Response to comments from referees on "Process-based diagnostics using atmospheric budget analysis and nudging technique to identify sources of model systematic errors" by C. Matsukawa et al.

We thank the referee for the valuable comments and constructive suggestions.
We have listed the comments by referee #2 below in black italics, followed by our responses below in blue.

Reply to Referee #2

The reviewer would like to thank the authors for their comprehensive study on model-based diagnostics to identify sources of model systematic errors, which could be useful for any NWP and climate models. The method successfully detects sources and origins of model forecast errors in the Met Office Unified Model, which could be common in the state-of-the-art operational and research atmospheric GCMs. I think The paper is comprehensive and well organized as a WCD paper, so that I would already recommend minor revision, though I have several comments below before acceptance.

Thank you so much for your reviewing and suggestive comments. We respond to your specific comments below.

Under the category "Specific comments"

1) The relaxation timescale of 6 hours could be practically useful to reduce/erase the initial errors that can grow up in subsequent forecasts. I am curious that how sensitive is "tau" (timescale) to your conclusion? For example, if you choose 12 or 24 hours as tau, you could obtain consistent results, that is, could detect the same origins causing the NH wind field errors?
The relaxation timescale might be similar with the evaluation time of the Forecast Sensitivity to Observations (Hotta et al. 2017, Prive et al. 2020). The shorter timescales might be more practical for forecasts.
(Hotta et al. 2017, MWR, DOI:10.1175/MWR-D-16-0290.1; Prive et al. 0220, QJRMS, DOI:10.1002/qj.3909)

We use the GLN experiment to make quasi-analysis data and use as a best estimate of truth of atmospheric budgets. In the nudging experiments, forced variables (i.e., U, V, and T) are relaxed towards the MetUM analysis data every model time step throughout the integration

up to 15 days in this study. The relaxation timescale "tau" determines how strong the model prognostic variables are relaxed towards forcing data rather than how long the nudging is applied to. Therefore, the GLN has small error against MetUM analysis as shown in Fig. 2(d) even at longer forecast lead time than 6 hours.
The choice of the relaxation timescale of nudging experiments is arbitrary. The shorter the relaxation timescale is, the stronger the model prognostic variables are relaxed towards forcing data (i.e. MetUM analysis data) throughout the model integration. As described in the cited previous study (Telford et al. 2008), too long relaxation timescale is ineffective to use a nudging experiment as a best estimate of truth, but too short relaxation timescale nudging could make the model unstable. We will add an additional explanation mentioned above to Section 2.2.
This paper doesn't focus on a difference in the relaxation timescale, but it sounds interesting and might provide an interesting result which could be useful for understanding a different timescale of error contribution of each component in the budget equations.

2) Could you comment in your conclusion part about two topics that could be useful as future studies?

Thank you so much for your suggestive comments.
These specific comments are really helpful to improve our manuscript and our response to two comments kindly given us are described below.

- Budget analysis based on conservation values, potential vorticity or angular momentum (TEM or MIM), could be useful for your further budget analysis. Currently I only refer to a paper by Kobayashi and Iwasaki (2016), though there may be other important studies to be cited.
(Kobayashi and Iwasaki 2016, JGR-A, DOI:10.1002/2015JD023476)

Thank you so much for your informative suggestions.
We will cite potential vorticity and angular momentum equations in the conclusion part.

- As a comparison with the experiment for 2018 boreal winter, additional experiment for an SH winter could be helpful to improve your conclusion (e.g., Yamazaki et al. 2023).
(Yamazaki et al. 2023, WAF, DOI:10.1175/WAF-D-22-0159.1)

We agree to your suggestive comments. Some conclusions in our study depend on seasons

and hemispheres. For example, gravity wave drag, which could be one of the sources of the wind error, is expected to be more active in NH because of the orographic features and also in winter because of the wind fields and atmospheric stabilities near land surface. Even if the temperature bias around tropical tropopause wouldn't depend on season, balanced situation of the compensating errors would change and result in different consequences. We are wondering that a focus on specific season and hemisphere makes this paper detailed and less confusing. However, we will implement some comments described above into Section 6 in a revised manuscript. This could help to improve our conclusion and discussion.