Response to Referee #1's (Dick Dee) comments

On behalf of all the co-authors, I would like to thank the reviewer, Referee #1, Dick Dee, for his thoughtful and constructive comments which helped us to improve our study. We have responded to the comments in the attached file.

This paper describes the East Asia Reanalysis System (EARS), a regional reanalysis covering all of East Asia, with 3-hourly products provided at a horizontal resolution of 12 km and 74 levels in the vertical. The paper covers the methodology, the use of observations, and a variety of performance aspects. The paper is well organized and well written, with explanations in clear language.

In recent years, CMA has made great strides in developing an ambitious reanalysis program, which has already delivered a global atmospheric reanalysis (CRA40) and now also a unique regional reanalysis product. As the authors point out, EARS is the first regional reanalysis covering all of East Asia. This fills an important gap.

According to the paper, all reanalysis data as well as many of the observations used will be accessible via the China Meteorological Data Service Centre at data.cma.cn. (I was not yet able to find the data when I tried during this review). It is very gratifying and good news for the global reanalysis community that CMA is making their data products and observation data available.

The work on observations that has been done in preparation of the ERAS production is significant and potentially very valuable. As the authors point out, many of the observations have not been used before, either for global numerical weather production or for reanalysis. It is very good news for the scientific research community if CMA is indeed able to share these data openly. It would be good to have more information (possibly in a separate paper) describing the observations and their quality control.

Overall, I think this is a good paper about an important dataset that can be highly valuable for large groups of users around the world. I have many questions and suggestions to the authors for additional work, but I don't think there is need for a major revision. My recommendation is therefore to publish after minor revision.

Answer: Thank you very much for your comments. We are glad to see that our dataset is beneficial to the community. We have thoroughly considered the concerns and will revise the manuscript accordingly.

Here are my comments and questions about the details:

If I understand correctly, the background fields used for the reanalysis are WRF short forecasts, which are initialized from ERA-Interim data, and using ERA-Interim data for lateral boundary conditions. There is a 6-h spin-up. Do you have any diagnostics (or have you investigated) the size of the spin-up for different variables, and whether this spin-up depends on the interval (e.g. 6-h vs. 12-h vs. 24-h)? Spin-up can be especially significant for precipitation and cloud, especially because the model used to generate ERA-Interim data is very different from the WRF model.

Answer: We agree with you that it is important to determine the spin-up period as it depends on model configuration, such as time-step, horizontal resolution, and boundary conditions. Following the WRF model user guide, a spin-up period was calculated on the basis of d(surface pressure)/dt or d(mu)/dt in the present work (Fig. 1). One can see, d(surface pressure)/dt decreases rapidly in the first 120 min, then tends to be stable after nearly 300 min. Given the result, a 6-h spin-up period was used.

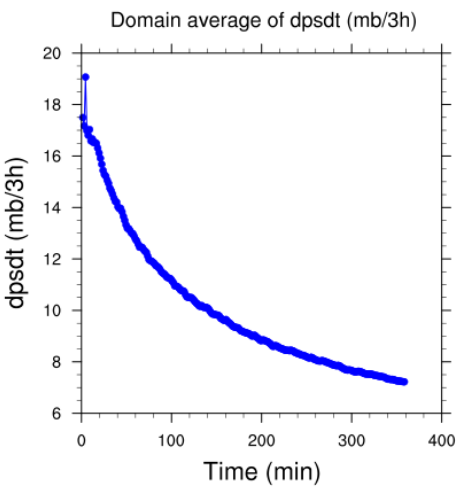

Fig. 1 Evolution of d(surface pressure)/dt with the WRF model integration time

As you stated, the WRF model bias increased with increasing integration time length (Fig.2) because of the gaps between the ERA-Interim data and the WRF model. Further sensitivity tests indicated that the mismatch of soil moisture and soil temperature resulted in the main bias.

In terms of the WRF model spin-up processes and bias, a 6-h spin-up period was cut and the WRF model integrates 12 h for each run.

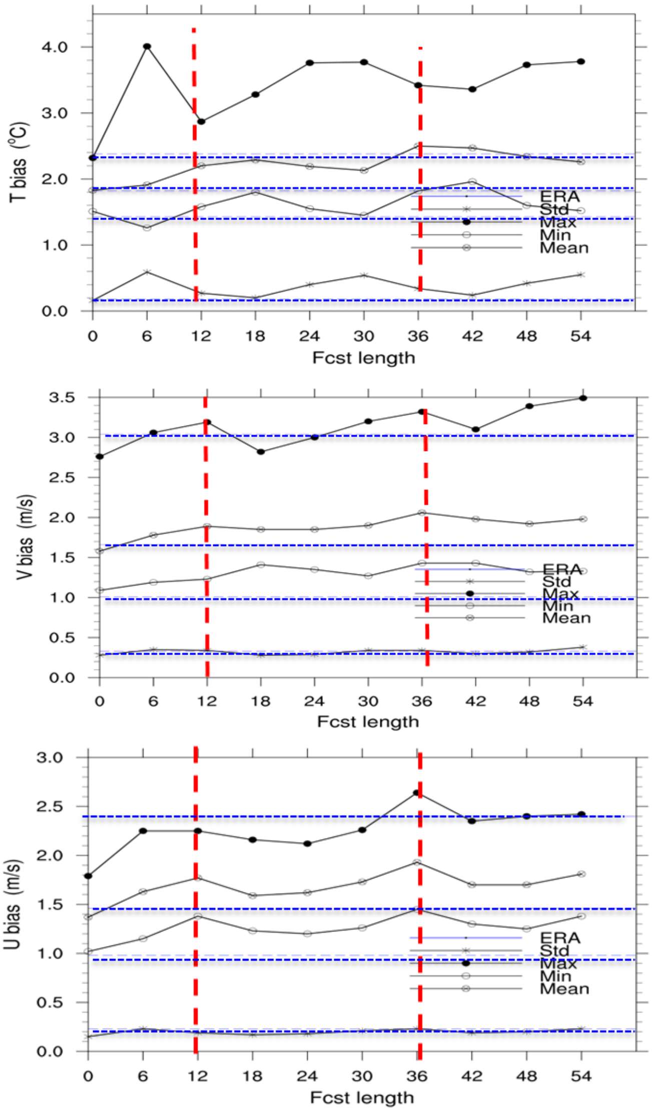

Fig. 2 Evolution of model absolute bias during the period from 01 to 30 June 2016. Mean, Min, and Max denote the mean, maximum, and minimum biases for different forecast length, respectively. The blue dashed lines represent the values of mean, maximum, and minimum biases of ERA-Interim during this period. Std denotes the standard deviation, which suggests the range of bias variation.

Can you provide statistics of the analysis increments (defined as: EARS analysis minus WRF forecast)? This will help to expose biases in the system, due to biases in the model and/or in the observations.

Answer: Thank you for providing these ideas. In early experiments before long-term runs, we made comparisons between WRF forecast and analysis. There were slight differences in surface fields because surface observations were nudged during model integration, rather than assimilated at the analysis moment. As for the upper-level variables, RMSEs were significantly reduced, owing to the assimilation of upper-level observations using a 3D-VAR method. After that, we paid no attention to the difference between the WRF forecast and analysis. Since the differences between the WRF forecast and analysis help to expose biases in the system, statistical analyses in this area will be carried out with special attention.

Can you provide more information about the nudging scheme used to introduce surface observations? Does the scheme depend on estimates of uncertainty of the observations?

Answer: Thanks for this suggestion. In the revision, we have added the suggested content to the manuscript in Subsection 2.1.

"More specifically, observation nudging is a type of four-dimensional data assimilation (FDDA) wherein artificial tendency terms are introduced during the model integration (Reen, 2016). Since it is applied at every time step, nudging is a continuous form of data assimilation. Therefore, observations in the model integration time window can be ingested. Generally speaking, the differences between the WRF model and observation are utilized to create an innovation. Then, the innovation is multiplied by various factors and added to model tendency equations. It should be noted that observation nudging is affected by the uncertainty of the observations. Therefore, surface observations are strictly quality controlled by the OBSGRID module (Wang et al., 2017). "

References:

Reen, B.: A brief guide to observation nudging in WRF, https://www2.mmm.ucar.edu/wrf/users/docs/ObsNudgingGuide.pdf (last access 17 February 2023), 2016.

Wang, W., Bruyère, C., Duda, M., and Dudhia, J.: User's Guides for the Advanced Research WRF (ARW) Modeling System, http://www2.mmm.ucar.edu/wrf/users/docs/user_guide_V3/contents.html (last access 17 February 2023), 2017.

Can you provide more information about the quality control steps used to prepare the input observations, especially the older observations recovered from analogue sources?

Answer: We already provided more information about the quality control steps in Subsection 2.2 as follows:

"Observations were greatly improved by combining datasets from various data sources, especially the observations over China. Firstly, the duplicate (in time and location) data reports were merged. Secondly, all the ground-based observations, were checked by climatic outliers and variation ranges. Besides, internal consistency between meteorological elements and temporal consistency were also carried out. Moreover, soundings were examined based on hydrostatic assumption, temperature lapse rate, and horizontal wind shear."

Can you provide details on any bias corrections applied to the observations?

Answer: In the present study, all observations were assumed to be unbiased like most data assimilation systems. Consequently, the observations were assimilated directly without any bias correction employed.

What kind of automated quality control is applied in GSI for the upper-air analysis? Do you have any statistics on the rejection rates etc.?

Answer: To the best of our knowledge, simple automated quality control is applied in GSI because of high-quality observations in BUFR/prepBUFR. Specifically, only the gross check for each data type is performed. Because of this, observations were quality controlled using OBSPROC (provided by WRF) before being written in prepBUFR format. As it is addressed in the WRF user guide (Wang et al., 2017), the purpose of OBSPROC is to:

- ✓ Remove observations outside the specified temporal and spatial domains;
- ✓ Re-order and merge duplicate (in time and location) data reports;
- ✓ Retrieve pressure or height based on observed information using the hydrostatic assumption;
- ✓ Check multi-level observations for vertical consistency and superadiabatic conditions;
- ✓ Assign observation errors based on a pre-specified error file;
- ✓ Write out the observation file to be used by WRFDA in ASCII or BUFR format.

After strict quality control, almost all the observations passed GIS gross check.

References:

Wang, W., Bruyère, C., Duda, M., and Dudhia, J.: User's Guides for the Advanced Research WRF (ARW) Modeling System, http://www2.mmm.ucar.edu/wrf/users/docs/user_guide_V3/contents.html (last access 17 February 2023), 2017.

Can you provide more information about the characteristics the background error covariances used in the GSI analysis?

Answer: It is well known that background error covariance (BE) plays an important role in three-dimensional variational assimilation. In this work, the NMC method was used to build BE based on one-month continuous WRF simulations in each season (Fig. 3). Compared with the default BE (Fig. 3d), more details can be obtained from the built BE. The finer the latitude band is used for statistics, the more detailed the features are (Fig. 3a-c). However, from idealized and real cases simulations, the newly built BE provided poorer performance than the default BE. Therefore, we used the NAM background error covariance with scale factor adjustment in this study.

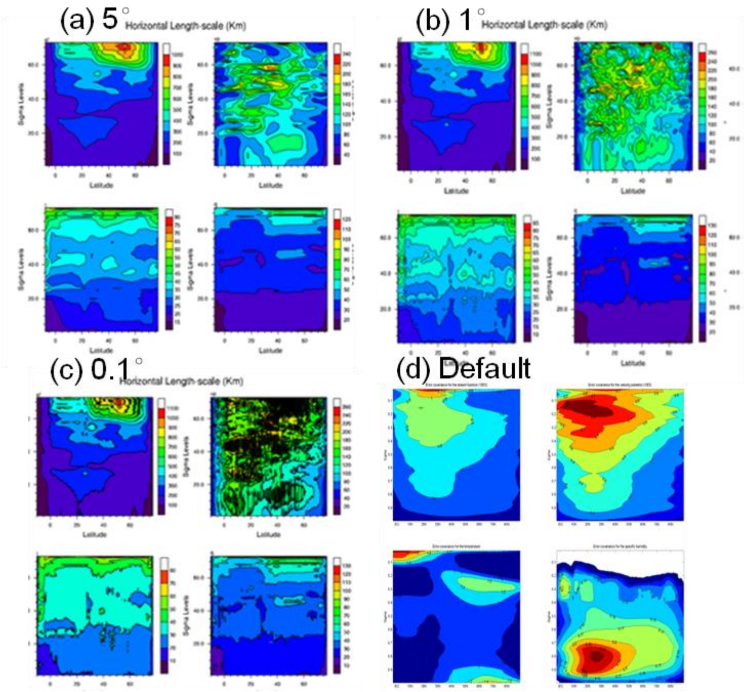

Fig. 3 Properties of background error covariance (BE). (a)-(c) are calculated from 5°, 1°, and 0.1° latitude bands, respectively. (d) is the default option in the GSI.

Can you provide more information about the assimilation of radar data? Has there been any pre-processing of the radar data?

Answer: We already provided more information about the assimilation of radar data in Subsection 2.2 as follows:

In the present work, radar reflectivity was ingested by the way of cloud analysis, while the radial wind was not assimilated at present as more work is required. The cloud analysis module in the GSI came from the Advanced Regional Prediction System (ARPS) (Hu and Xue, 2007), and can be further traced to the Local Analysis and Prediction System (LAPS) (Albers et al., 1996). The radar basic data was provided by CMA. All the radar basic data were quality controlled, such as removing isolated non-meteorological echoes and ground clutter. After quality control, all radar observations at the same time are utilized to generate mosaic products in BUFR format, which can be inserted into the GSI cloud reanalysis module.

References:

Hu, M., and M. Xue, 2007: Implementation and evaluation of cloud analysis with WSR-88D reflectivity data for GSI and WRF-ARW. *Geophysical Research Letters*, **34**, doi:10.1029/2006GL028847.

Albers, S. C., J. A. McGinley, D. L. Birkenheuer, and J. R. Smart, 1996: The Local Analysis and Prediction System (LAPS): Analyses of Clouds, Precipitation, and Temperature. *Weather and Forecasting*, **11,** 273-287, doi:10.1175/1520-0434(1996)011<0273:TLAAPS>2.0.CO;2.

Many of the validation results in the paper refer to the improvements in EARS relative to ERA-Interim. Those are mostly good results, but they are not very surprising given the higher resolution and use of many additional observations. I think that it would be very useful to show more diagnostics that focus on the use of observations specifically, such as time series of observation-minus-background statistics. These can be very informative and can be used to identify issues and problems with the observations and/or the data assimilation scheme, that could possible be addressed in a future reanalysis.

Answer: Thank you for this great suggestion. GSI outputs detailed diagnostic files which provide useful information to diagnose potential benefits and problems. To date, We have not yet carried out detailed statistical work and only sampled some basic characteristics, such as cost and gradient function (Fig. 4), and scatter plot of observation-minus-background (O-B). Keeping this suggestion in our mind, detailed statistical analyses will be carried out in the future.

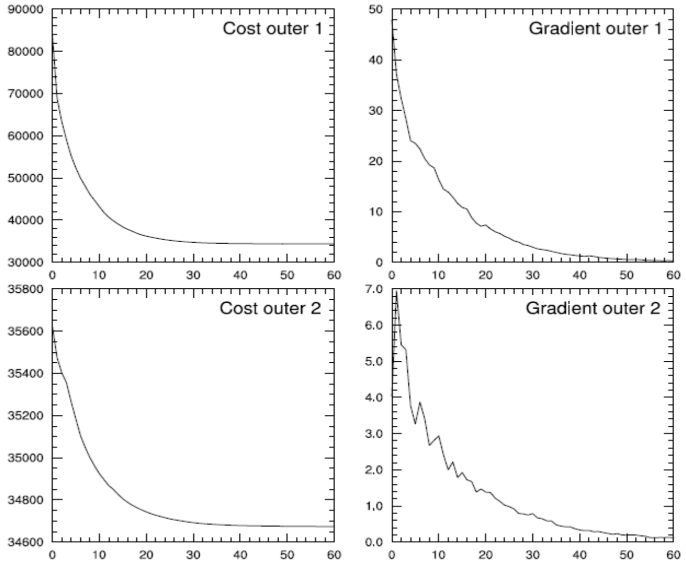

Fig. 4 Evolution of cost function (left) and the norm of gradient of cost function (right) in the first outer loop (top) and the second outer loop (bottom) plotted vs assimilation iteration number at 1200 UTC 01 July 2014.

Fig 5: Is this for a single sounding? What does the shift signify? There is no description of the x-axis.

Answer: Thanks for your kind reminder. Yes. We just took the sounding at Beijing station (54511) at 0000 UTC 1 July 2016 as an example. All the soundings in China were processed with the same approach. The two profiles are perfectly overlapped except for newly added observation points. To avoid overlaying the two data points, we have shifted the IGRAv2 profile slightly to the left. We further illustrated this in the figure caption.

Fig 11: I don't understand the grey shapes in this figure.

Answer: The grey fill is to highlight the box-percentile plots, although the black border is readable.