Review of nhess-2019-174:

"Simulations of the 2005, 1910 and 1876 Vb cyclones over the Alps – Sensitivity to model physics and cyclonic moisture flux"
by
Peter Stucki, Paul Froidevaux, Marcelo Zamuriano, Francesco Alessandro Isotta, Martina Messmer, Andrey Martynov

<u>Recommendation:</u> major revisions

The authors present three cases of extreme precipitation along the Alps associated with the classical synoptic situation Vb. They utilize downscaling of different reanalysis products for their analysis, where they employ WRF in a nested method to go down to 3 km resolution in the innermost domain. The authors test several sensitivities of their results with respect to the simulation setup, addressing both numerical as well as physical changes. They find that the lead time is among the most crucial parameters. While I find the manuscript well written, I struggle to see a clear motivation and conclusion for the study. The motivation is probably also somewhat difficult, because the manuscript tries to address many different questions at the same time: (1) What is the best downscaling setup for the cases in question. (2) What lead to the extreme precipitation. (3) Description of the three individual cases and comparison. The treatment of all these topics makes the paper sometimes difficult to follow. Regarding the conclusions, similar arguments apply and the authors sometimes appear to state opinions/speculations that are not necessarily solidly grounded in the material they presented. However, after responding to some of my major concerns, I believe this manuscript is acceptable for publication in NHESS.

<u>General Comments:</u>

In the title, I am not sure what the authors really mean with the "cyclonic moisture flux". Is it a moisture flux going in a cyclonic direction or the moisture flux associated with a cyclone? Given this ambiguity, I encourage the authors to further clarify this aspect in their title as well as throughout the manuscript to help the reader making a clear link.

Given that there are only three cases, and the fact that a majority of the reasoning for the downscaling is based on the most recent case, the authors should be more cautious about general statements on downscaling procedures, as the results are highly sensitive to the case(s) at hand. While the technicalities that were overcome by the group are certainly impressive, it is still not clear to me how generic these results can be treated. In order to make a more general claim about the downscaling for Vb situation, one would need to explore many more cases to arrive at a firm conclusion. The authors should thus make it clear that this study can at most give an indication what one might need to test in order to arrive at a more general conclusion.

What made the authors pick a 10-day spin-up time? It seems excessively long for the investigation of such a regional and meso-scale influenced precipitation event. At the end, the authors arrive at a 1-day spin-up time anyway, but the vastness of the parameter space is not sufficiently motivated, similar to some of the other sensitivity tests.

The list of sensitivities is extremely exhaustive, ranging from resolution to resolution ratios over spin-up time to parameterizations and model domains. The enormous parameter space is rather difficult to grasp and all results will primarily be in relation to the 2005 case, with general deductions being rather limited due to the specifics of the case. In general, it would aid the reader if the authors more clearly state their working hypotheses as well as the reasoning for their choices and expectations. This will make it more straight forward to follow the ensuing arguments.

The authors often refer to reproducing “correct” precipitation amounts. What is meant by correct? Presumably compared to observations, though the authors list several observations that are used. In addition, all of these “observations” also rely on some sort of downscaling and gridding, as data voids need to be filled. The authors, however, do not provide a detailed analysis of the representativeness of these observations. They refer to other studies that addressed these to some extent, but given the specifics of the case studies, the authors should also comment on the validity of the observations before comparing the model simulations to the data in order to claim “correctness”.

For the validation of, for example, precipitation, it has proved useful to use feature-based detections that consider location, shape, and timing. Why have the authors not considered more such verification tools for the study at hand? It appears the method referred to as EMD is in fact such a measure, though it appears confusing why the authors use a visual inspection for a quantitative comparison. The reasoning for the choices and omission of other tools should be clearly motivated.

For the philosophical concluding paragraph on page 25 not much hard evidence has been provided in the manuscript for the claims put forward. It thus reads more like a written piece of opinion than a well and quantitatively justified conclusions.

Specific Comments:

The page (P) and line (L) numbers refer to the ones in the manuscript.

P1 L23: “to the cyclonic” and see comment above about the ambiguity of “cyclonic moisture flux”.

P1 L28: “accurate directions” with respect to what? What is the reference?

P8 L1: The precipitation data is interpolated. Can the authors please clarify if the interpolation was carried out in such a way that the total precipitation was unaffected by the interpolation? Depending on the kind of interpolation, the results for the totals can deviate.

P11 L12: The authors speculate on the differences between ERA-Interim and 20CR in terms of moisture distribution, though the authors could provide direct evidence for their claim by investigating differences between ERA-Interim and 20CR fields for the case at hand in more detail.

P11 L22: Why did the authors chose to identify and track cyclones using geopotential at 500 hPa? This seems rather unconventional and needs further motivation.

Fig. 4: The cyclone tracks for ERA-Interim look very edgy. In order to compare them better to the other plots, a grid that is not fixed to the grid spacing of the data could be beneficial, which is most often done in other cyclone track algorithms, see also comment above about cyclone track determination in this manuscript.

Fig. 7, 8, and 9: I find these figures not very legible. Maybe this is due to the downgrading of the figure quality for the review process, but otherwise the readability of the information of these figures needs to be significantly improved. In particular the arrows are not very visible.

P22 L21: The authors should explain how PV is produced in the downscaling process, as this appears to be crucial in their arguments.

P23 L24: How can the authors conclude that “nudging smaller domains can still be beneficial”? Has any evidence been provided in this study to support such a claim?

P24 L1: The authors should be more specific what they are referring to with “traditional spatial verification scores”.