The paper by Andrianaki et al. deals with a topic of interest for HESS readers: the modelling of runoff in a glacierised catchments and projections of its evolution. The manuscript reads easily and is concise; I would like to thank the authors for that, as it is often not the case and readers are burdened with loads of not so useful information in many papers.

That said, I feel that there is room for improvement before the paper reads as a scientific paper. Here are my **main remarks**.

1) The main criticism is that <u>I failed to identify clearly what the readers could bring home from this manuscript</u>. Definitely not a new methodology, as the SWAT model is basically used as is, the sensitivity test is not detailed and the calibration and climate change exercises are classical. In my opinion, results are also not so remarkable. It is very interesting to see the validation exercise on a different period and then on a different catchment, but in the end we have results about one catchment and the calibration period is very short. As a consequence, we could wonder if we have the right answer for the right reason or not. I find it very difficult to extrapolate anything from results on this catchment for further works.

   If the main additional value of the paper is the fact that SWAT works for this area, then this should be better highlighted and put into perspective with relevant literature. This reflects on the objectives of the study, which are barely presented in the paper and makes it look like an application of the model rather than an actual research work. Only lines 51-52 give some elements on the interest of this work. Consistently, the conclusions only briefly highlight one novelty of the study (L. 354).

   In my opinion, the abstract, the introduction and the conclusions should be clear about the novelty of this work.

2) It is, if I'm not wrong, never clearly stated that calibration of SWAT is done compared to discharge observations only. Calibration is mentioned many times (abstract, end of introduction, section 3.3) but the used observation is not given. SWAT is physically based and snow observations are definitely an additional value to models calibration in snowy areas, so it is legitimate to wonder if the authors used any kind of snow data here.
3) The calibration set up is unclear and at some point flawed to me.
   First, we don't know exactly what the objective function is: authors introduce NS and R^2 but they don't specify how they used them: through a composite criterion? With a Pareto front? Then, the use of NS in snowfed basins is not advised. Indeed, this criterion relates the performance to the mean observed discharge, which is a bad predictor in such a seasonally-variable environment (see Schaefli and Gupta (2007)). It also underestimates discharge variability.
   Finally, we don't know how the parameters from the small basin are transferred to the larger one. Are some of these parameters time or scale dependent? It is just said that they are adjusted.
4) The structure of section 4.1 is not easy to follow. Some kind of sensitivity test is done to identify which parameters to calibrate. I failed to understand if it was done by the authors, and if yes I don't understand why it is mentioned only in the third paragraph, so after talking about the values of the calibrated parameters. Also, the word "set" is often used to refer to

parameters; as it is unclear what is meant since both a manual calibration and an automatic one are mentioned, I got a bit lost.
In addition, authors seem to infer that Table 1 shows the results of a sensitivity test. What I rather see here is how different the calibrated values are from the default ones, some of them being unrealistic maybe (I don't know where they come from). L. 239: which ones are the least sensitive ones?

5) The actual setup of this whole study is not justified by the authors. Why is the model calibrated on the small basin that has few data and validated on the large basin with a lot of data rather than the opposite?
6) L. 304: I thought that the black (reference) curve in Fig. 7 should be the same as the SWAT curve in Fig. 6, but it does not seem so. Did I get something wrong? The resolution of Fig. 7 could be improved, it is more difficult to read than Fig. 6.
7) L. 317: the authors state that the volume of the glacier reduces to half in 2070. I wonder how this is considered in the SWAT model. Indeed, I expect that the initial conditions of the model (due to the Delta method used for producing the climate projections a continuous hydrological projection cannot be done) had to be adjusted. How was that done? Also, please precise who estimated this reduction (authors? Literature?).

Minor remarks:

Title: The title is not very sexy… Also CZO is an acronym, is it well known enough to be used in a title?

L. 30, 32 and many other places: a space is missing after the semi-colon.

L. 31: I think that the lack of observed data of sufficient quality could also be mentioned.

Section 2: what is the surface area of the small watershed? It is only given for the larger one.

L. 60: after "(Fig. 1)" I think that "is" is missing.

L. 62: inconsistent (lack of) space between number and unit.

L. 69, 74…: why is "et al." suddenly in italics?

L. 77: I would add a comma after "interface"

L. 135: strange punctuation after "Climate change scenarios"

L. 149-150: are the parentheses necessary around Delta P and Delta T? "(Bosshard et al. 2011)" should be "Bosshard et al. (2011)"

L. 158: I would add "scenarios" after "SMHI"

L. 164: if I got it right, Delta P close to 1 mean no change. Is it correct?

L. 172: "extenT"

L. 211: what you have done is a proxy-basin sample test according to the well-known paper Klemes (1986). This is not done so often, I recommend citing this paper.

L. 220: “temperatureS”

L. 225: I would add a comma after “September”

L. 302: I also see a shift of the peak for the far future

L. 320: “snow-fre”

L. 323: using the future is a bit too categorical. There are some uncertainties in projections.

L. 360: any ideas about these other uses? I think this is of interest for the readers.

L. 428: Farinotti et al. (2012) is given twice.

L. 471: Viviroli et al. (2004) has been published, please update

L. 480: “SIMULATION1”: what is this “1”?

Table 1: space or no space between “mm” and “H2O”? In the caption, I would place “SWAT parameters” just after “sensitive”

Fig. 1 and 2: scale and north direction are missing. I would skip “The Damma Glacier CZO” on top of Fig. 1.

Fig. 3 and others: months are not given in English (“Dez”). I would also lie to see each time in the caption the catchment of interest and the period.

Fig. 5: panel (a) is too small for the long period given; it hides potential serious mismatches between simulation and observations.

Fig. 6: is it 1981 as in the text or 1987? Is that an interannual mean? Please comment why SWAT underestimates low flows.

References:

Schaefli, B. and Gupta, H. V. (2007), Do Nash values have value?. Hydrol. Process., 21: 2075-2080. doi:10.1002/hyp.6825