Answers to the comments of Reviewer 2

In the comparison several forecast chains are compared. In the experiment setup effects of initial conditions, boundary conditions (forcing) and model bias due to calibration dataset are mixed. From the experimental setup and description it is not clear which effects are taken into account or are studied in this manuscript. It is also not made clear or mentioned what the limitations are of the experimental setup. This leads to 'surprising' results and information that was not given or at least not highlighted as limitation in the material&methods section (see page 2311 where the effect of the model calibration is mentioned).

We thank the reviewer for pointing out this lack of information. We will include the information about the data used for model calibration in section two. We will also better focus the paragraphs pointing at the aims of the study and we will declare its relevance as incremental study with respect to previous work in our institution.
In Zappa et al. (2011) the superposition of different sources of uncertainties in the hydro-meteorological forecast chain is investigated and in Liechti et al. (2013) the radar ensemble product REAL and a parameter ensemble approach were tested for hydrological nowcasting. The present study is a follow up and goes beyond nowcasting. The coupling with COSMO-2 and the novel probabilistic radar-based forecast NORA allow us to investigate flash-flood forecasts with some hours lead time.

The authors need to rethink/write what can be learned/deduced from comparing chain x with chain y and clear write this down. I see two different experiments:

- The experiment with the three chains with same initial condition (RADAR) and different forecast products can say something about the performance of the different forecasts products (NORA, C2, PERS) relative to each other;

-The experiment with the three chains with different initial conditions (REAL, RADAR, PLUVIO) but same det. forecast product (C2) can be compared to say something about the importance/effect of different initial conditions (and model calibration data) using the same forecast product. Maybe using an approach as recently was put forward by Rakovec et al (2012) where based on pluviometer data a historic precipitation ensemble was created could be a recommendation or future line of research.

The authors need to separate the results and discuss the results in this manner to reach meaningful conclusions that can be generalized in same way. They should also attempt to relate to other similar work that is focused on the different effects of initial conditions and forecast products on forecast quality.
This will also require a reshape/write of the introduction and formulation of the aims of the manuscript. The effect of initial conditions is not mentioned in the introduction nor are relevant references provided in the introduction.

We see the need of separating the results in a different way and will present the different experiments in a clearer way. We also appreciate the suggestion of the reviewer. However, our study aims to investigate the potential of the probabilistic radar-based forecast products for flash-flood

early warning. Therefore we conduct the following experiment and will rearrange the results so that these come clear.

In comparing REAL-C2 and NORA to RAD-C2 we can see the effect of probabilistic vs. deterministic forecasts.
In more detail:

- REAL-C2 vs. RAD-C2: effect of ensemble of initial conditions.
- NORA vs. RAD-C2: effect of probabilistic forcing.
- NORA vs. REAL-C2: combination of the above.

The comparison with the persistence (PERS) additionally shows the effect of having a forecast in the first place.

A direct comparison of the radar-based forecast with the pluviometer-base forecast is not fair, as the model was calibrated with precipitation data from rain-gauges. Therefore the PLU-C2 forecasting chain is more seen as a reference, which from our point of view, has to be included into the study for completeness and transparency reasons.

The main focus of this study was not the investigation of the effect of initial conditions, but, as the results show, ensembles of initial conditions have the potential to improve forecasts significantly. Thus we see, that we need to address this issue already in the introduction.
It would also be interesting to test ensembles of initial conditions derived from probabilistic pluviometer-base nowcasts. Therefore we will address approaches as presented in Rakovec et al. (2012) or Moulin (2009) in the outlook of our revised manuscript.

Specific:
-The aim of our study is to explore the space between radar-based nowcasting and radar-based forecasting and, in particular, to investigate the potential of purely radar-based flash flood forecasting.
-The main objective of our study was to investigate a possible added value of NORA for flash-flood early warning.
What is the real objective? And how does the experimental setup matches/serves this aim?

A: We thank the reviewer for pointing out this inaccurateness. We will reformulate the aim of the paper to “What is the potential of radar-based hydro-meteorological forecasts for flash-flood early warning.” and may also change the title to “The potential of probabilistic radar-based forecasts for flash-flood early warning in the Southern Swiss Alps”.

-Is the using the persistence of the radar field at T0 a realistic forecast scenarios otherwise leave this out of the comparison/manuscript;

A: Yes, it is. It follows also the procedure used by Panziera et al. (2011) that determined that NORA is better than a persistence forecast. We needed to evaluate if such finding is still valid after propagating NORA through the hydrological model. And if the “persistence” (PERS) forecast is better than COSMO2. If PERS would yield similar quality as NORA, and better outcomes as COSMO2, then

such “expensive” advanced products would not have any future as components of flash-flood early warning systems in alpine areas.

-combine chapter 2/3/4/5 into one chapter 2 Materials & Methods

A: We address this in our revised manuscript.

- Is hourly time step enough for hydrological modeling/forecasting? What is the rational for choosing this model time step?

A: The model was chosen in the first place because it has been used for many studies already in the study region. Furthermore, it is used since 2007 for operational forecasts of the study catchments. The model is not set up to work for smaller time-steps than 1 hour so far. As the response time of the small catchments is about 2 hours a time step of 1 hour is still justifiable (also with respect to the lead time of the forecast system presented here). Moreover COSMO-2 forecasts are only available every 3rd hour. Also the radar-ensemble REAL, which is used during the nowcast period, is available hourly.
So the only data source, in terms of meteorological forecast, included in this study, which would be available for shorter time-steps (5 min) is NORA.

References

Liechti, K., Zappa, M., Fundel, F. and Germann, U., 2013. Probabilistic evaluation of ensemble discharge nowcasts in two nested Alpine basins prone to flash floods. Hydrological Processes, 27(1): 5-17.

Moulin, L., Gaume, E. and Obled, C., 2009. Uncertainties on mean areal precipitation: assessment and impact on streamflow simulations. Hydrology and Earth System Sciences, 13(2): 99-114.

Panziera, L., Germann, U., Gabella, M. and Mandapaka, P.V., 2011. NORA–Nowcasting of Orographic Rainfall by means of Analogues. Quarterly Journal of the Royal Meteorological Society, 137(661): 2106-2123.

Rakovec, O., Hazenberg, P., Torfs, P.J.J.F., Weerts, A.H. and Uijlenhoet, R., 2012. Generating spatial precipitation ensembles: impact of temporal correlation structure. Hydrology and Earth System Sciences, 16(9): 3419-3434.

Zappa, M., Jaun, S., Germann, U., Walser, A. and Fundel, F., 2011. Superposition of three sources of uncertainties in operational flood forecasting chains. Atmospheric Research, 100(2-3): 246-262.