Reviewer #2

#RC2.1 This very well written paper addresses the important question of how to model streamflow in small catchments with different precipitation products.

We thank the reviewer for this overall positive opinion on this study.

#RC2.2 Overall, the paper is a bit limited in terms of references to existing literature on the question of how input resolution interacts with model performance.

It is true that many papers have investigated the impact of input resolution for hydrological modeling, usually based on dense gauge networks (Dong et al., 2005; Meselhe et al., 2009; Bárdossy and Das, 2018; Zeng et al., 2018). For example, Xu et al. (2013) show that the increase in gauge network density can improve the performance of the hydrological model. For a large basin in China of 94,660 km^2, the sensitivity of the performances to the gauge density reaches a threshold when the number of gauges is high (greater than 100 gauges for this catchment). Similarly, Emmanuel et al. (2017) show that higher spatial resolution of rainfall leads to better hydrological model performance for 25 catchments with areas ranging from 42 km^2 to 1,855 km^2. Huang et al. (2019) also find that a higher temporal resolution of rainfall improves the model performance if the station density is high and that an increase in spatial resolution does not improve significantly the performance of the hydrological model for four German catchments with areas ranging from 417 km^2 to 1300 km^2. Using synthetic rainfall fields, Zhu et al. (2018) shows that the spatial resolution of precipitation has a larger impact than the temporal resolution for the simulation of floods but also depends strongly on initial soil conditions. Compared to the existing literature, a major difference in our study is that only small catchments (< 300 km^2) are selected and, in addition, which are located in mountainous areas where it is not known that precipitation measurements are more scarce and uncertain, particularly at high elevations. The interaction between input resolution and model performance is thus exacerbated in our study, compared to studies where catchments are often located in plains with larger areas (typically up to 1,000 km^2). A few papers studying the impact of rainfall resolution are focused on small urbanized catchments (Cristiano et al., 2016) or small lowland catchments (Terink et al., 2018; Hohmann et al., 2021). These references will be added to the introduction.

#RC2.3 Also, it does not discuss what different strategies actually exist to infer the hydro model meteorology at the appropriate resolution, from a meteorological product that has a different resolution and a different model topography.

Downscaling methods and conditional simulation can be applied to obtain meteorological forcings at the appropriate resolution for the hydrological model. These approaches can be applied to disaggregate precipitation data temporally (Parkes et al., 2012, Breinl and Di Baldassarre, 2019) and/or spatially (Bárdossy and Pegram, 2016) which is also a strategy applied in our study. Some references will be added to the manuscript in Section 2.2. An important challenge for the disaggregation is the need of a meteorological product at a fine resolution to establish the relationship between large-scale and small-scale meteorological statistics.

#RC2.4 It appears to me that an essential modelling choice is missing: how to combine the meteo product with the model? There is one option presented per hydro model but we do not know if this is a heuristic choice or the best option or what the literature says about this. In general, I do e.g. not think that retaining simply the meteo pixels within a catchment is the best option (but perhaps it is for rainfall?).

In our knowledge, taking the average of the meteorological inputs over the pixels covering the catchment in order to estimate areal meteorological forcings is the standard approach when lumped hydrological models are applied. There exist alternative approaches when meteorological data has a coarse resolution but there are not relevant here since the reanalysis considered in this study have a fine resolution.

#RC2.5 There is a short discussion on the absence of a precip gradient for one product but perhaps it would be good to have a more systematic discussion of how to create the hydro model meteo based on the input meteo.

Thank you for this comment. In this study, it is true that we mainly discuss the relevance of the meteorological products for hydrological modelling. How the different hydrological models process the meteorological products is a different point. In our case, MORDOR-SD assumes that precipitation and temperature have a linear relationship with the elevation, parameterized with orographic gradients parameterized by the parameters gpz and gtz, respectively. There are used to provide precipitation and temperature data for different elevation bands. SMASH is a distributed model and directly takes as inputs what is provided by the meteorological data. These aspects are really model specific, along with the choice of the parameters (fixed/inferred), the processes which are represented, the data used to estimate the free parameters. In this context, it seems difficult to provide a systematic discussion on these aspects.

#RC2.6 Also, I strongly recommend to make much clearer what take-aways are relevant beyond the studied catchments and how new they are (at the moment, there are two take-aways which do not reveal completely new insights).

It is true that some important and general messages could be provided in the conclusions. A first important message is that ERA5-Land or satellite-driven reanalysis are not likely to provide fine-scale precipitation dynamics which are necessary for the hydrological modelling of this type of small mountainous catchments. While the limitations of ERA-Land have been discussed by a few papers (l. 209-211 of the manuscript), the implications of these limitations for hydrological applications are absent from the literature in our knowledge.

The other key message concerns the added-value of radar measurements for the hydrological modelling of these small mountainous catchments. In our case, the underestimation of annual precipitation amounts at high elevations was a known deficiency of the COMEPHORE reanalysis. Our study reveals that despite this limitation, COMEPHORE provides better performances for some events because it provides additional information concerning the spatial extent of precipitation fields. However, this key message is really specific to this reanalysis and for the region of the study, since COMEPHORE strongly depends on the availability of radar measurements (l. 410-416). Beyond the studied catchments, a general message is that radar measurements are interesting to incorporate for the hydrological modelling of small mountainous catchments despite their limitations (radar signal attenuation by precipitation or beam blockage). While it has been shown for some flood events (Delrieu et al., 2005; Borga et al., 2007), our study provides a comprehensive evaluation of the added-value of radar data in this context, for many catchments and with two different hydrological models. The main conclusions will be rephrased in the revised version of the manuscript.

#RC2.7 What does all this for the modelling of ungauged catchments, for which no correction parameters can be calibrated? Would be cool to have some input on this question.

Thank you for this comment. This is a very important and difficult question. In this study, correction parameters are applied mainly to balance amounts of water at aggregated scales. Therefore, it can correct different aspects that affect the water balance: lack of precipitation or precipitation excess, groundwater exchanges, impact of the karst. As such, these parameters are very difficult to regionalize since they are related to different characteristics specific to each catchment and/or to the meteorological forcings for which we have incomplete and uncertain information. This issue remains the main unresolved challenge for the application of hydrological models (conceptual models or other types) to ungauged catchments, before the representation of the hydrological system.

#RC2.8 From the abstract, the actual innovation is not clear, it seems like another paper on a previously often studied topic; what is small?

We will precise that “small” refers to catchments with an area smaller than 300 km^2 in the abstract. We are not aware of similar studies comparing very different precipitation reanalysis for the hydrological modelling of many small mountainous catchments. Studies which apply hydrological models to small catchments are usually restricted to a few lowland or urban catchments, and usually consider meteorological forcings based on gauged data only (see response to comment #RC2.2).

#RC2.9 Very few references at the start of the intro, little reference to extensive literature on the role of spatial resolution of input (rainfall) forcing on the quality of hydrological simulations.

See our response to the comment #RC2.2.

#RC2.10 From the methods, I understand that the spatial input product is aggregated to the hydro model by taking the meteo values per pixel: is this a good strategy given that the meteo product does not represent the true topography of the catchment ? In particular for the model that uses elevation bands? How are the pixel of the meteo product matched with the pixels of the hydro model for the other model?

See our response to the comment #RC2.5.

#RC2.11 From the methods it seems like the two models are not calibrated with the same criterion, why? Is this a good idea? does this impact the analysis beyond what is mentioned in the paper?

We agree that the choice of different criteria for the optimization of the two hydrological models impacts potentially the modelled streamflows and the corresponding results. This was made in this study to have two different hydrological models applied with the choices made by the developers of the two models (the choice of the parameters (fixed/inferred), the processes which are represented, the data used to estimate the free parameters, the input processing, etc.

#RC2.12 The paper would benefit from a concise summary of how the two models differ in terms of process representation? How do they estimate evaporation? Evaporation is almost not mentioned in the entire paper? But it should have a key influence on the representation of the water balance?

We agree that the description of how each model estimates the evapotranspiration is missing form the current manuscript and that it should be included. For MORDOR-SD, potential evapotranspiration (PET) is estimated using the formula provided by Oudin et al. (2005). The actual evapotranspiration is estimated according to the PET and different model states. SMASH has a similar formulation.

#RC2.13 Section 3: how do the model topographies of the meteo products differ?

See our response to the comment #RC2.5.

#RC2.14 Section 4: do I understand correctly that the models are calibrated with criteria that are discussed in the sections dedicated to the models and that other criteria are used to evaluate the performance? Or are the criteria in the model section not relevant?

Yes, this is correct. Different criteria are used to calibrate the hydrological models and to evaluate their performance. The set of criteria used to calibrate the models is limited and has a primary objective to provide constrained estimates of the parameters and avoid parameter equifinality. The set of criteria used to evaluate their performances is a lot richer in order to obtain a comprehensive assessment of the performances.

#RC2.15 Are the models calibrated with each meteo product?

Yes, this is correct. This will be clarified in Section 4.2.

#RC2.16 And why is the error on the floods not simply computed as a square-error, is NSE appropriate for this kind of signal? Are the values comparable to those of an entire year?

Mean square errors depend on the magnitude of the streamflows, which can be very different from one catchment to another. The NSE is comparable to other criteria (R square, mean relative errors) and the aim here is to intercompare their values considering different meteorological forcings or hydrological models.

#RC2.17 Results: are the mNSE values a priori comparable for the different catchments? Since we do not see any streamflow time series and do not know if there are differences in the regimes, it is hard to judge.

We agree with the reviewer that the mNSE values are not comparable for the different catchments. Here, we want to intercompare the mNSE values with different meteorological forcings or hydrological models and see if systematic differences can be observed.

References

- Bárdossy, A., et T. Das. « Influence of Rainfall Observation Network on Model Calibration and Application ». *Hydrology and Earth System Sciences* 12, nº 1 (25 janvier 2008): 77-89. https://doi.org/10.5194/hess-12-77-2008.
- Bárdossy, András, and Geoffrey G. S. Pegram. 2016. "Space-Time Conditional Disaggregation of Precipitation at High Resolution via Simulation." Water Resources Research 52 (2): 920–37. https://doi.org/10.1002/2015WR018037.
- Borga, Marco, Paolo Boscolo, Francesco Zanon, and Marco Sangati. 2007. "Hydrometeorological Analysis of the 29 August 2003 Flash Flood in the Eastern Italian Alps." *Journal of Hydrometeorology* 8 (5): 1049–67. https://doi.org/10.1175/JHM593.1.
- Cristiano, Elena, Marie-claire ten Veldhuis, Daniel B. Wright, James A. Smith, and Nick van de Giesen. 2019. "The Influence of Rainfall and Catchment Critical Scales on Urban Hydrological Response Sensitivity." *Water Resources Research* 55 (4): 3375–90. https://doi.org/10.1029/2018WR024143.
- Delrieu, Guy, John Nicol, Eddy Yates, Pierre-Emmanuel Kirstetter, Jean-Dominique Creutin, Sandrine Anquetin, Charles Obled, et al. 2005. "The Catastrophic Flash-Flood Event of 8–9 September 2002 in the Gard Region, France: A First Case Study for the Cévennes–Vivarais Mediterranean Hydrometeorological Observatory." *Journal of Hydrometeorology* 6 (1): 34–52. https://doi.org/10.1175/JHM-400.1.
- Dong, Xiaohua, C. Marjolein Dohmen-Janssen, et Martijn J. Booij. « Appropriate Spatial Sampling of Rainfall or Flow Simulation/Echantillonnage Spatial de la Pluie Approprié pour la Simulation D'écoulements ». *Hydrological Sciences Journal* 50, nº 2 (1 avril 2005): null-297. https://doi.org/10.1623/hysj.50.2.279.61801.
- Emmanuel, I., O. Payrastre, H. Andrieu, et F. Zuber. « A Method for Assessing the Influence of Rainfall Spatial Variability on Hydrograph Modeling. First Case Study in the Cevennes Region, Southern France ». *Journal of Hydrology* 555 (1 décembre 2017): 314-22. https://doi.org/10.1016/j.jhydrol.2017.10.011.
- Fu, Suhua, Torben O. Sonnenborg, Karsten H. Jensen, et Xin He. « Impact of Precipitation Spatial Resolution on the Hydrological Response of an Integrated Distributed Water Resources Model ». *Vadose Zone Journal* 10, nº 1 (2011): 25-36. https://doi.org/10.2136/vzj2009.0186.
- Hohmann, Clara, Gottfried Kirchengast, Sungmin O, Wolfgang Rieger, and Ulrich Foelsche. 2021. "Small Catchment Runoff Sensitivity to Station Density and Spatial Interpolation: Hydrological Modeling of Heavy Rainfall Using a Dense Rain Gauge Network." *Water* 13 (10): 1381. https://doi.org/10.3390/w13101381.

- Huang, Yingchun, András Bárdossy, et Ke Zhang. « Sensitivity of Hydrological Models to Temporal and Spatial Resolutions of Rainfall Data ». *Hydrology and Earth System Sciences* 23, nº 6 (19 juin 2019): 2647-63. https://doi.org/10.5194/hess-23-2647-2019.
- Meselhe, E. A., E. H. Habib, O. C. Oche, et S. Gautam. « Sensitivity of Conceptual and Physically Based Hydrologic Models to Temporal and Spatial Rainfall Sampling ». *Journal of Hydrologic Engineering* 14, nº 7 (1 juillet 2009): 711-20. https://doi.org/10.1061/(ASCE)1084-0699(2009)14:7(711).
- Oudin, Ludovic, Frédéric Hervieu, Claude Michel, Charles Perrin, Vazken Andréassian, François Anctil, and Cécile Loumagne. 2005. “Which Potential Evapotranspiration Input for a Lumped Rainfall–Runoff Model?: Part 2—Towards a Simple and Efficient Potential Evapotranspiration Model for Rainfall–Runoff Modelling.” *Journal of Hydrology* 303 (1): 290–306. https://doi.org/10.1016/j.jhydrol.2004.08.026.
- Parkes, B. L., F. Wetterhall, F. Pappenberger, Y. He, B. D. Malamud, and H. L. Cloke. 2012. “Assessment of a 1-Hour Gridded Precipitation Dataset to Drive a Hydrological Model: A Case Study of the Summer 2007 Floods in the Upper Severn, UK.” Hydrology Research 44 (1): 89–105. https://doi.org/10.2166/nh.2011.025.
- Terink, Wilco, Hidde Leijnse, Gé van den Eertwegh, and Remko Uijlenhoet. 2018. “Spatial Resolutions in Areal Rainfall Estimation and Their Impact on Hydrological Simulations of a Lowland Catchment.” *Journal of Hydrology* 563 (August): 319–35. https://doi.org/10.1016/j.jhydrol.2018.05.045.
- Xu, Hongliang, Chong-Yu Xu, Hua Chen, Zengxin Zhang, et Lu Li. « Assessing the Influence of Rain Gauge Density and Distribution on Hydrological Model Performance in a Humid Region of China ». *Journal of Hydrology* 505 (15 novembre 2013): 1-12. https://doi.org/10.1016/j.jhydrol.2013.09.004.
- Zeng, Qiang, Hua Chen, Chong-Yu Xu, Meng-Xuan Jie, Jie Chen, Sheng-Lian Guo, et Jie Liu. « The Effect of Rain Gauge Density and Distribution on Runoff Simulation Using a Lumped Hydrological Modelling Approach ». *Journal of Hydrology* 563 (2018): 106-22. https://doi.org/10.1016/j.jhydrol.2018.05.058.
- Zhu, Zhihua, Daniel B. Wright, and Guo Yu. 2018. “The Impact of Rainfall Space-Time Structure in Flood Frequency Analysis.” *Water Resources Research* 54 (11): 8983–98. https://doi.org/10.1029/2018WR023550.