Response to Reviewer 3

We greatly appreciate your timely and insightful review. Your comments have been instrumental in guiding our revisions. We provide a detailed point-by-point response below, using black for your comments, blue for our responses, and red for proposed changes to the manuscript. We will finalize the manuscript and indicate specific locations of changes after receiving all reviews. Additional references are included at the end of this document.

The manuscript titled “Enhancing Single-Precision with Quasi Double-Precision: Achieving Double-Precision Accuracy in the Model for Prediction Across Scales-Atmosphere (MPAS-A) version 8.2.1” explores the use of the quasi-double precision algorithm by Møller et al. (1965) within a numerical modelling framework. By implementing the algorithm in MPAS-A across various case studies and idealized scenarios, the study highlights its potential to lower computational costs compared to double-precision approaches while enhancing accuracy over single-precision approaches.

Overall, the manuscript is well-organized and relatively clear, but there are several areas that could benefit from further clarification and elaboration. In particular, I have concerns regarding 1) the practical significance of the reported improvements, 2) the lack of detail and rationale in the model setup, and 3) the insufficient explanation or justification for the choice of metrics. Based on these issues, I recommend major revisions before the manuscript can be reconsidered for publication.

Thank you for your thorough review and valuable feedback on our manuscript, we appreciate your positive comments on the organization and clarity of the manuscript.

We understand your concerns regarding the practical significance of the reported improvements, we have carefully considered your feedback and addressed each of these points in the detailed responses below. We believe the revisions made in response to your comments significantly strengthen the manuscript and clarify the contributions of our work.

We are grateful for your insights and believe that addressing your concerns has resulted in a substantially improved manuscript. We look forward to your further assessment.

Below, you find a list of my main concerns:

1. The practical impact of the improvements achieved with the quasi-double precision algorithm compared to the single-precision approach remains unclear to me. While the authors demonstrate large relative accuracy gains compared to single-precision, it is not evident that these improvements are meaningful in absolute terms. For example, in the idealized scenario of Section 3.1, the RMSE of surface pressure improves by only 10^-2 Pa at longer lead times (12+ days), with negligible improvements at shorter times. Similar patterns emerge in Section 3.3. Given that these improvements are an order of magnitude smaller than the World Meteorological Organization's (WMO) recommended measurement uncertainty for atmospheric pressure (e.g., https://gcos.wmo.int/en/essential-climate-variables/pressure), it might be difficult to justify the additional computational costs. I encourage the authors to elaborate on the significance of their results in absolute terms from both a practical and forecasting perspective.

A related point: Are current operational implementations of MPAS-A defaulting to double- or single-precision for the variables discussed in this manuscript? Given the results, if double precision is currently the standard, the authors might want to emphasise the potential of quasi-double precision as a substitute for double precision, rather than focusing primarily on its benefits over single-precision.

Thank you for your insightful comments and suggestions. We appreciate your careful review and have revised the manuscript accordingly. We address your points as follows:

(1) This research is based on double precision testing as the benchmark experiment, we agree that highlighting the computational advantages of quasi-double precision compared to double precision is crucial. We apologize for the omission of this discussion in the original manuscript. As you suggested, we have now added a detailed analysis of computational costs in the revised manuscript, comparing SGL, DBL and QDP. This analysis demonstrates that QDP achieves accuracy very close to DBL while significantly reducing computational costs. We provide further details on this analysis in our response to your second point below.

(2) You correctly pointed out that the SGL errors already meet the WMO's target values, and our primary focus should be the comparison between QDP and DBL. We included the comparison with SGL for the following reasons: within the MPAS-A model, the quasi double-precision algorithm is applied only to the basic and tend processes, while other parts remain in their original single-precision implementation. In essence, the quasi double-precision algorithm is applied to a limited portion of the dynamical core. Since MPAS-A supports single-precision calculations, we aimed to investigate the impact of the quasi double-precision algorithm on both computational cost and accuracy relative to the existing single-precision implementation. A modest increase in computational cost coupled with a substantial improvement in accuracy would suggest the potential of quasi double-precision as a replacement for single precision in scenarios where storage limitations are a major concern. We have retained the comparison to SGL to provide this context but have strengthened the emphasis on the comparison with DBL throughout the manuscript as you recommended.

(3) You correctly pointed out that while the errors observed within the current 12+day simulations appear relatively small, these could accumulate over longer time scales. Therefore, we plan to conduct extended simulations (monthly to yearly) in future research to investigate the potential for error accumulation and to determine the feasibility of using quasi-double precision as a replacement for double precision in such long-term applications. As you suggested, our future work will prioritize direct comparisons between the QDP and DBL in these extended simulations to fully characterize the differences and evaluate the long-term viability of our proposed method.

2. The experimental setup lacks sufficient detail and justification. It is unclear why the authors chose these particular case studies and idealised scenarios, especially in the applications to real data. This makes it difficult to assess the generalisability of the results to other atmospheric conditions or variables. Additionally, a more detailed comparison of computational costs among the three algorithms (single-precision, quasi-double precision, and double precision) is necessary, as

improvements in accuracy may not be feasible in practice if they come with substantial increases in computational expense.

Thank you for your constructive feedback. We address your concerns regarding the case selection and computational cost as follows:

（1）Case Selection and Justification: We apologize that this rationale was not clearly articulated in the original manuscript. We will explain it as follows in Section 2.4. Thank you for your suggestions. We would like to ask whether you believe the descriptions and tables mentioned should be added to the main text of the manuscript. We look forward to your further input on this matter.

（2）Computational cost: We will add a dedicated section 3.4 in the paper to describe the computational performance. The size of the computational performance will be represented in terms of runtime, and we will discuss the runtime for each case in tabular form.

2.4 Additional content

The type of initial conditions in MPAS-A cases (Table 1) can be found in “MPAS-Atmosphere Model User’s Guide Version 8.1.0” (can be download on MPAS website). Only configurations for the idealized cases 2, 5, and 6 are provided on the official website. To avoid redundancy, the results for Case 6 are not shown, as they were similar to those obtained for Cases 2 and 5. For real-data simulations, case 7 as shown in Table. We explored different resolutions for real-data simulations to assess the impact of grid refinement on the performance of the quasi-double precision algorithm. Our focus is on numerical precision in global models, where boundary condition effects are absent. Thus, we do not consider regional models like case 8 and case 9 in this study.

Table 1. Cases in MPAS-A.

type of initial conditions to create:
1. Jablonowski and Williamson baroclinic wave (no initial perturbation)
2. Jablonowski and Williamson baroclinic wave (with initial perturbation)
3. Jablonowski and Williamson baroclinic wave (with normal-mode perturbation)
4. squall line
5. super-cell
6. mountain wave
7. real-data initial conditions from, e.g., GFS
8. surface field (SST, sea-ice) update file for use with real-data simulations
9. lateral boundary conditions update file for use with real-data simulations

3.4 Computational performance

In comparison with the SGL, although there is a slight increase in runtime, it is minimal, at only 6% (Jablonowski and Williamson baroclinic wave), 0.3% (Super-cell), 2% (Real data with resolution of 120km) and 18% (Real data with resolution of 240km) (Table 2). This slight increase is attributed to the addition of a small number of global variable arrays when using quasi double-precision. And compared to DBL, QDP demonstrated relatively better performance across different cases, reducing the runtime by 29% (Jablonowski and Williamson baroclinic wave), 29% (Super-cell), 21% (Real data with resolution of 120km) and 6% (Real data with resolution of 240km) (Table 2).

Table 2. Elapsed time of DBL, SGL and QDP test (unit:s).

Case name	DBL	SGL	QDP
Jablonowski and Williamson baroclinic wave	1768	1191	1263
Super-cell	1507	1073	1077
Real data with resolution of 120km	19126	14765	15092
Real data with resolution of 240km	1397	1118	1317

Abstract

The content ‘The round-off error of surface pressure is reduced by 68%, 75%, 97%, 96% in cases, the memory has been reduced by almost half, while the computation increases only 2%, significantly reducing computational cost.’ will be revised to ‘The bias of surface pressure are reduced respectively by 68%, 75%, 97% and 96% in cases, the memory has been reduced by almost half, while the computation increases only 6%, 0.3%, 2%, and 18% in cases, significantly reducing computational cost.’

3. The manuscript currently employs only grid-point level and "spatial RMSE" as comparison metrics, but the term "spatial RMSE" is not clearly defined. I recommend the following:

1) Clearly define "spatial RMSE," particularly when used as a summary statistic over a larger domain. For example, are the authors applying any latitude-based weighting?

2) Justify the use of this metric and consider comparing the algorithms using additional relevant metrics. If alternative metrics are not applicable, please explain why they were not used.

Thank you very much for your insightful comments regarding the metrics used in our study. Your feedback has been incredibly helpful in improving the clarity of our analysis. We address your concerns in three parts:

(1) We apologize for the omission of a clear definition of spatial RMSE in the original manuscript. As your suggestion, we have now included a detailed explanation of the spatial RMSE calculation in Section 3.1. This explanation includes the formula and a description of its application in our study, as detailed below.

(2) We selected spatial RMSE as a primary metric because it is widely used in the evaluation of

atmospheric numerical models and effectively reflects model performance improvements. Studies such as Vánˇa et al. (2016), demonstrate its utility in assessing the accuracy of model simulations. Therefore, we believe spatial RMSE is an appropriate and informative metric for evaluating the impact of the quasi-double precision algorithm in our study.

(3) We appreciate your suggestion to include additional metrics. We decide to add a new metrics of Mean Absolute Error (MAE) to improve our analysis in Section 3.1. To facilitate comparison across different cases, we have also included a table summarizing the spatial RMSE and MAE values for each case in the revised manuscript. The specific changes are detailed below.

3.1 Additional content

To quantify the difference between the simulations using SGL, QDP, and DBL, (used as the benchmark), we calculate the spatial root-mean-square error (RMSE). First, for each grid point, the temporal averages of the variables (e.g., surface pressure, 500hPa height) are computed across the entire simulation period for each experiment (SGL, QDP, and DBL). Then, the spatial RMSE is calculated as the root-mean-square difference between the temporally averaged fields of the control experiment (SGL or QDP) and the benchmark double-precision experiment (DBL), following (1):

$$\text{Spatial RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(M_i - C_i)^2}{N}} \qquad (1)$$

Where, N is the total number of grid points, M_i is the temporally averaged value at grid point i for the benchmark double-precision experiment, C_i is the temporally averaged value at grid point i for the control experiment (SGL or QDP).

In addition to the spatial RMSE, we also calculate the Mean Absolute Error (MAE) to assess the magnitude of the difference between the control experiments (SGL and QDP) and the benchmark double-precision experiment (DBL), irrespective of the direction of the difference. Like the spatial RMSE calculation, we first compute the temporal average for each grid point across the entire simulation period for each experiment. The MAE is then calculated as the average absolute difference between the temporally averaged fields of the control experiment and the benchmark experiment, following (2):

$$\text{MAE} = \frac{1}{N} * \sum_{i=1}^{N} |M_i - C_i| \qquad (2)$$

where N is the total number of grid points, M_i represents the temporally averaged value at grid point i for the benchmark double-precision experiment, and C_i represents the temporally averaged value at grid point i for the control experiment (either SGL or QDP).

Additional table

As shown in Table 4 (MAE) and Table 3 (spatial RMSE), the addition of the quasi-double precision algorithm consistently improves accuracy (compared to single precision) across all cases.

Table 3. The spatial RMSE values of surface pressure compared to DBL for cases, unit: Pa.

Case name	SGL	QDP
Jablonowski and Williamson baroclinic wave	$3.42 * 10^{-2}$	$1.09 * 10^{-2}$
Super-cell	$8.80 * 10^{-4}$	$2.27 * 10^{-4}$
Real data with resolution of 120km	$6.33 * 10^{-2}$	$2.25 * 10^{-3}$
Real data with resolution of 240km	$6.68 * 10^{-2}$	$2.25 * 10^{-3}$

Table 4. The MAE values of surface pressure compared to DBL for cases, unit: Pa.

Case name	SGL	QDP
Jablonowski and Williamson baroclinic wave	$1.29 * 10^{-2}$	$3.81 * 10^{-3}$
Super-cell	$8.79 * 10^{-4}$	$2.26 * 10^{-4}$
Real data with resolution of 120km	$5.38 * 10^{-2}$	$1.95 * 10^{-3}$
Real data with resolution of 240km	$5.52 * 10^{-2}$	$1.94 * 10^{-3}$

Minor concerns/technical corrections:

- Figure 6a: The higher error with the quasi-double precision algorithm compared to the single-precision in this subpanel is confusing. If the error with the single-precision algorithm is anyway so small that it does not matter in practice, why was this case study selected, and why include this figure?

Thank you for bringing this to our attention. We sincerely apologize for the error in our data processing, which was also noted by Reviewer 1. We have carefully re-examined our code and identified a mistake. We have corrected the error in our data processing code and have now generated the revised results. The corrected figure is shown below (fig. 1), Figure 6(a) in manuscript:

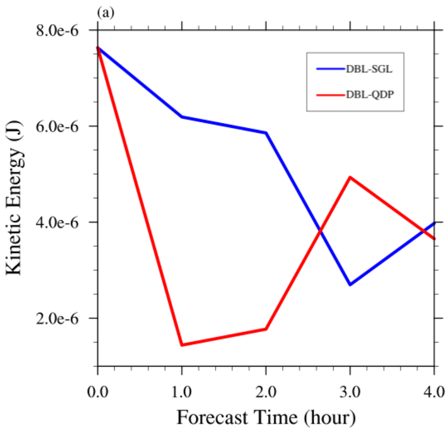


Figure 1. The temporal evolution of spatially averaged difference of kinetic energy between DBL and SGL, as well as difference between DBL and QDP in case of super-cell.

I apologize for needing to provide additional information regarding this section. In all cases presented in the manuscript, the time-evolution plots (Figures 4, 6, and 8 in the manuscript) currently utilize kinetic energy and surface pressure, which correspond to the conservation of energy and mass, respectively. However, I realize that directly using total energy and total mass would offer a more accurate representation. Please see the figures 2, 3 and 4 showing the temporal variations of energy and mass for all cases. If you allow it, I would be happy to replace the existing Figures 4, 6, and 8 (in the manuscript) with these updated versions (figures 2, 3 and 4). I want to emphasize that this change would not affect the overall results or conclusions of the study, which remain consistent with those currently presented in the manuscript. We have received feedback from two reviewers so far, and Reviewer 1 has approved the replacement. We are eagerly awaiting your response. If all four reviewers agree, we will update the figure in the final version of the manuscript.

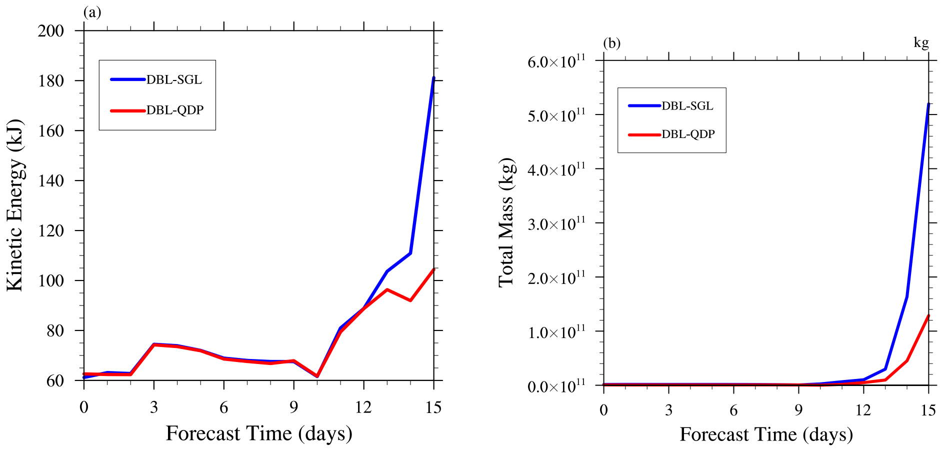


Figure 2. The temporal evolution of spatially averaged difference of (a) total energy, (b) total mass between DBL and SGL, as well as difference between DBL and QDP in case of Jablonowski and Williamson baroclinic wave.

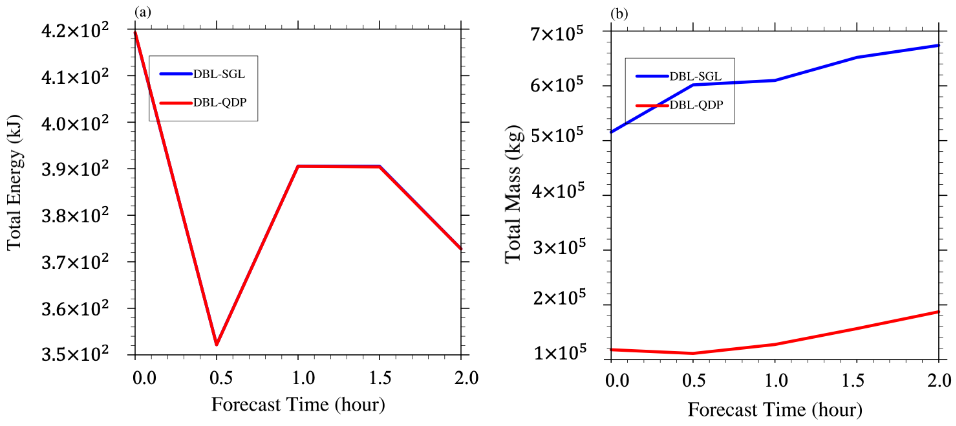


Figure 3. The temporal evolution of spatially averaged difference of (a) total energy, (b) total mass between DBL and SGL, as well as difference between DBL and QDP in case of super-cell.

It is important to note that, for enhanced clarity and to facilitate a better understanding of the trend, the x-axis unit for the figure representing Case 7 has been changed to hours (Fig. 4).

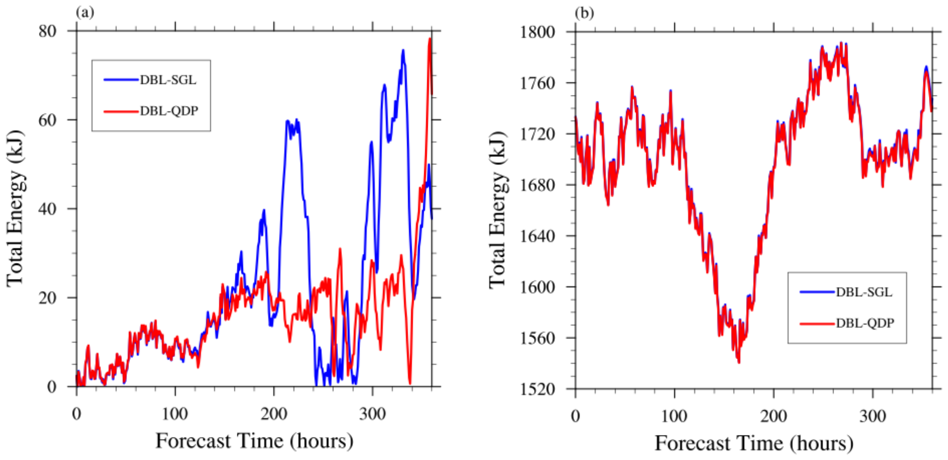


Figure 4. The temporal evolution of spatially averaged difference of total energy between DBL and SGL, as well as difference between DBL and QDP in case of real data, with resolution of (a) 240 km × 240 km, (b) 120 km × 120 km.

- Figure 7: The phrase "the circle represents the most clear error" lacks objectivity, especially since other areas in the figure show similar errors. Consider rephrasing or finding a more precise way to characterize the comparison between errors in this case.

Thank you very much for your careful review and valuable comments on my manuscript. Regarding the issue you raised in the figure 7. We decide to adjust the color bar, after adjusting, the two distinct areas of improvement are now evident, compared to only one in the original figure. I sincerely apologize that the previous color bar setting was not appropriate. and the revised figure is shown as Figure 5 (corresponding to Figure 7 in the manuscript).

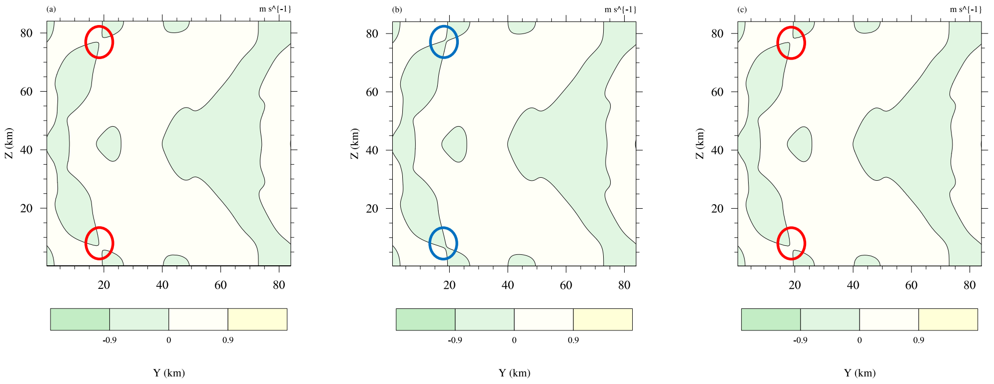


Figure 5. Perturbation theta in super-cell development at 5400s in the (a) DBL simulation, (b) SGL simulation and (c) QDP simulation (bias has reduced), unit: K, the circle represents the pattern bias.

- Please avoid using the term "significant" or "significantly" when not referring to statistically significant results to avoid potential confusion.

Thank you for this valuable feedback. We have corrected this oversight, and the revised text will be included in the final manuscript submission after all reviews have been addressed.

- Unclear phrase (line 95): "achieves basically consistent results comparable to those of double precision" – Do you mean that the accuracy is similar in a specific case, or that the quasi-double precision consistently performs similarly to double precision in most cases?

Thank you for raising this point. The phrase 'the single-precision version improved with the quasi double-precision algorithm achieves basically consistent results comparable to those of double precision. However, we acknowledge that the phrasing was ambiguous. We have revised the sentence to more accurately reflect this finding.   The revised sentence now reads:

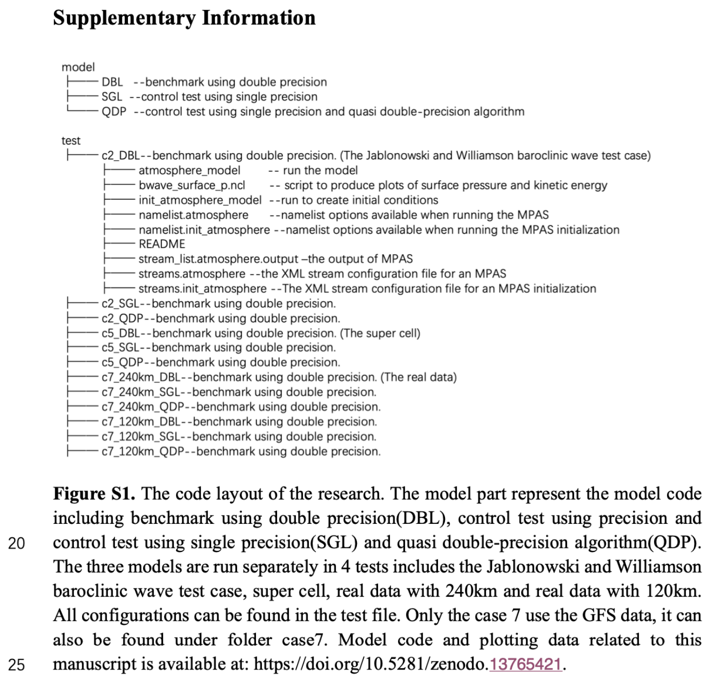
The quasi-double precision algorithm effectively mitigates the accuracy limitations of single precision, producing results largely consistent with double-precision calculations.

- Reproducibility note: Although the manuscript adheres to GMD's guidelines by providing a DOI to a permanent repository, no instructions are provided on how to use the code or what the various files and subfolders contain. I suggest including at least a brief README file that describes the content and organization of the repository, providing users with clearer guidance.

Thank you for your valuable suggestion regarding the code documentation. We have updated the supplemental material with a comprehensive description of the code structure and included a README file for each case, as you recommended. We apologize that this information was not initially referenced within the main manuscript. To ensure clarity and accessibility, we have also added explicit mentions of the code description and README files within the Code and data availability section of the manuscript, directing readers to the supplemental material for these details. Figure 6 shows a screenshot of the updated supplemental material, which now clearly includes these additions. Each case directory within the supplement contains a corresponding README file, an example of which is shown in Figure 7.

We believe these revisions significantly improve the transparency and reproducibility of our research. Thank you again for your helpful feedback.

Code and data availability. Model code and plotting data related to this manuscript is available at: https://doi.org/10.5281/zenodo.13765422. Details regarding the code structure and instructions for running the code are provided in the supplementary material, which can be downloaded and viewed in Fig. S1. This figure provides a visual overview of the code organization. The information of steps how to execute the simulations can be found in README file in each test case folder.



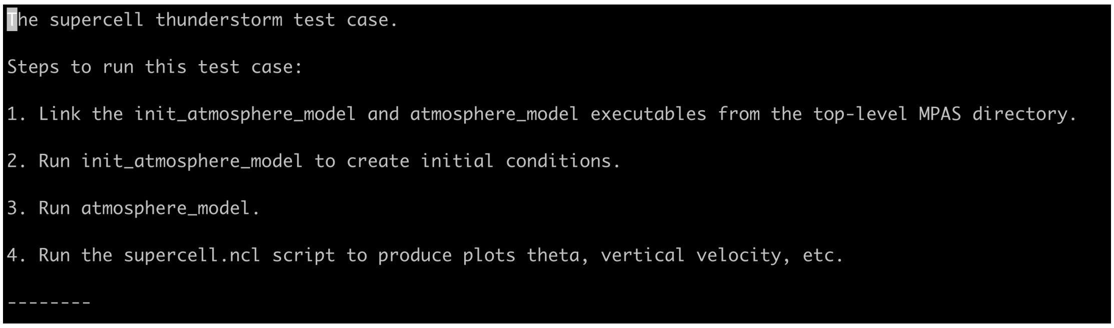
Supplementary Information

```
model
├── DBL   --benchmark using double precision
├── SGL  --control test using single precision
└── QDP  --control test using single precision and quasi double-precision algorithm

test
├── c2_DBL--benchmark using double precision. (The Jablonowski and Williamson baroclinic wave test case)
        ├── atmosphere_model        -- run the model
        ├── bwave_surface_p.ncl      -- script to produce plots of surface pressure and kinetic energy
        ├── init_atmosphere_model  --run to create initial conditions
        ├── namelist.atmosphere      --namelist options available when running the MPAS
        ├── namelist.init_atmosphere --namelist options available when running the MPAS initialization
        ├── README
        ├── stream_list.atmosphere.output –the output of MPAS
        ├── streams.atmosphere --the XML stream configuration file for an MPAS
        ├── streams.init_atmosphere --The XML stream configuration file for an MPAS initialization
├── c2_SGL--benchmark using double precision.
├── c2_QDP--benchmark using double precision.
├── c5_DBL--benchmark using double precision. (The super cell)
├── c5_SGL--benchmark using double precision.
├── c5_QDP--benchmark using double precision.
├── c7_240km_DBL--benchmark using double precision. (The real data)
├── c7_240km_SGL--benchmark using double precision.
├── c7_240km_QDP--benchmark using double precision.
├── c7_120km_DBL--benchmark using double precision.
├── c7_120km_SGL--benchmark using double precision.
├── c7_120km_QDP--benchmark using double precision.
```

Figure S1. The code layout of the research. The model part represent the model code including benchmark using double precision(DBL), control test using precision and control test using single precision(SGL) and quasi double-precision algorithm(QDP). The three models are run separately in 4 tests includes the Jablonowski and Williamson baroclinic wave test case, super cell, real data with 240km and real data with 120km. All configurations can be found in the test file. Only the case 7 use the GFS data, it can also be found under folder case7. Model code and plotting data related to this manuscript is available at: https://doi.org/10.5281/zenodo.13765421.

Figure 6. The code layout in supplement.



```
The supercell thunderstorm test case.

Steps to run this test case:

1. Link the init_atmosphere_model and atmosphere_model executables from the top-level MPAS directory.

2. Run init_atmosphere_model to create initial conditions.

3. Run atmosphere_model.

4. Run the supercell.ncl script to produce plots theta, vertical velocity, etc.

--------
```

Figure 7. The screenshot of README file in case suppercell.

Vánˇa, F., Düben, P., Lang, S., Palmer, T., Leutbecher, M., Salmond, D., and Carver, G.: Single Precision in Weather Forecasting Models: An Evaluation with the IFS, Monthly Weather Review, 145, 495-502, doi:10.1175/MWR-D- 16-0228.1, 2016.