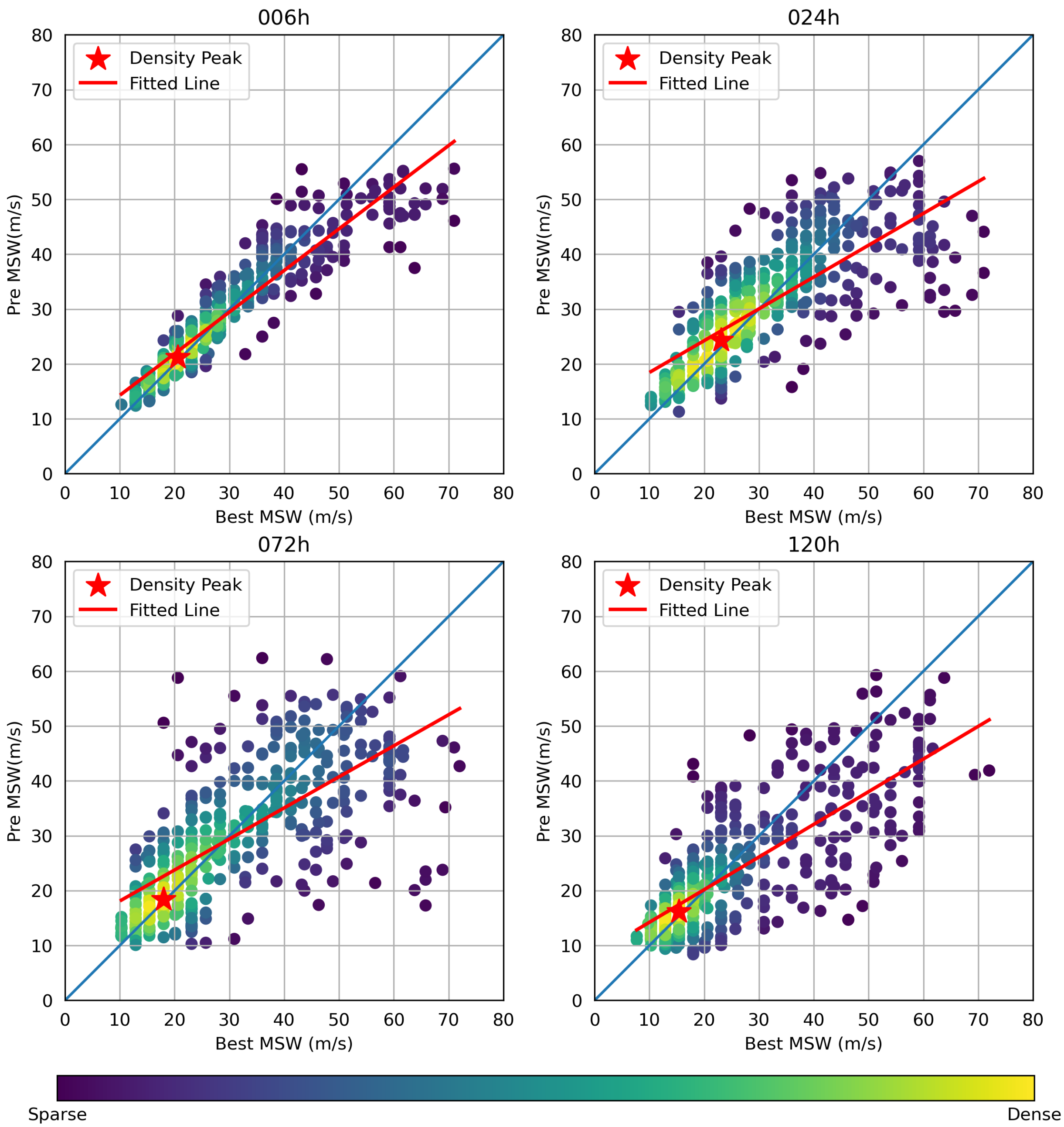

Figure 4: VQLTI's MSW forecast distribution in 2022 real-time forecasts. The figures show the MSW forecast distributions of VQLTI at 6-hour, 24-hour, 72-hour, and 120-hour lead times for 2022 real-time forecasts. The blue line represents the diagonal line, the red line is the least squares fit of the forecast data, and the red pentagram marks the location of maximum forecast density.

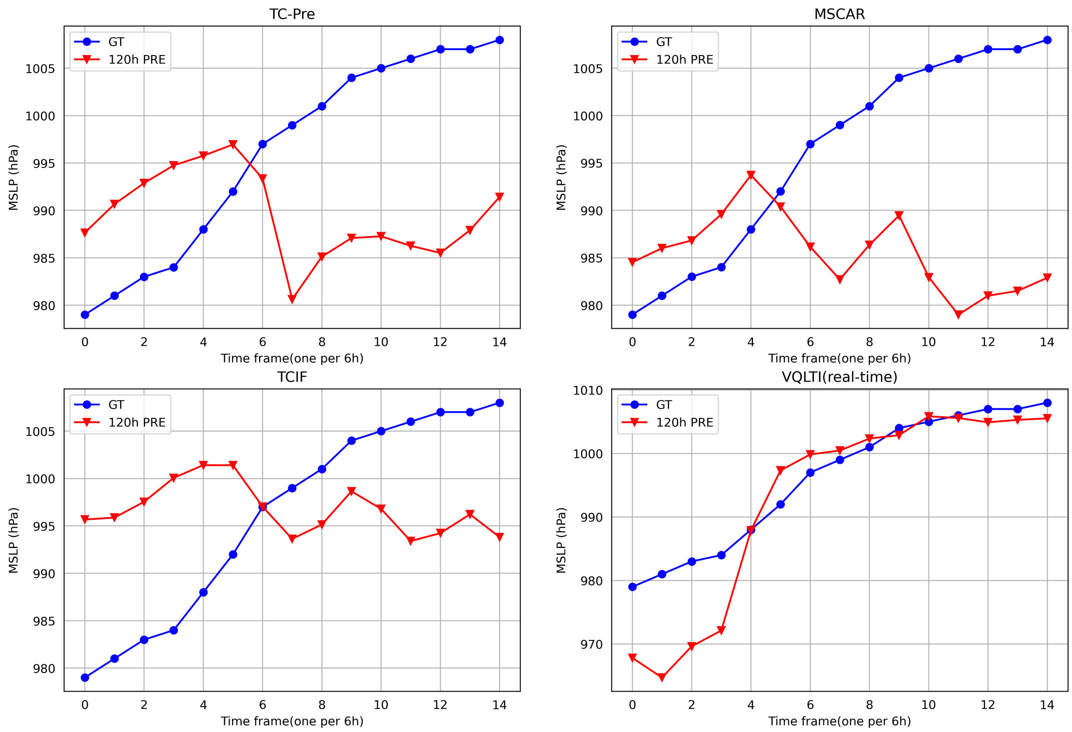

Figure 5: Comparison of 120-hour MSLP Forecasts for Hurricane FRANK in 2022.

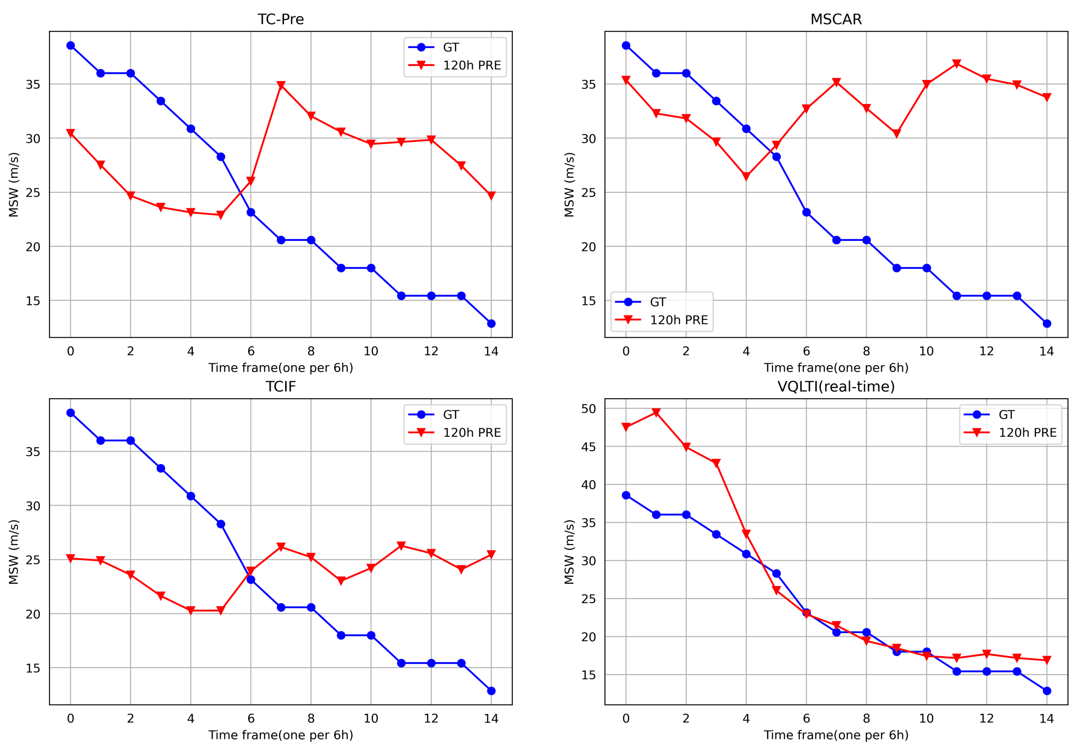

Figure 6: Comparison of 120-hour MSW Forecasts for Hurricane FRANK in 2022.