Anonymous Referee #1

General comments:

- ***Overall, the manuscript is well-written, the evaluation methods presented are sound, the manuscript appears to fit reasonably well within the "Model Evaluation" category of GMD, and it presents results that may be of use to future users of MPASv5.2. That said, I have a few significant concerns about the manuscript: it provides minimal discussion about the physical meaning of the results, it lacks discussion of some highly relevant areas of literature, and it lacks a discussion of uncertainty (or statistical significance) in comparisons across resolutions and between simulations and observations. I don't expect that these comments will require much change to the underlying analysis, but I do think they should result in a substantial amount of new or revised text. Based on this, I am recommending that the manuscript be returned to the authors for major revisions***

We thank the reviewer for the detailed and constructive comments. They are very helpful for improving the quality of the manuscript.

In the revised manuscript, we added a new "supporting material" document to include many new figures to support some statements in the text and to address the review comments. We added a discussion of the synoptic condition of this event during the Meiyu front in the introduction. This provides useful information to understand the precipitation differences among the various simulations. Notably, the precipitation biases are related to the shift of circulation pattern in the simulations (Figure S4, S7). In addition, the results of resolution-dependent precipitation and updraft are further discussed in relation to previous studies. Particularly, Fig. 13 on the relationship between precipitation and the upward moisture flux is added to reveal the mechanism of resolution-dependence following Rauscher et al. (2016) and O'Brien et al. (2016). Following the reviewer's suggestion, the bootstrapping statistical analysis is used to test the significance of statistical difference among multiple experiments. Table 2 with the summary of statistical metrics is added with more analysis. Other text and figures have also been revised as the reviewer suggested.

Specific comments:

- ***Lack of discussion of physical meaning of results***

  Overall, the manuscript reads more like a technical report than a scientific manuscript; it focuses much more on questions of 'what' than questions of 'why'. In my opinion, this severely limits the usefulness of the paper. In its current form, I suspect that the only readers who might find the manuscript interesting would be users of the MPAS-Atmosphere model, since it essentially only focuses on describing how precipitation and vertical velocity characteristics depend on resolution and microphysics. Instead, if the manuscript had a stronger emphasis (even speculative) on why, the manuscript might be of interest to other model users

facing similar questions about the effects of resolution and parameterization.

For example, in Section 3.2.2, the authors present an intriguing result: the GFS model (which are used as initial conditions!) has precipitation that is shifted far too much to the north, whereas the MPAS simulations have the rain band much closer to where it is observed. But the authors provide no speculation on why this might be, nor do they even comment that this is interesting that the MPAS model is able to 'correct' an error in the GFS starting condition. Could it be because of better-resolved topography? Is the northward propagation of the rainband perhaps less rapid in MPAS than in GFS? Are there possibly eddy-mean-flow interactions that MPAS resolves that could cause the rain band to be shifted relative to GFS?

That is just one example; this lack of exploration of 'why' is pervasive in the manuscript. A symptom of this is that almost all of the paragraphs in Section 3 have a fairly repetive structure in which they (1) introduce a new figure, (2) synthesize information contained in that figure, and (3) report some set of model performance metrics for each run (e.g., spatial correlation coefficients). That said, the authors do explore the effects of resolution on updraft velocity, which does start to get at questions of 'why', but their analysis of this is somewhat superficial, and as discussed in the section below, it misses some key literature that could enrich their analysis and discussion of this.

In summary, the authors should dig quite a bit more deeply in the analysis of their results. I would hope to see mini-hypotheses and hypothesis tests for some of the interesting intra experimental differences that they show.

We thank the reviewer for the suggestion to include more analysis and discussion of the underlying reasons of the differences across multiple experiments. Now we added more analysis and discussion in the manuscript. A new "supporting material" document is added with substantial amounts of figures to explain the model performance. The difference in simulating the precipitation distribution among the experiments is mainly due to the difference in wind shear structure simulated during the Meiyu front of the East Asian summer monsoon. Now more discussion about this is added. For resolution-dependence analysis, Fig. 13 is added to show the relationship between precipitation and upward moisture flux following Rauscher et al. (2016) and O'Brien et al. (2016). The mechanism underlying the simulated resolution-dependent precipitation and updraft is discussed.

Now our discussion of the model results has been substantially enhanced. Here, we list some of the added text in the main manuscript as follows:

"The first precipitation peak was generated by the southwest-northeast wind shear line formed over Central East China along with a vortex over the Southwest at 00 UTC of 26 June. The shear line gradually extended eastward, leading to the second precipitation peak around 00 UTC of 27 June (Fig. S4 in the supporting material). All four experiments generally simulate the Southwest vortex and wind shear during the event, although the strength and location do not match perfectly with the reanalysis. As the large-scale environment is quite well represented in the model, the simulations also generally capture the two peaks of precipitation along 31°N as observed. However, both U15km and V16km simulate a broader rain belt, resulting in positive biases of precipitation south of 30°N (Fig. S5 in the supporting materials). Both simulations shift the first peak precipitation southward. In addition, the simulations extend the first peak precipitation period and shorten the second one to some extent (Fig. S5). The lower averaged total precipitation around 31°N from the simulation with the NTD parameterization (Fig. 3) is mainly due to the lower rainfall before 26 June compared to the one with the GF parameterization (Fig. S5). For the two precipitation peaks, the simulation with NTD is comparable to the one with GF. Although the two convective parameterizations lead to significant difference in simulating total precipitation before 26 June, both simulations generate consistent wind circulations at 700 hPa before 26 June with spatial correlation coefficients above 0.9 (over the domain as shown in Fig. S4). Although the two convective parameterizations lead to different total precipitation, they have negligible impact on the consistency in modeling precipitation propagation using uniform and variable resolutions during this event."

"The northward shift of rain belt during the event (shown in Fig. 6 and 7) is related to the GFS forecast that only produced the second peak of precipitation around UTC 0000 of 27 June while totally missing the first peak. In addition, the GFS forecast overestimates the second peak and shift it towards the north by about 4°. The timing and location shift of the rain belt in the GFS forecast are mainly because of the bias in simulating the wind shear in this event. The GFS forecast failed to produce the southwest-northeast wind shear line around UTC 0000 of 26 June, and generated too broad vortex over the west. Around UTC 0000 of 27 June, GFS simulated the wind shear line but locating it further north (Fig. S8 in the supporting material)."

"It is interesting to note that MPAS and GFS forecasts, sharing the same initial condition, simulate different large-scale circulation particularly the wind shear structure with the system evolving (Fig. S8). The model capability in successfully capturing the wind shear structure during this event determines the performance in generating the rain belt evolution. The formation and evolution of wind shear during the Meiyu front over East China have been found interacting with multiscale processes and systems, including terrain and convective latent heat (Yao et al., 2017). Different representation of the terrain over East China in various resolutions may impact the simulated wind shear structure. Previous studies also found that convective latent heat may vary with resolutions and physics (Hagos et al., 2013; Zhao et al., 2016), which can further affect the simulation of wind shear structure. Therefore, the difference in resolution and physics between MPAS and GFS may result in their difference in simulating the formation and evolution of wind shear structure during the event. A more detailed exploration of the differences between the MPAS and GFS simulations is beyond the scope of this study."

- ***Missing discussion of key literature***

  The authors devote a significant portion of their analysis and discussion to the connection between vertical velocity and precipitation. This is good, but considering how significant this discussion is to the paper, the authors should discuss how these results relate to a number of recent papers on this connection.

  Specifically, there are currently 3 theories in recent literature for why vertical velocity depends on resolution (with the subtext in these manuscripts that these theories can help explain the resolution dependence of precipitation):

  Rauscher, S.A. et al. "A Multimodel Intercomparison of Resolution Effects on Precipitation: Simulations and Theory." Climate Dynamics 47, no. 7–8 (October 27, 2016): 2205–18. doi:10.1007/s00382-015-2959-5.

  Jeevanjee, N. "Vertical Velocity in the Gray Zone." Journal of Advances in Modeling Earth Systems 9, no. 6 (October 2017): 2304–16. doi:10.1002/2017MS001059.

  Herrington, A.R., and K.A. Reed. "An Explanation for the Sensitivity of the Mean State of the Community Atmosphere Model to Horizontal Resolution on Aquaplanets." Journal of Climate 30, no. 13 (July 2017): 4781–97. doi:10.1175/JCLI-D-16-0069.1.

  Rauscher et al. suggest that the resolution dependence results from an interaction between the constraint of fluid continuity and macro-scale turbulence. Jevanjee suggests that the resolution dependence is related to the aspect ratio of ascending parcels, which he argues scales with resolution. Herrington and Reed suggest that the resolution dependence is related to the horizontal-wavelength-dependent growth rate of buoyancy wave instabilities.

  At a minimum, this manuscript should discuss these theories, and it would be interesting if the authors provided some sort of analysis that attempts to evaluate these theories in this model. I would also suggest that the authors refer to two other relevant manuscripts: O'Brien et al. (2016) and Fildier et al. (2018), who quantitative descriptions of the connection between vertical velocity and extremes (which the authors refer to qualitatively at the end of Section 3).

  O'Brien, T.A. et al. "Resolution Dependence of Precipitation Statistical Fidelity in Hindcast Simulations." Journal of Advances in Modeling Earth Systems 8, no. 2 (June 2016): 976–90. doi:10.1002/2016MS000671.

  Fildier, B. et al. "Prognostic Power of Extreme Rainfall Scaling Formulas Across Space and Time Scales." Journal of Advances in Modeling Earth Systems 10, no. 12 (2018): 3252–67. doi:10.1029/2018MS001462.

We thank the reviewer for pointing us to previous studies on the resolution dependence of precipitation simulations. We added more discussion about these studies in the result and discussion parts of the manuscript. We also added Fig. 13 to show the relationship between precipitation and upward moisture flux following previous studies of Rauscher et al. (2016) and O'Brien et al. (2016). The mechanism underlying the simulated resolution-dependent precipitation and updraft is discussed.

Changes in the text are highlighted as follows:

"Previous studies have proposed some mechanisms underlying the resolution impacts on modeling vertical velocity (e.g., Rauscher et al., 2016; Jeevanjee et al., 2017; Herrington and Reed, 2017; O'Brien et al., 2016; Fildier et al., 2018). Among these mechanisms, Rauscher et al. (2016) argued that the resolution-dependent vertical velocity is caused by the interaction between the constraint of fluid continuity and macro-scale turbulence. They suggested that the vertical velocity should be more intense at higher resolution because the horizontal velocity increment follows approximately a power law of resolution. Therefore, the resolved vertical transport must increase as grid spacing decreases. Assuming atmospheric moisture is relatively insensitive to resolution, the upward moisture flux should increase as grid spacing decreases, hence producing more precipitation.

Figure 13 shows the PDFs of the upward moisture flux and the relationship between hourly precipitation versus upward moisture flux at 850hPa during the event from the MPAS simulations at 60km, 30km, 16km and 4km. It is evident that the simulations at higher resolutions produce more frequent intense upward moisture fluxes at 850hPa, consistent with Rauscher et al. (2016) and O'Brien et al. (2016). Rauscher et al. (2016) found a linear relationship between precipitation and upward moisture fluxes at lower level. The relationship lines from this study as shown in Fig. 13 parallel the 1:1 reference line for all resolutions. However, the lines are consistently below the reference line for convection-permitting simulations (4km) and above the reference line for hydrostatic simulations with convective parameterization (e.g., 16km, 30km, 60km). The simulated precipitation can be larger than the lower level upward moisture fluxes at hydrostatic scale because part of the precipitation is contributed by the convective parameterization rather than contributed by the resolved upward moisture flux (Rauscher et al., 2016). On the contrary, the precipitation could be lower than the upward moisture flux at convection-permitting scale (e.g., 4km) as moisture is removed from the cloud updrafts due to detrainment (e.g., O'Brien et al., 2016). Overall, our results of the resolution-dependent updraft and precipitation are consistent with Rauscher et al. (2016) and O'Brien et al. (2016)."

"Previous studies (Xue et al., 2007; Clark et al.) noted the importance of ensemble simulations in predicting heavy precipitation. Due to computational limitation, only one set of experiments with different physics and resolutions are evaluated in this study. The MPAS simulations of heavy precipitation with different initial conditions and refinement sizes deserve more evaluations. Finally, some studies noted that convection-permitting modeling does not always add values in simulating heavy precipitation compared to hydrostatic scale modeling (e.g., Kain et al., 2008; Rhoades et al., 2018; Xu et al., 2018). Rhoades et al. (2018) found that the improvement by increasing resolution may depend on cloud microphysics parameterization. Increasing horizontal resolution alone sometimes can even lead to worse model performance. The impacts of increasing horizontal resolution on the overall model performance in simulating extreme precipitation may also be affected by the model structure and coupling among model components and processes (Jeevanjee et al., 2016; O'Brien et al., 2016; Herrington et al., 2017, 2018; Gross et al., 2018). This study also found some sensitivity of modeling extreme precipitation to cloud microphysics, particularly at convection-permitting scale."

- ***Lack of statistics***
  The authors make a variety of quantitative statements comparing across simulations or between simulations and observations: e.g., "As a result, the correlation coefficients between the observations and the MPAS experiments at the resolutions of 60 km, 30 km, 16 km, and 4 km are 0.20, 0.21, 0.29, 0.50 (WSM6), and 0.42 (Thompson), respectively" (lines 422-424). However, the authors do not provide any estimates of uncertainty in these quantities, which makes it difficult to assess whether they are significant. I would expect that many of them are, but if a core goal of this paper is to assess how model skill changes with resolution, the authors should be certain that their claims are statistically robust. I see two straight-forward ways to assess uncertainty: bootstrap confidence intervals (e.g., bootstrap sample from spatial points), or running ensembles. Ideally, the authors would run more ensemble members, but I recognize that computational constraints may prohibit that. At the very least, a bootstrap analysis would allow the authors to state the sampling uncertainty in the correlation coefficients.
  Related to this, it does concern me that all of the conclusions in this manuscript are based on single-member ensembles of a single event. Would these results hold if the authors simulated another event, perhaps in another season, or even if the authors ran another ensemble member? The authors should at very least acknowledge this limitation of their study, and at best run a few additional simulations to explore whether new simulations qualitatively alter their conclusions.

We thank the reviewer for the suggestion to test the statistical significance. Due to the large computing cost and data storage, particularly for the U15km and V4km experiments, we cannot afford to perform ensemble simulations in this study. Instead, we take the suggestion by the reviewer to use the bootstrap sampling method to test the statistical significance. The statistical test confirms that the difference among the experiments is statistically significant. We added Table 2 with the summary of statistical metrics and more analysis in the revised manuscript. The conclusion of this study does not change. Now we acknowledge this in the method as follows:

"Due to the large computing cost and data storage of the experiments conducted, particularly for the U15km and V4km experiments, this study does not perform ensemble simulations. Instead, the bootstrapping statistical analysis is used to test the statistical significance of the difference among multiple experiments investigated in this study. The bootstrap method uses resampling technique to extract certain samples, called bootstrap samples, within the range of the original data. Statistical metrics such as averages, variances, correlation coefficient, can be calculated for each bootstrap sample. For a given confidence level (e.g., 95%), bootstrap confidence intervals of specific statistical metric can be estimated (e.g., Efron, 1992; Efron and Tibshirani, 1994)."

We added Table 2 to summarize the statistical metrics and more discussion about the statistical significant tests with the bootstrap method has been added in the text:

"In order to test the statistical significance of the difference among the experiments, the 95% confidence intervals of spatial correlation are estimated based on the bootstrap analysis. Although the correlation coefficients estimated above have an uncertain range, at the 95% confidence level the results still indicate that the V16km simulation produces better spatial pattern of precipitation than other hydrostatic-scale simulations. In addition, the simulation at the convection-permitting scale is comparable to, if not better than, the V16km simulation. The results are summarized in Table 2."

"At the 95% confidence level (Table 2), the experiments at 16 km and 4 km are comparable in terms of simulating the propagation of the rain belt and better than the experiments at other resolutions."

"The statistical significance tests based on the bootstrap analysis indicate that at the 95% confidence level the model performance at 16 km and 4 km in terms of simulating vertical structure of winds are comparable and better than the simulations at coarser resolution (Table 2)."

"The statistical significance tests based on the bootstrap analysis indicate that at the 95% confidence level the model performance at 16 km and 4 km in terms of simulating vertical structure of winds are comparable and better than the simulations at coarser resolution (Table 2)."

"The statistical significance test based on the bootstrap analysis indicates that at the 95% confidence level the simulations at 4 km can better capture the spatial distribution of heavy precipitation than the simulations at resolutions of hydrostatic scale (Table 2)."

This study focuses on evaluating the MPAS simulations of heavy precipitation over East China (Yangtze River Delta Region). Most of heavy precipitation events over East China occurred in summer, so we focused on the events in summer instead of other seasons. The event we selected in this study was reported as one of the most influential precipitation events in summer of 2012 over East China. See our response to your other comments about the reason selecting this event.

Due to the computational limitation, we only run one set of experiments with different physics and resolutions. We agree that it may not represent all the cases. We acknowledge that one case study is not enough to fully evaluate the MPAS performance over East China. Now, we added the discussion in the summary section:

"Due to the computational limitation, only one set of experiments with different physics and resolutions are evaluated. The MPAS simulations of heavy precipitation over East China with different initial conditions and refinement sizes deserve more evaluations."

"more events of heavy precipitation over East China should be investigated in the future to more systematically evaluate the MPAS variable-resolution modeling framework and the impacts of resolution and physical parameterizations."

- ***lines 100-101: "regional models limit feedback to global scale". This is true, but is that really a good point to make in this paper? The simulations only run for ~5 days, so there is very little time for feedback--e.g., from rossby wave propagation--to feedback onto global scales. In my opinion, that makes this point a bit irrelevant, and almost misleading for this paper, since it could be read to imply that this gives these variable-resolution simulations an advantage in this experimental design over limited-area model approaches.***

Thanks for this suggestion. Since this study is about forecasting of an extreme event, the feedback we care about is the feedback that would affect the forecasting of the extreme precipitation event. In this context, the feedback is through condensational heating that generates potential vorticity and excites Rossby waves that influence the storm. 5-day is long enough to see the impacts of the storm on mesoscale and large-scale circulation that influence the storm. Now we correct the sentence in the text as "…and also limit regional feedback to large-scale circulation".

- ***lines 129-131: Why were these dates chosen? Presumably it is because it is a representative, strong event; or perhaps it was chosen randomly. But a cynical version of that answer could be 'because it was a date for which the model looked very good'--I truly doubt this is case, but without any discussion of the motivation for choosing this date, a reader could wonder if this date was cherry-picked.***

This study focuses on evaluating the MPAS simulations of heavy precipitation over East China (Yangtze River Delta Region). Most of heavy precipitation events over East China occurred in summer. The event we selected in this study was reported as one of the most influential precipitation events in summer of 2012 over East China. We agree that a single event may not represent all the cases. As we respond to one comment above, we acknowledge that one case study is not enough for fully evaluating the MPAS performance over East China. Now, we added the clarification in the introduction section as "A heavy precipitation event that occurred on June 25-27 of 2012 over the YRD of East China, one of the ten heaviest precipitation events in 2012, is selected. This rainfall event was triggered by a typical southwest vortex in the middle and high troposphere and wind shear in the lower layer of Meiyu front over East China during the East Asian summer monsoon (e.g., Xiang et al., 2013; Yao et al., 2017), initiated around 1200 UTC of 25 June. During this period, a heavy precipitating system propagated along the Yangtze River and produced as much as 244 mm of precipitation in 24 hours at some locations. The continuous precipitation led to 17 deaths and about RMB 3.68 billion in total damage, and affected more than 685 million people in the provinces of Central and East China."

We added the discussion in the summary section:
"Due to the computational limitation, only one set of experiments with different physics and resolutions are evaluated. The MPAS simulations of heavy precipitation over East China with different initial conditions and refinement sizes deserve more evaluations."
"more events of heavy precipitation over East China should be investigated in the future to more systematically evaluate the MPAS variable-resolution modeling framework and the impacts of resolution and physical parameterizations."

- ***lines 235-241: The authors provide very little discussion on the meteorological conditions preceding the event, or of the initial condition. This limits the author's and the reader's ability to interpret differences between the simulations and observations. For example, was the Meiyu front already present and propagating in the initial condition, or did it form in the day or two preceding (presumably it was already present)?***

The event studied occurred mainly on June 25-27. This event was triggered by a typical southwest vortex in the middle and high troposphere and wind shear in the lower layer of Meiyu front over East China during the East Asian summer monsoon. The wind shear structure formed on June 25. The simulation started on June 23. Therefore, the wind shear structure is not present in the initial condition. More discussion about the synoptic condition of this event is added in the text:
"A heavy precipitation event that occurred on June 25-27 of 2012 over the YRD of East China, one of the ten heaviest precipitation events in 2012, is selected. This rainfall event was triggered by a typical southwest vortex in the middle and high troposphere and wind shear in the lower layer of the Meiyu front over East China during the East Asian summer monsoon (e.g., Xiang et al., 2013; Yao et al., 2017), initiated around 1200 UTC of 25 June. During this period, a heavy precipitating system propagated along the Yangtze River and produced as much as 244 mm of precipitation in 24 hours at some locations. The continuous precipitation led to 17 deaths and about RMB 3.68 billion in total damage, and affected more than 685 million people in the provinces of Central and East China."

We added the discussion about the formation and evolution of the system:
"The first precipitation peak was generated by the southwest-northeast wind shear line formed over Central East China along with a vortex over the Southwest at 00 UTC of 26 June. The shear line gradually extended eastward, which led to the second precipitation peak around 00 UTC of 27 June (Fig. S4 in the supporting material)."

- ***line 246: The authors should use ERA5 instead. It covers this date, it has a significantly higher resolution, and the data are very easy to obtain either directly from ECMWF or from the NCAR RDA.***

Thanks for your suggestion. Now we change the reanalysis data from ERA-interim to ERA5 in the figures and revise the related discussion. The results and conclusion are similar.

Anonymous Referee #2

General comments:

- ***Zhao et al. in "Modeling extreme precipitation over East China with a global variable resolution modeling framework (MPASv5.2) Impacts of resolution and physics" recreate an extreme weather event that occurred over East Asia between 25-27 June 2012. This study presents a comprehensive look at the performance of the Model for Prediction Across Scales (MPAS) across both hydrostatic and non-hydrostatic scales (e.g., 60km to 4km), uniform and variable-resolution grid-spacing, and three different microphysics schemes (one of which is "scale aware" for convective/resolved precipitation). The authors assess MPAS skill compared with CMA observations in East Asia and tersely compare results to GFS forecasts over a single-member, sub-weekly simulation period.***

  Overall, I think the paper is well written and fits within the scope of GMD and could be, given a bit more work, a valuable contribution to the scientific community, particularly due to its emphasis on evaluating the use new variable-resolution global climate models for extreme event recreation and sub-weekly weather forecasting. However, I think there are still several major revisions that need to happen prior to this paper being accepted. I would suggest that the editor assign major revisions to this manuscript.

We thank the reviewer for the detailed review. The comments help a lot on improving the quality of the manuscript.

In the revised manuscript, we provided a more nuanced assessment of simulations. The difference between observations and simulations are highlighted to provide a more balanced discussion of model skill. Due to the large amounts of experiments conducted in this study, the computational and storage demands are very large, preventing the ensemble simulations. Following the other reviewer's suggestion, the bootstrapping statistical analysis is used to test the statistical significance of the difference among multiple experiments. The statistical metrics are now summarized in Table 2 in the revised manuscript. In terms of resolution-dependence analysis, more discussion about previous related studies are also added. Other text and figures have also been revised as the reviewer suggested. A new document on supporting material is added in the revised manuscript with substantial amounts of figures to support some statements and address the comments.

Major comments:

- ***1) Given that this paper centers around the recreation of one weather event, why did the authors not perform an ensemble of simulations with slightly perturbed initial conditions to highlight internal variability impacts on precipitation intensity and spatial distribution? Was the computational demand too high to do so? If so, as mentioned below, it would benefit the reader to know this type of information***

explicitly. If not, why not perform, at least, a small ensemble of simulations (as the authors state is needed for GFS in Section 4 – line 529).

a) Line 213-215 – This might be a good time to bring up real-time computational demand (e.g., nodes used, simulated years per actual day, etc.) and physics/dynamics timesteps across "U" and "V" cases.

Due to the large amounts of experiments conducted in this study, the computational and storage demands are very large, preventing the ensemble simulations. Instead, the bootstrapping statistical analysis is used to test the significance of statistical difference among multiple experiments investigated in this study. Now it is clarified in the text "Due to the large computing cost and data storage of the experiments conducted, particularly for the U15km and V4km experiments, this study does not perform ensemble simulations. Instead, the bootstrapping statistical analysis is used to test the significance of statistical difference among multiple experiments investigated in this study."

Now we also added more information about the configurations of multiple experiments in the text:

"The difference in the number of mesh cells leads to a difference in computational and storage demand. With the TH-2 supercomputer of National Supercomputer Center in Guangzhou (NSCC-GZ), it takes ~9000 CPU hours and ~240 CPU hours to finish one-day simulation for U15km and V16km resolutions, respectively. In addition, with the standard MPASv5.2, the sizes of output data per one-day simulation for U15km and V16km are 0.5 T and 0.02 T, respectively. The same time step of 60 second is used for physics and dynamics for both U15km and V16km simulations."

"The numbers of grid cells in the U60km, V30km, V16km, and V4km meshes are ~0.16 million, 0.10 million, ~0.11 million, and ~0.8 million, respectively. Difference in the number of cells and minimum cell size also leads to the difference in computational and storage demand. With the TH-2 supercomputer of NSCC-GZ, it takes ~200 CPU hours, ~150 CPU hours, ~240 CPU hours, and ~1800 CPU hours to finish one-day simulation for U60km, V30km, V16km, and V4km meshes, respectively. In addition, with the standard MPASv5.2, the sizes of output data per one-day simulation for the four meshes are 0.03 T, 0.02 T, 0.02 T, and 0.15 T, respectively. The time steps used for physics and dynamics for the four meshes are 300 seconds, 120 seconds, 60 seconds, and 20 seconds, respectively."

- ***2) The authors should highlight other variable-resolution modeling efforts to give readers a sense that there are a community of models now available.***

  a) Line 106 – Given that the authors examine both hydrostatic and non-hydrostatic configurations of MPAS, the authors should also point to the literature of hydrostatic variable-resolution global climate models such as variable-resolution CESM (etc.) as these variable-resolution options have been used extensively for

various applications (e.g., Rauscher et al., 2013; Zarzycki et al., 2014, 2015; Rhoades et al., 2016; Huang et al., 2016; Wu et al., 2017; Gettelman, et al., 2018; Wang et al., 2018; Burakowski et al., 2019).

Although this study investigated some MPAS simulations at the hydrostatic scales, the MPAS model used in all the experiments in this study is the fully compressible non-hydrostatic version (MPAS-A v5.2) as described in the manuscript. We did not examine the hydrostatic configurations of MPAS. However, we thank the reviewer for pointing out previous literatures using other variable-resolution models such as CESM. Now these previous studies are mentioned and the literatures are cited in the text "Although global hydrostatic variable-resolution climate models, such as the variable-resolution version of Community Earth System Model, have been used extensively for various applications (e.g., Rauscher et al., 2013; Zarzycki et al., 2014, 2015; Rhoades et al., 2016; Huang et al., 2016; Wu et al., 2017; Gettelman, et al., 2018; Wang et al., 2018; Burakowski et al., 2019), so far few studies used global non-hydrostatic variable-resolution models to investigate weather or climate simulations, particularly at convection-permitting scales (e.g., Prein et al., 2015)."

- ***3) This manuscript would also benefit from the discussion of previous studies that have shown that solely refining horizontal resolution alone has led to differing results in simulated precipitation bias across various models (including variable-resolution approaches).***

  a) Line 83-87 – There also have been studies showing that solely refining horizontal resolution alone can lead to unexpected "oscillations" between positive/negative simulated bias in daily-to-seasonal average precipitation too. For example, refining horizontal resolution from 55km to 28km has been shown to improve various assessments of simulated bias (e.g., orographic precipitation, hurricanes, atmospheric rivers, etc.), however refining resolution from 28km to 14km has shown an enhancement of bias (Rhoades et al., 2018; Xu et al., 2018). These differences have been shown to be bounded in theory and model structure decisions (Jeevanjee et al., 2016; O'Brien et al., 2016; Herrington et al., 2017, 2018; Gross et al., 2018). The authors should discuss these studies as well to give the readers clear perspective that resolution alone will not be the sole solution to better representations of extreme precipitation.

Thanks for the reviewer in pointing out the interesting and related studies. We agree that resolution alone will not always improve the simulations. Now we added more discussion about previous studies to highlight the resolution impacts. We clarified in the text in the introduction:

"Although not a panacea for weather and climate modeling (NRC, 2012), previous studies suggested that increasing grid resolution could significantly improve modeling of extreme precipitation because the impacts of topography, land-use, land-atmosphere interaction, and other important processes are better resolved (e.g., Giorgi and Mearns, 1991; Giorgi and Marinucci, 1996; Leung et al., 2003; Bacmeister et al. 2014; ECMWF2016)."

We added the text in the discussion:

"Finally, some studies noted that convection-permitting modeling does not always add values in simulating heavy precipitation compared to hydrostatic scale modeling (e.g., Kain et al., 2008; Rhoades et al., 2018; Xu et al., 2018). Rhoades et al. (2018) found that the improvement by increasing resolution may also depend on cloud microphysics parameterization. Increasing horizontal resolution alone sometimes can even lead to worse model performance. The impacts of increasing horizontal resolution on the overall model performance in simulating extreme precipitation may also be affected by the model structure and coupling among model components and processes (Jeevanjee et al., 2016; O'Brien et al., 2016; Herrington et al., 2017, 2018; Gross et al., 2018). More events of heavy precipitation over East China should be investigated in the future to more systematically evaluate the MPAS variable-resolution modeling framework and the impacts of resolution and physical parameterizations."

b) Figure 4 and 8 – These two plots somewhat prove the point made above that if CMA is used as the simulation skill benchmark and by eye, another reason Figure 4 and 8 should be difference plots, V16km seems to get the timing of the two locations of precipitation maxima and magnitudes the most correct over the storm track (i.e., 30 N +/-2 deg), whereas at V4km the precipitation magnitudes seem too positively biased over a greater area and longer time. I think these plots also highlight the potential need for an ensemble of simulations given that the time-space structure of precipitation in each of the simulations is quite different and could simply be due to using one realization of the atmospheric internal variability.

Thanks for your comment, and we agree that the simulated timing and magnitudes do not perfectly match the observations. We show the simulated temporal-spatial distribution of precipitation instead of the difference against the observation, because in this way the simulated propagation of rain belt can be better demonstrated although there are biases. From both Fig. 4 and 8, we can see clearly some simulations can capture the propagation better than others, particularly for V16km.WSM6 and V4km.WSM6. Now we added Fig. S5 and Fig. S9 in the supporting materials to show the difference between the simulations and observations. We also added more discussion about the difference in the text:

"All four experiments generally simulate the southwest vortex and wind shear during the event, although the strength and location do not match perfectly with the reanalysis. As the large-scale environment is quite well represented in the model, the simulations also generally capture the two peaks of precipitation along 31°N as observed. However, both U15km and V16km simulate a broader rain belt, resulting in positive biases of precipitation south of 30°N (Fig. S5 in the supporting materials). Both simulations shift the first peak precipitation southward. In addition, the simulations extend the first peak precipitation period and shorten the second one to some extent (Fig. S5 in the supporting materials). The lower averaged total precipitation around 31°N from the simulation with the NTD parameterization (Fig. 3) is mainly due to the lower rainfall before 26 June compared to the one with the GF parameterization (Fig. S5). For the two precipitation peaks, the simulation with NTD is comparable to the one with GF. Although the two convective parameterizations lead to significant difference in simulating total precipitation before 26 June, both simulations generate consistent wind circulations at 700 hPa before 26 June with spatial correlation coefficients above 0.9 (over the domain as shown in Fig. S4 in the supporting material). Although the two convective parameterizations lead to different total precipitation, they have negligible impact on the consistency in modeling precipitation propagation using uniform and variable resolutions during this event."

"The northward shift of rain belt during the event (shown in Fig. 6 and 7) is related to the GFS forecast that only produced the second peak of precipitation around UTC 0000 of 27 June while totally missing the first peak. In addition, the GFS forecast overestimates the second peak and shift it towards the north by about 4°. The timing and location shift of the rain belt in the GFS forecast are mainly because of the bias of GFS in simulating the wind shear in this event. The GFS forecast failed to produce the southwest-northeast wind shear line around UTC 0000 of 26 June and generated too broad vortex over the west. Around UTC 0000 of 27 June, GFS simulated the wind shear line but locating it further north (Fig. S8 in the supporting material)."

"All simulations roughly produce the two peaks of precipitation as observed during the event. However, the experiment at 60 km simulates the first precipitation peak southward and the second peak northward of the observations, while the experiment at 30 km simulates the second peak further south and a few hours earlier. The time and location shift corresponding well to biases in simulated wind shear (Fig. S8). The spatial correlation coefficients of precipitation are 0.30 and 0.32 between the observations and the simulations at 60 km and 30 km, respectively. The experiments at 16 km and 4 km with the WSM6 cloud microphysics scheme can better capture the timing and latitude of the observed precipitation event than U60km and V30km (Fig. S9 in the supporting materials), however both V16km and V4km overestimate the first peak precipitation and underestimate the second peak. The experiment at 4 km with the Thompson scheme overestimates the precipitation amount of both peaks."

In terms of ensemble simulations, due to the large amounts of experiments conducted in this study, the computational and storage demands are very large. Instead, the bootstrapping statistical analysis is now added to test the significance of statistical difference among multiple experiments investigated in this study. Table 2 is added to summarize the statistical analysis. Now it is also clarified in the text "Due to the large computing cost and data storage of the experiments conducted, particularly for the U15km and V4km experiments, this study does not perform ensemble simulations. Instead, the bootstrapping statistical analysis is used to test the significance of statistical difference among multiple experiments investigated in this study."

- ***4) Given the emphasis on microphysics choice (e.g., three different ones used in this manuscript), a more in-depth discussion of each of the microphysics schemes should be presented in Section 2. In my opinion, the reader should be able to glean some of the tradeoffs of each of the microphysics schemes within the text and not just be referred to other publications.***
  a) Line 186 – Highlight a bit more detail about the microphysics schemes used as they can impact the spatial distribution of extreme precipitation that you discuss later on. For example, one-moment vs two-moment schemes, diagnostic vs prognostic, which hydrometeor species are represented in each scheme (i.e., rain, snow, graupel, etc.), what are their assumptions in drop velocity, horizontal advection, etc. A new book chapter has been published that could be a good lead on this as well (Gettelman et al., 2019).

Thanks for your suggestion. This study used two cloud microphysics schemes available in MPAS, i.e., WSM6 and Thompson. Now more details about the two schemes are added in Section 2 as "For cloud microphysics, the WSM6 (Hong and Lim, 2006) and Thompson (Thompson et al., 2008) schemes, both of which are bulk microphysical parameterizations, are selected. Both schemes include six hydrometeor species: water vapor, cloud water, rain, cloud ice, snow, and graupel. The WSM6 scheme is a one-moment prognostic parameterization, while the Thompson scheme includes a two-moment prognostic parameterization for cloud ice and the single-moment parameterization for the other hydrometeor species. The two schemes apply the same formula of gamma distribution of hydrometeor species: $N(D) = N_0 D^{\mu} e^{-\lambda D}$, where D is the particle diameter, N_0 is the intercept parameter, μ is the shape factor, and λ is the slope parameter, although the parameter values or functions vary in the two schemes. The mass-size relationship in WSM6 and Thompson is also expressed in the same formula as $m(D) = aD^b$. The mean falling speed is calculated as $V(D) = cD^d (\frac{\rho_0}{\rho})^{0.5}$ in WSM6 and $V(D) = cD^d (\frac{\rho_0}{\rho})^{0.5} exp\,(-fD)$ in Thompson, respectively (Hong and Lim, 2006; Thompson et al., 2008). In the formula, the WSM6 scheme assumes a power-law fit between terminal velocity and particle size as Locatelli and Hobbs (1974), while the Thompson scheme incorporates an exponential decay parameter to allow for a decrease in falling speed with increasing size (Molthan et al., 2012)."

- ***5) The authors could give a more nuanced assessment of MPAS skill. Based on***

how the text has been written, MPAS never seems to perform poorly, yet correlation coefficients and spatial structures of the storm events in the figures clearly show important differences compared with observations. In addition, the authors should be very clear over which area the correlation coefficients are being computed given that they are used throughout the text. I presume these correlations are computed over the entire domain. Given that this study is evaluating performance over a single weather event, shouldn't the correlation coefficients be computed over the "mean" track of the event (i.e., using CMA as reference, 30 N +/- 2 deg lat)

Now we added more discussion about the difference between the observations and MPAS results, particularly for Fig. 4 and 8 (see our response to your other comments). We added clarification in the text to indicate that the spatial correlation coefficients were calculated for the entire regions shown in the specific figures. Besides the entire region of East China, we did evaluate the performance over the rain belt region (27°N-32°N and 110°E-122°E) as shown in Fig. 10 and include the discussion in Section 3.2.3.

a) ***Shouldn't ERA5 rather than GFS be used for forecast comparison skill? ERA5 resolution is much more closely aligned with MPAS horizontal resolutions used in this study.***

The publicly available ERA5 forecast is only for 18-hour, so it cannot be used as the reference to compare with the 5-day MPAS forecast starting from 0000 UTC of 23 June 2012. However, now we changed the reanalysis data from ERA-interim to ERA5 in all the related figures and analysis.

b) ***Line 289-290 – A correlation coefficient of 0.48 and 0.42 for the GF scheme simulations doesn't indicate to me that these simulations reproduce the observed precipitation propagation. I agree with your later assessment that the differences between "U" and "V" simulations are small, especially for NTD, (which is an interesting result), but to say this compares well with observed is a bit misleading. I think this statement (and others like it) must be a bit more caveated and highlight the negatives/positives of the simulations more clearly.***

In the revised manuscript, now we provided a more nuanced assessment of simulations. Now more discussion about the difference between the observations and simulations are added as "All four simulations show that the rain belt started from the South and eventually stayed around 31°N. The simulations also generally show the two peaks of precipitation along 31°N as observed. However, both U15km and V16km simulate broader rain belt, resulting in positive biases of precipitation south of 30°N (Fig. S5 in the supporting materials). Both simulations shift the first peak precipitation southward. In addition, the simulations extend the first peak precipitation period and shorten the second one to some extent (Fig. S5 in the supporting materials)." Now Fig. S4 is also added in the supporting materials to show the difference between the observations and simulations.

c) ***Figure 8 – This plot (as indicated below) should be remade into a difference plot. If one uses CMA as reference, it appears that the precipitation maxima during this event occurs too soon (i.e., one-day) across all of the simulations.***

Thanks for your comment, and we agree that the simulated timing and magnitudes do not perfectly match the observations. See our response to your comment above. Now we added Fig. S9 in the supporting materials to show the difference between the simulations and observations. We also added more discussion about the difference shown in Fig. 8 and Fig. S9 in the text:

"The northward shift of rain belt during the event (shown in Fig. 6 and 7) is related to the GFS forecast that only produced the second peak of precipitation around UTC 0000 of 27 June while totally missing the first peak. In addition, the GFS forecast overestimates the second peak and shift it towards the north by about 4°. The timing and location shift of the rain belt in the GFS forecast are mainly because of the bias of GFS in simulating the wind shear in this event. The GFS forecast failed to produce the southwest-northeast wind shear line around UTC 0000 of 26 June and generated too broad vortex over the west. Around UTC 0000 of 27 June, GFS simulated the wind shear line but locating it further north (Fig. S8 in the supporting material)."

"All simulations roughly produce the two peaks of precipitation as observed during the event. However, the experiment at 60 km simulates the first precipitation peak southward and the second peak northward of the observations, while the experiment at 30 km simulates the second peak further south and a few hours earlier. The time and location shift corresponding well to biases in simulated wind shear (Fig. S8). The spatial correlation coefficients of precipitation are 0.30 and 0.32 between the observations and the simulations at 60 km and 30 km, respectively. The experiments at 16 km and 4 km with the WSM6 cloud microphysics scheme can better capture the timing and latitude of the observed precipitation event than U60km and V30km (Fig. S9 in the supporting materials), however both V16km and V4km overestimate the first peak precipitation and underestimate the second peak. The experiment at 4 km with the Thompson scheme overestimates the precipitation amount of both peaks."

Minor comments:

- ***Line 217-221 – The authors may want to include this analysis of minimal precipitation difference in the supplemental for reader clarification.***

Now we added the figures in the supplemental materials showing the total and resolved precipitation from the V16km simulations with both WSM6 and Thompson cloud microphysics. We also added clarification in the text as "The impact of cloud microphysics (WSM6 and Thompson) on the consistency in modeling total precipitation is also examined and is found to be negligible (Fig. S1 and S2 in the supporting materials), although there are some impacts on the simulated grid-resolved precipitation

(Fig. S3 in the supporting materials)."

- ***Line 238 – In MPAS, could reanalysis data also be used to replace the coarse resolution portion of the simulation outside of the refinement regions (i.e., akin to a conventional regional climate model)? This could be an interesting next step for a future publication (unless it has already been done) to limit model drift due to the large-scale boundary conditions.***

This is a good point. It deserves effort in future, but this capability is not in the current released version of MPAS.

- ***Line 263 – Given that you point to the Meiyu front a few times in this text, you may want to spend a bit of time in the introduction to discuss the importance of the Meiyu front for shaping East Asian precipitation and cite some studies for further reading.***

Thanks for your comment. Now more discussion about the synoptic condition of this event is added in the introduction as "A heavy precipitation event that occurred on June 25-27 of 2012 over the YRD of East China, one of the ten heaviest precipitation events in 2012, is selected. This rainfall event was triggered by a typical southwest vortex in the middle and high troposphere and wind shear in the lower layer of Meiyu front over East China during the East Asian summer monsoon (e.g., Xiang et al., 2013; Yao et al., 2017), initiated around 1200 UTC of 25 June. During this period, a heavy precipitating system propagated along the Yangtze River and produced as much as 244 mm of precipitation in 24 hours at some locations. The continuous precipitation led to 17 deaths and about RMB 3.68 billion of total damage, and affected more than 685 million people in the provinces of Central and East China."

- ***Line 283-284 – Again, the authors may want to offer this analysis in the supplemental material to allow for readers to determine how "negligible" the results were between microphysics schemes.***

Now, we added the figures in the supplemental materials showing the total and resolved precipitation from the V16km simulations with both WSM6 and Thompson cloud microphysics. We also added clarification in the text as "The impact of cloud microphysics (WSM6 and Thompson) on the consistency in modeling total precipitation is also examined and is found to be negligible (Fig. S1 and S2 in the supporting materials), although there are some impacts on the simulated grid-resolved precipitation (Fig. S3 in the supporting materials)."

- ***Line 324 – I think the authors should guide the readers intuition on GFS skill at 1 deg and 0.5 deg (seems pretty poor compared with CMA) as a comparison to MPAS. This seems to be a central point of the study that MPAS can offer enhanced***

skill (in some measures) for extreme precipitation forecasts, but seems to be a bit muted in the text. Line 341 – Again, the authors may want to put this analysis that is not shown in the supplemental.

As we mentioned in the manuscript "Since the focus of this study is not to investigate the difference between MPAS and GFS and to evaluate the performance of GFS, details about the GFS are not discussed here but can be found on the website listed above.", we do not want to guide the readers about the performance of GFS. The GFS product is only used as a reference in this study, and the central point of this study is evaluating the MPAS performance against observations. Now the comparison of global distributions is put in the supporting material.

- ***Line 355 – What does "fairly well" indicate? This is an example of the larger major comment that MPAS skill is not evaluated in a nuanced fashion.***

Thanks for your suggestion. "fairly well" is deleted now.

- ***Line 366-368 – If I'm reading the space-time plots correctly, it appears that the peak precipitation of the weather event at 30 N +/- 2 deg consistently occurs a day earlier than expected compared with CMA (save for V4km.Thompson which seems to overly precipitate over several days). Therefore, if this is true, the use of the word "roughly similar" is a bit misleading.***

"roughly the same time" is deleted. Now we added more discussion about the difference between the observations and simulations. See our response to your other comments.

- ***Line 489-490 – Another example of a slightly misleading statement about the MPAS 4km simulation skill. The heavy precipitation is captured much better than other coarser resolution MPAS simulations, but precipitation magnitudes, especially maxima, are high biased and precipitation durations are biased over a much longer time than other resolutions (e.g., Figure 8). In my opinion, this warrants a Table that explicitly states the summary statistics for each of the MPAS simulations (and CMA as well).***

Now we revised the statement and added more discussion in the text as "Among the MPAS experiments with multiple resolutions, the simulations at 4 km can better capture the observed locations of heavy precipitation than the ones at hydrostatic scales, however, the results also show that the simulations at 4 km overestimate the first peak precipitation and underestimate the second one. The simulations at 4 km are also very sensitive to cloud microphysics, which deserves more investigation in future." Table 2 is also added to summarize the statistics among the multiple experiments.

- ***Figure 2 – The authors may want to make the color labels from 0-1 mm/day white instead of blue. This might make it easier to see the vector winds.***

Revised. Thanks for your suggestion.

- ***Figure 4 and 8 – Given that these plots are meant to just show the spatial distribution of precipitation (and not include vector winds) compared with the observed CMA product, the authors may want to provide these plots as difference plots to help readers locate mismatch. This would also highlight how GFS is suboptimal for forecasting extreme precipitation events in East Asia at the moment. Also, in Figure 8, each title should be consistent with the explicit use of WSM6 or Thompson microphysics.***

Now the difference plots are added as Fig. S5 and Fig. S9 in the supporting materials. More discussion is also added. See our response to your other comments. The titles of Fig. 8 are changed as suggested.

- ***Figure 5 – Given that this plot is purely meant to show the partitioning of resolved and parameterized precipitation across resolutions spanning hydrostatic/non-hydrostatic scales, could the authors change the units from mm/day and instead use a % of total precipitation? Again, I would suggest that a white/transparent color be used at the lower end of the color bar as well.***

The fraction plot will show large values over the small total precipitation region and make the result misleading. Now we revised Fig. 5 to adjust the colorbar and added Fig. S6 in the supporting material to show the fraction of resolved and parameterized rain in the total. We also added discussion in the text as "The fraction of parameterized precipitation in the total decreases significantly from the simulations at 16 km to the ones at 4km over the heavy precipitation region (Fig. S6 in the supporting materials). It is also interesting that the fraction of parameterized precipitation increases from the simulations at 60 km to the ones at 16 km to some extent."

- ***Figure 6 – I would suggest that a white/transparent color be used at the lower end of the color bar to focus reader attention and more clearly present vector winds.***

Revised. Thanks for your suggestion.

- ***Table 1 – The table should be standalone; therefore, "U" and "V" should be defined in the caption (i.e., Uniform and Variable Resolution) and WSM6, NTD, GF, etc. should be as well.***

Thanks for your suggestion. Now the note is added with the table.

Supporting materials for "Modeling extreme precipitation over East China with a global variable-resolution modeling framework (MPASv5.2): Impacts of resolution and physics"

[1]Chun Zhao, [1]Mingyue Xu, [1]Yu Wang*, [1]Meixin Zhang, [2]Jianping Guo, [3]Zhiyuan Hu, [4]L. Ruby Leung, [5]Michael Duda, [5]William Skamarock

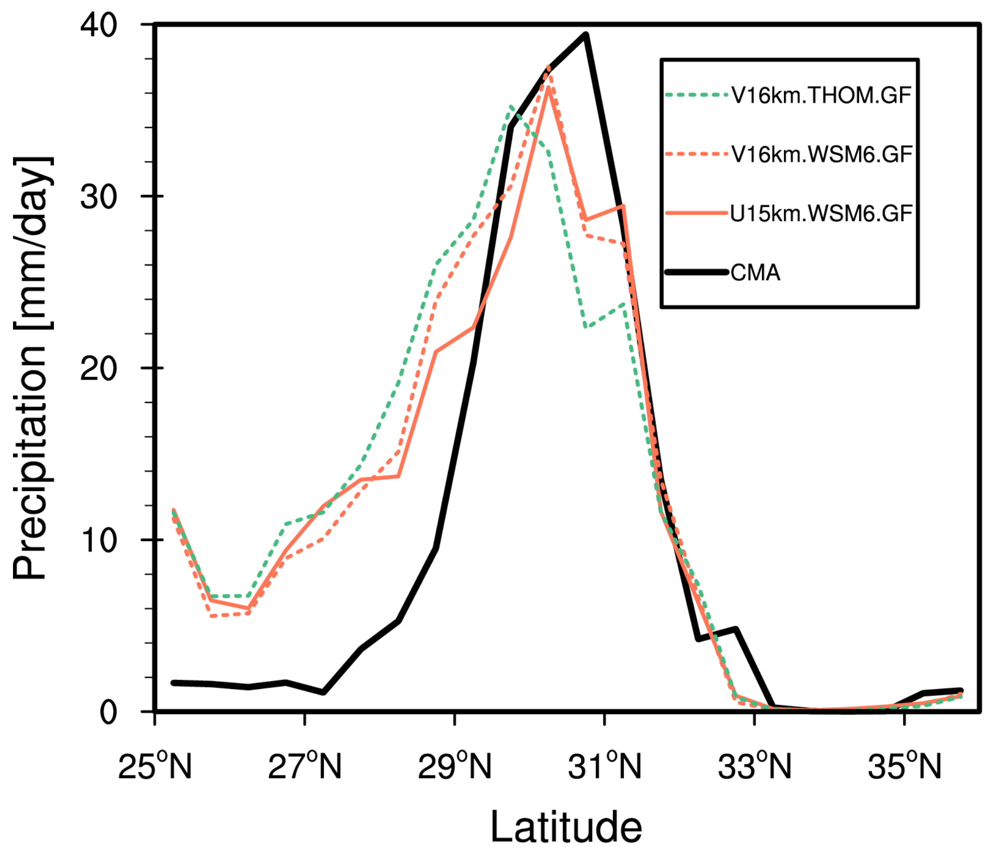

Figure S1 Zonal distributions of precipitation averaged during the event (June 25 00:00 to June 27 12:00 UTC time) over East China (denoted as the black box in Fig. 2) from the CMA station observations and the simulations with the global uniform (15 km, solid lines) and variable (16 km over the refined region as shown in Fig. 1c, dash lines) resolutions with two cloud microphysics parameterizations (WSM6, red dash lines; Thompson, green dash lines). The modeling results are sampled at the CMA station.

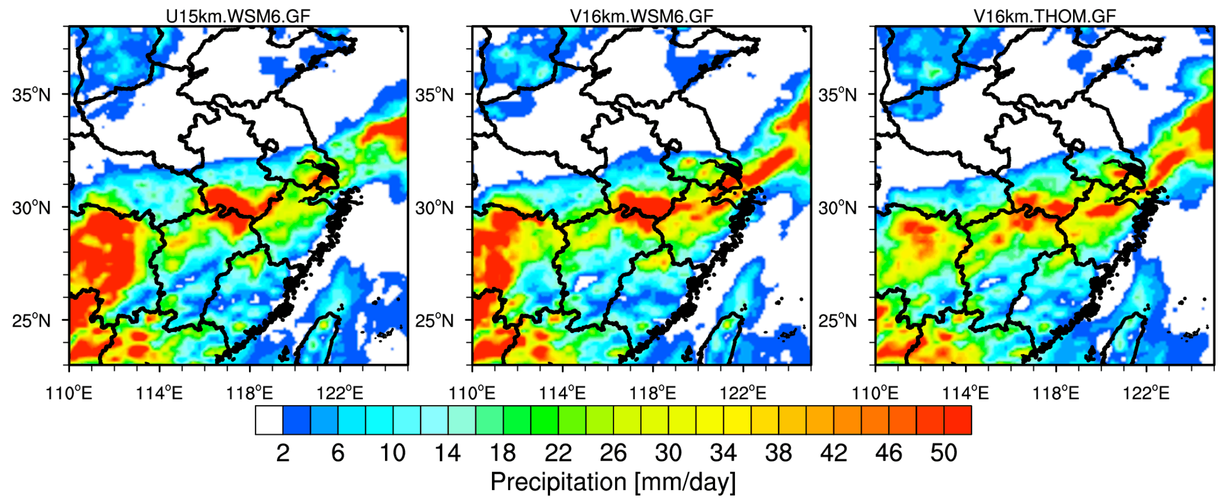

Figure S2 Spatial distributions of precipitation averaged during the event (June 25 00:00 to June 27 12:00 UTC time) from the simulations with the global uniform (15 km) and variable (16 km over the refined region as shown in Fig. 1c) resolutions. The V16km simulations with the WSM6 and Thompson cloud microphysics schemes are shown. The spatial correlation coefficient between the V16km simulations with two different cloud microphysics is 0.85.

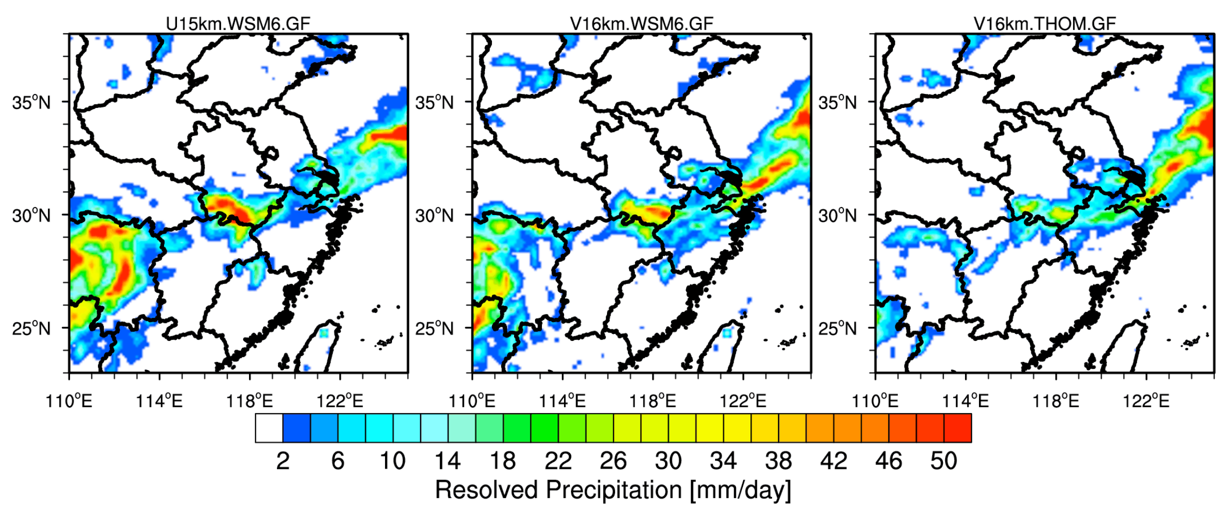

Figure S3 Spatial distributions of grid-resolved precipitation averaged during the event from the simulations with the global uniform (15 km) and variable (16 km over the refined region) resolutions. The V16km simulations with the WSM6 and Thompson cloud microphysics schemes are shown. The spatial correlation coefficient between the V16km simulations with two different cloud microphysics is 0.65.

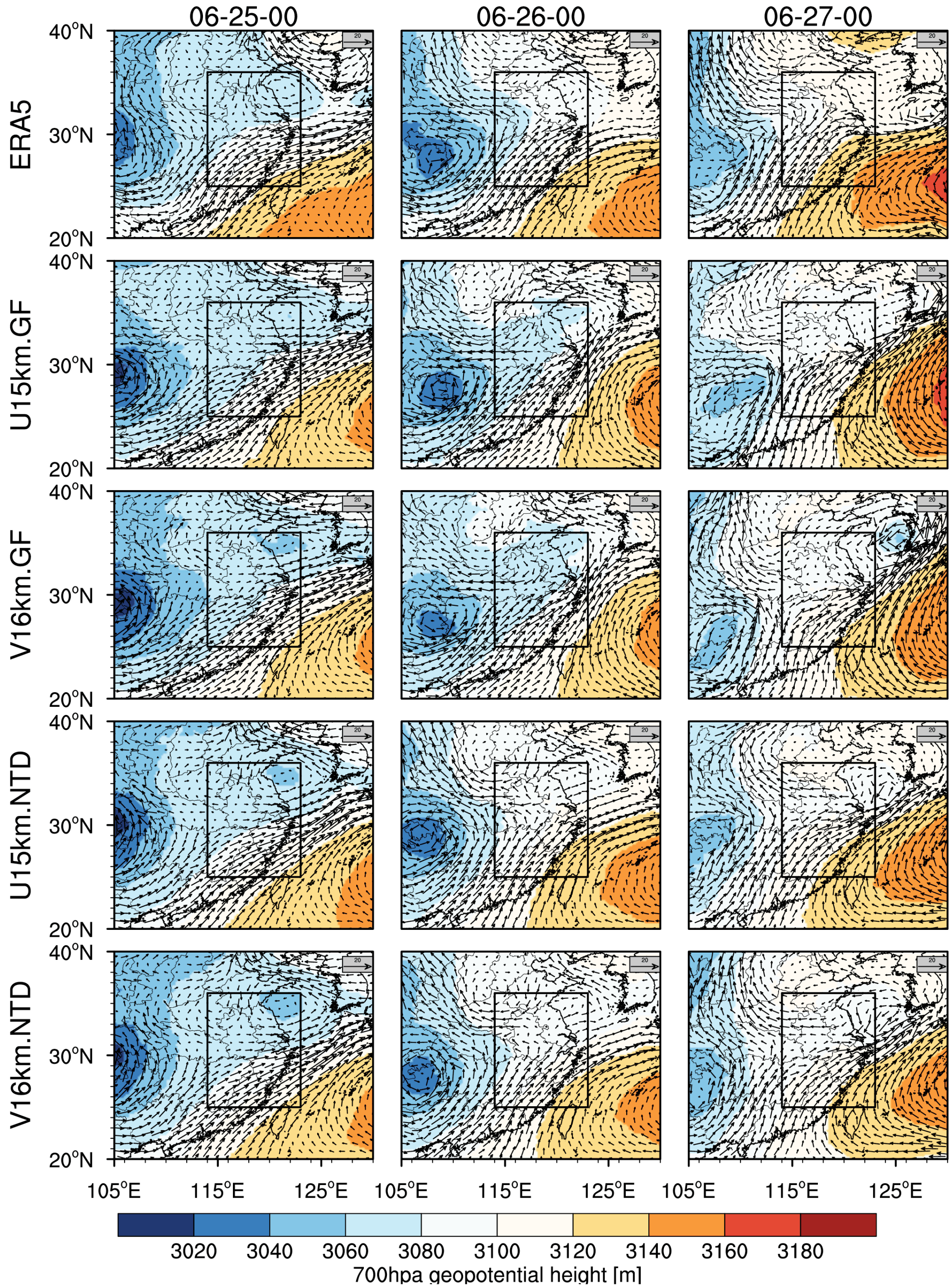

Figure S4 Spatial distributions of geopotential height and wind fields at 700 hPa at UTC 0000 of each day during the simulation (June 23 00:00 to June 27 00:00 UTC time) from the simulations with the global uniform (15 km) and variable (16 km over the refined region as shown in Fig. 1c) resolutions. The black box denotes the region for the analysis in this study.

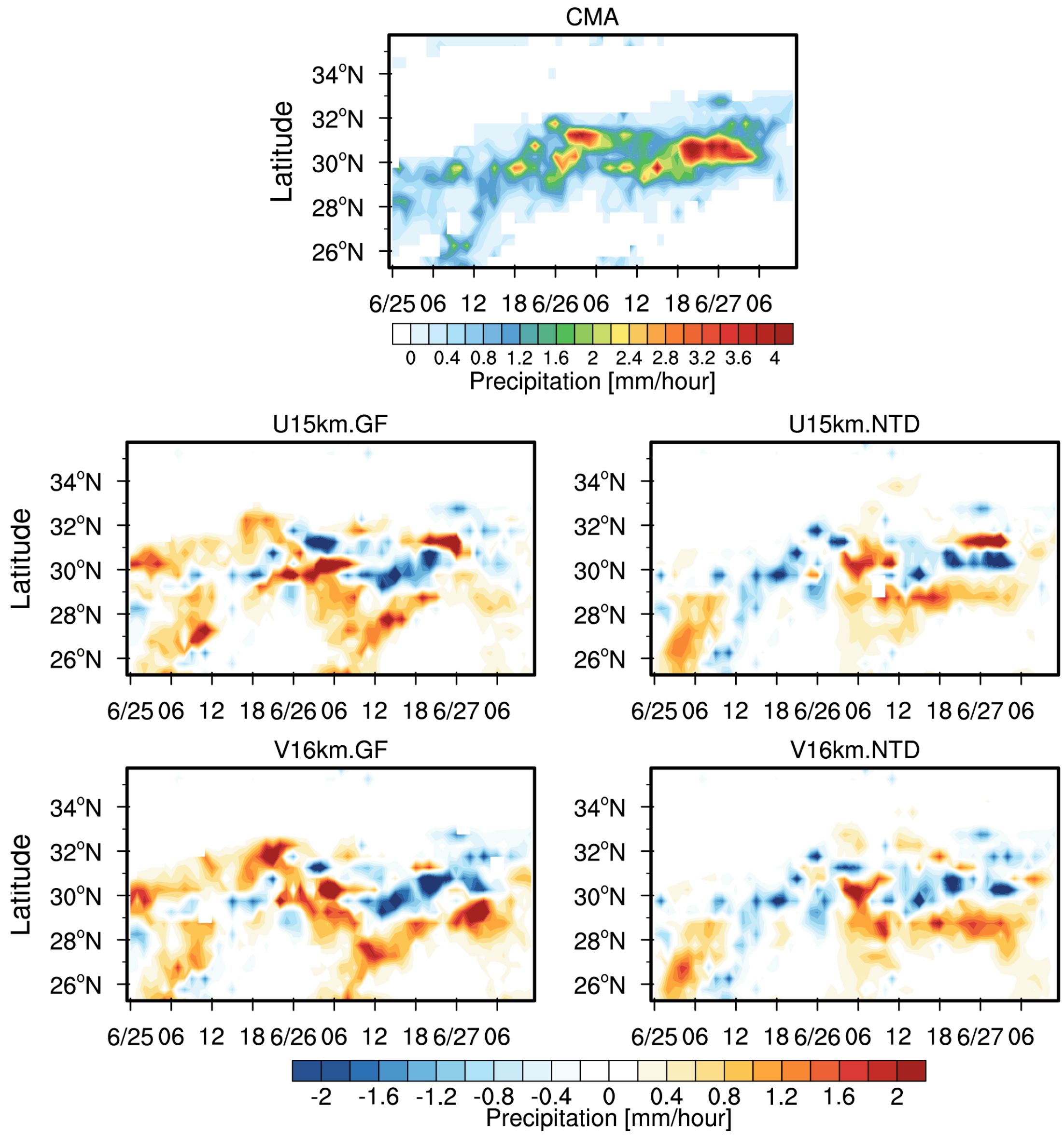

Figure S5 Time-Latitude cross section of precipitation during the event over East China from the CMA station observations and difference between the CMA and the simulations (simulation minus observation) with the global uniform and variable resolutions with two convective parameterizations. The modeling results are sampled at the CMA stations.

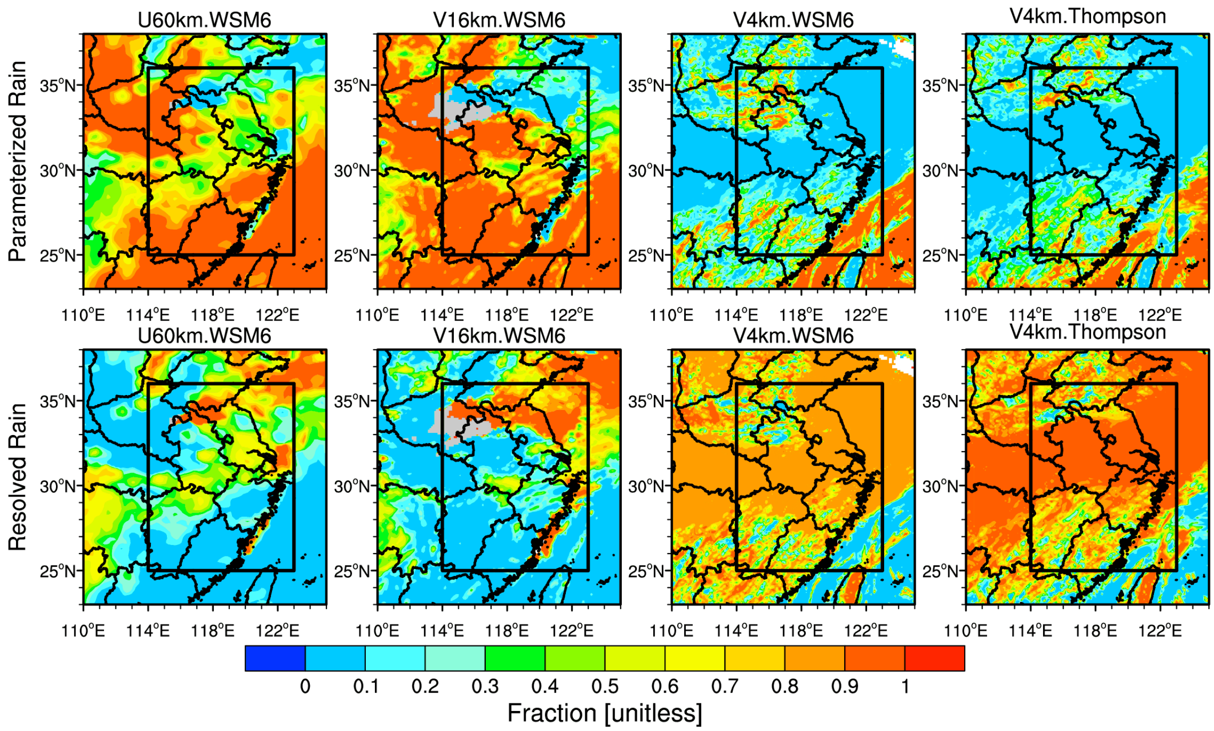

Figure S6 Spatial distribution of fraction of averaged parameterized and resolved precipitation in total precipitation during the event over East China from the simulations with the resolutions of 60 km, 16 km, and 4 km.

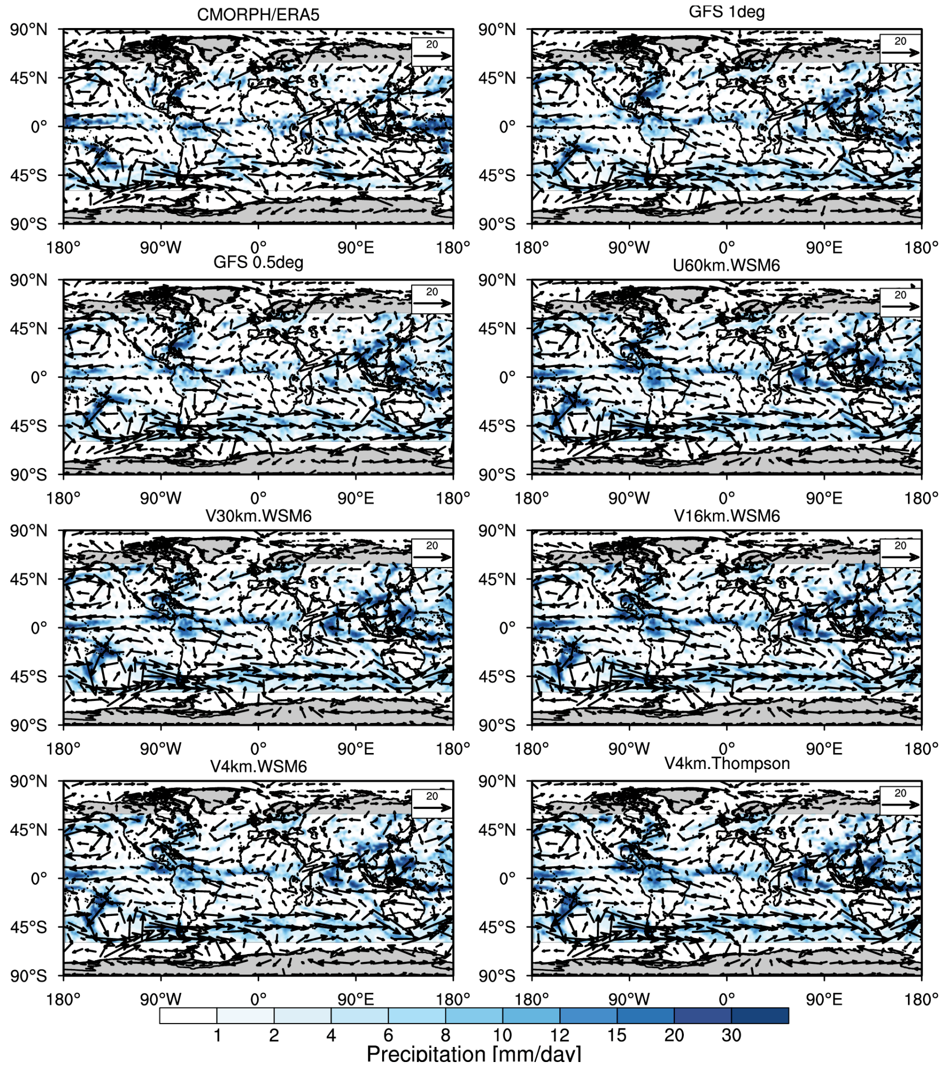

Figure S7 Global distributions of precipitation and wind fields at 850 hPa averaged during the event from the MPAS simulations at the resolutions of U60 km, V30 km, V16 km, and V4 km. The observed mean precipitation from the CMORPH satellite retrievals (downloaded from https://climatedataguide.ucar.edu/climate-data/cmorph-cpc-morphing-technique-high-resolution-precipitation-60s-60n) and the wind fields from the ERA5 reanalysis are shown as well. The black box denotes the region for the analysis in the following. For comparison, the GFS forecasts at 1 degree and 0.5 degree resolutions are also shown.

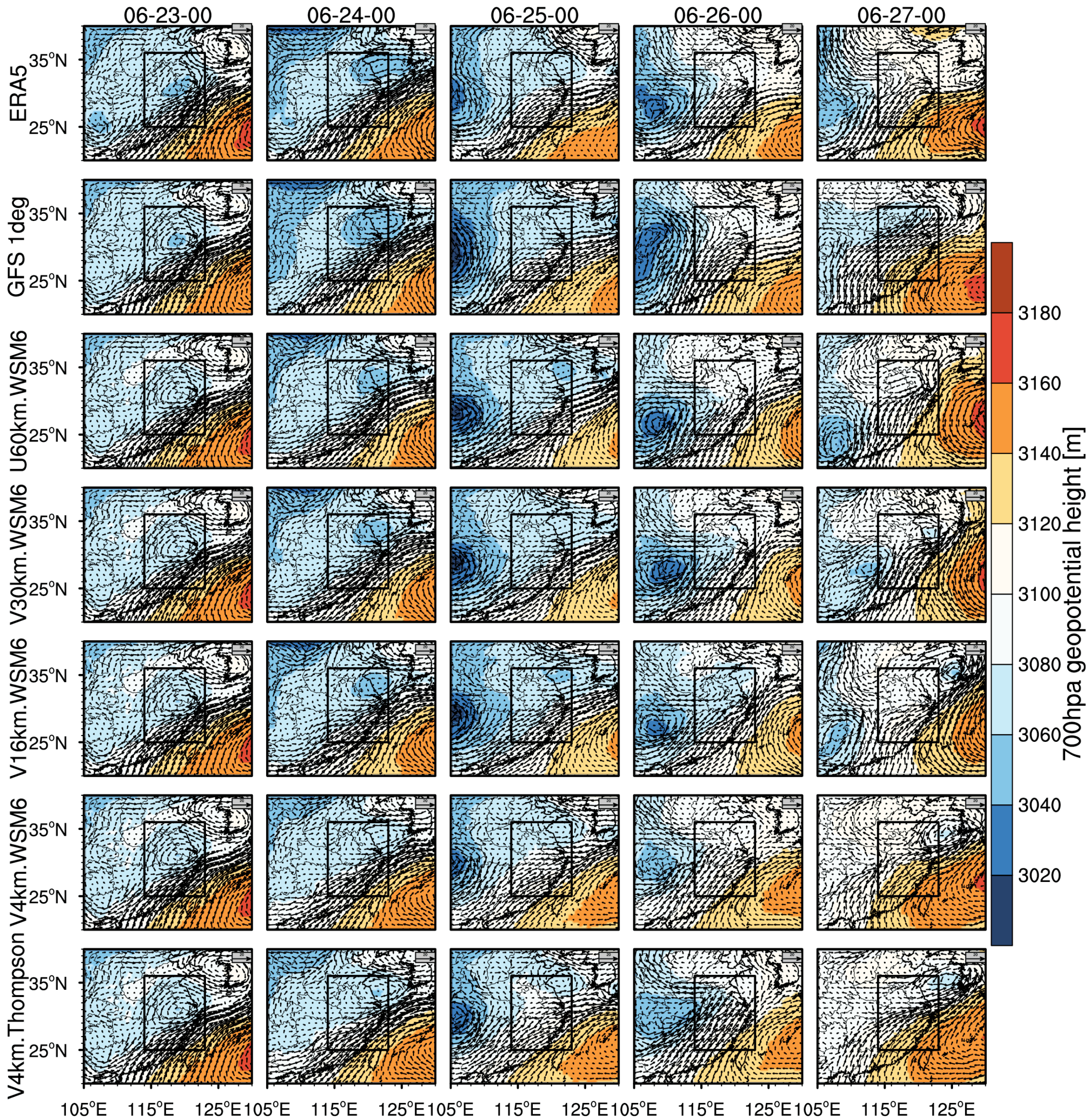

Figure S8 Spatial distributions of geopotential height and wind fields at 700 hPa at UTC 0000 of each day during the simulation (June 23 00:00 to June 27 00:00 UTC time) from the MPAS simulations at the resolutions of 60 km, 30 km, 16 km, and 4 km. The black box denotes the region for the analysis in this study. For comparison, the GFS forecast at 1 degree resolution is also shown.

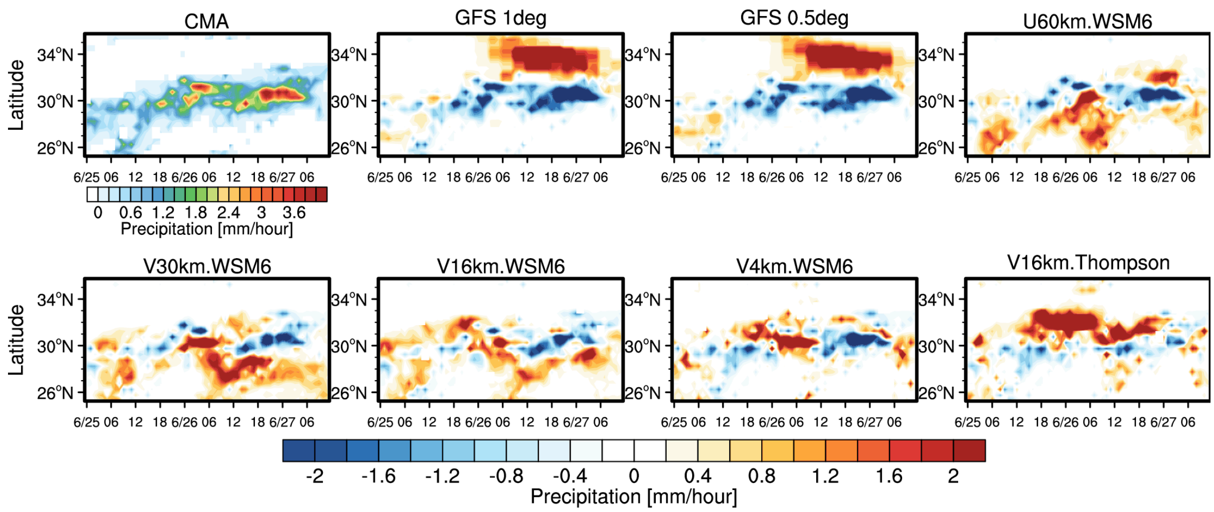

Figure S9 Time-Latitude cross section of precipitation during the event over East China from the CMA station observations and difference between the CMA and the simulations (simulation minus observation) with the global uniform and variable resolutions with two convective parameterizations. The modeling results are sampled at the CMA stations.