ESSD-2021-217
Response to RC2 (Referee 2)

The authors thank referee 2 for a thoughtful review of the manuscript. We agree with many of the referee's points and have made the necessary changes. The responses for the referee's specific comments are as follows.

1. Title. Suggest the authors change it. If it is the development of EARR, the information of input observations data quality or evaluations of more variables/indexes during the whole 10 years 2010-2019 should be included in the figures, but not only 201701 and 201707 mainly in Fig 2-4, 7-13 and only 2017-2018 in Fig 4-5. If it is the development of AdvHG, the innovation contents from your own group should be included in 2.2.2 (page5-8), otherwise, they are all the approaches you could adopt, but not develop. The main contents of the paper are evaluation in 2017-2018, including the method, results and usage in EARR, so maybe it is more suitable to call the title like "Evaluation of EARR based on AdvHG", for your reference. If more figures of longer time series results could be replaced here, it is better. Anyway, the results are not enough, the period is short. The representativeness of the result is limited, comparing with the ERRA (2010-2019).

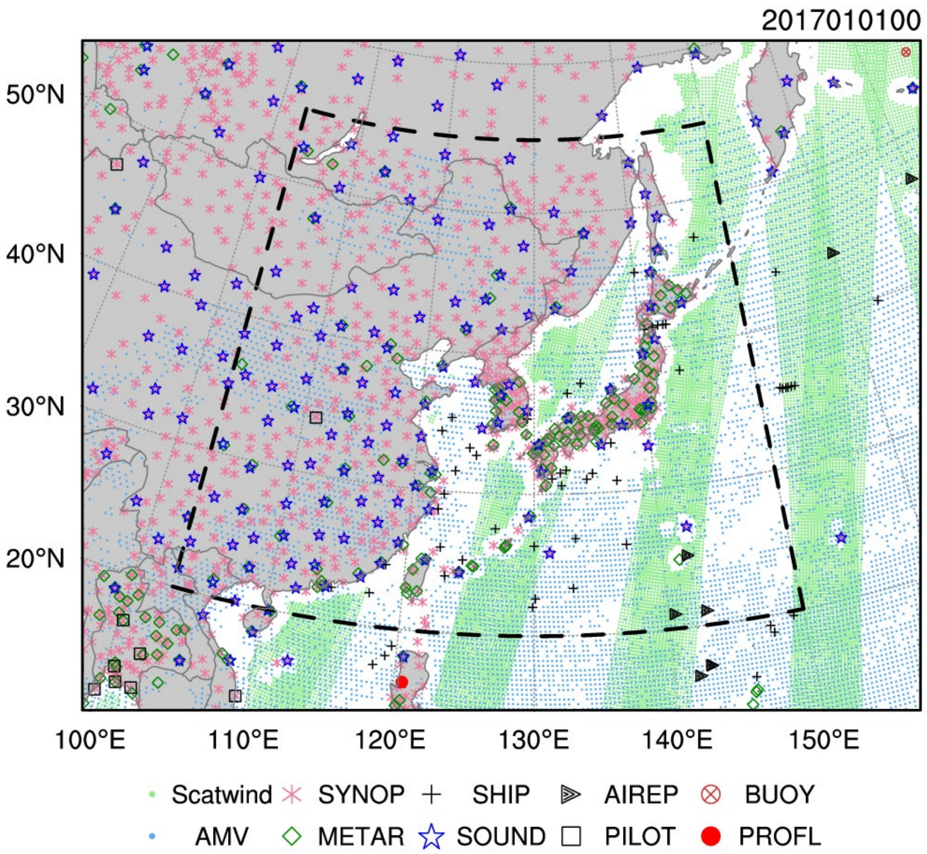
Authors' response: We agree with the referee's point, so we have added or changed the contents according to the referee's suggestions in order to keep the original title for this manuscript.

Firstly, as the referee recommended, we have added the information on observations data quality to the revised manuscript as Table 2 (Table_rev1 below). Figure 1 is also modified to show spatial distributions of observations used in this study in the revised manuscript (Figure_rev1 below). Furthermore, the explanation of observation quality control procedure applied to this study is added to section 2.3 in the revised manuscript (underlined below).

Table_rev1. Summary of observations used in this study. The default observation error statistics provided in WRFDA system are used for assimilation in this study. The variables u, v, T, RH, Ps, and TPW denote zonal wind, meridional wind, temperature, relative humidity, surface pressure, and total precipitable water, respectively.

Observations	Descriptions	Variables	Observation errors (depending on vertical levels)
SOUND	Upper-air observation from radiosonde	u, v	1.1-3.3 m/s
		T	1 K
		RH	10-15%
PROFILER	Upper-air wind profile from wind profiler	u, v	2.2-3.2 m/s
PILOT	Upper-air wind profile from pilot balloon or radiosonde	u, v	2.2-3.2 m/s

AIREP	Upper-air wind and temperature from aircraft	u, v	3.6 m/s
		T	1 K
Scatwind	Scatterometer oceanic surface winds	u, v	2.5-3.8 m/s
SHIPS	Surface synoptic observation from ship	u, v	1.1 m/s
		T	2 K
		Ps	1.6 hPa
		RH	10%
SYNOP	Surface synoptic observation from land station	u, v	1.1 m/s
		T	2 K
		Ps	1 hPa
		RH	10%
BUOY	Surface synoptic observation from buoy	u, v	1.4-1.6 m/s
		T	2 K
		Ps	0.9-1 hPa
		RH	10%
GPSPW	Precipitable water vapor from global positioning system (GPS)	TPW	0.2 mm
METAR	Aviation routine weather report from automatic weather station (AWS)	u, v	1.1 m/s
		T	2 K
		Ps	1 hPa
		RH	10%
AMV	Conventional atmospheric motion vector data from geostationary satellite	u, v	2.5-4.5 m/s

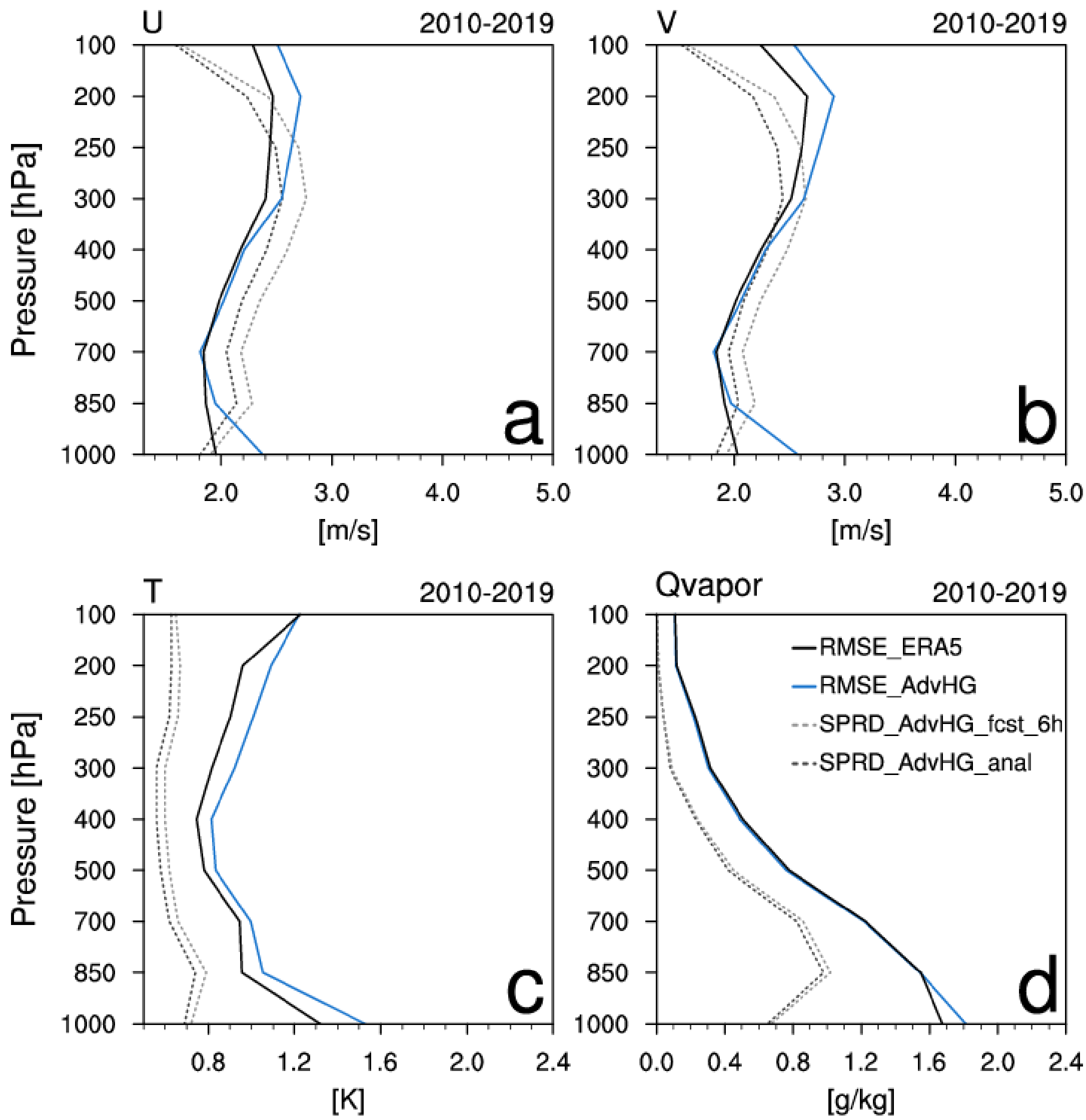


Figure_rev1. The East Asia Regional Reanalysis domain with different types of NCEP PrepBUFR observations available for assimilation at 00 UTC on 1st of January in 2017. The black dashed box denotes a verification area.

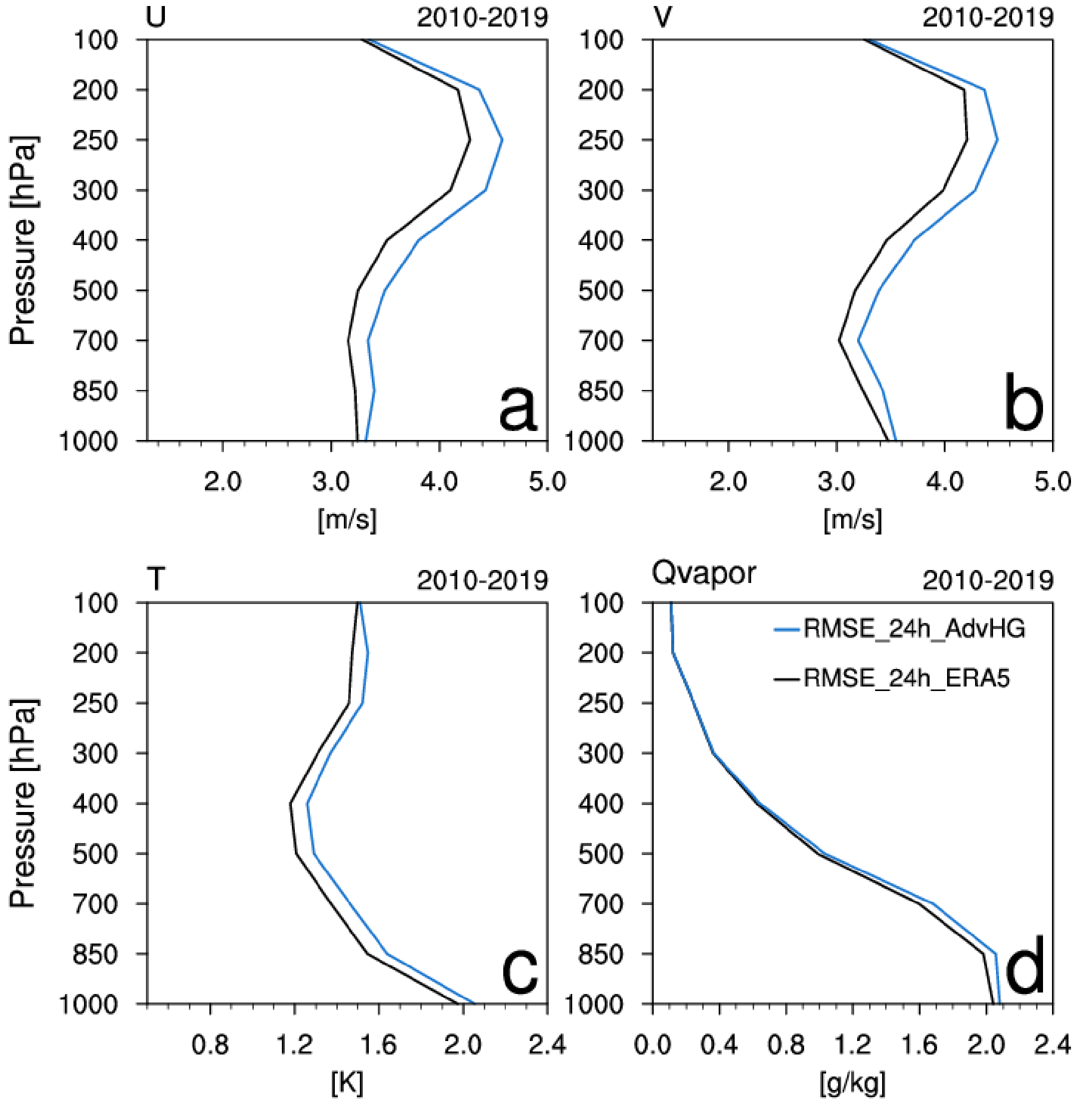
(L178-196) "The NCEP PrepBUFR [Prepared or QC'd data in BUFR (Binary Universal Form for the Representation of meteorological data) format] conventional observations (global upper air and surface weather observations, NCEP/NWS/NOAA/U.S.DOC 2008) are used every 6 h (00, 06, 12, and 18 UTC) for an assimilation by E3DVAR and AdvHG methods (Fig. 1). The PrepBUFR is the output of the final process for preparing the observations to be assimilated in the different NCEP analyses. For observations, rudimentary multi-platform quality control (QC) and more complex platform-specific QC were conducted (e.g., surface pressure, rawinsonde heights and temperature, wind profiler, aircraft wind and temperature) in NCEP (Keyser 2013). Furthermore, if the innovations (i.e., observation minus background) of some observations are greater than 5 times the observational error, then that observation is rejected during assimilation procedure in this study.

The assimilated observations are as follows: the surface observations (SYNOP, METAR, Ship, and Buoy), radiosonde observation (SOUND), upper-wind report (PILOT), wind profiler, aircraft, atmospheric motion vector (AMV) wind from a geostationary satellite (GEOAMV), scatterometer oceanic surface winds (Scatwind), and precipitable water vapor from global positioning system (GPSPW). The observation errors depending on each observation platform, variable, and vertical levels are assigned based on the default observation error statistics provided in WRFDA system (Table 2). All observations are spatially thinned by 20 km except for AMV thinned by 200 km as done by Warrick (2015), Cotton et al. (2016), and Shin (2016)."

Secondly, as referee proposed, we have evaluated longer-term datasets of EARR and ERA5 that are able to be verified for the whole 10-year period and replaced the results of two-year period with those of ten-year period (Figs. 6 and 7 in the revised manuscript) (Please see Figs_rev2 and 3 below). Because the aim of our study is to investigate EARR (AdvHG) performance with ERA5, it is worth evaluating reanalysis and (re)forecast fields of EARR and ERA5 for 10-year period, as referee pointed out. However, it seems unfeasible to compare EARR performance with various experiments (e.g., E3DVAR, ERA-Interim) for the whole period due to the high computational costs producing those datasets, especially ensemble-based one (i.e. E3DVAR). The updated results for the period of 2010-2019 (Figs_rev2 and 3) are almost the same as the previous results with two-year period, except for the water vapor mixing ratio (Qvapor). Although Qvapor RMSEs of reanalysis and (re)forecast of EARR and ERA5 for ten-year period are greater than those for two-year period, both of EARR and ERA5 Qvapor RMSEs increase and the RMSE differences between EARR and ERA5 for ten-year period are similar to those for two-year period. Thus, the longer-term evaluation reveals a large variability of atmospheric humidity and consequent predictability variability over East Asia for ten-year period of 2010-2019. We have revised the manuscript accordingly.

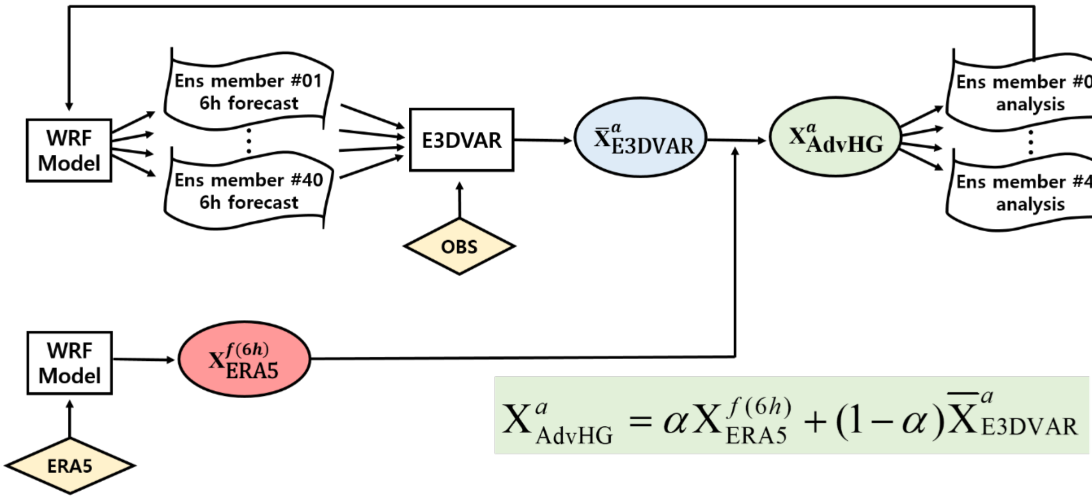


Figure_rev2. RMSEs of analysis of (a) zonal wind, (b) meridional wind, (c) temperature, and (d) Qvapor (water vapor mixing ratio) from ERA5 (black solid) and AdvHG (blue solid) and spreads of analysis (black dashed) and 6 h forecast (gray dashed) of AdvHG depending on pressure levels averaged over the ten-year period of 2010–2019.



Figure_rev3. Same as Fig_rev2 except for RMSE of 24 h forecast.

Lastly, we have added a new content from our group which is a schematic diagram of Advanced Hybrid Gain (AdvHG) method as Fig. 2 in the revised manuscript (Fig_rev4 below). And the section 2.2.2 for an explanation of AdvHG method is divided into two sections 2.2.2 and 2.2.3 to differentiate AdvHG method we newly developed in this study from the existing Hybrid Gain (HG) method.



Figure_rev4. The schematic diagram of the advanced hybrid gain data assimilation method in the East Asia regional reanalysis system.

*2. Horizontal resolution, 12km. It is suggested to mention the raw description of model/DA like other reanalysis, for it is not the same resolution anywhere in the global. Add the information only once in 2.1, like in line 84 (540*432 grid points), it is suggested.*

Authors' response: The model used in this study is the WRF model, which is a regional model based on a grid-point model not a spectral model. Thus, 12-km horizontal resolution for the WRF model is a reasonable way to express a horizontal grid spacing of a model. Meanwhile, for ERA-Interim and ERA5 models' resolution, as the referee suggested, we have added the information on spectral truncation (underlined) to the revised manuscript as follows.

(L205-206) "The horizontal resolutions of ERA-I and ERA5 are approximately 79 km <u>(TL255)</u> and 31 km <u>(TL639)</u>, respectively."

3. 1. introduction. The motivation is described well enough here, like a full story, while the scientific background introduction is not enough, not like an excellent scientific introduction in a paper.

Authors' response: As the referee suggested, we have revised introduction to have the scientific background (underlined) in the revised manuscript.

(L50-63) "<u>The long-term high-resolution datasets are essential to investigate the past extreme weather events which might be associated with mesoscale features such as heavy rainfall events with high spatial and temporal variability which coarser-resolution model cannot represent. The dynamical downscaling approaches can be a solution for generating high-resolution dataset, but they have some issues with insufficient spin-up (Kayaba et al. 2016). Moreover, Fukui et al. (2018) demonstrated that regional reanalysis over Japan assimilating only the conventional observations had the potential to reproduce precipitation fields better</u>

than the dynamical downscaling approaches. Ashrit et al. (2020) also found that the high-resolution regional reanalysis over India showed substantial improvements of regional hydroclimatic features during summer monsoon for the period of 1979-1993 compared to the global reanalysis ERA-Interim (ERA-I, Dee et al. 2011) from ECMWF. Furthermore, He et al. (2019) revealed that the pilot regional reanalysis over the Tibetan Plateau was able to represent more accurate precipitation features as well as atmospheric humidity than the global reanalyses of ECMWF (i.e., ECMWF's fifth-generation reanalysis (ERA5, Hersbach et al. 2020) and ERA-I)."

4. 2. system. Line 83-85, the sentence is not right, Fig 1 is the domain.

Authors' response: As the referee pointed out, we have revised the sentence (underlined) in the revised manuscript as follows.

(L101-103) "In this study, the Advanced Research Weather Research and Forecasting (WRF, v3.7.1) model is used with 12-km horizontal resolution (540 x 432 grid points) and 50 vertical levels (up to 5 hPa) for East Asia domain shown in Fig. 1."

5. Line 134-140, what is alpha in EQ 7?

Authors' response: Alpha in Eq. (7) is a tunable parameter the same as in Eq. (2). To elucidate it, we have added a sentence (underlined) in the revised manuscript as follows.

(L156-158) "where $X_{ERA5}^{f(6h)}$ denotes the 6 h forecast of ERA5 reanalysis based on WRF model and $\overline{X}_{E3DVAR}^{a}$ denotes the analysis of E3DVAR (Fig. 2). In Eq. (7), α is a tunable parameter and is assigned to be 0.5 in this study."

6. The authors make great effort in the DA approach, while what is the characteristics in East Asia of the EARR, comparing with other regional reanalysis, considering of the terrain, climate state like monsoon. In this scope, 50% is cover by the ocean, how is the ocean-land-atmospheric coupled here or just simply depends on all in WRF?

Authors' response: There are a variety of regional reanalysis datasets particularly focusing on the impact of terrain like the Tibetan Plateau (He et al. 2019) and regional hydroclimatic features during monsoon over India (Ashrit et al. 2020). However, the main aim of this stage of this study is to develop a regional reanalysis over East Asia with newly proposed DA method and investigate the uncertainties and characteristics of general meteorological variables of the reanalysis such as temperature and precipitation with existing global reanalysis. Furthermore, we consider it more valuable to make our dataset open to the public at the earliest possible time, so that it could benefit more people using this dataset. As the referee pointed out, more investigation from different perspectives would be conducted in the future.

Ocean, land, and atmosphere are not coupled in WRF model, because WRF model is an atmospheric model producing atmospheric simulations. In this study, sea surface temperature (SST) obtained from ERA5 is used to be updated in WRF model and Unified Noah Land Surface Model (Tewari et al. 2004) is used as a land surface model.

7. Line 160, it is wrong here to mention QuikSCAT which is 199907-200911, it is not in 2010-2019. For your reference: (1) Coriolis/WindSAT (20070813-20120804) from CFSR prepbufr, (2) Oceansat-2/OSCAT (20091215-20140220) KNMI reprocessed but not in CFSR, (3) MetOp-A/ASCAT (20070101-20140331 KNMI reprocessed , GTS data till present in GDAS), (4) MetOp-B/ASCAT (20140408-present in GDAS). You may not use reprocessed ASCAT wind, but it is used in ERA5.

Authors' response: We appreciate the referee's correction. Even though QuikSCAT no longer collects ocean wind data, the convention of naming the scatterometer oceanic surface winds as a group of "QSCAT" still remains in WRFDA system, which made us mistaken. Because the scatterometer oceanic surface winds (Scatwind) are assimilated in EARR, we have rectified the error in the original manuscript by revising the sentence (underlined).

(L189-193) "The assimilated observations are as follows: the surface observations (SYNOP, METAR, Ship, and Buoy), radiosonde observation (SOUND), upper-wind report (PILOT), wind profiler, aircraft, atmospheric motion vector (AMV) wind from a geostationary satellite (GEOAMV), scatterometer oceanic surface winds (Scatwind), and precipitable water vapor from global positioning system (GPSPW)."

8. 4. Result, it is suggest to shorten the results to 60%. The emphasis is how good is EARR but not ERA5. There are many sentences/paragraphs with the subject of ERA5 but not your reanalysis. Like line 235-256, 260-262, 290-291, 328-331, 410-412, 449-452, 520-521. And the difference between ERA5 and ERA-I could be shorten like line 265-270. Pay more attention in how good EARR but not how is ERA5 like line 357-363. The order of the results from different reanalysis is also important. Line 312-314 is good in expression.

Authors' response: Following the referee's suggestion, we have shortened the results and revised the manuscript to pay more attention to EARR rather than ERA5.

9. Line 324, except for strong thresholds, how is strong? >4 mm/6h? how is week?

Authors' response: As the referee pointed out, we have revised sentences (underlined) in the revised manuscript as follows. For a more objective and specific description, the adjectives for threshold such as 'strong' and 'weak' are replaced by 'high' and 'low' and specific threshold values are presented in the revised manuscript.

(L353-356) "For January 2017 (Fig. 8a), ETS of ERA5 (i.e., WRF-based ERA5) is higher than that of ERA5_fromECMWF for all thresholds, whereas ETS of ERA-I (i.e., WRF-based

ERA-I) is lower than that of ERA-I_fromECMWF except for high thresholds (8 and 16 mm (6 h)$^{-1}$).”

(L360-362) “Regarding FBI in winter (Fig. 8b), for 4, 8, and 16 mm (6 h)$^{-1}$ thresholds, all the results show the FBI smaller than 1, implying the underestimation of frequency of precipitation for high-threshold events.”

(L374-376) “With respect to FBI in July 2017, the WRF-based results show the FBIs greater than 1, whereas reforecast from ECMWF show the FBIs greater than 1 for 0.5, 1, and 4 mm (6 h)$^{-1}$ thresholds and smaller than 1 for higher thresholds (8 and 16 mm (6 h)$^{-1}$) (Fig. 8d).”

(L417-420) “At 0.5, 1, and 4 mm (6 h)$^{-1}$ thresholds, E3DVAR BSS is the greatest, which is similar to ETS. At 8 and 16 mm (6 h)$^{-1}$ thresholds, ERA5 ETS is the highest, followed by AdvHG and E3DVAR, whereas overall E3DVAR BSS is the highest, followed by AdvHG and ERA5.”

(L440-442) “During July in 2017, ERA5 and ERA-I simulate heavier precipitation than AdvHG (not shown), which is consistent with larger FBI of ERA5 and ERA-I at higher thresholds.”

(L456-460) “Moreover, although all the results similarly represent overall features of precipitation in January (Fig. 13), ERA5_fromECMWF (Fig. 13g) simulates the overestimated precipitation over South China, which is consistent with the results in the previous section as well as its larger FBI at lower thresholds (0.5 and 1 mm (6 h)$^{-1}$) shown in Fig. 8b.”

(L468-470) “This is consistent with the result in Fig. 8d, in which FBIs from WRF-based results are generally greater than 1 for higher thresholds (8 and 16 mm (6 h)$^{-1}$), whereas those from ECMWF are smaller than 1.”

(L538-542) “In addition, the ETS differences between the results are not distinctive in July. For higher thresholds (8 and 16 mm (6 h)$^{-1}$) in July, AdvHG ETS is greater than E3DVAR ETS and smaller than ERA5 ETS, whereas E3DVAR ETS is the greatest followed by ERA5 and AdvHG for lower thresholds (0.5 and 1 mm (6 h)$^{-1}$).”

10. Fig 7, Line 346-348, results of AdvHG in Jan is better than in Jul, FBI closer to 1. Different FBI results in Jul are larger than 1 (over-forecast) more than in Jan, more difficult to improve for summer than winter. Index ETS and FBI are more difficult to handle and analysis than POD and FAR which is better when it is larger and smaller, separately.

Authors’ response: POD and FAR seem to be straightforward to deal with. However, as mentioned in section 3.2, POD can be artificially improved by systematically over-forecasting the events (Wilson 2010), so FAR should be used with POD. Moreover, ETS is a more balanced score than POD and FAR, because it is sensitive to both false alarms and misses (Wilson 2010). To elucidate this, we have added the explanation (underlined) about ETS in the revised manuscript.

(L223-225) "The ETS range is from -1/3 to 1 and the value 1 for ETS is a perfect score. ETS is a more balanced score than Probability of Detection (POD) and False Alarm Ratio (FAR), because it is sensitive to both false alarms and misses (Wilson 2010)."

11. Line 354, (Figs. 8a and b), is it right? 8b is FAR.

Authors' response: As the referee pointed out, we have revised the sentence (underlined) in the revised manuscript as follows.

(L383-384) "For January in 2017, AdvHG POD is the greatest among the WRF-based results, followed by E3DVAR, ERA5, and ERA-I (Fig. 9a)."

12. 6. Summary, it is not concise in this paragraph.

Authors' response: As the referee pointed out, we have revised the summary in the revised manuscript to make it concise.

13. Reference, line 608-614, it is repeated, please delete. Line 621-622 seems with larger font size.

Authors' response: As the referee suggested, we have deleted the repeated reference and reduced the font size of the reference the referee pointed out in the revised manuscript.

14. Fig 2, is it the result of all EARR domain? Jan and Jul shown in Fig 2-6 while YYYYMM or YYYYMMDD shown in Fig 7-11, it is suggested to use unified expression. Fig 6, Temp revised to T like Fig 2.

Authors' response: Figure 2 in the original manuscript (Fig. 3 in the revised manuscript) is the result of verification domain (dashed box in Fig. 1), not all EARR domain. To make it clearer, we have added the explanation about verification domain in the revised manuscript as follows (underlined).

(L272-276) "The analysis and forecast RMSEs of E3DVAR, AdvHG, the WRF-based ERA-I, and WRF-based ERA5 are calculated for zonal wind, meridional wind, temperature, and Qvapor (water vapor mixing ratio in WRF) variables against sonde observations at 00 and 12 UTC in verification domain (dashed box in Fig. 1) for January and July in 2017 and averaged over each month (Figs. 3, 4, and 5)."

Furthermore, as the referee suggested, we have modified Figs 3-7 to use unified expression (e.g., YYYYMM) and have changed 'Temp' to 'T' in Fig. 7 in the revised manuscript.

References

He, J., F. Zhang, X. Chen, X. Bao, D. Chen, H. M. Kim, H.-W. Lai, L. R. Leung, X. Ma, Z. Meng, T. Ou, Z. Xiao, E.-G. Yang, and K. Yang, 2019: Development and evaluation of an ensemble-based data assimilation system for regional reanalysis over the Tibetan Plateau and surrounging regions, *Journal of Advances in Modeling Earth Systems*, **11**(8), 2503-2522.

Kayaba, N., T. Yamada, S. Hayashi, K. Onogi, S. Kobayashi, K. Yoshimoto, K. Kamiguchi, and K. Yamashita, 2016: Dynamical regional downscaling using the JRA-55 reanalysis (DSJRA-55). *Sola*, **12**, pp.1-5.

Keyser, D., 2013: An Overview of Observational Data Processing at NCEP (with information on BUFR Format including "PrepBUFR" files), *GSI tutorial*, August 6, 2013.

Tewari, M., F. Chen, W. Wang, J. Dudhia, M. A. LeMone, K. Mitchell, M. Ek, G. Gayno, J. Wegiel, and R. H. Cuenca, 2004: Implementation and verification of the unified NOAH land surface model in the WRF model. *20th conference on weather analysis and forecasting/16th conference on numerical weather prediction* (Vol. 1115). Seattle, WA: American Meteorological Society.

Wilson, L., 2010: Verification of severe weather forecasts in support of the ''SWFDP Southern Africa'' project. Rep. for the WMO, 21 pp., www.wmo.int/pages/prog/www/BAS/documents/Doc-7-Verification.doc.