Response to Reviewer 4

We are incredibly grateful for your efficient review process. Your insightful comments have provided valuable guidance for revising the manuscript. We have revised the manuscript according to your suggestions and will respond to your comments paragraph by paragraph. The comments are given below in black, our responses are in blue, and proposed changes to the manuscript are in red. Additional references are provided at the end of this document. The final revisions and specific locations corresponding to the manuscript will be marked uniformly after receiving feedback from other reviewers.

Major Comments

The abstract states: “Low precision computations can significantly reduce computational costs, but inevitably introduce rounding errors, which affect computational accuracy.” However, it is not clearly defined what is meant by “low precision computations,” and the statement that such computations inevitably affect accuracy may not always hold true.

We thank the reviewer for the valuable feedback. In the revised manuscript, we have added a clear definition of "low precision computations".

Regarding the assertion that "low precision computations inevitably affect accuracy," we appreciate the reviewer’s insight, and we will revise the description in the abstract. The specific modifications are detailed below.

1 Introduction

Specifically, we define low precision computations as those that utilize a limited number of significant digits (less than 64 bits) during numerical operations, which can significantly reduce the computational resources required while potentially introducing rounding errors.

Abstract

While low precision computations can significantly reduce computational costs, they may introduce rounding errors that can affect computational accuracy under certain conditions.

Both the abstract and conclusions mention the computational impact of the proposed methods; however, these effects are not discussed further within the main body of the manuscript.

We appreciate your suggestion, following your insightful comment, we will:

1. 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. Revise the description of computational performance in the abstract to reflect these updated and more accurate measurements.

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 1). 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 1).

Table 1. 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.'

Additional context would be beneficial in distinguishing which differences are relevant. Including an uncertainty analysis would help place the magnitude of the errors into perspective. For example, Figure 8 shows significant differences in errors between low- and high-resolution grids, with these discrepancies appearing more impactful than those arising from precision changes alone.

Thank you for your insightful comments. As you mentioned, there are multiple factors contributing to the errors observed in our study. In addition to round-off errors associated with floating-point arithmetic, the choice of grid resolution also has a significant impact on bias. We appreciate your suggestion regarding the inclusion of an uncertainty analysis, and we will consider incorporating this into the revised manuscript to provide further context. We will add a description in Section 3.3, you can see as follows:

3.3 Additional content

In this research, we focus on the processes of summing the basic field and trends. When the resolution is increased, the basic field remains relatively unchanged; however, the trends become smaller. This characteristic aligns with the nature of adding large and small numbers, making the advantages of the quasi double-precision algorithm more pronounced. Thus, it is evident from Figure 8 that as the resolution increases, the improvement achieved by quasi double-precision

algorithm also enhances.

On the other hand, it is important to note that the propagation of rounding errors is not immediately apparent over short time scales. However, as the number of iterations increases, these errors can become more significant. The quasi double-precision algorithm employs compensation mechanisms that help mitigate the propagation of these errors.

Section 2.3 would benefit from further detail on which parts of the code were modified and how these sections were selected for modification.

Thank you for your valuable feedback. To address this, we will:

1. Add the clear description of the solution method for the equations including temporal integration scheme and spatial discretization scheme in section 2.2. See 2.2 additional content below.

2. Specify at which step the quasi double-precision algorithm is applied within the computation process in section 2.3 and replace the figure (corresponding to Figure 3 in the manuscript) and explain this process using formulas and explanations. See 2.3 additional content below.

2.2 Additional content

The MPAS-A solves the fully compressible, nonhydrostatic equations of motion (Skamarock et al. 2012). The spatial discretization uses a horizontal (spherical) centroidal Voronoi mesh with a terrain-following geometric-height vertical coordinate and C-grid staggering for momentum. The temporal discretization uses the explicit time-split Runge–Kutta technique from Wicker and Skamarock (2002) and Klemp et al. (2007).

The algorithm applied here primarily addresses the rounding error compensation between large and small numbers in addition. Currently, it is only applicable to the time integration process and has not been implemented in the spatial discretization process. Therefore, this section will provide a detailed introduction to the time integration scheme. For the spatial discretization scheme, please refer to Skamarock et al. (2012), and it will not be introduced upon here.

The formulation of the scheme can be considered in on dimension as equation Wicker and Skamarock (2002):

$$\frac{\partial \emptyset}{\partial t} = RHS_{\emptyset} \tag{1}$$

The variable Ø represents any prognostic variable in the prognostic equations, while RHS represents the right-hand side of the prognostic equations (i.e., the spatial discretization equation). In MPAS-A, a forward-in-time finite difference is used, and it can be written as Eq. (2):

$$\frac{{\emptyset_i}^{n+1} - {\emptyset_i}^{n}}{\Delta t} = RHS_{\emptyset} \tag{2}$$

Where superscript represent the time step, and subscript represent the position of grid zone.

The two-order Runge-Kutta time scheme is used in MPAS-A as described in Gear et al. (1971):

$$\emptyset^* = \emptyset^t + \frac{\Delta t}{2} * RHS(\emptyset^t) \tag{3}$$

$$\emptyset^{**} = \emptyset^t + \frac{\Delta t}{2} * RHS(\emptyset^*) \tag{4}$$

$$\emptyset^{t+\Delta t} = \emptyset^t + \Delta t * RHS(\emptyset^{**}) \tag{5}$$

2.3 Additional content

According to Equation Eq. (3), (4) and (5), it can be observed that in the time integration scheme, each step involves the process of adding tends on the basic field $\emptyset^t$. In numerical models, the basic field is generally much larger than the tends, which aligns with the principles of numerical computation regarding the addition of large and small numbers, as well as the time integration process. It is important to note that the quasi double-precision algorithm currently only addresses time integration and has not been validated during the spatial discretization process. The spatial discretization primarily involves subtraction, specifically the subtraction of a small number from a large number or the subtraction of two close values. Whether this algorithm is applicable in spatial discretization remains uncertain, therefore, we will not apply it in this context.

Based on the application principles of the algorithm, which involve the processes of adding large and small numbers as well as the time integration process, we have established a strategy for applying the quasi double-precision algorithm within the MPAS-A. Specific improvements are provided based on the predictive equations:

$$\frac{\partial \boldsymbol{V}_H}{\partial t} = -\frac{\rho_d}{\rho_m}\left[\nabla_\zeta\left(\frac{p}{\zeta_z}\right) - \frac{\partial z_H p}{\partial \zeta}\right] - \eta \boldsymbol{k} \times \boldsymbol{V}_H - \boldsymbol{\nu}_H \nabla_\zeta \cdot \boldsymbol{V} - \frac{\partial \Omega \nu_H}{\partial \zeta} - \rho_d \nabla_\zeta K - eWcos\alpha_r - \frac{\nu_H W}{r_e} + \boldsymbol{F}_{\boldsymbol{V}_H} \tag{6}$$

$$\frac{\partial W}{\partial t} = -\frac{\rho_d}{\rho_m}\left[\frac{\partial p}{\partial \zeta} + g\tilde{\rho}_m\right] - (\nabla \cdot \boldsymbol{\nu} W)_\zeta + \frac{uU+vV}{r_e} + e(Ucos\alpha_r - Vsin\alpha_r) + F_W \tag{7}$$

$$\frac{\partial \Theta_m}{\partial t} = -(\nabla \cdot \boldsymbol{V}\theta_m)_\zeta + F_{\Theta_m} \tag{8}$$

$$\frac{\partial \tilde{\rho}_d}{\partial t} = -(\nabla \cdot \boldsymbol{V})_\zeta \tag{9}$$

The meaning of each variable in the equations exactly follows Skamarock et al. (2012), so that we don't repeating explanation. For a numerical model, the most crucial variables are the prognostic variables. Therefore, In the MPAS-A model we applied the quasi double-precision algorithm to the time integration process of these prognostic variables, including horizontal momentum ($\boldsymbol{V}_H$) , dry air density ($\tilde{\rho}_d$), potential temperature (Θ_m) and vertical velocity (W) , that is , the process in red of Eq. (6), (7), (8) and (9). (Only the predictive equations for the dynamic core are presented here, without the scalar transport.)

In Section 3.1, the differences between cases only emerge after 10 days of integration. It would be valuable to contextualize these differences with the error growth from other potential sources of uncertainty.

Thank you for your valuable suggestion. In response, we will add an analysis of other sources of error in Section 3.1. The content will include the following:

3.1 Additional content

The sources of unpredictability, as noted by Bauer et al. (2015), include instabilities that inject chaotic 'noise' at small scales and the upscale propagation of their energy. For the cases examined, both SGL and QDP begin to exhibit errors after 10 days of integration. These errors arise from factors such as rounding errors due to reduced numerical precision and energy loss during the propagation process. The quasi double-precision algorithm can reduce the impacts of these errors.

While we acknowledge other potential sources of uncertainty, such as initial condition errors, we have not conducted an in-depth study on them in this research. Our primary focus remains on evaluating the improvements provided by the compensation algorithm in addressing rounding errors.

Section 3.2 states that "the errors are very small and can be ignored." More context is needed here to help determine which differences are meaningful.

Thank you for your insightful comment regarding Figure 6(a) of section 3.1. According to RC1, we carefully reviewed our code based on your comments and identified that the issue was a problem with the data processing and plotting. We have corrected this issue, and the revised Figure 1 is provided below, it can be seen, the average bias between DBL and QDP is smaller than DBL and SGL.

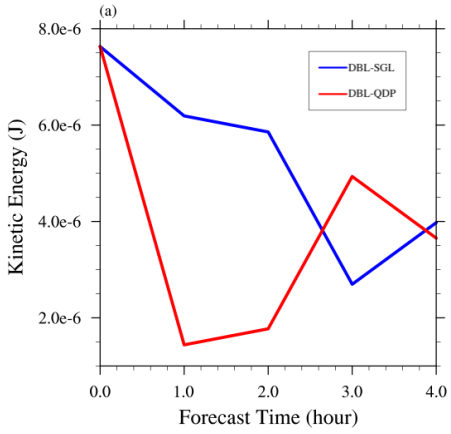

Figure 1. The temporal evolution of spatially averaged difference of kinetic energy between DBL and SGL, as well as difference between DBL and QDP in case of super-cell.

In Section 3.3, the authors note that “Differences in error begin to emerge after 500 steps.” This could be strengthened by comparing this error growth to that of other sources of uncertainty, some of which may become relevant earlier in the integration process.

Thank you for your insightful suggestion. We will enhance our discussion in Section 3.3 by adding a comparison of error growth with other sources of uncertainty. The added content will be as follows:

3.3 Additional Contents

Consistent with the analysis presented in Section 3.1, errors are relatively small in the early stages and begin to emerge after 500 steps. This increase is attributed to the accumulation of round-off errors and energy loss over time. The effects become more pronounced beyond 500 steps. Overall, the quasi double-precision algorithm demonstrates a certain level of improvement in addressing these errors.

Minor Comments

Line 19 – The authors reference a 2015 source to indicate that systems are expected to grow. While this is still valid, the “current systems” referenced in 2015 are no longer today’s current systems.

Thank you for your suggestions, we have revised it by modifying the "current systems" to “2015’s systems".

Line 47 – The mixed-precision reference for NEMO notes that “95.8% of the 962 variables could be computed using half precision,” though the publication itself refers to single precision.

Thank you for your suggestions, we have revised it by modifying the "half precision" to “single precision".

Line 143 – It is unclear if this version of MPAS-A is indeed the only one capable of utilizing single-precision.

We appreciate your comment. Upon reviewing the official documentation and user guide, we found that MPAS-A can be complied and run in single-precision since version 2.0. We apologize for the oversight and will revise the description accordingly.

Figures 5, 7, 9, 10, 11, and 12 – Alternative colormaps are suggested for all figures containing maps.

Thank you for your suggestion regarding the colormaps. We have made the requested modifications to the colormaps in all relevant figures. The updated figures are provided as Figure 1(corresponding to Figure 5 in the manuscript), Figure 2(corresponding to Figure 7 in the manuscript), Figure 3(corresponding to Figure 9 in the manuscript), Figure 4(corresponding to Figure 10 in the manuscript), Figure 5(corresponding to Figure 11 in the manuscript), Figure 6(corresponding to Figure 12 in the manuscript).

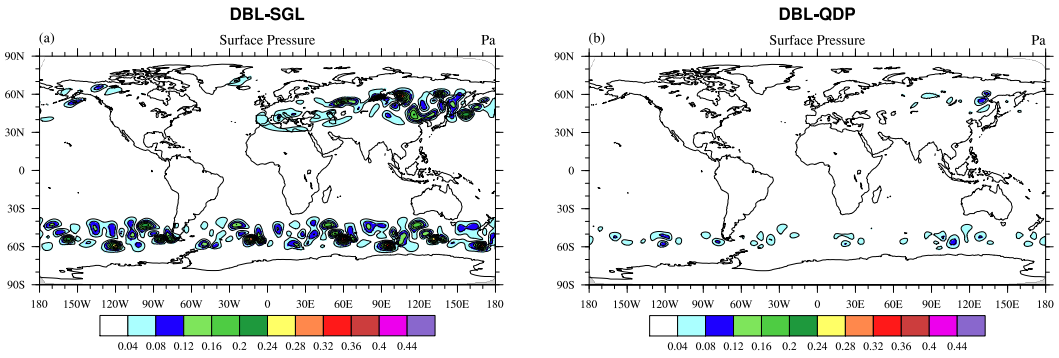

Figure 1. Spatial distributions of averaged (1-15days) difference of surface pressure (units: Pa) between DBL and (a) SGL simulations, (b) QDP simulations (round-off error has reduced) in case of Jablonowski and Williamson baroclinic wave.

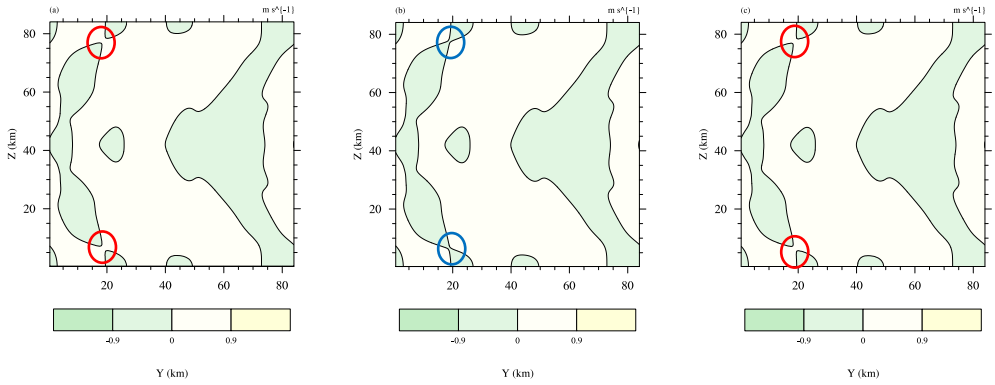

Figure 2. Perturbation theta in super-cell development at 5400s in the (a) DBL simulation, (b) SGL simulation and (c) QDP simulation (bias has reduced), unit: K, the circle represents the pattern bias.

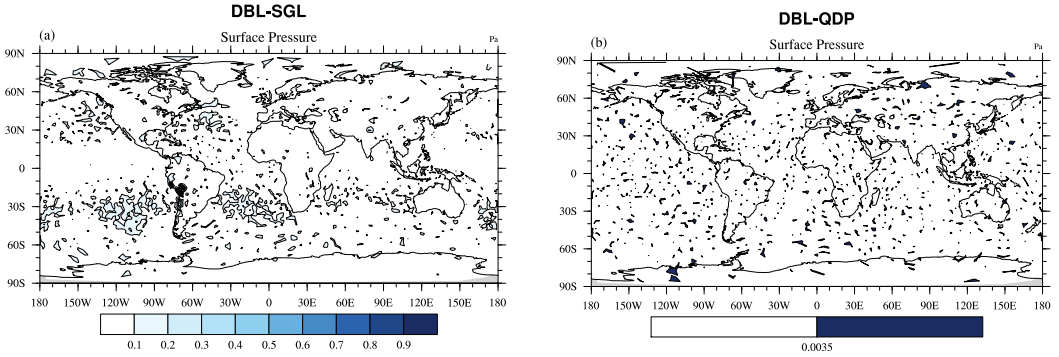

Figure 3. Spatial distributions of averaged (1-15days) difference of surface pressure (units: Pa) between DBL and (a) SGL simulation, (b) QDP simulation (resolution: 240 km × 240 km). The RMSE of surface pressure between DBL and (a) SGL simulation is 6.68×10^{-2} Pa, (b) QDP simulation is 2.25×10^{-3}Pa. (The color bars of (a) and (b) are different)

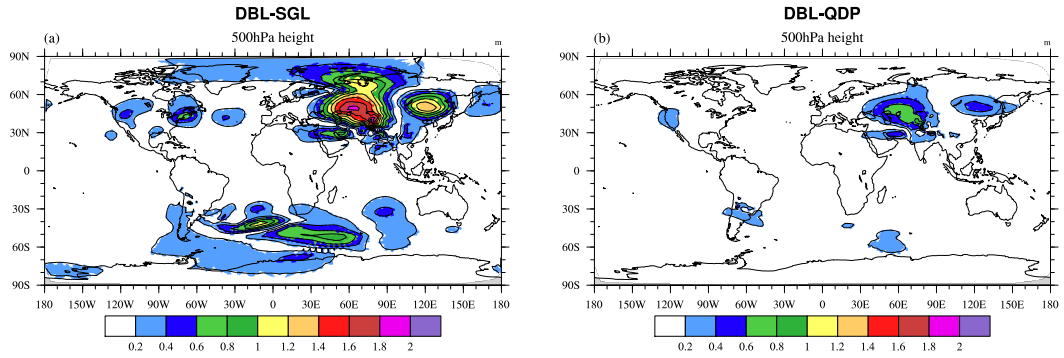

Figure 4. Spatial distributions of averaged (1-15days) difference of 500hPa height (units: m) between DBL and (a) SGL simulation, (b) QDP simulation (resolution: 240 km × 240 km). The RMSE of 500hPa height between DBL and (a) SGL simulation is 2.80×10^{-1} m, (b) QDP simulation is 1.40×10^{-1}m (round-off error has reduced).

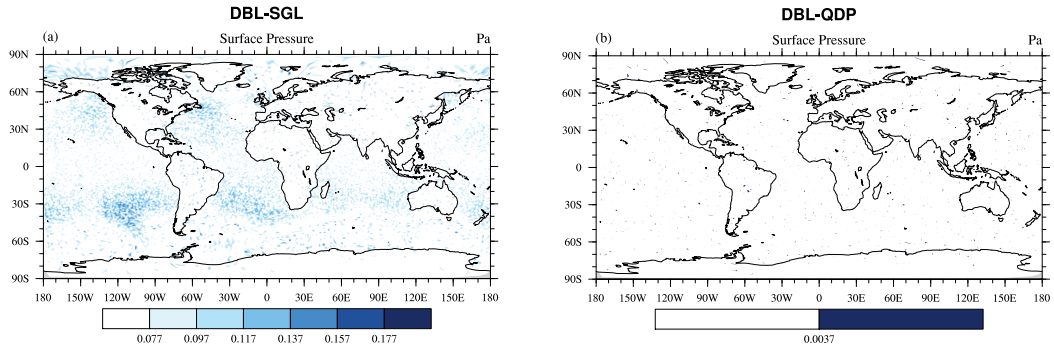

Figure 5. distributions of averaged (1-15days) difference of surface pressure (units: Pa) between DBL and (a) SGL simulation, (b) QDP simulation (resolution: 120 km × 120 km) (round-off error has reduced). The RMSE of surface pressure between DBL and (a) SGL simulation is 6.33×10^{-2} Pa, (b) QDP simulation is 2.25×10^{-2} Pa. (The color bars of (a) and (b) are different)

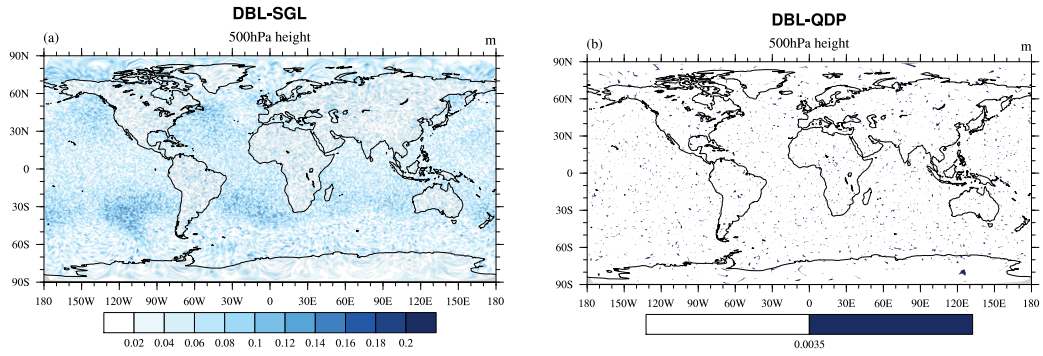

Figure 6. Spatial distributions of averaged (1-15days) difference of 500 hPa height (units: m) between DBL and (a) SGL simulation, (b) QDP simulation (resolution: 120 km × 120 km). The RMSE of 500 hPa height between DBL and (a) SGL simulation is 4.35×10^{-3} m, (b) QDP simulation is 1.90×10^{-3}m. (The color bars of (a) and (b) are different)

Line 241–242 – The phrase "but the process is more sensitive for the precision" lacks clarity and would benefit from rephrasing.

Thank you for your valuable suggestion. I will rephrase this statement for improved clarity in the revised manuscript. See bellows:

Currently, the quasi double-precision algorithm is only implemented in the time integration scheme of the dynamic core of the MPAS-A model, without factoring in tracer transport. However, the tracer transport process involves numerous operations where large and small numbers are added together, making it more sensitive to precision requirements.

Bauer, P., Thorpe, A., and Brunet, G.: The quiet revolution of numerical weather prediction, Nature, 525(7567):47-55, doi:10.1038/nature14956, 2015.

Skamarock, W. C., Klemp, J. B., Duda, M. G., Fowler, L. D., Park, S. H., and Ringler, T. D.: A Multiscale Nonhydrostatic Atmospheric Model Using Centroidal Voronoi Tesselations and C-Grid Staggering, Monthly Weather Review, 240(9):3090-3105, doi:10.1175/MWR- D-11-00215.1, 2011.

Wicker, L. J., & Skamarock, W. C.: Time-splitting methods for elastic models using forward time schemes. Monthly Weather Review, 130(8), 2088. https://www.proquest.com/scholarly-journals/time-splitting-methods-elastic-models-using/docview/198148677/se-2, 2002.

Gear, C. W.: Numerical initial value problems in ordinary differential equations[M]. Englewood Cliffs, N.J: Prentice-Hall, 1971.