

The comments below come from someone from the atmospheric modelling community, and I know some of the co-authors are also atmospheric model users. Hence, some comments may come from a somewhat different angle.

Overall, I find the paper quite easy to read, but there are some issues in the earlier part of the paper that the authors seem to have forgotten readership may include non-hydrologists. It will be nice to put some of the results in the broader context – like the need of high resolution models to deal with complex topography and what the snow-prescribed simulations may mean in climate change.

Major comment:

1. Be aware who might be reading the paper, and make sure they understand your terminology. Just in the abstract and the Section 1, there are few cases that the authors use acronyms and terminologies that may be readily understandable by hydrologists but not to other readers; proper definitions somehow get mentioned later in the paper, but not on the first incidence that they are used. WRF is a tool commonly used by the NWP and climate community, so it is important to make sure all members of the WRF community to understand what you are talking about. See minor comments for details.
2. As the authors indicate that one needs the high-resolution WRF model to get the precipitation right around the complex topography in their region of interest. One way to show that is to demonstrate how different the precipitation looks like with the lower-resolution 9km model. I hope it is not too much to ask, but how would Figure 5 looks like for the 9km WRF simulation?
3. Figure 9: Is this due to the short spin up runs have soil dry enough to absorb additional precipitation into the ground during the initial phases of the event? It will be good thing to check within the context of what you see for Figure 7 and will show what the WRF land surface scheme is doing to your results.
4. If I understood Section 4.5 correctly is that the authors' sensitivity simulations with snow generally falls into two scenarios ("no or nearly-no snow" (control or 0.1 m snow) or "a lot of snow" (0.5+m). In between 0.1 and 0.5m, there will be cases that we may see snow making an impact to the discharges in between those two above binary regimes. This is important in climate change impact – that the snow feedback to river flow will depend on which snow regime we will be in (no snow, a lot of snow, or somewhere-in-between regime). Of course, this is just one event and the authors have given context how rare the event is (50-year return), and it occurred in October (which would have no or little snow). Hence, it will be nice to comment on the following issues:
 - a. Some general comments about the seasonal cycle of extremes in Norway: How often do you see comparable precipitation extreme events during late winter, spring, early summer in which snow would become a factor? Are there any reasonable reason to believe the probability of such events during the snow-relevant months to change in the future? I hope this is just some additional literature review.
 - b. What are the typical snow depths that you get by the end of winter? How is that expected to change? I would speculate snow depth would likely to be in "a lot of snow" regime, but it would be good to see some actual numbers and how they compare with the prescribed snow depths. It will also justify your choices of snow depths for the sensitivity simulations.

Minor Comments:

Consistency with the use of Øvstedal/Svartavatn, Vossevangen/Bulken and various basin/station name pairs: There are quite a few places in the figures and the text that the names for the name pairs are used interchangeably. Would it be possible to keep both names together at all times (i.e. always say “Øvstedal-Svartavatn” together, but not one without the other)?

Abstract line 18-20: Define HBV and NSE.

Line 36: How much that is in € or USD terms? (~ €13 million, 14 million USD?) Also, it may be helpful to quote human casualties as well (if there are any).

Line 65: The meaning here is somewhat unclear here; are you trying to ask if it is the snow or the rain that causes flooding during a rain-on-snow event?

Line 110-111: Given the broad readership of the journal, the authors may have to briefly discuss what the Nash-Sutcliffe Efficiency is. Are NSE = 0.8 and 0.27 good or bad (this is only briefly mentioned much later in the manuscript in section 3.3)?

Line 134: I presume “offline” refers specifically to what happens to the hydrological physics (the hydrological model that is coupled to WRF just responds passively to WRF precipitation and radiative flux forcings). Otherwise, “offline” makes little sense to a reader from the NWP or regional climate modelling community. One way to avoid this problem is to properly define what “offline” means and it only applies to the hydrological model and has nothing to do with WRF.

Figure 1: If possible, add a horizontal legend how big your domain is (like the ones you commonly see in maps and atlases).

Line 168: “snow observations from NVE” → “NVE snow observations”

Lines 282-283: Perhaps I am not familiar with the calibration of hydrological models, but this comment about computational costs strikes me as a bit unusual as km-scale WRF should be quite expensive; I would imagine that a month or two to run all those WRF setups that are described in Table 2. It will be educational to put in context how much wall clock time is needed to calibrate the hydrological model for one basin and to do the WRF simulations.

Line 393: “doesn’t” → “does not”

Figures 12, 13: the figures will be easier for the readers to follow if the average heights of the basin are given in the subtitle of each panel.