Dear reviewer,

We thank you very much for your constructive and instructive comments on our manuscript. We are very well aware that this review has been a significant time investment and therefore especially appreciate the reviewer's feedback and commitment. Please find below a detailed answer to all the points of this thorough review. We hope that these answers together with the corresponding proposed modifications of the manuscript will provide the requested clarifications and consider the useful suggestions.

Reviewer comments are repeated in *italic*, author replies in regular font.

Best regards,

Adrien Michel, on behalf of the authoring team.

The manuscript by Michel et al details a study that uses a chain of models to simulate river temperatures in 12 Swiss catchments that encompass two main landscape types: alpine and lowland. They used downscaled climate change projections in a suite of models that simulate glacier, snow, soil temperature, runoff, and river temperature dynamics. They assess the ability of the modelling chain to simulate historic conditions using observations from around 2005-2018. They conclude that river temperatures will increase for both catchment types under climate change; however, the magnitude and seasonality of increases will differ. The lowland (Plateau) catchments are expected to see relatively uniform river temperature increases throughout the year, with slightly higher increases in summer. In contrast, alpine catchments are expected to see small increases in winter river temperature, but large increases during summer due to an earlier shift in peak discharge associated with lower amounts of snow and glacier melt contributions.

Overall, the manuscript is relatively well written (although a number of typos and grammatical errors need to be corrected), and the logical development and presentation of the study generally makes sense. This was clearly a significant logistical effort to build and work with this chain of models. I can also appreciate the challenge in synthesizing these efforts. Because of the amount of work presented, and that much of it builds on previous efforts (i.e., the specific model components), I found it an overwhelming piece of work to digest. I appreciate the amount of supplementary material included, but the decision to not always reference that material directly in the text (supposedly to 'alleviate the text'?) makes it very difficult for the reader to navigate the immense number of results shown. Despite the considerable amounts of results included, the key findings from these modelling exercises don't strike me as overly complicated; therefore, I think there is a great opportunity to streamline the manuscript by reducing the amount of results presented while still providing support for the key findings in this study.

We are aware that this manuscript is rather long and extensive. It has been discussed at some point to maybe split it into two papers (one about the dataset and one about the analysis), but we preferred to keep it as one comprehensive story rather than chopping it down into fragments. We agree that there is potential for shortening and we will try our best during the review process to make the story more concise. Some suggestions for shortening are presented below and in the answer to the second reviewer. At the same time, the reviewer also suggested some additions and clarifications. We will do our best to include this information without further extending the paper. Regarding more frequent references to the supplementary material, we will add these to the text.

1. Describing the model chain and key asumptions/limitations

Keeping track of the various model components, what they do, and how they were individually and collectively calibrated is challenging. Perhaps a flow chart or a diagram detailing how the models interact, what components are tested against observations and/or what parameters are calibrated would help readers? For example, I was really confused about the time periods used to calibrate the individual models. Alpine3D was calibrated to 2012-2018? StreamFlow was calibrated to 2012-2014, but then validated to 2015-2018? So this is not an entirely independent validation since Alpine3D was already calibrated to that period? In addition the 'Validation over climate change period (Section 4.2)' was validated for 2005-2015 - which encompasses the period used for calibrating the model. So again, this is not an independent validation? I'm likely misunderstanding this workflow, which is why a better description of the steps taken might help the reader.

Thank you for this suggestion. Please find below a first draft of a figure summarizing the models' interactions and the various steps (Figure 1). To answer your question, there is no general calibration of Alpine3D. We perform a few model runs of Alpine3D over the whole period to adjust the vertical precipitation lapse rate in order to achieve a correct mass balance over the year. We agree that this choice is not validated afterwards. A similar procedure was used in Gallice et al. (2016) and Brauchli et al. (2017).

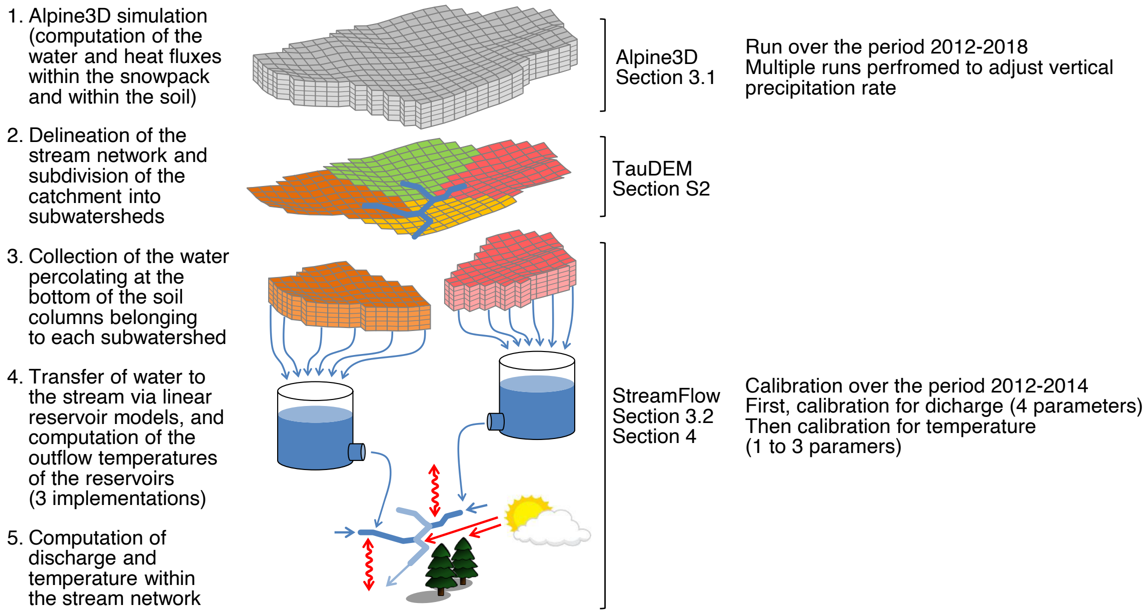

Figure 1 – Details of the models' workflow. The calibration and validation periods indicated holds for all catchments except for the Eulach catchment (where the periods 2015-2016 and 2017-2018 are used). Figure adapted from Gallice et al. (2016)

Regarding Section 4.2 "Validation over climate change period", the objective here is to compare the model output for runs forced with climate change scenarios compared to runs forced with measurements, and consequently we do not consider it an issue if a few years overlap with the calibration period. Maybe the section could be renamed "Evaluation of climate change scenarios", because this is the real purpose of this section. Indeed, climate change scenarios have been downscaled to perform this work. Both the “base” climate change scenarios CH2018 and the downscaled version come with a few shortcomings and warnings (detailed in Section 2.4 and in Michel et al., 2021), and the purpose of this section is to ensure the correct usage of the models when forced with these time series (which are not the time series used for calibration). In order to shorten the manuscript, we propose to move this section into the Supplementary.

In addition, key details or assumptions made by the various model components are not really addressed - although these assumptions can be critical for interpreting the model results. Instead the reader is often referred to other studies and publications to get these important details. I recognize that the authors don't want to duplicate information already contained in other papers; however, I think the key aspects and assumptions of those models relevant to this study need to be presented in this manuscript.

Indeed, we tried not to extensively repeat the existing literature for the sake of length of the manuscript. Alpine3D is entirely physics-based and thus no particular assumptions are made, however there are still some relevant issues (e.g., no snow transport by wind or no partially snow-covered pixels, we will mention this in Section 3.1). Regarding StreamFlow, the main assumption concerns the soil water temperature and we think this is already largely discussed in Sections 3.3, 4.1.2, and 4.4. The two-linear reservoir approach used is also an approximation; submitting to a hydrology journal we omitted the details on this well-known concept but we will refer explicitly to a key reference such as Perrin et al. (2003). The incoming shortwave radiation would deserve more attention. We discuss it further below and make a suggestion on how to integrate it in the paper.

Also, were any 'spin-up' or 'warm-up' periods used for any of the models? If not, why?

Yes, there is a spin-up period. Thank you for having noticed this missing information. Alpine3D is always started for the month of July, with a temperature profile approximated from previous simulations, and model results are used in StreamFlow only from October on. There is thus a 3 months period for the soil temperature profile and water content to equilibrate. StreamFlow always uses a period of two years as a spin-up. This is achieved by running the model over the first two years of the time period, and then re-starting from the beginning of the time period from either the real calibration or simulation. We will clarify this point by adding the details above in Section 4.1.1 where we present the time periods used.

2. Clarifying the advection term associated with runoff

Probably a consequence of the information overload outlined in my point #1 above, I'm struggling to understand if and how the advection term associated with hillslope/land runoff is treated in these models. Alpine3D simulates spatially distributed soil temperatures and water available for runoff. StreamFlow sums the water available for runoff for all the Alpine3D cells draining to a stream reach of interest?

Yes, each stream reach is associated with a sub-watershed. Figure 1 above, that will be added to the manuscript, clarifies this.

This water available for runoff is assigned a temperature by one of three methods: 1) the energy balance approach in Comola et al (2015), 2) the HSPF algorithm, or 3) a soil temperature value (I am assume taken from Alpine3D?). Very little detail is provided in the text or in the supplementary material about the details and differences in these approaches (other than different RMSE reported in Table S7) and the authors conclude that the HSPF is the most consistent, so they use that for all subsequent model runs. However, I'm confused by the statement that 'in the HSPF scheme, the soil temperature has a less important impact than in the other schemes (soil temperature is only needed for heat conduction between water and river bed).' In this statement, does this mean that the soil temperature output from Alpine3D is only used for the bed conduction term, but in the other schemes it is used for something else? I get that the soil temperature from Alpine3D is used to set the runoff temperature in scheme 3, but how is it used in scheme 1 (the Comola approach)? Also, it sounds like Alpine3D simulates soil temperatures at different depths - so which depth is used for scheme 3?

Yes, in the approach 3), the soil temperature at a given depth (obtained from Alpine3D) is used in StreamFlow to set the temperature of the water entering the river. In the approach 1), the temperatures of the two reservoirs are obtained as follows (see full details in Comola et al., 2015 and a summary in Appendix A2 in Gallice, 2016):

$$dT_U/dt = I_U/S_U * (T_S-T_U) + (T_S-T_U)/k$$

$$dT_L/dt = I_L/S_L * (T_S-T_L) + (<T_S>-T_L)/k$$

Where k is a calibration parameter corresponding to the characteristic time of thermal diffusion, T_S the soil temperature at a chosen depth (obtained from Alpine3D), T_U and T_I the water temperature in the upper and lower reservoirs, respectively, I_U and I_L the infiltrating water to the corresponding reservoir, S_U and S_I the current reservoir storage, and <…> denotes the average of the value over the simulation period. In both equations, the first term corresponds to the energy flux from the inflow water to the reservoirs, and the term on the left-hand side to the diffusive heat exchange between the water and the soil. We will add this mathematical description in the supplementary. By model design, the soil depth used to take the soil temperature in the approaches 1) and 3) is the same as the depth used to compute the heat exchange with the river bed, which can be interpreted as a limitation. However, since we went to the HSPF approach where the soil temperature is used only for heat exchange with the riverbed, this problem disappears.

For all 3 approaches, the model is calibrated using soil depths between 0.5 and 3 meters, at intervals of 0.5 m. The soil depth leading to the best results is retained. The temperature calibration being computationally intensive, an iterative approach is used, i.e., only some depths are tested and additional depths are added if necessary, explaining why different depths are shown in the figures below. Figure 2 shows the details over the calibration period for two non-alpine catchments, and Figure 3 for two alpine catchments. From these figures we see 1) that the difference resulting from the choice of the soil depths is almost negligible for the HSPF approach, and 2) that the HSPF approach is performing better. The particular situation of the alpine catchments is discussed below when answering to your remark concerning the choice of HSPF instead of the approach 1) of Comola et al. (2015). An important note for the interpretation of the plots below is that all results come from a separate calibration. So, if most of the difference seen can be attributed to the chosen soil depth, part of the difference can also be due to the calibration, and this ratio is hard to quantify. The discussion above along with Figures 2 and 3 below will be added in Supplementary Section S5. The main message will also be better explained in Section 4.1.1.

Also, is it appropriate to use the Alpine3D soil temperature for the channel bed temperature? The channel bed has entirely different upper boundary conditions than the terrestrial parts of the catchment and it seems inappropriate to use the Alpine3D soil temperatures to represent channel bed temperatures (especially since the authors note that some of these catchments experience substantial flow losses along their network; therefore, the stream bed temperature will likely be influenced by river water infiltrating the subsurface).

Indeed, this is a rather rough approximation. However, as shown in Figures 2 and 3, for the HSPF approach, the effect of the chosen depth is almost negligible, despite of the soil temperature cycle having a very different annual amplitude at 0.5 meter or at 2.5 meters depth, suggesting that the actual effect of the riverbed heat exchange is in general not a major energy flux. In addition, there is a calibration parameter (the amount of energy transferred between the river and the riverbed by degree kelvin of temperature difference) that plays an important role and this calibration parameter might also be able to correct for the fact the we simplify the system by taking the soil temperature for a terrestrial part and ignoring the upper boundary condition imposed by the river. The values of this calibration parameter are indicated in Figures 2 and 3, and the wide range of values obtained along with the trend observed with soil depth support this hypothesis of the calibration parameter correcting for the simplified soil temperature used. As for the point above, in the revised version this will be detailed in Section S5. In addition, in Section 3.3, where StreamFlow is presented, will discuss this aspect.

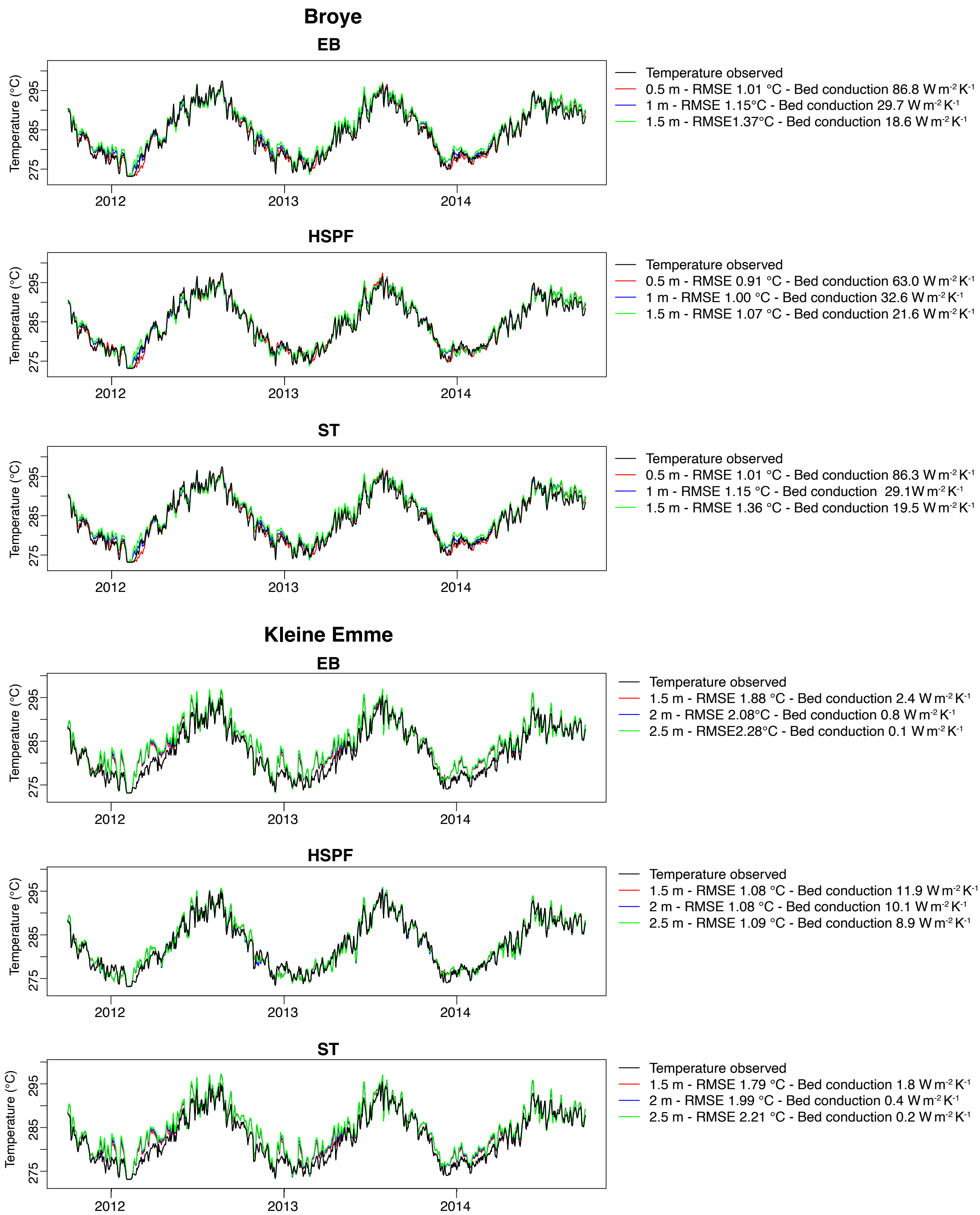

Figure 2 – Details of the water temperature calibration phase with different soil depths used for the soil temperature. Calibration for the Broye catchment (top) and the Kleine Emme catchment (bottom), both located on the lowlands. The three approaches available to compute the water temperature in the soil are used: 1) the energy balance formulation (EB, Comola et al., 2015), 2) the HSPF formulation (Bicknell et al., 1997), and 3) the soil temperature approach (ST, Gallice et al., 2016). The root mean square error (RMSE) and the calibrated value obtained for the river-streambed heat transfer coefficient are indicated in the legend.

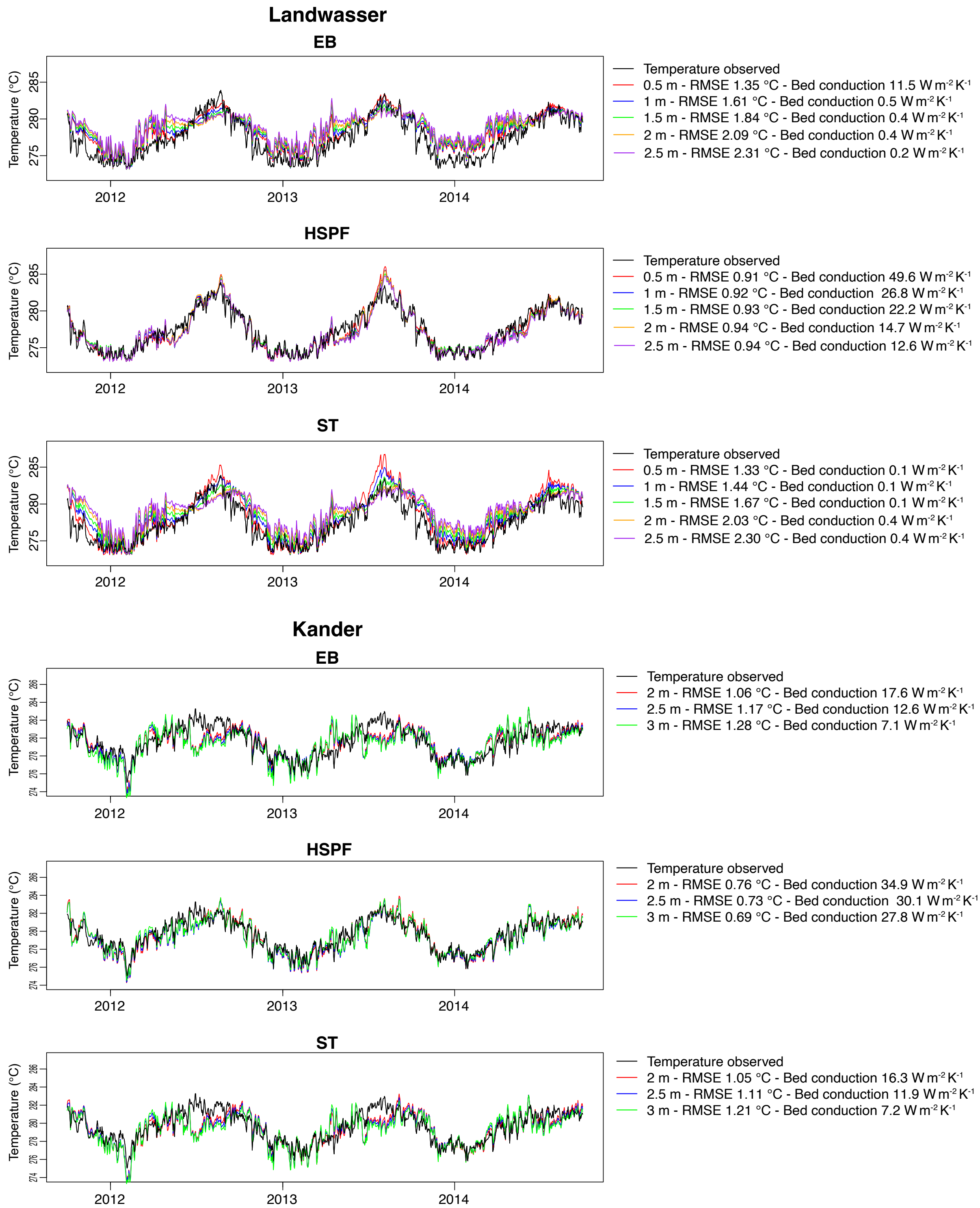

Figure 3 – Details of the water temperature calibration phase with different soil depths used for the soil temperature. Calibration for the Landwasser catchment (top) and the Kander catchment (bottom), both located on the Swiss Alps. The three approaches available to compute the water temperature in the soil are used: 1) the energy balance formulation (EB, Comola et al., 2015), 2) the HSPF formulation (Bicknell et al., 1997), and 3) the soil temperature approach (ST, Gallice et al., 2016). The root mean square error (RMSE) and the calibrated value obtained for the river-streambed heat transfer coefficient are indicated in the legend.

The decision to use the HSPF approach seems like it could have important implications for the climate change modelling, particularly for the alpine catchments that see a decrease in snow cover. As the work by Yan et al (2021) highlights (and some of the calibration issues in this study also suggest), this runoff advective flux can be important in snow-dominated catchments. Based on a quick search, I see that the HSPF algorithm doesn't account for the presence of snow and therefore may be unsuitable for looking at climate change impacts in snow influenced catchments (see Leach and Moore, 2015). Assuming this is the same HSPF algorithm (I suspect it is), could this be partly why the modelling is under-estimating spring/summer temperatures for the alpine sites? Why not use the Comola or Alpine3D approach which, I presume (but no details on these schemes are provided) can account for the influence of snow cover on runoff temperatures? If it can account for the snow influence, I would argue that would be a preferable approach even if the calibration metrics aren't as good as the HSPF values, since it should better extrapolate to future conditions.

The question of snow cover is indeed relevant. As presented in Section 4.4, we think this is the main issue leading to the overestimation of the water temperature in summer. As you correctly mention, the HSPF approach does not take into account snow cover, and Leach and Moore (2015) state that: *"The HSPF model, original CEQUEAU function, and LARSIM-WT were the most successful approaches at estimating throughflow temperatures at Griffith Creek with prediction rates of 53%, 33%, and 31%, respectively. However, a major drawback of these approaches is that they do not account for snow, which limits their use as a predictive tool to under- stand stream temperature response to changes in climate and land cover, and corresponding impacts on snowpack dynamics.".* However, among the 6 different approaches tested for a snow-covered area, they also show that: *"Of the models considered here, the HSPF interflow equation was the most successful at predicting throughflow temperatures. HSPF predictions fell within the range of observed throughflow temperatures 53% of the time and only overestimated throughflow temperatures by a maximum of 1.38 °C.".*

As discussed in the manuscript in Section 4.1.1, the only influence of snow on water temperature when using the HSPF approach, in addition to the mass added, is the fact that snow cover influences the mean sub-catchment soil temperature, which is then used to compute heat transfer (advection) to the river. During the melt season (when advective flows are important) it leads to good results since air and soil temperature are still low, as shown in Figure 3 above, but not in summer.

In the energy balance approach of Comola et al. (2015) detailed above, the soil temperature plays a more important role. Since the soil temperature in Alpine3D is heavily influenced by the presence of snow, we can expect it to lead to better results. However, in this approach, there is no direct accounting for the water originating from snow melt or from precipitation (this would not be straightforward to implement in the current versions of the models). In the lowlands catchments (Figure 2), performances are similar between the HSPF and the energy balance approaches, except that the energy balance approach exhibits a too marked temporal variability compared to measurements. In alpine catchments (Figure 3), this approach shows clear problems in correctly reproducing the water temperature. In the Landwasser catchment, the water temperature in winter is clearly overestimated and the variability during the melt season is largely overestimated. In the largely glacierized Kander catchment, the summer stream temperature is underestimated. This is probably explained by the usage of the mean soil temperature over the whole sub-catchment to compute the energy balance, i.e., all the water originating from the glaciated area enters the rivers mostly at 0°C, while in reality the water already warms while flowing around the rocks. Overall, if the HSPF approach underestimates the snow cover effect, the approach of Comola et al. (2015) seems to overestimate it. This is also observed by Gallice et al. (2016). Figure 3 also shows that the EB approach or the simple approach of taking the soil temperature as proxy for soil water temperature gives very similar results.

Despite the quality metric is only slightly lower than obtained with HSPF, we see in Figure 3 that the EB approach of Comola et al. (2015) presents major issues, which we estimate to be more important than the problem of summer overestimation of the HSPF approach, explaining why the HSPF approach has been used. In the current version of the manuscript only the RMSE is discussed. In order to fully clarify the choice of the HSPF approach, we will add the present Figure 3 into the supplementary.

As shown in the two studies you suggest (Leach and Moore, 2015; Yan et al., 2021), this question of subsurface water temperature is not a straightforward task and the HSPF approach used has some obvious weaknesses. However, we believe that we sufficiently emphasize this in the manuscript and we are very clear about the associated limitation. In addition, the snow cover effect accounted in the model through streambed conduction has an impact, as shown by the increased warming in Alpine catchments expected with climate change compared to the lowlands ones. Nevertheless, we agree that both the abstract and conclusion do not emphasize enough on these modelling aspects, and we will revise these accordingly (see comments from Reviewer 2 and our answer).

Thank you also for pointing us to the recent work of Yan et al. (2021), which we were not aware of at the time of submission. Indeed, that paper shows that snowmelt has an impact on water temperature. However, in our interpretation of their results, Yan and colleagues were primarily considering the upper and lower bound of the contribution of snow melt on water temperature. For infiltrating water temperature, they use either 0°C for water originating from snowmelt (which is clearly a lower bound and neglects any warming before entering the river), or mean annual air temperature otherwise (which is clearly an upper bound during the melt season). Indeed, when they turn on or off the value of 0°C for snowmelt the difference is significant, as one could expect using such bounds for infiltrating water, but we note that the actual effect must be in between those two scenarios and therefore likely to be smaller.

We also think that the analysis in the work of Yan et al. (2021), does not confirm that the approach would lead to more accurate temperatures in the melt season, since the model is calibrated using summer temperature only, and the water temperature is not validated for the calibration station. In addition, the model is calibrated for stations situated rather low in the catchment and when we go to the stations upstream (see Figures 3 and S3 in the paper), we clearly see that the simulated water temperature is underestimated in winter and fall and overestimated in spring (similar to the issues we have for the Landwasser when using Comola's approach shown in Figure 3 here). The metrics used for calibration are not directly comparable (they use MAE, we use RMSE, and MAE <= RMSE). The MAE obtained by Yan et al. (2021) for calibration centered in one season, leads to MAE which are similar or larger than the RMSE we obtain (and RMSE is highly penalized by the summer overestimation in our results). Based on this, we conclude that the model used in Yan et al. (2021) is not expected to show a big advantage in the catchments in our study.

Their main results regarding the impact of snowmelt are shown in Figure 3 in their paper. As we argued before, the values obtained in Yan et al. (2021) represent upper and lower bounds and we also would argue that the effect is overestimated, since the model does not seem to have been recalibrated for the case without the temperature effect from snowmelt. Since we do calibrate our model chain with only a weak effect of snowmelt on water temperature, the calibration may partially compensate for that.

We agree that the snowmelt temperature is an important parameter and its neglection is a drawback leading to a summer overestimation that we amply discuss in the paper. We will include the study by Yan et al. (2021) in our manuscript and briefly summarize the elaborate explanation given above. However, we think that the study by Yan et al. (2021) is not directly applicable to our study approach, as detailed above.

3. Other modelling studies from mountainous snow-dominated environments

The authors primarily reference other Swiss studies throughout the manuscript. There have been other studies looking at hydrology and river temperature response to potential climate change scenarios conducted in mountainous snow-dominated environments. I'm familiar with some of the work from western North America. Some examples include: Null et al 2013, Leach and Moore 2019, Yan et al 2021. In particular, Yan et al (2021) seems highly relevant here. I think it would enrich this manuscript to incorporate the findings from some of these studies in the introduction and discussion (there are some interesting similarities and differences between the findings from those studies and the results presented here).

Thank you for these suggestions. By the time of our manuscript submission, we were not aware of the very recent publication of Yan et al. (2021), discussed above. The work of Leach and Moor (2019) was in fact one of our motivations to use physics-based models for our study and unintentionally got forgotten to me mentioned. This will be of course corrected. We will include more comparisons with results outside Switzerland (e.g. Morrison, 2002; Null et al., 2013; van Vliet et al., 2013; Du et al., 2019; Leach and Moore 2019).

4. Key assumptions on river temperature modelling

Maybe these details are contained in the StreamFlow references, but I was surprised by the lack of discussion on potentially key assumptions around some of the river temperature modelling. In particular, there is almost no details or discussion about the role of riparian vegetation and its influence on radiation exchange and the sensible and latent heat fluxes. The manuscript mentions that topographic shading is taken into account (at least for Alpine3D, it's not clear if this is also the case for StreamFlow) - is that the only source of shading for these rivers? Maybe that is the case? If so, I would recommend clarifying this point. If not, it seems prudent to discuss the potential issues that ignoring the role of riparian vegetation might *have on the modelling. Along these lines, I also wonder if a discussion on potential land cover changes in these catchments over the next decades, and how they might also influence river temperatures, might be worth including? This is touched on a bit, but could be expanded.*

Thank you for bringing up this point. Indeed, riparian vegetation is a topic absolutely worth investigating. Let us first answer the questions related to how the model takes riparian vegetation into account. We will add these clarifications into the sections where the models are described (3.2 and 3.3).

In Alpine3D, the topographic shading is takeninto account to compute the radiation. In addition, a two-layer canopy module is used to compute, among other, the vegetation shading. The forcing grids used for the hydrological simulation in StreamFlow take both of these aspects into account (note that the canopy does also modify the wind speed). However, this has some limitations. Both topographic and vegetation shading are computed from elevation or land use grids at 500 m spatial resolution. The local small-scale topography of the river is ignored (and in alpine areas, rivers can be some meters below the surrounding terrain and experience some shading). Regarding vegetation, the local riparian vegetation might also be ignored since pixels at 500m resolution from the CLC dataset are used. In mountainous area, this might also lead to underestimated shading, while in the lowlands the shading might be overestimated when a large river crosses a forested pixel (and will thus be considered as totally shaded since the canopy module in Alpine3D does not compute any shading projection).

We initially thought of a possible overestimation of radiation to explain the overestimation of water temperature we observe in summer in alpine catchments. However, the detailed analysis presented in Section 4.4, in particular, the energy fluxes show that the warming events are quite sudden and not related to the radiative fluxes, explaining why we do not consider it as the first candidate for the error of the model. An extended discussion on this matter is presented below when answering another reviewer question.

Given the current extent of our manuscript, a longer discussion on riparian vegetation and potential impacts of land use change was considered beyond the scope of this paper. However, we will add a paragraph about this in Section 5.4 in which we address this limitation of the model. The limitations of the topographic and vegetation shading using 500m grid cells will be mentioned, together with the fact that this study assumes that the land cover will not change throughout the 21st century. This will also be mentioned in Section 4.4 when discussing the water temperature overestimation in summer (again see discussion below).

Overall, both the impact of snow cover and of the vegetation are primordial questions to be addressed in future studies and we hope that the models' setup established for this work will serve as starting point for related discussions.

Specific comments:

P1L3: Perhaps expand or give an example why rivers are important socio-economical factor.

We can add the following in the abstract (paraphrasing P1L20-22): "The literature clearly identified several sectors which are vulnerable: agriculture, tourism, electricity production, and drinking water supply and quality (e.g., Hock et al., 2005; Barnett et al., 2005; Schaefli et al., 2007; Bourqui et al., 2011; Viviroli et al., 2011; Beniston, 2012; Hannah and Garner, 2015)."

P3L5: I would replace 'attributed to the' with 'associated with an', since it is fairly well established that although air temperature is often correlated with water temperature, air temperature itself, via the sensible heat flux, is not often a key control.

We will change this.

P6 Section 2.3: How were data from various met stations used as inputs to the models? Lapse rates? Thiessen polygons? Some other adjustment? Ok - I see this is provided in Section 3.2.

Yes, we can add a cross-reference to Section 3.2 in Section 2.3.

P9 Section 3.3: How are energy exchanges at the stream-atmosphere interface dealt with? Is radiation exchange adjusted for riparian conditions? Are the land-based meteorological measurements adjusted for above-stream conditions for the sensible and latent heat flux calculations? The reader is directed to Gallice et al 2016, but some general overview on this aspect should be included here.

The point about vegetation is already discussed above. For both latent and sensible heat fluxes, empirical equations are used (see Hannah et al., 2004; Haag and Luce, 2008; Magnusson et al., 2012, and Comola et al., 2015). These equations use the air temperature without local correction. Although these equations are common and widely used, we will add a short paragraph giving a general overview and the relevant references.

P18 Section 4.4: The model's inability to reasonably simulate the extremely warm 2003 period seems to be a critical issue, particularly since this model is being used to simulate climate warming scenarios (the model seems to be clearly missing an important heat sink). The authors do a reasonable job of discussing this modelling error, but the justification for continuing with the climate change predictions is a bit confusing to me. It seems like the checks (by comparing the 2014 and 2015 summer periods) doesn't really get at the heart of the matter in that it seems to be checking whether the model gets the right answer, but doesn't care if it is for the right reason or not.

First, we want to clarify an important point: the problems with the year 2003 is only for the Alpine catchments. Indeed, Figure S3 shows that for the lowland Broye catchment the extremely warm year 2003 is very well simulated. We agree that the discussion should be clarified and we put additional emphasize on this in the revised version. Here, we provide a summary of the main points. First, by looking at all energy fluxes, we investigate the origin of this error. We show that the warming occurs in small upstream reaches and that the missing heat sink there is probably the absence of direct cold advection due to snow and glacier melt. In the future, the snow cover in summer is expected to greatly decrease or disappear, and the glacier melt will be reduced compared to nowadays, so this issue might be of lesser importance.

To ensure that the issue is not caused by another factor and that the overestimation will not grow in the future, we use the comparatively cold summer of 2014 (for which the model performs well) and the warm summer of 2015 (when the model overestimates the water temperature). Since the method used to downscale the climate change scenario keeps the interannual variability of past time series, the summers e.g., 2084 and 2085 will exhibit the same relative difference (but on a warmer baseline). We can then again compare the relatively cold and warm years to see whether the difference grows (suggesting that the overestimation would grow). The results of the comparison show that this is not the case.

In summary, we show strong evidence that the cause is the missing snow melt water (see also discussion below) and that the error does not increase in simulations of future scenarios. Since snowmelt will become less important in the future, we decide to go ahead with climate change simulations. For the sake of transparency, we provide an in-depth analysis (not present in some other published studies with similar issues), present arguments to support our statements, and in the conclusion mention that confirmation of these results by further studies would be beneficial.

P32 Section 5.3: I think this section and analysis can be removed. The physics-based modelling exercise already highlights the differences in discharge and stream temperature response to climate change for the alpine and plateau catchments. I'm not sure what the statistical analysis adds and the hypotheses being tested with these analyses are likely not what is intended (see Greenland et al 2016 for a discussion on this topic). In particular, the conclusion made on P32L16 that 'changes in discharge have no impact on water temperature change' is clearly wrong when considered from first principles (except for very unique cases that would not occur in reality).

We will shorten this section and move most of the content into the Supplementary. However, we still think that the statement on P32L16 is relevant. Discharge is a factor influencing water temperature, but it is not clear to which extent. In many models, the larger influencing factor is the air temperature, which is clearly the best explanatory variable for water temperature, while the discharge plays a more minor role (see e.g., Feigl et al., 2021).

Here, we do not discuss the direct link between discharge and water temperature, but the link between discharge trends and water temperature trends in a warming climate. In Michel et al. (2020) we show from historical data across Switzerland that observed changes in discharge are not related to observed changes in water temperature on the long term. On the short term, heat waves in central Europe are usually linked to very dry periods (Fischer et al., 2007), and assessing the interplay between low discharge and high-water temperature is thus like a chicken and egg problem.

What we show in this paper is that discharge trends have a weak to inexistant ability to explain future change in water temperature in the lowlands, while it is a more important factor in Alpine catchment. This is what we state in P32L16: “changes in discharge have no impact on water temperature change” when talking about the lowland catchments. Correlations and linear regressions do not mean causality, but to have causality we expect to see some correlation, especially for discharge where we would expect the impact to be linear (i.e., the link between mass, energy exchange and temperature would be linear). We find this result very interesting and worthwhile investigating. Indeed, not all energy fluxes will be influenced by discharge and these results raise interesting questions regarding the dominating energy fluxes governing water temperature. Again, this result is also supported by historical measurements (Michel et al. 2020). Section 5.3 is specifically devoted to this question of the link between the trends in the variable, rather to just the difference in the response to climate change between the lowland and Alpine catchments.

In summary, the complexity behind stream temperature leads to a situation where less water would not necessarily mean warmer water (at least not as the dominant factor), in contrast to what we could expect from simple thermodynamic laws. In order to clarify and remove any ambiguity, we suggest to reformulate P32L16 to: “This confirms the hypothesis that changes in discharge is weakly correlated with water temperature change in the lowlands regions”.

The question of the interplay between changes in discharge and in water temperature could be the core of a separate publication in the future with a focus only on this question, using historical measurements, i.e., an in-depth analysis of what we started in Michel et al., 2020, and the simulations results produced here.

P19L8-9: Missing relatively cold runoff inputs seems like a plausible reason for the model overpredictions (see my general point #2 above); however, would we expect the mechanism proposed in the previous paragraph (snow and glacier melt flowing over frozen or saturated soils) to be occurring during summer periods? Wouldn't it be more likely that HSPF is simply simulating warmer runoff temperatures than is actually occurring? Or maybe cold groundwater inputs (perhaps from a more regional source) are not being accounted for in the models? Or could not accounting for riparian vegetation shading be a factor here?

This question has partially been discussed above. The starting point of our reasoning was the fact that the summer temperature overestimation only occurs in Alpine catchments while the model gives very good results for the lowland catchments. This allowed us to already exclude some candidates (e.g., programing errors in the model source code). The groundwater hypothesis is discussed in Section 5.4 as one of the shortcomings of the model (by groundwater we mean here water below the water table, i.e., the saturated zone, not subsurface runoff). The fact that the problem with summer temperature overestimation occurs only in Alpine catchments, regardless of a known or unknown interaction with groundwater (e.g., the Landwasser, see Epting et al., 2021) and not on the lowland areas, again regardless of any groundwater interaction, shows that the groundwater estimation is not a plausible explanation to the problem of stream temperature overestimation during summer.

Riparian vegetation has already been discussed above. For alpine rivers, which are usually smaller than those of the lowlands, not considering the local vegetation may have a stronger impact on stream temperature than on the lowlands. However, for example for the very small catchment of the Rietholzbach on the lowlands, explicit consideration of riparian vegetation would also be important. Figures 4 and 5 show orthoimages of the Rietholzbach catchment and of the Alpine Lonza catchment, and Figure 6 shows the plot of the water temperature during the calibration and validation phases of the model.

From the orthoimages, we see that riparian vegetation is more present in the Rietholzbach catchment. However, Figure 6 shows that in this catchment the model does not produce the sudden water temperature overestimation simulated in the Lonza catchment. These results suggest that the lack of riparian vegetation shading is not responsible for the overestimation we have in alpine catchments. However, in the Rietholzbach catchment there is indeed a slight overestimation of the water temperature during summer and fall (but all over the period and not as peaky as in the Alpine catchments) which could be attributed to the lack of vegetation shading. As mentioned above, riparian vegetation will be discussed in Section 5.4 as one of the limitations of the model, and also in Section 4.4 as a factor probably contributing to the summer overestimation in Alpine region, but not as the main driver of the sudden peaks observed.

Topography and its treatment have also been already discussed. For completeness, we show in Figure 7 the same orthoimage for the Lonza catchment as in Figure 5 but adding also the DEM used in Alpine3D to compute the shading. Despite the coarse resolution of the DEM, we see from the figure that the topography is still reasonably well represented to allow for a correct shading on the main reaches of the river (for the large-scale topography).

In order to push further the question of the radiation impact (and the potential errors arising from topographic and riparian vegetation shading), we made a new run of the model for the Landwasser catchment with the artificial situation of zero incoming shortwave radiation. The results are shown in Figure 8. Note that StreamFlow has not been re-calibrated for this run, but that it uses the same calibration parameters as the other runs performed for the Landwasser catchment. The goal here is to only assess the model sensitivity to radiation and to see whether the high-water temperature peaks observed in summer disappear when incoming solar radiation is removed.

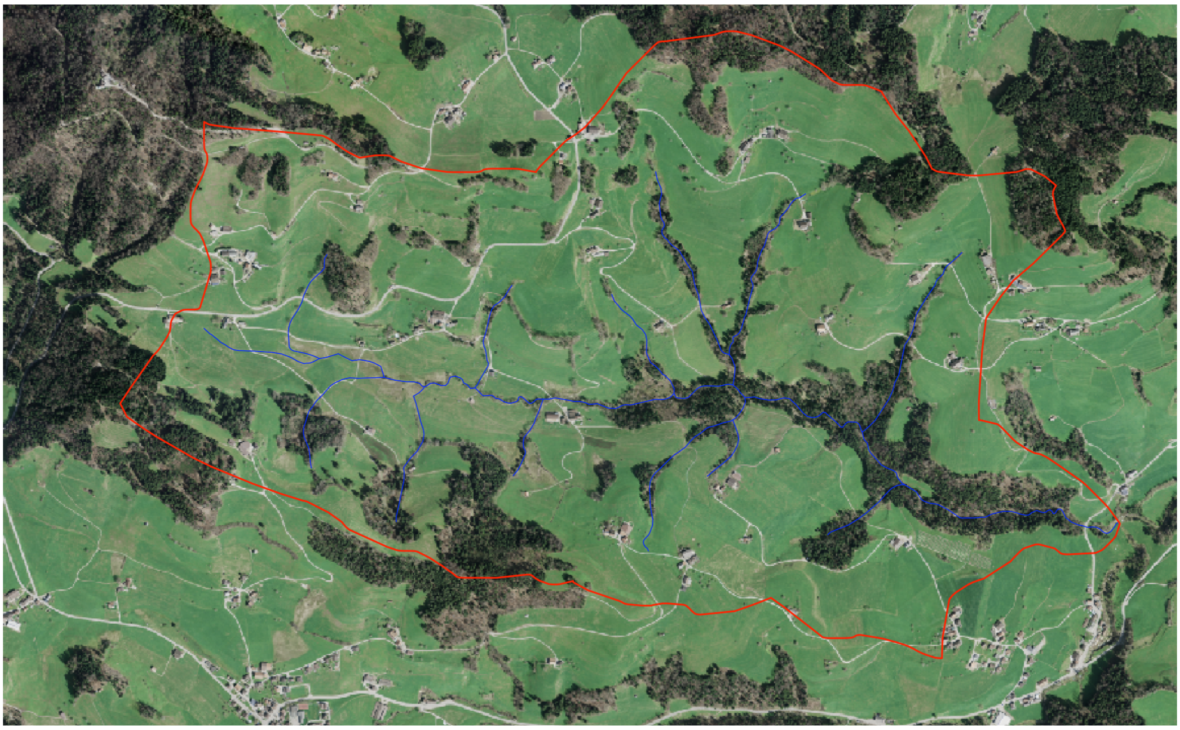
Figure 4 – Rietholzbach catchment (red) and river course (blue). Source: Swisstopo

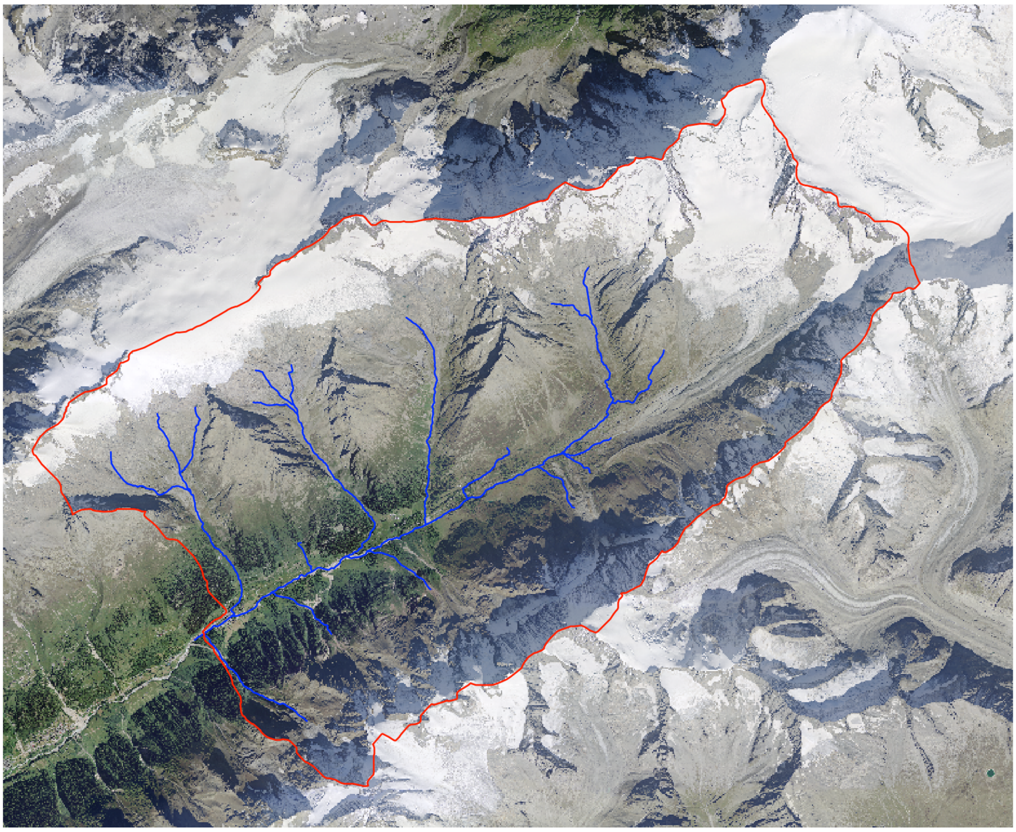
Figure 5 – Lonza catchment (red) and river course (blue). Source: Swisstopo

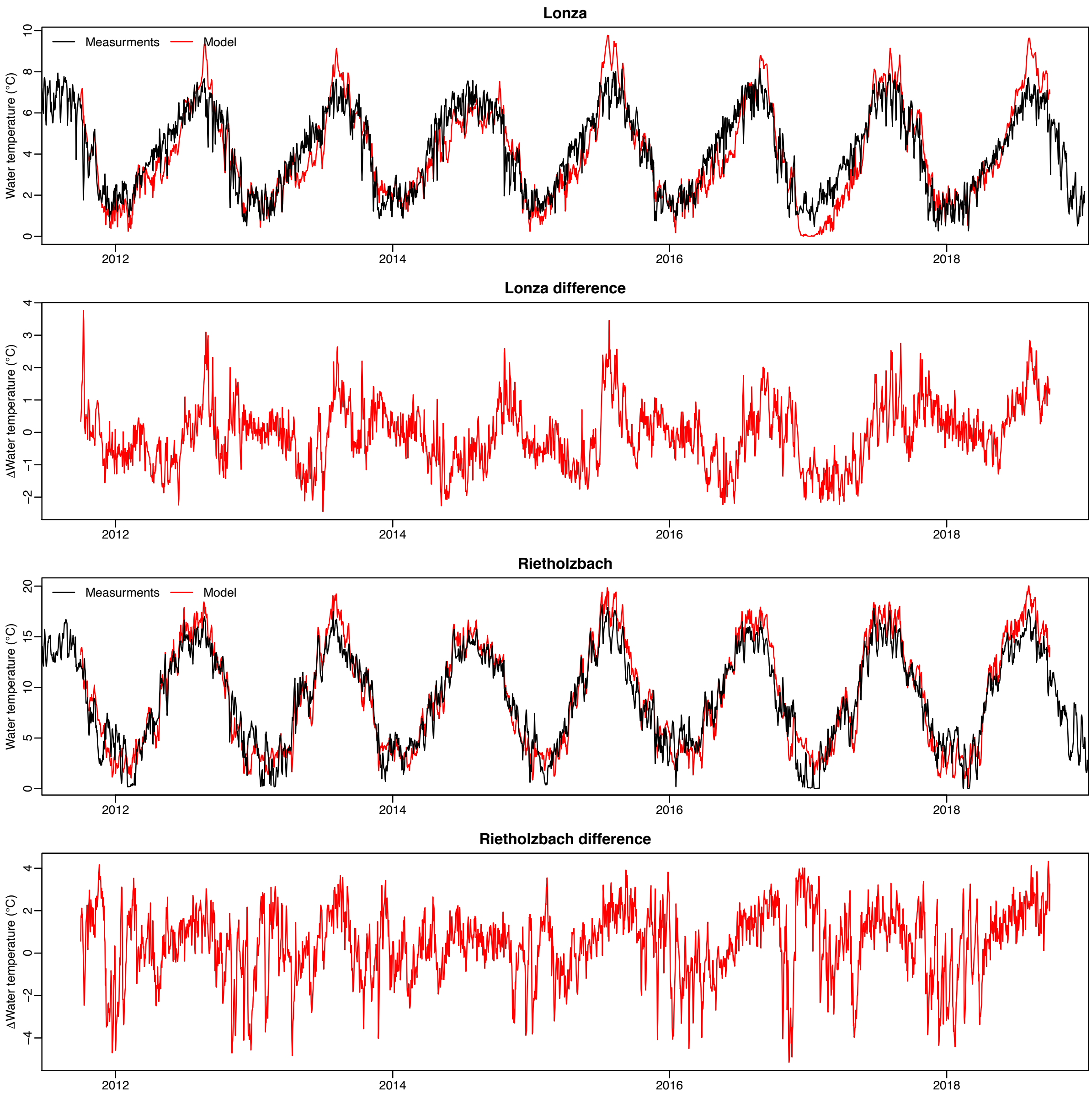

Figure 6 – Top two panels: Measured (black) and simulated (red) water temperature for the Alpine Lonza catchment (first plot) and difference between the measured and simulated water temperature (second plot). Bottom two panels: Same as top, but for the lowland Rietholzbach catchment.

As expected, there is a general bias toward colder temperature when removing solar radiation. A first look at Figure 8 could suggest that the summer overshoot might be corrected by removing radiation, but when considering the summer 2014 we see that we have now a negative bias in relatively cold summers. The second panel of Figure 8 is more informative. Here we see that most of the time the model error in the summer (black line) is not at all correlated with the temperature difference we see between model runs with and without radiation (red line). This suggests again that radiation issues are not the main driver of this error.

In summary, we agree that uncertainties related to solar radiation should be more clearly mentioned and we will adapt the manuscript accordingly by summarizing the discussion above in the Supplementary and referring to it in Sections 3.2, 3.3 and 4.1.

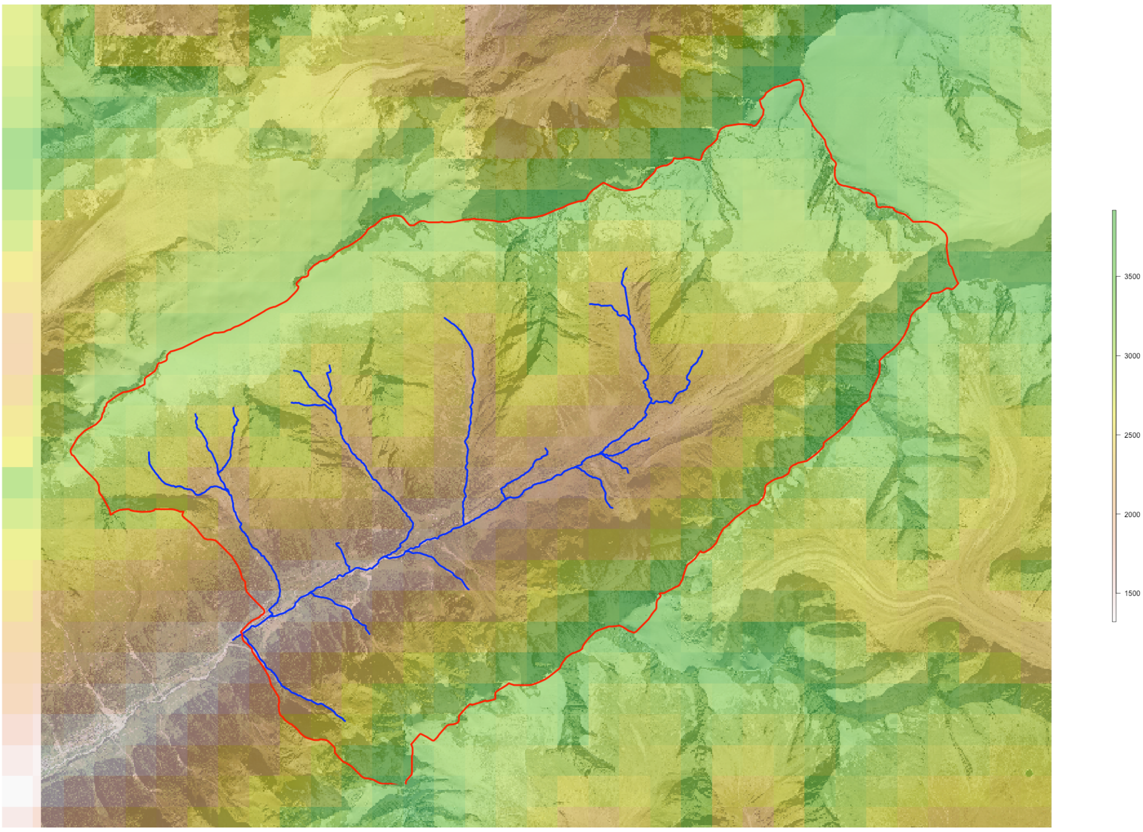

Figure 7 – Lonza catchment (red) and river course (blue) with DEM added on top. Source: Swisstopo

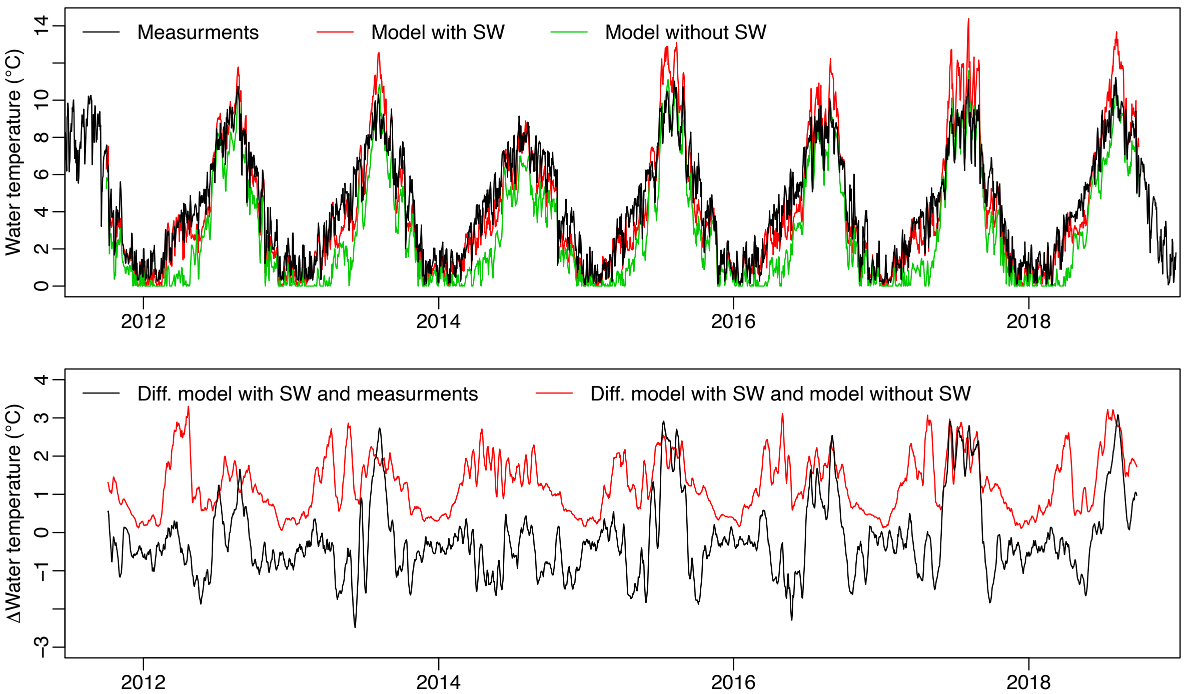

Figure 8 – Top: Measured (black) and simulated water temperature (red with shortwave radiation, green without shortwave radiation) for the Alpine Lonza catchment. Bottom: Difference between measured and modeled (with shortwave radiation) water temperature (black), and difference between water temperature modeled with and without shortwave radiation (red).

Finally, the hypothesis suggested by the reviewer of the HSPF approach being the cause of the summer overestimation of the water temperature in Alpine catchments remains. Expressed in other words, this is already what we state in the paper. Indeed, when considering that the main explanation of the problem is the missing cold-water advection, and knowing that this is not considered in the HSPF approach, we realize that this could be an explanation for (a part of) the problem. To show this, the HSPF solver has been rewritten in R outside of the model and run over the Landwasser catchment using the mean air temperature over the catchment (computed from the grid output of Alpine3D). The results are shown in Figure 9. We clearly see that the peaks of high-water temperature in the river in summer correspond to peaks in the temperature of the HSPF outflow. This figure will be added in the Supplementary Material in Section S7 where we present the energy fluxes.

The issue of high-water temperature is extensively discussed in this review response. We thank the reviewer for the various hypotheses suggested. We performed an additional analysis for better identifying this issue for the purpose of this reply, completing the already extensive description in the paper. The new hypothesis tested here brings us however to the same conclusion as in the paper: the issue arises from not accounting for some cooling processes in the headwater regions for the water in the soil reservoirs. Snow and glacier melt (likely also melting permafrost) are the main candidates after having excluded many other hypotheses. This problem is well recognized in the paper and discussed in a transparent manner. We also provide some evidence of the robustness of the climate change simulations performed despite this error. It is shown in our work that the development of a new, more accurate scheme for dealing with water temperature in the soil reservoirs is important for Alpine regions, but this is beyond the scope of this paper.

We agree that several clarifications are needed in the manuscript; most of them have been suggested above and will be implemented in the manuscript as described.

Table 1: I recommend including some metrics of dominant land cover in this table (e.g., %forest, %agriculture, %urban, %lake, %rock/meadow).

Thank you for the suggestion, we'll add it.

Figure 1: Perhaps distinguish between 'lowland' and 'alpine' catchments using colour?

Thanks for the suggestion, we'll use two different colors and re-arrange the numbering to be able to split the catchment list on the right between lowland and Alpine catchments.

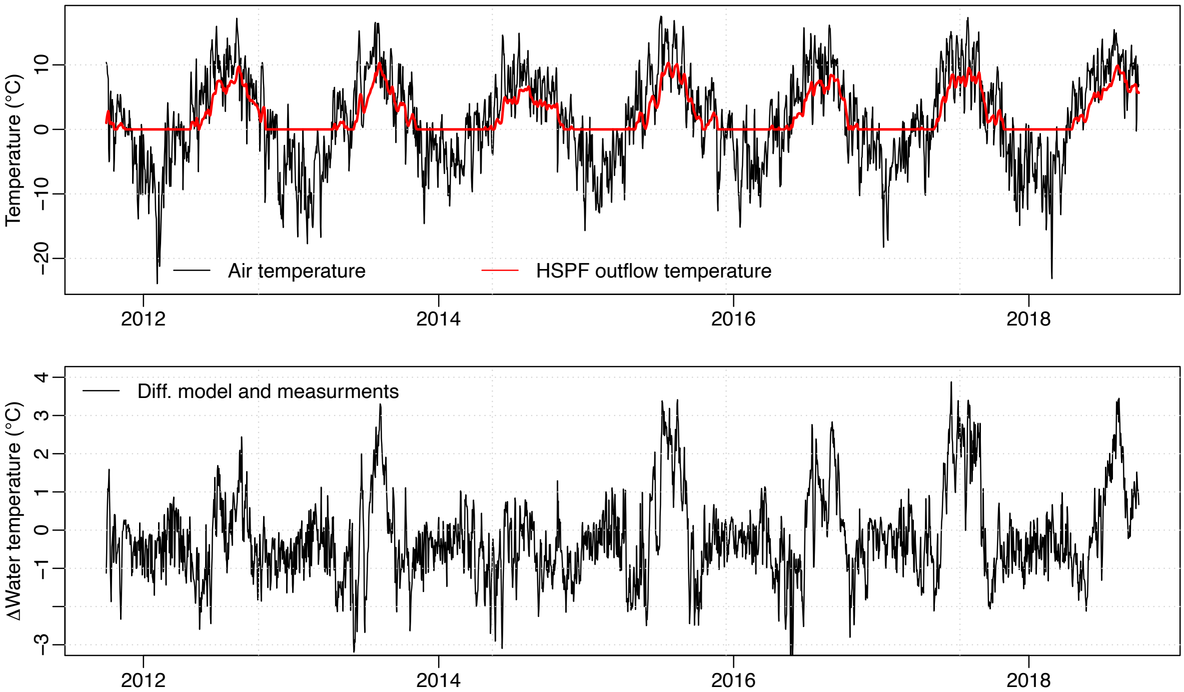

Figure 9 – Top: mean air temperature over the Landwasser catchment and outflow water temperature from the HSPF scheme when forced with the mean air temperature (and the parameters calibrated over the Lanswasser catchment presented in Table S8 of the Supplementary). Bottom: difference between simulated and measured water temperature for the Landwasser catchment.

References

Brauchli, T., Trujillo, E., Huwald, H., & Lehning, M. (2017). Influence of Slope-Scale Snowmelt on Catchment Response Simulated With the Alpine3D Model. Water Resources Research, 53, 10723-10739.

Du, X., Shrestha, N. K., & Wang J. (2019). Assessing climate change impacts on stream temperature in the Athabasca River Basin using SWAT equilibrium temperature model and its potential impacts on stream ecosystem. Science of The Total Environment, 650(2), 1872-1881.

Epting, J., Michel, A., Affolter, A., & Huggenberger, P. (2021). Climate change effects on groundwater recharge and temperatures in Swiss alluvial aquifers. Journal of Hydrology X, 11, 100’071.

Feigl, M., Lebiedzinski, K., Herrnegger, M., & Schulz, K. (2021). Machine-learning methods for stream water temperature prediction. Hydrology and Earth System Sciences, 25, 2951-2977.

Fischer, E. M., Seneviratne, S. I., Lüthi, D., & Schär, C. (2007). Contribution of land-atmosphere coupling to recent European summer heat waves, Geophysical Research Letter, 34, L06707.

Gallice, A., Bavay, M., Brauchli, T., Comola, F., Lehning, M., & Huwald, H. (2016). StreamFlow 1.0: an extension to the spatially distributed snow model Alpine3D for hydrological modelling and deterministic stream temperature prediction. Geoscientific Model Development, 9, 4491-4519.

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. European journal of epidemiology, 31(4), 337-350.

Haag, I. & Luce, A. (2008). The integrated water balance and water temperature model LARSIM-WT, Hydrological Processes, 22, 1046-1056.

Hannah, D. M., Malcolm, I. A., Soulsby, C., & Youngson, A. F. (2004). Heat exchanges and temperatures within a salmon spawning stream in the Cairngorms, Scotland: seasonal and sub-seasonal dynamics, River Research and Applications, 20, 635-652.

Leach, J. A., & Moore, R. D. (2015). Observations and modeling of hillslope throughflow temperatures in a coastal forested catchment. Water Resources Research, 51(5), 3770-3795.

Leach, J. A., & Moore, R. D. (2019). Empirical stream thermal sensitivities may underestimate stream temperature response to climate warming. Water Resources Research, 55(7), 5453-5467.

Magnusson, J., Jonas, T., and Kirchner, J. W. (2012). Temperature dynamics of a proglacial stream: Identifying dominant energy balance components and inferring spatially integrated hydraulic geometry, Water Resources Research, 48, W06510.

Michel, A., Brauchli, T., Lehning, M., Schaefli, B., & Huwald, H. (2020). Stream temperature and discharge evolution in Switzerland over the last 50 years: annual and seasonal behaviour. Hydrology and Earth System Sciences, 24, 115-142.

Michel, A., Sharma, V., Lehning, M., Huwald, H. (2021). Climate change scenarios at hourly time-step over Switzerland from an enhanced temporal downscaling approach. International Journal of Climatology, 41, 3503-3522.

Morrison, J., Quick M., & Foreman, M. (2002). Climate change in the Fraser River watershed: flow and temperature projections. Journal of Hydrology, 263(1-4), 230-244.

Null, S. E., Viers, J. H., Deas, M. L., Tanaka, S. K., & Mount, J. F. (2013). Stream temperature sensitivity to climate warming in California's Sierra Nevada: impacts to coldwater habitat. Climatic Change, 116(1), 149-170.

Perrin, C., Michel, C., & Andréassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. Journal of Hydrology, 279(1-4), 275-289.

Yan, H., Sun, N., Fullerton, A., & Baerwalde, M. (2021). Greater vulnerability of snowmelt-fed river thermal regimes to a warming climate. Environmental Research Letters, 16(5), 054006.

van Vliet, Franssen, W., Yearsley, J., Ludwig, F., Haddeland, I., Lettenmaier, D., & Kabat, P. (2013). Global river discharge and water temperature under climate change. Global Environmental Change, 23(2), 450-464.