Dear Anonymous referee, thank you for your review and very constructive comments.

General remarks

Reviewer: Even though the paper is about important issues in hydrology (model complexity, impact of climate change), the current version has several flaws. As pointed out by Guillaume Thirel, its main goal is not clearly stated. You state that SWAT "has rarely been used for high alpine areas" and imply to study the suitability of SWAT for such environment. This is not completely true, as SWAT has been widely used in mountainous regions during the last decade (see for example Rahman et al. 2013, references within and papers citing it). The authors should carefully streamline the main goal of the paper.

Authors: You are right that the main goal of the paper is not clear. In this manuscript, we wanted to show not only the applicability of SWAT on a glacierised watershed but also to assess its transferability in different spatial and temporal scale and subsequently to test whether it can be applied on a high altitude glacierised ungauged watershed for runoff simulation and climate change simulations. This is something that hasn't been done before with SWAT but can be quite useful in water management considering the fact that SWAT is a widely used model, used even in large scale simulations (Pagliero et al, 2014).

It is true that in the last years there is an increasing interest in the application of SWAT in high mountainous areas and since a big part of the watersheds in these regions is ungauged, we believe that our study can contribute towards this direction. In our site we have the opportunity to test the upscaling of the model, because we have a quite unique situation; a small well gauged watershed monitored through the CZO projects, which is part of a larger watershed and for which we have hydrological data thanks to its use forthe hydroelectric power plant. This gives us the opportunity to verify the model with independently collected data on the large watershed.

We have rewritten the abstract and conclusions, and extended the introduction focusing on the points mentioned above in order to make our objectives clearer. We also added relevant literature to put them into perspective.

Reviewer: A second major problem is the lack of references or justifications throughout the text. You make strong statements without justifying them or explaining why you made that choice. Here are a few examples:

• The calibration and validation periods are both very short (line 181-183). Why have you chosen such a limited period?

Authors: The reason why the calibration and validation periods are short is that for the Damma glacier watershed we had runoff data for the period 2009-2013. Probably it would be best if we had used these runoff data only for calibration and omitted the validation step, but the performance of the calibration is the same as the calibration and therefore we think that it wouldn't make any difference. In addition, this period is short, but it still includes a relatively large variability in the weather conditions and precipitation amounts. For example, it includes a rather wet year and hot summer and dry and warm autumn.

Reviewer: You estimate the glacier retreat during the last 90 years (line 63-64) without any reference. Where does it come from?

Authors: Damma glacier watershed is a well studied site. Glacier retreat was estimated in previous studies described in Bernasconi et al., 2008 and Bernasconi et al., 2011 (already cited in the paragraph) using systematic recordings.

Reviewer: • Climate models (line 147-151): why have you chosen these 3 models out of the 10 available in CH2011? is there any reason?

Authors: We used these three models, because they were theones in common with both Alpine 3D and PREVAH.

Reviewer: To the best of my knowledge, the CH2011 scenarios (based on the delta change method) were not suitable for assessing changes in extreme events. Based on which element, are you stating an increase in extreme events (Line 342-343)?

Authors: What we meant is that predicted runoff of the far future period T2 shows higher fluctuations from year to year than that of the near future period especially from September to October. Sentence is rephrased.

Reviewer: You are making strong assertions based on the Nash-Sutcliffe model efficiency throughout the paper (line 197-198, 250-251, 259, 268), but be careful, because this indicator strongly depends on the hydrological regime (Schaefli and Gupta, 2007). In alpine basins where you have a strong annual cycle, a NSE coefficient of 0.49 is rather bad and not satisfactory as you state. When comparing averaged models results (Figure 6, line 284-292), based on which elements (objective/subjective) can you say that the performance of SWAT is comparable to PREVAH and Alpine3D? I personally do not agree based on the NSE coefficients you provided.

Some of the SWAT parameters seem to be scale-dependent (in time and space), which could partly explain the model performance deterioration. You should somehow discuss which parameters are the most sensitive in space (validation over the Göschneneralpsee) and in time (with regard to climate scenarios). In addition, you are using different soil and landuse maps in the Damma and Göschneneralpsee catchments (Line114-122). For me, this choice is a bit risky as you upscale your parameters and could bring some inconsistency

Authors: In response to your comment and the comment by the Editor, we investigated further the predictive power of the model for the greater catchment by comparing the observed data with the model results for the spring snowmelt timing, timing of highest flow, autumn recession period and the centre of mass (COM). To do this analysis we used the 15-day average of the daily runoff. Results are presented in Fig. 1 and 2 and the Table 1 given below.

The model predicts efficiently the spring snowmelt timing and the autumn recession period. The difference between the COM of the observed and the simulated runoff, which is given in Table 1, is low and for some years close to zero, which is also satisfactory. The main inconsistencies between measured and simulated data are observed for the general timing of the highest peak, Fig. 2.

One of the reasons for the deterioration of the model is that it doesn't differentiate between snow and glacier dynamics and only one parameter for both snowmelt and glacier meltrates is applied. This becomes more important in our study, since there is a difference between the glacier coverage of the two catchments. The Damma glacier is 50% covered by the glacier while the greater catchment is 20%.

One more reason is the difference in the response of the Damma glacier watershed in comparison to the greater area. Damma is characterised by very steep slopes (even up to nearly 80 degrees) and runoff originates mainly from snowmelt, glacier melt and rainfall (Magnuson et al., 2012). For this reason, the small watershed is characterised by very fast response, which led to the high value of ALPHA_BF and the low value of the GW_Delay parameters. On the other hand, the Göscheneralpsee feeding area is less steep on average and for the two out of the four of its watersheds, runoff is drained through tunnels into the reservoir.

The most sensitive parameters are the ones related to the snowmelt, like SFTMP, SMTMP and TIMP. During the manual calibration we checked many of the parameters related to landuse and soils and we think that we do not have an inconsistency. The parameter values set during the delineation of the watershed and initial parameterisation should be adequate. Finally, because our site is above tree line evapotranspiration parameters are not significant.

It is true that comparing SWAT with Alpine3D and PREVAH is tricky since they were calibrated for different catchments and different periods of time. The NS efficiency and the benchmark efficiency BE (added in the revised text)for the calibration period only are: 0.85 and 0.19 respectively for ALPINE3D, 0.91 and 0.49 for PREVAH and 0.84 and 0.22 for SWAT. These efficiencies of Alpine 3D and SWAT are in good agreement, with the efficiencies of PREVAH being slightly higher.

We have rewritten the entire paragraph for the comparison of the models. We focused less on comparing the efficiency of the model and mainly on what we can conclude from the comparison between the three models.

Minor remarks

Reviewer: Some typos are visible throughout the paper, the authors should carefullyproofread it.Here are some minor comments:

Authors: We will proofread the paper

Reviewer:1. Line 44: what do you mean by "its structure is physically based"? For me, Alpine3D is a physically based model, SWAT is not. Please clarify!

Authors: It really depends on how you define the term “physically based”. Some researchers consider SWAT to be a physically based model and others don’t since not all of its parameters can be defined directly by measurements. Since it wasn’t adding to the context, the sentence was deleted.

Reviewer:2. Line 98: what do you mean exactly by this statement?

Authors: Fontaine et al., (2002) revealed the importance of improving SWAT algorithms to include in the model the influence of elevation and season on the dynamics of the snowpack.

Reviewer: 3. Line 104: "basic input" is a subjective statement.

Authors: “basic” is deleted

Reviewer: Line 124: the new MeteoSwiss network is named SwissMetNet not ANETZ anymore.

Authors: Corrected

Reviewer: 5. Line 1341-134: you are right, lapse rate are critical in mountainous regions, so tell the reader which values you have used in you study!

Authors: precipitation lapse rate PLAPS was set to 5 (mm/km) and temperature lapse rate was set to -5.84 (^{o}C/km).

Reviewer: 6. In figure 1, what is the added value of the inset for the present study? There is aninconsistency in the orientation (North) between figure 1 and 2. You should justcombine them into a single figure.

Authors: You are right. The Figures will be combined.

Reviewer: 7. Figure 3a, is it really useful to show the uncalibrated time series?

Authors: We wanted to show that SWAT cannot be used here without calibration.

Reviewer: 8. We can hardly see the difference between the two curves in figure 5a. Consequently, the reader cannot really assess the quality of the model

Authors: A better version of Fig. 5a is given below in Fig. 3. As you can see in this Fig. as well as in Fig. 5b in the manuscript, that shows the accumulative runoff, there is an overestimation of the streamflow by the model during the years 2000 to 2002. This overestimation must be related to the runoff melt rate but we need to investigate this further by looking into the weather data. Furthermore, the simulated runoff peaks are higher and narrower than the observed ones, which must be related to the differences in the response and groundwater interactions between the small watershed and the greater area, as discussed above.

Reviewer: 9. In figure 6, it is somehow hard to make the difference between the lines. Try different colors.

Authors: Is Fig. 4 here clear enough?

References

Bernasconi, S. M., Christl, I., Hajdas, I., Zimmermann, S., Hagedorn, F., Smittenberg, R. H., Furrer, G., Zeyer, J., Brunner, I., Frey, B., Plotze, M., Lapanje, A., Edwards, P., Venterink, H. O., Goransson, H., Frossard, E., Bunemann, E., Jansa, J., Tamburini, F., Welc, M., Mitchell, E., Bourdon, B., Kretzschmar, R., Reynolds, B., Lemarchand, E., Wiederhold, J., Tipper, E., Kiczka, M., Hindshaw, R., Stahli, M., Jonas, T., Magnusson, J., Bauder, A., Farinotti, D., Huss, M., Wacker, L., Abbaspour, K., and Biglink Project, M.: Weathering, soil formation and initial ecosystem evolution on a glacier forefield: a case study from the Damma Glacier, Switzerland, Mineral. Mag., 72, 19-22, 10.1180/minmag.2008.072.1.19, 2008.

Bernasconi, S. M., Bauder, A., Bourdon, B., Brunner, I., Bunemann, E., Christl, I., Derungs, N., Edwards, P., Farinotti, D., Frey, B., Frossard, E., Furrer, G., Gierga, M., Goransson, H., Gulland, K., Hagedorn, F., Hajdas, I., Hindshaw, R., Ivy-Ochs, S., Jansa, J., Jonas, T., Kiczka, M., Kretzschmar, R., Lemarchand, E., Luster, J., Magnusson, J., Mitchell, E. A. D., Venterink, H. O., Plotze, M., Reynolds, B., Smittenberg, R. H., Stahli, M., Tamburini, F., Tipper, E. T., Wacker, L., Welc, M., Wiederhold, J. G., Zeyer, J., Zimmermann, S., and Zumsteg, A.: Chemical and Biological Gradients along the Damma Glacier Soil Chronosequence, Switzerland, Vadose Zone J., 10, 867-883,http://dx.doi:10.2136/vzj2010.0129, 2011.

Kobierska, F., Jonas, T., Zappa, M., Bavay, M., Magnusson, J., and Bernasconi, S. M.: Future runoff from a partly glacierized watershed in Central Switzerland: A two-model approach, Advances in Water Resources, 55, 204-214, http://dx.doi.org/10.1016/j.advwatres.2012.07.024, 2013.

Pagliero, L., Bouraoui, F., Willems, P., and Diels, J.: Large-Scale Hydrological Simulations Using the Soil Water Assessment Tool, Protocol Development,and Application in the Danube Basin, Journal of Environmental Quality, 43, 145–154, http://dx.doi:10.2134/jeq2011.0359, 2014.

Table 1 Difference of the centre of mass (COM) and autumn recession period in days, calculated from the 15-day average.

Year	**COM**	**Autumn recessionperiod**
1997	6.8	1
1998	4.2	1
1999	1.0	0
2000	3.0	16
2001	0.6	1
2002	7.8	19
2003	0.6	5
2004	2.4	4
2005	4.3	0
2006	4.1	1
2007	8.1	1
2008	3.1	0
2009	4.6	0
2010	6.0	0

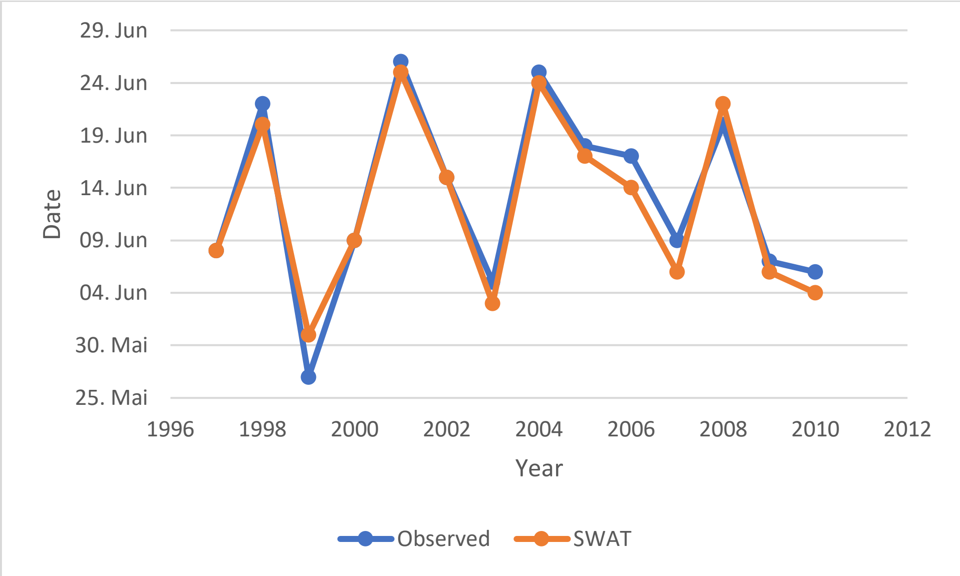

Figure 1 Comparison between the observed and simulated spring snowmelt timing. A 15-day average filter was applied on daily measurements.

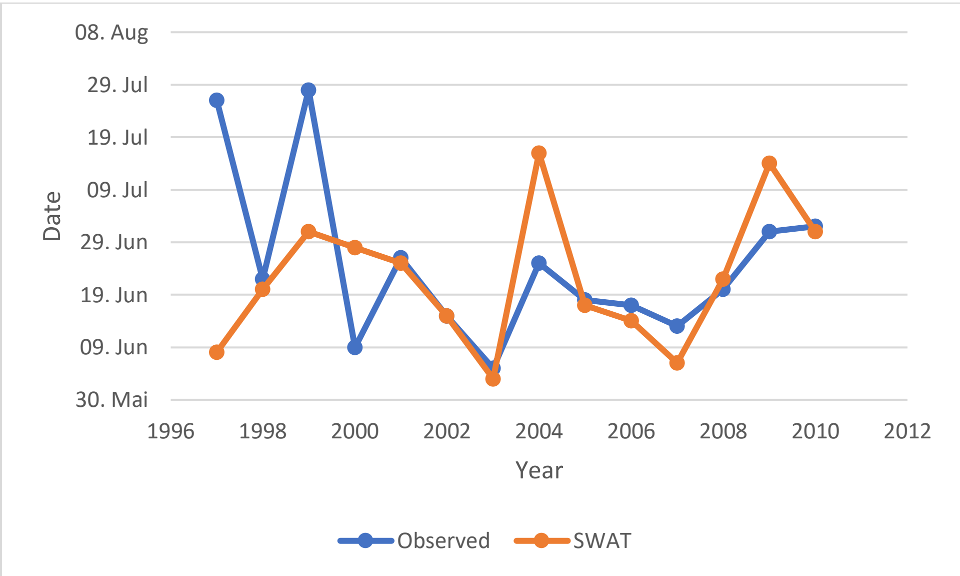

Figure 2 Comparison between the observed and simulated spring snowmelt timing. A 15-day average filter was applied on daily measurements.

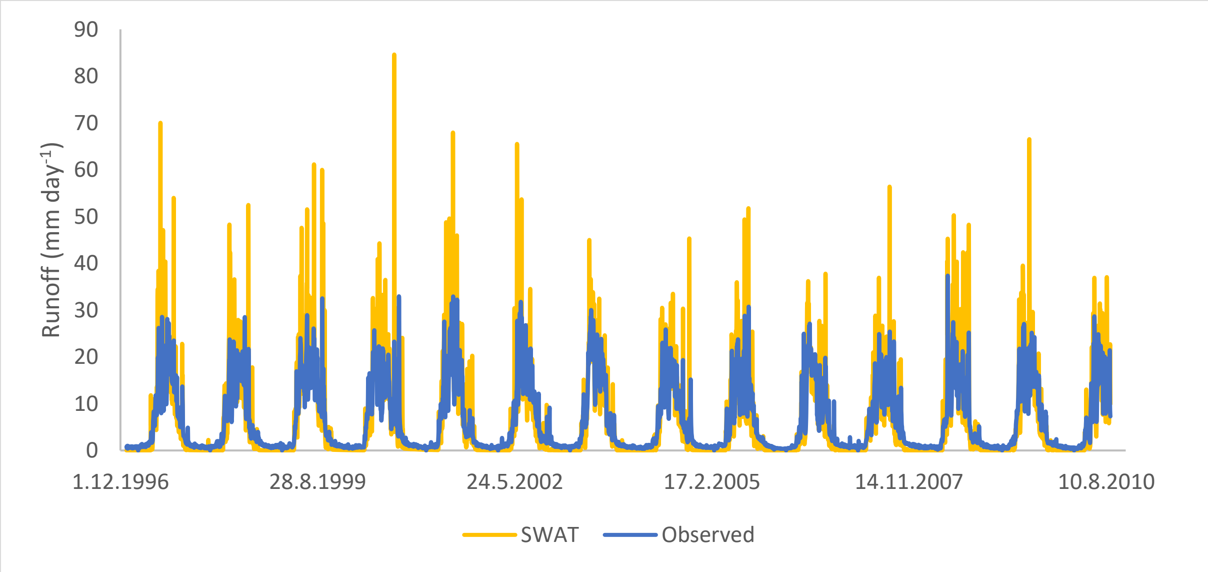

Figure 3(Figure 5a in manuscript) SWAT results and measured runoff values of the feeding catchment of the Göscheneralpsee for the period 1997-2010

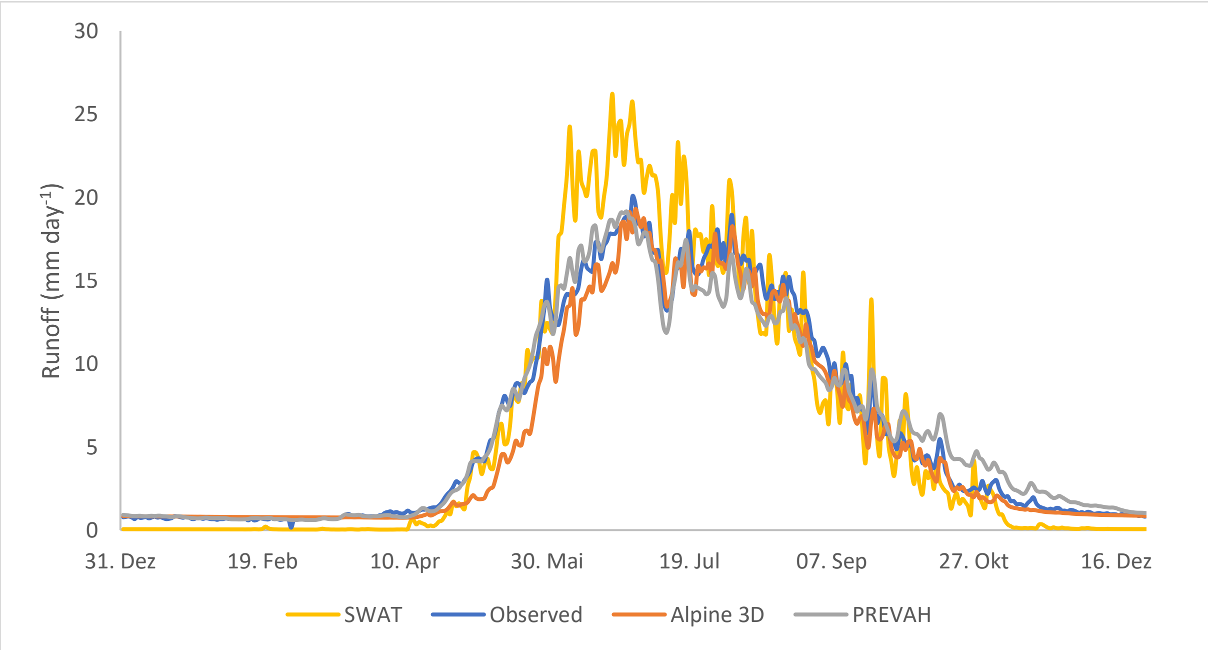

Figure 4 (Figure 6 in manuscript) Interannual mean of the results of SWAT, Alpine3D and PREVAH models and the measured runoff of the Göscheneralpsee feeding catchment for the 1997-2010 period.