Response to Reviewer 2:

We thank the anonymous referee for his/her valuable comments and suggestions that have helped us improve the paper quality. Our detailed responses (**Bold**) to the reviewer's questions and comments (*Italic*) are listed below.

Anonymous Referee #2:

Overview:

The manuscript investigates the influence of spin-up and restart in a global weather forecast system GRAPES. Such a topic is important, as careful handling of those technical issues can greatly improve the accuracy of weather prediction. By comparing different spin-up and restart methods, the authors gain important knowledge of the forecasting system, such as that the GRAPES with its own analysis field performs better than the one using NCAR final reanalysis (FNL) data for the cold start in the spin-up. The paper contains useful information for model development and usage. I recommend its publication with GMD, pending on some minor comments below.

We highly appreciate the reviewer's positive evaluation about this study. We also thank the reviewer for the valuable and detailed comments and suggestions which have helped us improve the paper quality.

About the experiment setup. To better illustrate the differences between three experiments, which are of great importance to this paper, can the authors use a schematic plot to show how the three runs were performed and what are the key input data. Also, it should be explicitly stated in the Section 1.2, why those three experiments were conducted, or in other words, what we expect to learn by comparing them. The three-hour lag confuses me a little bit.

Fig. R1 shows the 4D-Var cycle assimilation system and the experiment setup. In fact, we have listed the experimental settings in Table 1. Since the contents of hydrometeor are not analyzed and updated in the 4D-Var system and the cloud information simulated by the G21 experiment is not retained during the restart in the G00 experiment, the input variables of the three tests are the same, we no longer specified the input variables for the three experiments.

To more clearly state why three experiments were conducted in the current GRAPES_GFS operational system, we rearranged the fourth paragraph in the Introduction section and added the following sentences: "Then another question is what advantages the new 4D-VAR assimilation analysis fields have in spin-up process compared with the cold start simulation with FNL.", "Actually, for numerical weather prediction model's users (especially forecasters), they are usually accustomed to referring the forecast productions of model staring to integrate from 00 UTC or 12 UTC (or more time, for example 06 UTC, 18 UTC). " , and "The reasons for the unretained cloud-field variables were mainly based on the following considerations: the hydrometeor contents are very small amount relative to water vapor and they can be quickly created in the spin-up process when the model restarts. Moreover, this treatment can save storage space and input/output (IO) time. However, its impacts on the spin-up process and model forecast performance have not yet been carefully analyzed and evaluated. Therefore, we need to fully diagnose and analyze

the necessity of the repetition of GRAPES_GFS spin-up during the re-integration, and the impact of the lost cloud-field information on the later forecast. ".

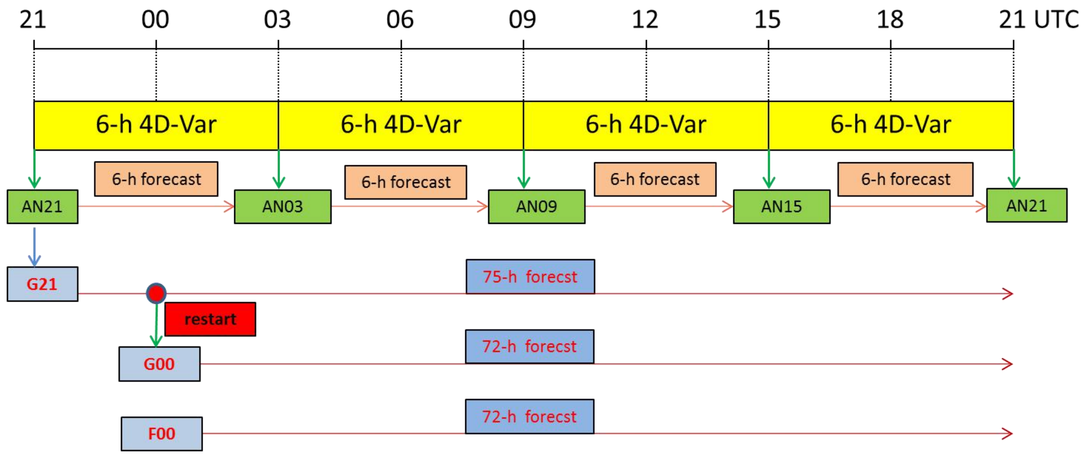


Figure R1. The schematic diagram of 4D-Var cycle assimilation system and experiment setup in this paper.

It is unclear to me what is the current protocol for spin-up and restart strategies used by CMA that uses GRAPES_GFS to conduct the daily weather prediction as well as the extreme weather prediction like typhoons. According to this study, is there any modification needed on the protocol?

It's a good question. In current protocol of operational GRAPES_GFS it still adopts analysis fields from the 4D-Var system as the model's initial field at 09UTC/21UTC and then restarts after 3hr-integration without the information of cloud fields, that is to say, same to the G21 experiment in this paper. Our research results show that it could lead to systematic biases for height, temperature and precipitation fields as well as typhoon track if the restart of the model does not include the information of cloud fields. We have told the results to the managers of NWPC/CMA, which attracted their considerable attentions. As we all know, the adjustment of numerical weather predication protocol has strict specifications, which needs to carry out parallel experiments for a period of time and evaluate the results before its operation application. The parallel experiment has been listed in the operational testing plan. If there are further results, we will be willing to share with you.

Is the total grid number of cloud (TGNC) related with the total cloud fraction? The latter is a more common term. Also, 1.0 e-4 g kg-1 threshold of cloud sounds an arbitrary choice. Are the results sensitive to this threshold definition?

It's a good question. TGNC is related with the total content of all hydrometeors (THC, that is to say, THC=cloud water + raindrop + cloud ice + snow + graupel). We define the grid with cloud when its THC is greater than 1.0 e-4 g kg-1 according to our results. We have tried three thresholds (THC= 1.0 e-5, 1.0 e-4, and 1.0 e-3, respectively) to compare the spinup time, and the results we got are basically the same. In other words, teh results seem not very sensitive to the threshold definition. Meantime, to clarify this part more clearly, we added the description for the definition of equilibrium in section 3.1.3 as

follows "Note that the statistical equilibrium state is defined when the difference of TGNC with respect to the 24-hour integration is insignificant (the difference is less than 20% of TGNC at 24-hour). " combing with the comments of the first reviewer.

Fig. 11a, why g00 and g21 are identical before 42 Hour and then become different abruptly?

The forecasted track errors before 42 Hour simulated by G00 and G21 experiments are not identical. Actually, their tracks are very close before 42 Hour, which are shown in Fig. R2. For the abrupt track difference after 42 Hour, it was caused by the continuous accumulation of the direct cloud-radiation processes, temperature difference, and even re-undergone spin-up process in the typhoon cloud area and their transmissions to the typhoon eye through dynamic processes with the integration. As stated in this paper, the restarted model (G00 experiment) with lost cloud-field information in initial field needs to re-undergo a spin-up process and causes systemic biases of cloud, temperature and geopotential height and precipitation fields at the model early forecast. These biases mainly exist in areas with clouds. For a typhoon, the differences of temperature and geopotential height of the G00 experiment initially exist in the cloud belt around the typhoon eye compared with G21 experiment. With the model integration, the peripheral system difference gradually affects the typhoon center (track) through the dynamic process. These changes can be confirmed from Fig. R3 and Fig. R4. For example, the large value area of temperature difference between G00 and G21 experiments at the early stage of integration is mainly located in the spiral cloud belt around the typhoon eye and its value can reach 2k, while its value over the typhoon central is only -0.25~0.25K. With the integration of the model, the temperature difference of the typhoon eye gradually becomes larger, and its value reaches 0.5-1K at 50 hours of integration, which is bound to affect the track of the typhoon by dynamic process.

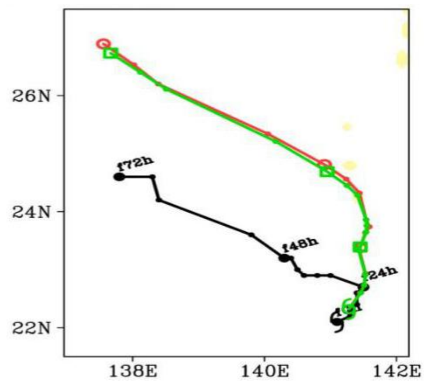
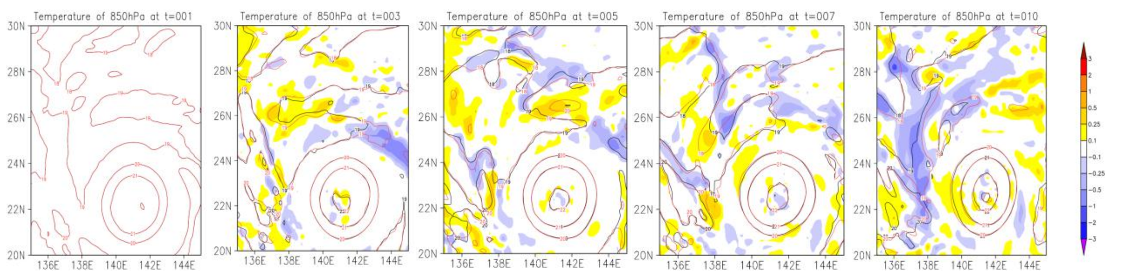


Figure R2. The track of typhoon "KROSA" observed (the black line) and simulated by G21 experiment (the green line) and G00 (the red line) experiment from 0000 UTC on August 9 to 0000 UTC on August 12, 2019.



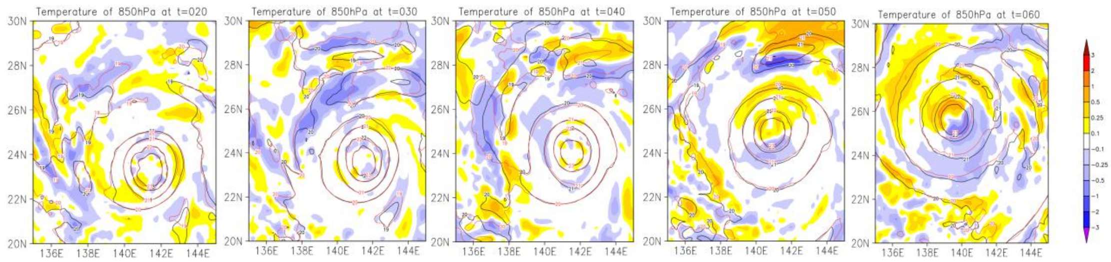


Figure R3. The temperature of 850 hPa simulated by G00 (the red line) and G21 (the black line) experiments and the difference of G00-G21 (shad) at different integration time, the unit is K.

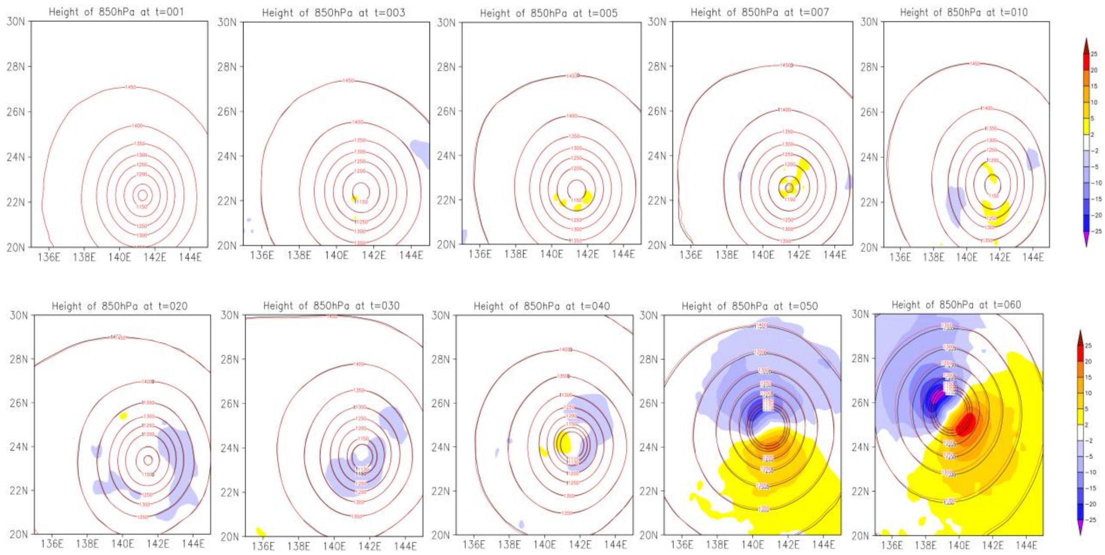


Figure R4. The geopotential height of 850 hPa simulated by G00 (the red line) and G21 (the black line) experiments and the difference of G00-G21 (shad) at different integration time, the unit is gpm.