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.

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?
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.
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).

Specific comments

1. Page 1, line 22: What is 'moderate' spectral nudging?
2. Page 2, line 20: How is the moist air 'let' around the Eastern Alps? Do you mean advected?
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?
4. Page 5, line 5. It would be useful to know if any of the assimilated surface pressure observations were located in Switzerland.
5. Figure 1: The numbers on the colour bar have been cut off.
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.
7. Figure 2: The right-hand axis does not have any units. Also, it is not clear what the red numbers represent.
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.
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?
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?
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.
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?
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?
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?
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.

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?
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?
18. Figure 4: I do not know what figures 4g-i are showing. Please explain these figures in the text.
19. Figure 4: These are quite complex figures. Are the country outlines important? Perhaps they could be removed? Or only Switzerland included?
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.
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?
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.
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?
24. Figure 10: As far as I can tell both the colours and size of dots represent the precipitation intensity. Are both methods needed?
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.
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?
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.
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.
29. Page 23, line 20: What are slow-reacting features?
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?

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.
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.
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.
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?