

The manuscript evaluates the use of assimilating dD measured by the satellite IASI in addition to traditional meteorological variables like q and T (also measured by IASI) in an Observation System Simulation Experiment (OSSE) with the isotope-enabled climate model IsoGSM. In general, dD adds little but notable skill to the analysis. Since dD carries information about the phase change history of air masses, the assimilation of dD particularly improves the analysis in cases of strong condensation or evaporation (identified by high latent heating/cooling rates). These extreme events usually involve strong precipitation and are therefore societally relevant, but are poorly captured in most analyses. Therefore the improvement for these events is very promising.

The manuscript is well-written and the figures are appropriate. I appreciate that everything is nicely structured and carefully documented. It is easy to follow the authors' explanations throughout the manuscript. I have a few comments that can hopefully further improve the manuscript and/or clarify some aspects.

Many thanks to the referee for reading our manuscript and for their efforts in evaluating the design and presentation of our study. The comments help in deed to further improve and clarify the manuscript.

General comments

- 1) As far as I know IASI does not see through clouds. Were observations in cloudy conditions filtered out before the assimilation? I would expect that strong latent heating events are almost always associated with clouds. If IASI cannot measure dD in these cases, is the improvement of the analysis thanks to dD still realistic?

This is an important comment and it has also been made by referee 1. MUSICA IASI data are only available when there are no clouds or when IASI's field of view is only very weakly affected by clouds. In the revised manuscript, we fully consider this limitation and calculated all skill values again. This required changes of Figs. 2-4, 6, 7 (numeration as in the revised manuscript). Some quantitative adjustments result from this, while qualitatively the results remain similar. In order to document the kind of atmospheric states that are represented by the observable events, we add a new figure (Fig. 1). This figure shows how the cumulative distribution functions of the different atmospheric variables change, if we only look on cloud-free events. For instance, the cloud-free observations represent only events with latent cooling (or low latent heating) and vertical downward transport (or weak vertical upward transport). For more details on the involved modifications and for an explanation/justification of the setup used in the version of the discussion phase please see our reply to referee 1.

- 2) The latent heating rate is defined as the change of specific humidity in an air parcel (material derivative of q) times the latent heat of net condensation (equation 9). However, mixing processes can also lead to a change in specific humidity in air parcels and could

therefore bias the analysis. Is there a way to separate mixing from latent heating/cooling, e.g. by diagnosing Q2 directly in the model?

To our understanding, the changes of q by horizontal or vertical transport are considered in Eq. (9) by the two terms “ $v \nabla q$ ” and “ $\omega \partial q / \partial \omega$ ”. Or is there a misunderstanding? In any case and in line with a recommendation of referee 1, in the revised manuscript we now stratify the results with respect to vertical velocity (ω) instead of latent heating (Q2).

- 3) I think it might be nice to show also the horizontal spatial distribution (i.e. maps) of the differences in the skills to see the regions where the assimilation does or does not work well. Or is it pretty uniform?

We agree that the spatial context of skill changes is an important aspect. We analyse only 40 days of data with a 6 hourly resolution, and performing skill calculation for each location independently is very uncertain, because there are only $40 \times 4 = 160$ data points available for calculating the RMSD values. For small RMSD, the respective skill uncertainty is then often larger than the obtained skill value. In our study, we use large data amounts when calculating skills, i.e. use analysis data from many different locations and time steps together. This assures that the obtained skill values are above their estimated uncertainties. For spatial patterns of RMSE values, we would also like to refer to the paper of Toride et al. (2021).

- 4) Is there a reason why you use daily mean values for the analyses? I would expect that especially the strong latent heating/cooling events are rather short-lived and the improvement could be better on shorter time scales, e.g. 6 hours.

IASI makes observations at approximately the same location every 12 hour. If we furthermore account for the lack of observations when there are cloudy conditions, the observation frequency of a certain geolocation is generally less than 12 hours. For this reason our study in the version of the discussion phase was made with daily mean analysis, i.e. we focus on the time scales that are representative for the observations.

On the other hand, your argument is valid and using an OSSE, it should be possible to reconstruct part of the daily cycle in the analyses even though the observation frequency is sparser. For this reason, in the revised manuscript we present the evaluation of 6 hourly data, which, however, causes no qualitative change of the results.

Specific comments

L24: Suggestion: „where latent heat is released or consumed“

Ok, thanks!

L25: impacting on > impacting

Ok.

L39: As far as I know, it should be D and H in the equation (instead of HD16O and H216O).

There is also H218O or HD18O, so we suggest to leave as is.

L50: There is also an isotope version of NICAM (Tanoue et al., 2023), which might be worth adding here.

Ok, we add this reference.

L53: clouds or precipitation involving processes > processes involving clouds or precipitation

Ok, thanks.

L83: such assimilation > such an assimilation

Thanks for the comment, we will double check with a native speaker.

L108: What is different in the 96 initializations? Later you write the initial conditions. What exactly is different in the initial conditions?

The initial conditions for 96 ensemble members are chosen from the nature run from 0000 UTC June 1, 2016 with a 6-h time step, i.e. we apply a time shift in the initial conditions of one month. This makes the initial conditions practically independent of the nature data one month later, but similar climatological conditions remain (for more details see Toride et al, 2021). We will add here the reference to Toride et al. (2021)

L155+: Add/Explain somewhere what is a good skill and what is a bad skill? E.g. 100% means perfect, 0% means same bad as the reference.

Ok.

L157 do do > to do

Ok, thanks.

L162+: This has been said many times already. I think it could be removed here or somewhere else.

Ok, we shorten here.

L171: „Therefore...“: I don't see how this sentence follows from the previous sentence.

Ok, we remove “therefore”.

L179: ... and 17 vertical levels?

Yes, the analysis output is provided at 17 vertical levels, but we do not think that this is important here.

L208: 2x between

Thanks.

L257: close the > close to

This sentence has been modified in the revised version.

L281: uncertainty > error

Ok.

L292: this uncertainties > these uncertainties

Thanks.

L303: Remove „an“

Thanks.

L304: independent on > independent of

Thanks for the comment, we will double check with a native speaker (colleague).

L306: quantitatively document > quantify?

Thanks.

L309: How did you define these bins? Why not the same 60 bins as before?

In Fig. 3 (revised Fig. 4) we show the distribution of RMSD values for many equidistant ω bins. This overview shows that RMSD values are often small at ω close to zero. For Fig. 5 (revised Fig. 6) we use less bins with large bins for ω close to zero. The reason is that for ω close to zero, the skills are small and if we used smaller bins the respective skill uncertainty would be much larger than the obtained skill value, i.e., it would not be informative.

L343: in particularly > in particular / particularly

Thanks.

L346: by the additional assimilating of > by additionally assimilating / by the additional assimilation of

Thanks.

L359: These subset > This subset

This sentence has been modified in the revised manuscript.

L392: I would add NICAM here again.

Ok.

Figure 1, caption: remove „skill“ after (e). Is „only q“ not also „only one type of observation“ (like „only T“ or „only dD“)?

Thanks.

Reference

Tanoue, M., Yashiro, H., Takano, Y., Yoshimura, K., Kodama, C., & Satoh, M. (2023). Modeling Water Isotopes Using a Global Non-Hydrostatic Model With an Explicit Convection: Comparison With Gridded Data Sets and Site Observations. *Journal of Geophysical Research: Atmospheres*, 128(23), e2021JD036419.