

## Answers to the comments of Reviewer 1

1.a) Dependence of results on hydrologic model parameters – The authors devote no more than 2 sentences regarding the details of hydrologic model calibration process toward the end of section 2. Later in the paper, we're told that the interpolated rain gauge fields were the forcing dataset to calibrate the model. Several details are lacking here. This is critical information because the model parameters can dictate the apparent skill in the forcing datasets, which are being compared.

A: We thank the reviewer for pointing out this lack of information and include the very important information on the use of interpolated rain-gauge data for model calibration in section 2 where the hydrological model and its calibration are described.

Concerning the calibration procedure we adopted the same procedure we developed in previous applications of PREVAH, which is detailed in Zappa and Kan (2007) and Viviroli et al. (2009b). In few words, citing Viviroli et al. (2009b): "With the aim of calibrating a large number of catchments for a semi-distributed, process-based conceptual hydrological model, we introduce a straightforward yet robust automatic calibration procedure. Since identification of a global parameter optimum is not feasible in practical terms, the procedure presents a trade-off between computational time and algorithm complexity to identify, with reasonable effort, a parameter set that is well representative of the catchment's dynamics. In its standard mode, the calibration combines three efficiency scores which are evaluated both for the entire calibration period and in terms of their annual and monthly variations. These scores are furthermore assessed only in their relevant value range, producing a comprehensive overall acceptability score which is used to guide an iterative parameter search algorithm". The basic scores are the Volume error and the Nash-Values (both linear and logarithmic expression).

Finally, all predictions in the forecast chain start with initial conditions (day -5, Figure 4) stemming from the simulation with interpolated pluviometers. In this respect we apply PREVAH with different forcings in a "event-mode". The difference in the outcomes do not rely on long-term model history, they stem only from different precipitation input short before the considered events that include orographic forcing activity.

1.b) How many parameters does the model have?

A: The number of tunable model parameters varies between the study catchment, for Verzasca it is 14, for Calancasca and Ticino 16 (presence of glacier).

We cite Viviroli et al. (2009a): "While some model parameterisations are assigned a priori through digital representations of the physiogeographical basin characteristics (...) and relevant values from the literature (...), a number of tunable parameters need to be adjusted in PREVAH to the specific modeling site. Depending on module specifications (e.g. evapotranspiration modelling scheme or presence of glaciers within the investigated area), this number typically ranges between 14 and 19." We cite Zappa and Kan (2007): "The most sensitive tunable parameters are the adjustment factors for scaling snowfall and rainfall, the parameters of the snowmelt module, the non-linearity factor controlling the soil moisture recharge and the parameters of the runoff generation module."

1.c) Are they spatially distributed?

A: The model is semi-distributed and operates with hydrological response units. This is stated in the first two sentences of section 2. The parameters are concentrated and valid for all spatial units of the basin (Viviroli et al., 2009b).

1.d) : Are the estimated separately for each basin?

A: We thank the reviewer for pointing out this information gap. The calibration was first done by Ranzi et al. (2007). They calibrated PREVAH for the Verzasca and mainly adopted the parameters for the Ticino and Calancasca, while the bias correction parameter (see 1.f)) has been adjusted specifically for each catchment . This information will be included in the present section 2.

1.e) Is the parameter estimation method automatic or manual?

A: The parameter estimation method is automatic and guided by different objective functions. See answer to 1.a) and Viviroli et al. (2009b) or Zappa and Kan (2007).

1.f) If the model parameters are estimated with rain gauge forcing, then the hydrologic model forecasts will be most skillful when the same inputs are used, or at least inputs that resemble the rain gauge dataset used during calibration. If something different is used as forcing, then the forecasts will inevitably become biased even if the forcing is closer to the truth.

In other words, the model parameterization process can obscure or rather account for errors (especially bias) in the rainfall forcing.

A: Yes, we agree with the reviewer that the model chain forced with interpolated rain-gauge data is by construction expected to perform best. There is a bias correction for the interpolated precipitation input. We call it water balance correction. We cite Viviroli et al. (2009a): “[Two parameters] are used to adjust the precipitation input in order to reduce the total discharge volume error of the model as observed at a catchment outlet. With this, a series of systematic errors in the modelling chain are addressed: a) the wind-dependent gauge error correction (...), b) spatial interpolation errors, c) errors arising by the insufficient representativity of the available gauge networks and d) errors in the estimation of evapotranspiration”.

Since we don't have enough long homogeneous time series to perform a calibration based on weather radar data, we use the “raw” input without water balance adjustment for the all radar-based products. Since we start here with identical initial conditions at “Day -5” the bias in the radar input refers only to each specific event addressed.

The reviewers note made us make some more thoughts and we see now that we have to clearly separate the results for the comparison of the different radar-based forecasts and the pluviometer-based forecast. However, as the radar-based forecast products all base on the radar QPE of time  $t_0$  the comparison of the different radar-based forecasts is still valid, as a potential bias will be the same or very similar for all of these products, as they base on the same original radar QPE. The pluviometer-based forecast can be seen more as a reference in this study.

Additionally we provide the ROC-area, which is an indicator of the potential value of the forecast for decision making if the model was calibrated correctly.

1.g) Another issue here is the statistical objective used during model calibration. The authors state on pg. 1295 that the best parameter set is chosen which simulates the average flows best and has the

smallest volume error. The authors need to report on the specific modeling objective and the statistical scores during calibration and validation. Also, if the model parameters are tuned to simulate average flows, then they're not necessarily guaranteed to produce forecasts that are skillful for flooding flows (i.e., q95).

A: The model system PREVAH which is used in this study is an operationally working hydrological forecast system, it is calibrated to get the smallest volume error for average flows. The calibration procedure and the statistical scores (NSE, logNSE and volume error) used for calibration are presented in detail in Viviorli et al. (2009b) and Zappa and Kan (2007), see also answer to 1.a). For catchment specific calibration results we refer here to Ranzi et al. (2007). For the Verzasca catchment the verification showed an NSE of 0.842 and a volume error of -2.5%. The Calancasca shows an NSE of 0.773 and a volume error of 35%. Ranzi et al. comment this rather large volume error with the findings of Margot et al (1992) who estimated that, on average, only 60% to 80% of the natural discharge from the Calancasca is effectively reaching the gauging station. The remaining water is diverted to neighbouring basins and used for hydropower generation. Thus Ranzi et al. (2007) state that accounting for this, the large volume error can be looked at from another perspective and considered as a fair estimate of the natural discharge from the Calancasca basin. In that publication no scores for the Ticino river basins have been presented. At that time we had no module for accounting for the hydropower management in the basins, which has been realized in summer 2007, before using PREVAH in this area during the MAP D-PHASE Project. In the revised manuscript we will summarize this.

It is of course true that the performance for the forecasting chains presented in this study is expected to be better if the model was calibrated for high flows like q95 of the climatology. However, in this case there is very little data available for calibration and validation. It has also to be seen that the q95 used in this study is the q95 of the sample climatology, that is the hours which are covered by at least one NORA forecast. In practice this results in a q95 which is still below the 2-yearly event.

The reviewer seems very much focused on the assumption that we want to explore if the hydrological model has quality in reproducing the flash-flood, while we want to explore whether the radar-based forecasts have added-value for flash-flood early warning within an operational chain established in 2007. We will make our focus clear in the revised manuscript.

2.a) The authors go to great lengths to develop ensembles based on the radar data combined with an NWP model. However, the resulting ensembles of hydrographs assume there is a perfect hydrologic model with perfect initial soil conditions and parameters. Hydrologic model uncertainty is significant, especially for extreme events in complex terrain, yet it isn't considered nor even mentioned in this study. Justification for this simplification is needed.

A: The study we present here is an incremental study to Zappa et al. (2011) and Liechti et al. (2013). In Zappa et al. (2011) uncertainty cascading has been investigated, combining uncertainty of the hydrological model parameters and uncertainty stemming from weather radar information and numerical weather forecasts. We found out that for "the operational simulation of peak runoff events the hydrological model uncertainty is less pronounced than the uncertainty obtained by propagating radar precipitation fields (...) and NWP forecasts through the hydrological model (...)." In Liechti et al. (2013) we investigated whether uncertainty from weather radar and from model parameters have influence on the quality of simulating peak runoff in a basin nested within the basin for which we realized the calibration. We found out that "the evaluation of the 4-year nowcast shows that pluviometer-based nowcasts outperform radar-based nowcasts in the gauged and calibrated

catchment and that there is added value in the application of parameter ensembles. For the small, ungauged catchment, the results achieved by the radar-based nowcasts are superior to the pluviometer-based nowcasts. Especially the radar ensemble proves to be of significant advantage for flash flood nowcasts in such catchments”.

In the present study we aim at finding out if new “off-the-shelf” weather radar forecast might yield added value as compared to other meteorological forcing in the context of flash-flood early warning. This is why we decided not to focus on the value of the hydrological model and omit to test the propagation of parameter uncertainty here again.

We will make this clear in the revised version and also refer to Zappa et al. (2011) for information about model uncertainty. We thank the reviewer to point out that this is missing.

3.a) Interpretation of results – First, interpreting the results have issues in regards to my first major comment. But, when I view the overall results in Fig. 6 it appears as though REAL-C2 is the best followed by PLU-C2. PLU-C2 has an advantage of being the forcing data set during calibration, so it should be relatively immune from any deleterious effects due to bias. However, it has the disadvantage of being fairly simple. The authors state on pg. 1298 that an inverse distance weighting scheme is used for the spatial interpolation. Why not consider orographic enhancement in the spatial interpolation scheme like you did with the radar data? Why not use kriging?

The interpolation method used in our study is robust for an operational application, where the availability of stations can vary. A new product which combines radar and rain-gauge information is available for testing since shortly (Erdirin et al., 2012; Sideris et al., accepted). There they work with spatiotemporal co-kriging with external drift in the alpine terrain of Switzerland. However in situations where one of the stations fails the errors can be very large. So for an operational context the method used in our study seems to be more robust.

3.b) Also, since this product is used as forcing during calibration, the reader really needs to see some of the gridded gauge fields. Perhaps the density is sufficient enough to resolve orographic effects already?

A: The aim of the paper is to investigate the potential of radar-based flash-flood early warnings in a mountainous region. It is not our intention to make a choice between radar-based and rain-gauge-based forecast chains. So even if the precipitation maps resolve the orographic forcing this would not change the results nor the interpretation. Therefore we do not see any added value or insight for the reader in providing precipitation maps from interpolated rain-gauge data. Also the calibration was done for the entire years of 1993 to 1996 and not specifically for events with orographic forcing and therefore we do not see the link suggested in the comment above.

3.c) Regardless, the skill of this product being second only to REAL-C2, which gets the advantage of providing an ensemble, isn't mentioned in the conclusions. I hypothesize that PLU-C2 would be the overall winner in the present experimental design if the authors created an ensemble of PLU-based inputs, which is readily possible if they used kriging.

Such an experiment has already been conducted in a preceding study (Liechti et al. 2013) where a parameter ensemble was used to initialize the nowcasts. The performance of the ensembles was mostly better than the one of deterministic nowcasts. However, the results of the comparison of radar-based and pluviometer-based nowcasts (with 0 lead time) depended on the catchment (size, gauge equipment) and for specific events on the characteristics of the storm.  
See also answer to 2.a)

3.d) Would it be possible that the best flash flood forecast chain is accomplished using the ensemble gauge estimates for nowcasting combined with the NWP for forecasting? This result would be rather problematic for the radar community I would think.

We don't think that such a result would be problematic. As we see it and what also our results show, is that one should always make use of rain-gauge data if they are available and of good quality. However, it is important to investigate the potential use of radar-based forecasts for regions where there is no or only a poor rain-gauge network available.

In our study we present one possible way to prolong the lead time of discharge nowcasts for such regions. That is, to make use of existing radar archives, and search for analogues of weather situations, which are likely to produce flash flooding. The resulting ensemble of analogues performed better than the deterministic radar-based forecast (RAD-C2). So this procedure presents a possible approach to prolong the lead time of discharge forecasts also for regions where there are no or only poor rain-gauge measurements available but where repetitive weather situations often lead to flash flooding or flooding. In this respect the model chain using interpolated rain-gauge data should more be seen as a reference.

4.a) The title in its present form is far too broad and does not adequately describe the study. It is quite specific to the radar ensemble method, orographic considerations, COSMO-2 forecasts, and the hydrologic model specifically calibrated to data on a few Alpine basins in Switzerland. The regionality and Alpine specificity needs to be included in the title.

A: We will make the title more specific and include the regionality into the title. The tentative new title is: "The potential of radar-based forecasts for flash-flood early warning in the Southern Swiss Alps"

4.b) Moreover, I would argue that this paper is a hydrologic evaluation of different rainfall inputs that are radar-estimated, forecast, deterministic, and probabilistic. It is doubtful the hydrologic model and its parameterization will apply elsewhere. Even the rainfall forcing products will have limitations in other complex terrain flooding situations where there is severe beam blockage, overshooting, and convective initiation in the mountains, rather than more straightforward quasi-linear propagation and orographic enhancement. This latter situation is more typical in mid-latitude synoptic systems that encounter very large mountain chains, such as the Alps in Europe or Sierras in California.

A: Yes, this is what the paper is about. The hydrological model is here an instruments for propagating the forcing and find out if efforts in advanced radar-based forecast applications are worthwhile in regions affected by orographic forcing. This specific parameterization of PREVAH cannot be transferred straight-forward to other areas. Although the main focus of PREVAH applications is Switzerland, the modelling system or parts of it have also been used successfully in mountainous regions of Austria, China, Germany, Italy, Kenya, Kyrgyzstan, Russia, Sweden, the United States, Peru and Uzbekistan (Viviroli et al., 2009a, updated).

Minor comments:

1. Pg. 1290, Line 18 - Please just report the skill rather than a qualitative description.

A: We will do this in the revised manuscript

2. Pg. 1294, Line 19 – An hourly time step is incommensurate with the model grid cell resolution (and flash flooding). Please comment on why they didn't run the hydrologic model at a shorter 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 for operational forecasts of the study catchments. The model is not set up to work for smaller timesteps 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. 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 timesteps (5 min) is NORA.

3. Figure order (throughout) – The figures should be called in sequential order, so that Fig. 1 is cited first, then Fig. 2, etc.

A: We will pay attention to that in the revised version of our manuscript.

4. Pg. 1296, line 10 – The 3rd criterion relates to my major comment #4. That is, these types of flash-flood producing systems are rather specific to this region.

A: See answer to major comment 4.a).

5. Section 3.2.1 (and throughout) - This is not coupling unless there are 2-way feedbacks between REAL and COSMO-2. Instead, the two forcings are merely provided to the PREVAH hydrological model. This needs to be corrected in several instances in the text.

A: To our understanding this is a coupling, in the sense that we directly connect the nowcast forced by REAL with the forecast forced by COSMO-2. For clarity we will exchange the verb "couple" with "connect".

6. : Pg. 1298, line 15 - This is an important detail that needs to be expanded. Were the bias factors computed for each input, or just the PLUVIO input? Why would a bias factor be needed during calibration? Typically, a systematic bias in the precipitation is readily dealt with by adjusting a model parameter (e.g., making the soil profiles either deeper or shallower).

A: See the answers to 1.f). Here we think there is a difference in general dealing of rainfall uncertainty. The reviewer argues here that if rainfall is uncertain and not representative because of station density and location or gridding by interpolation the best option is to make soils deeper or shallower. In PREVAH soil depth is a free tunable parameter; it is assigned a priori to each hydrological response unit. The data are obtained from soil maps (Viviroli et al., 2009a).

7. Pg. 1300, line 16 – Subject-verb disagreement.

A: ... one of the main transit routes that crosses the Alps.

8. Section 6.2 - Results are presented out of order in relation to Figs. 5-6. Also, the order of the basins in the figures does not match the sub-section ordering.

A: We will rearrange the results also according to comments from reviewer 2 and we will pay attention to bring the order of subsections and figures into agreement in the revised manuscript.

9. 1308, line 1 – How to explain this different bias behavior for the basins?

A: Ticino, Pg 1312. lines 11-14: “..... extreme events are overforecast for the Ticino river, possibly due to the influence of several storage lakes for hydropower production. The precipitation that actually falls in the catchment is not then recorded at the catchment outlet at the estimated time, but is stored in the lakes.”

Calancasca: It can be seen that the bias is mainly an issue for the radar-based forecasts in this catchment. So we assume that this is the effect of having a bias correction factor for the rain-gauge data (see 1.f)) but not for the radar data due to reasons pointed out in answer 1.f).

10. Pg. 1312, line 1 - Can this statement be true if there is no orographic enhancement?

A: We assume that the reviewer is referring to Pg. 1313, line 1. The answer then would be no, because NORA is only providing forecasts if orographic forcing is present.

## References

- Erdin, R., Frei, C. and Kuensch, H.R., 2012. Data Transformation and Uncertainty in Geostatistical Combination of Radar and Rain Gauges. *Journal of Hydrometeorology*, 13(4): 1332-1346.
- 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.
- Ranzi, R., Zappa, M. and Bacchi, B., 2007. Hydrological aspects of the Mesoscale Alpine Programme: Findings from field experiments and simulations. *Quarterly Journal of the Royal Meteorological Society*, 133(625): 867-880.
- Sideris, I.V., Gabella, M., Erdin, R. and Germann, U., accepted. Real-time radar-raingauge merging using spatiotemporal co-kriging with external drift in the alpine terrain of Switzerland. *Quarterly Journal of the Royal Meteorological Society*.
- Viviroli, D., Zappa, M., Gurtz, J. and Weingartner, R., 2009a. An introduction to the hydrological modelling system PREVAH and its pre- and post-processing-tools. *Environmental Modelling & Software*, 24(10): 1209-1222.
- Viviroli, D., Zappa, M., Schwanbeck, J., Gurtz, J. and Weingartner, R., 2009b. Continuous simulation for flood estimation in ungauged mesoscale catchments of Switzerland - Part I: Modelling framework and calibration results. *Journal of Hydrology*, 377(1-2): 191-207.
- 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.
- Zappa, M. and Kan, C., 2007. Extreme heat and runoff extremes in the Swiss Alps. *Natural Hazards and Earth System Sciences*, 7(3): 375-389.