Simulations of the 2005, 1910 and 1876 Vb cyclones over the Alps – Sensitivity to model physics and cyclonic moisture flux
Peter Stucki
et al. nhess-
2019-174

There are 3 main objectives of this paper (i) to find a setup of the WRF model that is adequate for dynamical downscaling from 20CR, (ii) to investigate sensitivity of heavy precipitation to cyclonic moisture flux and (iii) to assess the uncertainty along the downscaling steps and among the ensemble members for historical cases. The paper is well written but I'm concerned about the balance of the paper. The paper is dominated by the technical aspects of performing downscaling for historical events (with poorly-defined motivation for performing model tests and testing ensemble suitability) and contains too brief analysis of the cyclonic moisture flux to achieve objective (ii) (see general comments below). If the points below are addressed this paper would be suitable for publication in NHESS.

We thank the reviewer for pointing to some important aspects and potential flaws of the manuscript. We think that while addressing the concerns, we provide now a more balanced manuscript regarding the technical and dynamical parts of the article. The motivation, or rather the rationale, of doing the series of experiments is now described in a more explicit and clear way. Considering also comments by Reviewer 1, we have better defined and extended the analyses of cyclonic moisture flux with new figure panels, a new figure and extended text.
In all, we are grateful for the review, which supported us in improving the quality of the manuscript regarding some crucial aspects and clarity.

General comments

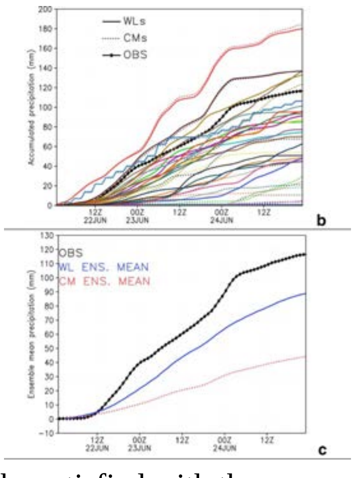
1. I'm concerned that there is no motivation given for testing sensitivity to the convection, microphysics schemes or nesting in WRF. What was the reason for performing these simulations? Did the authors have a hypothesis that they wanted to test? What are the differences between the schemes? The authors conclude that there is no difference in performance when changing the cumulus scheme or nesting but do not analyse this result. Did they expect to see a difference? If so, why are the results insensitive to these choices? What is the conclusion of these experiments
and how general are they, i.e. would the same hold for other historical cases or are they specific to these cases? If the conclusions are case specific, then perhaps this analysis could be reported in an appendix?

   We see the need for a better reasoning about the experiments. To begin with, we state more clearly why we started with ample spin-up time, then why are not satisfied with the standard setup, and then derive why we go with decreasing spin-up. Regarding the specific question of 'sensitivity' to a number of schemes and settings, we now stat more clearly that "The goal of these last experiments is not to achieve a thorough sensitivity assessment for each tuning option, but to make sure that we have not chosen a sub-optimal setup. This is checked by further modifying a number of configurations of the WRF model which may have an influence on the simulation performance according to literature and from our experience**." In turn, this means that we**

do not aim to analyze each tuning option in detail.
Some of this concern could also stem from the fact that we use the term ‘sensitivity’ in the title, which may provoke expectations of a thorough testing of many schemes and options. In a narrow sense, the term might only be acceptable for the more detailed tests regarding the spin-up. We use ‘sensitivity’ in a wider sense, where we also allude to checking the other tuning options in less detail as well as to the dynamic sensitivities addressed by contrasting members, for instance. In fact, we have considered replacing the term ‘sensitivity’ in the title and elsewhere (e.g. Implications from …, the role of …, etc.). We have not found a concise alternative term that would include all these meanings, and we think that the meaning in this article should become clear when reading the abstract.

2. There is some confusion in the paper over what constitutes a ‘good’ ensemble spread. The answer to this depends on the hypothesis being tested. At some points in the paper the authors claim the ensemble is good or bad by examining spread in the precipitation totals (figure 2, table 2) concluding that the 10-day forecast runs ‘too freely’ because the precipitation accumulation spread is large. However, later in the paper they examine the spread in cyclone tracks (figure 4) and conclude that there is a ‘good’ spread in storm track position for the 2005 and 1910 cases but not the 1876 case (meaning smaller track position differences in the ensemble). If the focus of the paper is to test the sensitivity of precipitation accumulation over Switzerland to cyclonic moisture flux, some spread in precipitation accumulation is surely necessary? However, spread in precipitation that occurs due to factors such as cyclone position presumably need to be minimised? Is this the rational for the later measure of ensemble suitability? If so, why is precipitation spread used in the early analysis of ensemble spread? There appears to be some inconsistency in the analysis of ensemble suitability in the paper which needs to be clarified.

We agree that we need to address in more detail what we mean by a ‘good spread’, and the comment has helped a lot in this respect. We discussed the trade-off between ‘plausible runs’ and ‘necessary spread’. In the end, we think that we can justify our line of thoughts well in an additional paragraph in Sect. 3.1.



In addition, we include a reference (Coppola et al., 2018, https://doi.org/10.1007/s00382-018-4521-8) to a thorough test of multi-model ensembles, where the inclined reader can actually look up how state-of-the-art models perform in a comparable Alpine setting, see their Figure 5 to the right.

The new text reads now as follows: “In all, we cannot be satisfied with these results yet. On the one hand, we aim to investigate particularly flood-inducing features of Vb-cyclones. For this, a certain variability in the ensemble is helpful and necessary. For example, we can find and assess (non-)decisive features by means of opposing ensemble members. On the other hand, we also need to ensure that our downscaling experiment delivers plausible results, especially regarding precipitation intensities and patterns. For this, the deviations from the observations must not become too large in the ensemble. A somewhat smaller spread of the simulated precipitation for the 2005 case would also increase our confidence that

the simulation of historical events will produce reasonable and valid results.
In short, we would have expected less underestimation and smaller deviations with this downscaling configuration (cf. Coppola et al., 2018, their figure 5, for estimations of accumulated precipitation over the Alps from a multi-model ensemble)."

3. Analysis relating to the sensitivity of precipitation to cyclonic moisture fluxes (figures 7-10) is described in just 19 lines. This is rather brief for 4 figures, especially given that one of the major objectives of the paper is to examine 'sensitivity to cyclonic moisture flux' (title). The analysis must be expanded to provide a better balance between the technical aspects and the scientific hypothesis testing analysis and to achieve objective (ii).

We agree on this comment, which is similar to concerns by Reviewer 1, and it has made us review and revise the manuscript in several ways.
The discussion of Figure 10 is actually done later in the text (see also reply to specific comment 25). We have introduced more references to Figure 10, for instance, to enhance the context of our findings.
To be more balanced between technical and dynamic aspects of sensitivity, and to address the innermost domains better (see comments by Reviewer 1), we have now redrawn Figure 5 and added two panel rows showing the contribution of convective precipitation in the 9-km domain, where precipitation is parameterized. In addition, we add a new Figure 11. It shows surface weather simulated for a historical maximum-precipitation member at three instances in time. With this, we can better show how shifts in the cyclonic flow and moisture transport translate into regional to local surface weather and precipitation patterns. An according clause is added in the Abstract.

<u>Specific comments</u>

1. Page 1, line 22: What is 'moderate' spectral nudging?
   Moderate refers to wavelengths of 1500 km; this is only defined in the text. We see that the term is not commonly used and decided to remove the information from the abstract, and instead focus on the most robust result, the initialization period. This is also in accordance to the comment by Reviewer 3.

2. Page 2, line 20: How is the moist air 'let' around the Eastern Alps? Do you mean advected?
   The sentence is rephrased in active voice with 'flowed over and around the Eastern Alps'. Note that in agreement with Reviewer 3, the dynamics of 'cyclonic moisture flux' is now better defined here.

3. Page 4, line 28: The authors state that they use a 'consistent part of the calibration period, which is accordingly slightly reduced'. I'm unclear what the consistent part of the calibration refers to. Please could the authors expand on this?
   We have modified the sentence to improve clarity. The stations used are the one that were available in the days of interest in 1910 or 1876 and at the

same time during most of the calibration period. We repeat the calibration period (1981- 2010, given also two sentences above). The calibration period is slightly reduced to the days when all chosen stations were available (the method needs to have the same set of stations in the calibration period, in our case from 1981-2010, and the reconstructed period, i.e. the days in 1910 or 1876).

4. Page 5, line 5. It would be useful to know if any of the assimilated surface pressure observations were located in Switzerland.
   We thank the reviewer for the hint. We have introduced the numbers here, and come back to the total number of assimilated stations in the conclusions. This is also in line with a suggestion by Reviewer 1.

5. Figure 1: The numbers on the colour bar have been cut off.
   We thank the reviewer for the remark. We have replaced Figure 1.

6. Page 6, line 26 and page 16, line 18: The authors refer to 'two peak episodes' but in figures 2 and 5 the CombiPrecip dataset does not show 2 peak episodes. Instead there is continuous high precipitation rates over a 30hr period.
   We were taking the perspective from the simulation, but agree that this is a misleading phrasing. We rephrased on Page 6 to 'Furthermore, CombiPrecip shows high precipitation rates over most of the analyzed period. In contrast, the downscaled ensemble has only two periods of very high precipitation rates (at around 12 and 36 hours after initialization), and obviously, there is too little precipitation between these two simulated peak episodes**'. A similar phrasing is employed on Page 16. We think that it fits also better with the revisions suggested by Reviewer 3.**

7. Figure 2: The right-hand axis does not have any units. Also, it is not clear what the red numbers represent.
   We thank the reviewer for pointing this out. We have redrawn Figure 2, and have added a # to mark the quartile members, and have also added it in the caption for clarity.

8. Page 7, line 1: It is not surprising that a 10-day forecast exhibits large spread in the ensemble. However, this is not necessarily a bad thing if the cyclone tracks are similar but with differing moisture flux as they would still be able to test the sensitivity of precipitation totals to moisture flux. Therefore, I don't think it is sensible to examine the suitability of the ensemble by looking at the spread in precipitation as is done in table 2.
   We suppose Figure 2 is meant instead of Table 2. This comment helped a lot for clarifying our rationale for the downscaling, and the follow-up series of experiments. Following also the comment on this topic by Reviewer 3, we now explain this in more detail in the analysis and interpretation of Figure 2. Later in the text, we also explain why we are not satisfied with the results and how we came up with the follow-up experiments. See also our reply to comment 10.

9. Page 7, line 16: Can the authors be more specific about the section containing the full evaluation. Currently they say is it ‘below’, below where?
   We have introduced ‘Sect. 3.2’.

10. Page 10, lines 11-12: There is no motivation given for testing sensitivity to the convection, microphysics schemes or nesting in WRF. What was the reason for performing these simulations? Did the authors have a hypothesis that they wanted to test? What are the differences between the schemes?
   We suppose it is Page 8 instead of 10. There, we now give the reasoning for the last series of experiments: ‘The goal of these last experiments is not to achieve a thorough sensitivity assessment for each tuning option, but to make sure that we have not chosen a sub-optimal setup. This is checked by further modifying a number of configurations of the WRF model which may have an influence on the simulation performance according to literature and from our experience.**’ We thank the reviewer for this remark; we think we are much clearer now about the purposes of the three series of experiments. See also our response to Reviewer 1 on this issue.**

11. Page 10, line 27: The authors claim that there is a systematic improvement with decreasing lead time. However, this is difficult to detect in the spatial verification statistics shown in table 2.
   We see that detection by eye might be difficult. Therefore, we have included the calculated Theil-Sen slopes in the text for better documentation. The new text reads now as follows: “Theil-Sen slope estimates are calculated over sp10, sp7, sp5, sp3, and sp1 for all measures; they are all negative (MAE 24 h: -0.2; MAE 48 h: -0.70; BOX 24 h: -0.03; BOX 48 h: -0.03; VIS 24 h: -0.75; VIS 48 h: -1.13; EMD 24 h: -18.5; EMD 48 h: -17.9). Although the trends are not significant in Mann-Kendall tests (or not clearly attributable, due to the small sample), the negative slopes indicate that performance generally increases with decreasing spin-up time**”**

12. Page 11, line 4: The authors conclude that there is no difference in performance when changing the cumulus scheme or nesting but do not analyse this result. Did they expect to see a difference? If so, why are the results insensitive to these choices? What is the conclusion of this experiment?
   We agree that the explanations are not very detailed in the manuscript. This is mainly because, as we state in the text now, the goal of these last experiments is not to achieve a robust sensitivity assessment for each tuning option, but to make sure that we have not chosen a sub-optimal configuration of the WRF model. We have inserted some explanations of why a different setup might change the precipitation pattern and indicate some more literature. Some of our reasoning is given below in more detail.
   The microphysics parameterization describes the hydrometeors (cloud water, raindrops, ice, snow, graupel, etc.), the number concentration, size, fall speed etc. These descriptions differ between the chosen parameterizations. These differences play a role in cloud development and hence, also in precipitation amounts and patterns. As the microphysics parameterization is responsible for all the precipitation in the innermost domain (no cumulus parameterization), changes in this parameterization can certainly affect precipitation amounts and patterns. Compared to

Thompson, the ice particles (ice, graupel and hail) are described in less detail in the Ferrier parameterization. The reason, why the difference is rather small might be that we only look at summer events, where ice particles are less important. We have included a sentence about this argumentation in the manuscript.
The two cumulus parameterizations that have been used show differences in scale-dependence. The Kain-Fritsch scheme is one of the most commonly used parameterizations over Europe. The scale-dependent scheme is designed to improve the realizations in the so-called grey zone between 10 and 5 km.
Since we do not employ a cumulus parameterization in the inner nest, the changes in the outer domains seem to be relatively small, when changing the cumulus scheme. This might be because the description of the convection in this case is not triggered by fine scale structures, but mainly by lifting along the orography. We have included a short description of the differences in the two cumulus parameterizations.
Two-way nesting can be helpful, if an event can be modified by the small-scale features that are resolved in the innermost domain. In such a case, the result of the innermost domain can be transported to the coarser domains as well. In a Vb event one or two-way nesting might not make a big difference, as the cyclone is steered by large-scale atmospheric flow.

13. Page 11, line 28: Here the authors present figures 4d-f and 4g-I but do not analyse these figures. If they are not referred to in the text should they be removed?
The references can be found on p11 bottom and p12 top, then again on p12l18 for Figure 4a, d and g, on p12l26 for Figure 4b, e and h, and on p14L8 for Figure 4c, f and i. We see that more references could be helpful, though, and introduced some more in the text where appropriate.

14. Figure 3: The right-hand edge of the figure has been cut off. It is also not clear what cross-section figures 3i-k are for. Could the cross-section be added to figures 3f-h respectively?
We do not see the cut in our submission. However, we have redrawn Figure 3, adding all cross-sections to simplify referencing, and have rephrased the caption for more clarity.

15. Page 14, line 18 and elsewhere: The authors conclude that the ensemble spread becomes increasingly larger when going back in time. Although this is an intuitive result, it is not possible to conclude this from 3 points only. More case studies would be needed to confirm this.
We did not intend to make a general claim here. The whole paragraph is meant as a summary / discussion of the synoptic analyses before we go on. This is now clearly stated at the top of the paragraph.

16. Page 14, lines 12-22: The authors do not refer to any figures in this analysis section. Which figures are used? Is this where the analysis of 4d-f and 4g-l is performed?
See also the reply to comment 14. We include now more specific references to the respective Figures, and add some more text to guide the reader.

17. Figure 4: Why is a different domain used in figures d-f? Is the ensemble track position agreement in the North-Atlantic relevant? It appears as though the track agreement over Switzerland is similar for all 3 cases, is this correct?
The larger area is chosen to show the relatively good performance of 20CR in the region of interest in comparison to other regions in this area. We have extended the phrasing to 'Overall, the analyses at synoptic scales (Figures 3 and 4) show that differences among the 20CR members are substantially smaller over the region of interest (Southern and Central Europe) than over other regions of the North Atlantic / European sector (Figure 4 d, e and f); this corresponds to the relatively high density of assimilated stations over Central Europe (not shown; see Compo et al. 2015).**'**

18. Figure 4: I do not know what figures 4g-i are showing. Please explain these figures in the text.
The Figure panels have been redrawn, according to a comment by Reviewer 1, and the caption has been extended. As requested, we now analyze the three panels in more detail. We thank the reviewer for this comment; we think the revisions make our points much clearer now.

19. Figure 4: These are quite complex figures. Are the country outlines important? Perhaps they could be removed? Or only Switzerland included?
We would like to keep the country borders as is because they can be useful for localizing the features. We used some lighter grey shades where appropriate, though.

20. Page 16, line 18: The authors say that the model 'agrees' with the CombiPrecip precipitation. How did they come to this conclusion? The time evolution of the CombiPrecip appears to lie outside the ensemble spread for a large part of the timeseries implying poor agreement.
Please see the reply to comment 6.

21. Page 18, line 13: Why is the fact that the storm track for max precipitation in 20CR and downscaled simulations is similar 'remarkable'? Did the authors expect to see large differences in the position of the storm track? Doesn't the similarity indicate that the track of the cyclone is the primary control on precipitation accumulations over Switzerland?
We have removed the word in the revised manuscript.

22. Figures 7, 8 and 9: These figures are of very poor quality. They do not contain lat/lon, a colour bar or continent outlines. This makes the analysis impossible to follow.
Similar comments are made by all reviewers. We think that the loss of quality was introduced when converting from vector format to PNG, and during the upload process. In the revised version, we have a better resolution of the PNG file, and readability of the PDF for print is good in our view. Note also that the color key had been missing.

23. Page 19, lines 5-15: Analysis of figures 7-9 is described in just 13 short lines. Is it therefore justified to include all 18 figure sub-panels?

This is an eye-opening comment. We have realized that the analyses needs more details, and we have extended the paragraph accordingly. We think that it links also much better with the additional analyses suggested by Reviewer 1.

24. Figure 10: As far as I can tell both the colours and size of dots represent the precipitation intensity. Are both methods needed?
    Yes, we prefer to keep both elements. The co-authors found in the internal review that comprehension of the Figure content is easier.

25. Page 20, lines 8-13: These lines describe figure 10. This is a complex diagram and the analysis of it is rather brief (6 lines). Given that one of the major objectives of the paper is to examine 'sensitivity to cyclonic moisture flux' (title) the analysis should be expanded.
    We agree that the text is rather short at this point. In fact, we come back to Figure 10 in a later discussion paragraph. Following the suggestions also earlier in the review, we introduced a specific reference. Note that according to the comment by Reviewer 1, the analysis of 'cyclonic moisture flux' in the two innermost model domains is extended, including a new Figure 11.

26. Page 21, line 12: The authors describe the act that one of the ensemble members produces higher precipitation for the 1910 event than those observed as 'remarkable'. I'm not sure why this is remarkable. The purpose of the ensemble is to represent the range of plausible situations given the large-scale flow conditions so if all of the ensemble members underpredicted the observed precipitation totals then this would be a poor ensemble. Perhaps I have misunderstood something here?
    We can follow this argumentation. Hence, the word 'remarkable' may not be appropriate here, and we omit it in the revised version of the manuscript.

27. Page 23, lines 1-8: While the discussion of PV streamers is interesting, it is not a result of this paper, so it should not be in the results section.

Throughout the article, discussion and interpretation follow the presentation of results. However, we have rephrased the discussion without PV streamers. In accordance to a similar comment by Reviewer 3, we have rephrased the discussion without PV streamers. The aim of the paragraph is to mention the underlying physical processes and interaction between the Alpine orography and upper-level troughs. These physical processes can be referred to without introducing the notion of PV, which is indeed not introduced in our study. In fact, this has allowed to focus more on explaining the untypical results from the 1876 case.

The new text reads as follows: "This behavior of the model can be explained by the interaction of the Alpine orography with atmospheric circulation. Indeed, Vb cyclone trajectories are typically initiated by deepening upper-level troughs, which finally cut off from the westerly flow when passing over the Alps (e.g. Awan and Formayer, 2017). The interaction of upper-level troughs with the Alpine orography have been described in detail (Buzzi and Tibaldi 1978; Aebischer and Schär 1998; Kljun et al. 2001); the underlying processes include flow splitting and lee cyclogenesis, with further amplifications of the cyclone formation by frontal retardation and latent heat release due to orographic lifting. The combination of these processes implies that the cyclones are formed on the lee of the right side of the Alps, typically over the Ligurian Sea. In 20CR, however, the Alpine orography is very coarse, smoothed and reaches only about 1000 m a.s.l. (cf. Stucki et al., 2012). Hence, the influence of the Alps on the large-scale flow is limited in 20CR. Given also that the 1876 case is least confined by pressure observations, this allows untypical cyclone tracks in many 20CR members. Once accounting for a more and more realistic orography throughout the downscaling steps with WRF, the high-resolution runs may thus end up in a compromise simulation - driven both by the WRF model physics and by the 20CR input flow. In other terms, the large-scale flow forced from 20CR might not be compatible with the orography of the high-resolution domains**.”**

28. Page 23, line 19: Whether short spin-up periods are ‘preferable over long spin-up periods’ depends on what you are trying to optimise and is not a general result. I think the objective in this study was to minimise spread in the ensemble tracks so as test sensitivity of precipitation to moisture flux rather than track position. Another objective may well have resulted in a different optimal spin-up period.
We agree. Accordingly, we introduced ‘for our purposes’.

29. Page 23, line 20: What are slow-reacting features?
We changed the line to ‘slow-reacting variables like soil moisture’. Note that we come back here to something we introduced in Section 2.3.

30. Page 23, line 20: Again ‘good results’ depends on what you are trying to achieve. Small differences in the ensemble will occur if the cyclones are already present in the outermost model domain. Is that the point?
We agree that we should be more specific here and expand on the suggestion of the reviewer in the revised manuscript.

31. Page 24, line 7: I do not think you can conclude that uncertainty increases gradually when going back in time using 3 case studies only.
In fact, we do not think that we generalize here, given the context of the paragraph, and the first word of the sentence being ‘This’, that is the uncertainty in our cases. Note however the added sentence in the conclusions: ‘Although going back far in time, we have only analyzed a very small number of events – many more cases would be needed to reach robust recommendations on how to configure a model for Vb cases.’

32. Page 24, line 9: Similarly, concluding that dynamical downscaling is less accurate going back in time is difficult using 3 case studies only. There are many other factors that would increase the uncertainty for specific case studies.
See the reply to comment 31.

33. Page 24, lines 14-23: This is an excellent summary and it would be nice to see a more in-depth analysis in the main body of the text to support these conclusions.
We tried to corroborate these conclusions with the extended analyses of the ‘cyclonic moisture flux’, including the new Figure 11 and an additional clause in the Abstract.

34. Page 24, line 26: How do you conclude that the 20CR tracks are not ‘realistically located’ for the 1876 case? Are you stating that 20CR produces unrealistic tracks, or simply that the uncertainty in the position of the track is large for this case potentially because it is a complex situation?
    Realistic is indeed the wrong term here; we thank the reviewer for pointing this out. We rephrased it to ‘variability becomes very large’, and we add that some cyclone tracks do not follow the classical Vb path anymore. Note also the extended explanations of uncertainty.

Further changes made

- **Additional references: Hofstätter and Blöschl, 2019; Zbinden, 2005 (Annals of MeteoSwiss); Cioni and Hohenegger, 2019; Coppola et al., 2018; Compo et al, 2015; Joliffe and Stephenson, 2012; Wernli et al., 2008.**
- **Figure A2 redrawn with the maximum- and minimum-precipitation members. This makes more sense since we focus on these in the text.**