Review of Villiger et al, part 1

April 17, 2023

This article presents an evaluation of high-resolution isotope-enabled simulations of the North Atlantic trades against observations from the EUREC4A campaign. This evaluation serves a a basis for the part 2 of the study, which investigates the role of the atmospheric circulation at different spatial scales on the isotopic variability. This article also presents interesting elements in itself, such as the sensitivity of the results to horizontal resolution and the isotopic signatures of clouds and of dry-warm patches.

The article is well written, although it is quite long. It is well illustrated. I have only minor comments.

- l 16: "mesoscale cloud base feature" is a bit enigmatic at this stage. Maybe be more specific, e.g. isotopic characteristics were investigated under precipitating clouds, non-precipitating clouds, clear-sky conditions and dry-warm patches.
- l 18: I don't understand what these numbers represent: what do the bounds represent? And the numbers in brackets? Maybe clarify with one or two additional sentences.
- l 313: clarify in this sentence if this is a daily value, or a temporal-mean? Or maybe it doesn't matter because the closest level to the mid-level is constant in time?
- p10, Fig 3:
  - How were cloud-base and subcloud levels defined in observations? A few sentences on this would be helpful.
  - Why is the subcloud level defined as mid-level between cloud base and the surface in simulations? COSMO overestimates the cloud-base level but systematically underestimates the subcloud level. Why not defining the subcloud level in COSMO in a way that is more consistent with observations?
  - To what extent do the variations in cloud-based altitudes contribute to the variations of observed and simulated variables at this altitude?
  - To what extent do the model-data differences in cloud-based altitudes contribute to the model-data differences in variables at this altitude?
- l 325: unclear sentence. Do you mean "... is assessed, without upscaling outputs from the higher resolution simulations to the lowest resolution"?
- l 418: "derive from it" -> "conclude"?
- l 420: "A behavior also" -> "This behavior is also"
- l 462: "10‰" -> "10‰ in $\delta^2 H$"?
- l 464: "promoting too" -> "promoted by too"? I understand from section 3.1 that the number of cloud droplets and their size is determined by some tunable parameter, so the overestimated evaporation is a consequence of this, rather than the reverse?
- l 460-467: What is the difference between the two mechanisms? It sounds like they are equivalent. In both cases, overestimated cloud evaporation leads to weak moistening and strong enrichment. The underestimated precipitation efficiency in (1) sounds like just a bulk view to simply formalize this effect,whereas (2) is the detailed mechanism?

- l 477-479: the impact of the low-level dry bias on the high bias in d-excess could easy be quantified: the slope of d-excess as a function of surface relative humidity is easy to estimate under the Merlivat and Jouzel 1979 approximation. If the low-level dry bias is enough to explain the high bias in d-excess, then it is not necessary to invoke more complicated explanations.
- l 636: the role of "partial evaporation of falling hydrometeors" is discussed, and then the conclusion is on "mesoscale transport processes". Doesn't the variability in d-excess rather reflect a combination of microphysical and mesoscale transport processes?
- l 668: "the higher the" -> "at higher"?
- l 706: "linked to shallow convective upward transport": and partial evaporation of falling hydrometeors?
- Fig 5:
  - To what extent is the simulated cloud-base cloud fraction sensitive to the threshold on condensed water content? Is there a possibility that the cloud fraction in observations and in the simulations represent two different things?
  - Is it possible that the sensitivity of the cloud fraction to the resolution is just an artifact of the different resolutions? Cloud fraction can be very sensitive to the resolution at which a cloudy pixel is defined. If all simulation outputs were coarse-grained on the same grid, would they exhibit such a large difference in cloud fraction? In other words, is the large difference related to different physical states, or to different definitions of a cloud? What does the observed cloud fraction represent, in terms of horizontal resolution of what is defined a cloud?