Author response to reviewer comments

I. Anonymous Referee #1

Received and published: 9 October 2015

In the manuscript “Semi-automated calibration method for modelling of mountain permafrost evolution”, Marmy et al. present simulations of the future ground thermal regime at six instrumented sites in the Swiss Alps. In its scope and effort, this work is virtually unparalleled and deserves publication in The Cryosphere. However, I am not at all convinced that this work can re-define the state-of-the-art for such studies, and I recommend major revisions before publication. From the material presented, it does not become clear to me that the method can increase the confidence in future predictions compared to much simpler methods.

I.1 Major Comment:

In a certain way, the authors treat the COUP model as a “black box” for which the calibration procedure produces an optimal set of parameters. However, at least in some cases, these parameter sets are not really physical realizations, i.e. some parameters would most likely not be confirmed if independent measurements (e.g. of surface or ground thermal properties) were available. In previous studies dedicated to future projections of the ground thermal state, the authors have chosen a sufficiently simple model (e.g. based only on heat conduction), estimated the parameters according to field knowledge/physical constraints and then compared the results to measurements e.g. in boreholes for validation. While the match with measured data is in general not as good as in this study, it will generate the right results for the right reasons, or at least the limitations will become more obvious. In particular for future simulations, which cannot be validated, the “black box calibration” approach chosen for this study has the potential to produce artifacts in the future simulation. The authors should therefore make the link between the fitted parameters and observable/observed processes much clearer (wherever this link exists). If a parameter set is clearly unphysical (see below for more specific comments), I suggest not to show the future simulations since artifacts are highly likely.

Response to the reviewer: we are thankful to the reviewer for this critical comment, which made us re-formulate our rationale behind the chosen approach in a more explicit way. As indicated in the reviewer’s comment, this semi-automated calibration approach targets the challenge of permafrost modelling at sites, where no independent measurements of ground thermal properties, porosity, snow parameters etc exist. This is in fact usually (almost always) the case, even for monitoring sites with a high number of observed parameters (such as Schilthorn or Murtèl rock glacier, which were simulated in the preceding paper by Scherler et al. 2013). Also in this case, ground thermal properties had to be assumed and were only calibrated against borehole temperatures such as in this paper.

The differences in the approaches lie in the way this calibration is performed: (1) manually, as in Scherler et al. (2013), or (2) automatically, as in the present study. Whereas in the former case, the user can manually make sure that no unphysical or implausible values are used for the unknown soil and snow parameters, this has to be assured by plausible parameter ranges in the latter procedure. However, as the reviewer correctly pointed out, a bad combination of (physically plausible) values could still lead to a correct calibration but for the wrong reasons, which might yield unrealistic projections for the future. If for example a too high thermal conductivity of the snow cover in winter is balanced by an unrealistic soil albedo in summer, this combination of parameters may lead to good calibration results during a 10-year calibration phase, but it will be wrong as soon as the snow cover will diminish in a future (warmer) climate, as the soil albedo will not change in a compensating way.

However, to avoid that, we made sure in our original manuscript that (1) all parameter used for semi-automated calibration stayed in a physically plausible range; and (2) that we identify those parameter with a large influence on calibration (Fig. 5) in order to spot potential combinations of unrealistic parameters, which may lead to implausible projections.

Author changes in revised manuscript: In our revised manuscript we tried to explain this rationale in a much clearer way in order to point the reader to the potential disadvantages of this calibration approach (e.g. in new paragraphs in section 7.1, 7.3 and 5.1). We also followed the suggestion of the reviewer and make the explicit link between the obtained parameter values during calibration and real observed processes (see section 5.1); we also enlarged our analysis of the importance of the various parameters for calibration to determine if there are sites, where implausible combinations of parameter values might lead to erroneous projections for the future. We identified one site (Ritigraben) where this is most probably the case, and consequently removed the long-term simulation of this site from the paper (revised Figures 7 and 8). For two other sites, where (a) missing processes (Murtèl-Corvatsch) or (b) missing input data (Muot da Barba Peider) influence the long-term results, we enlarged the discussion within chapters 5-7, and included additional Figures (new Figure 11 and supplementary material).

Finally, we tried to be more clear and detailed throughout the manuscript with respect to the above mentioned points by the reviewer and hope that the manuscript improved significantly. In the following we will address each minor comment of the reviewer in detail:

I.2 Minor Points:

-p.4788, l. 27: Explain what is meant by "GCM-RCM chain" and "ENSEMBLES data set".

Response and changes in manuscript: changed to *"…from the different data sets produced by coupled simulations of Global and Regional Climate Models (GCM-RCM) within the EU-ENSEMBLES project."*

-p. 4789, l. 2: I don't think that "Langer et al. (2013)" is not a good reference in this context. It would be much more appropriate in l. 20ff.

Response and changes in manuscript: *changed accordingly* (*Langer et al. 2013 in the second paragraph of the Introduction, and Romanovsky et al. 2010 in the first part instead of the original Langer reference*)

-p. 4789, l. 20 ff: The classification in 1D-2D- and 3D models does not follow strict and logical criteria, or at least it does not become clear which variable or process are 1D, 2D or 3D. To me, the mentioned 1D and 3D approaches have a lot in common, since they explicitly account for energy exchange processes in a more or less physically-based way. The mentioned 2D approaches, however, are more (semi-)empirical schemes which are aimed at estimating averages of the target variables of the 1D/3D schemes in a simplified way (except Hartikainen et al., which stands out in that it focuses on much longer timescales than the other studies). In addition, there is the class of spatially distributed 1D-models, sometimes referred to as 2.5D. Examples are the later mentioned Westermann et al. (2013), but also Jafarov et al. (2012) and Zhang et al. (2012).

Response and changes in manuscript: *that is correct, thanks for this clarifying comment (see also reviewer 2). We rewrote the entire paragraph and included the mentioned and additional new references.*

p. 4792, l. 1: the statement “potential scenarios of possible. . .” contains some redundancy. “scenarios of. . .” is enough, in my opinion.

Response and changes in manuscript: *Changed accordingly*

p.4795, l. 19: Wicky (2015) refers to a master thesis (in German) which has not undergone the normal review process. While this is generally problematic, the statement seems to be sufficiently backed up by another reference. I would therefore leave it up to the authors to decide whether to remove this reference or not.

Response and changes in manuscript: *We deleted the reference accordingly and rewrote sentence as follows: “These 2-dimensional (or potentially 3-dimensional) processes cannot be explicitly simulated with the COUP-model, however, their effect on the thermal regime has been indirectly confirmed by specific 1-d distributed COUP simulations at this site (Staub et al. 2015).”*

p. 4798, l. 10, p. 4800, l. 14: it is not directly clear to me why wind speeds play only a minor role in the modeling. From my experience, there are some cases (e.g. high global radiation, but cold air temperatures), where wind speeds have a pronounced effect on active layer thickness and ground temperatures using similar model approaches. It is quite possible that this is not the case for the investigated sites, but it should become clear whether this was checked, and to what extent the role is “minor”.

Response and changes in manuscript (see also comment by reviewer 2): *We apologise for the misunderstanding! The reviewers are correct in pointing out that wind speed may have a pronounced effect, what we wanted to say is that the wind speed scenarios of the different GCM-RCM chains are very similar so that the choice of the specific chain does not play a major role in this case. The sentence: “As the wind plays only a minor role in the long-term trend of soil, it was satisfactory to use the median series of the seven other chains” was accordingly changed to: “As the wind speed scenarios of all available GCM-RCM chains are very similar, we consider it acceptable to use the median of these as a substitute for the seven chains with missing wind speed scenarios”.*

p. 4799, l. 2: What is meant by “virtually all”?

Response and changes in manuscript: *we changed the sentence to: “While MAAT is predominantly negative in present-day climate, all six sites are subject to a significant increase in temperature and the majority of climate models indicate at four of the six sites positive mean annual temperatures by the end of the 21st century”*

p. 4800, l. 8ff: This is obviously a huge limitation for MBP, and the effect on the results is not clear at all. If there is a strong bias in the global radiation forcing the model, the optimization procedure would tend to correct this by adjusting parameters in a potentially unphysical way, making the simulated future ground temperatures more or less useless. How well can the model estimate global radiation based on latitude/air temperature? Has this been checked for the other sites, where global radiation was available?

Response: *We are thankful for this comment – and the reviewer is of course right, that this absence of radiation data is a huge limitation for the modelling of MBP. We checked the model estimates for the example of COR, which is the nearest station to MBP, and added this comparison as supplementary data. The results show an overestimation of global radiation values by the CoupModel, which may lead (in this case) to maximum surface temperature biases of up to 10°C in summer, which would of course be a serious overestimation. We further checked how the model is potentially compensating this overestimation (as we have no measured data we cannot be sure that the bias is similar for MBP) and found a very low value for the critical snow height parameter ΔS_{crit} which decouples the soil already for very low snow height parameter and could be a sign for model compensation as it affects the albedo calculation. The albedo values for dry and wet soil, however, were calibrated with average values (α_{dry} = 24.1%, α_{wet} = 19.1%), which rather points to no large radiation bias. Finally, the calibrated thermal conductivity values for the near-surface layer were about average as well (around 2-4 W/m*K) and do not indicate a large bias towards too warm radiation based surface temperatures.*
In conclusion, we see no real evidence for a large problematic compensation of a potential radiation bias, and therefore think that it will not induce a comparatively large uncertainty for future scenarios at MBP. If yes, then it would be a slight underestimation of the estimated warming.

Author changes in revised manuscript: The text in section 3.1.2 was modified accordingly: *"Global radiation for MBP has therefore been estimated by CoupModel based on potential global radiation (depending on latitude and declination) and atmospheric turbidity (Jansson 2012). Independent comparison between measured, reconstructed and CoupModel estimated global radiation values for COR showed an overestimation of global radiation by the CoupModel leading to near-surface maximum temperature biases of up to 10°C in summer (cf. supplementary material). However, the calibration technique applied (see sections 4 and 5) would compensate potential biases in the temperature simulations by adjusting related parameters in the model, e.g. snow cover parameters or the albedo. Corresponding uncertainties arising from a potential compensation in the MBP results will be further discussed below."*

In section 5.1. a corresponding paragraph was added: *"As mentioned above, a potential radiation bias could be present in the case of MBP due to the absence of on-site measured global radiation. A compensation of a potential bias would be expected either in the near-surface thermal conductivities or in the albedo values. Although the critical snow height parameter ΔS_{crit} for MBP is very low which could be a sign for model compensation as it affects the albedo calculation, the albedo values themselves were calibrated with average values (α_{dry} = 24.1%, α_{wet} = 19.1%), which rather points to no large radiation bias. Similarly, the calibrated thermal conductivity values for the near-surface layer are about average (around 2-4 W/m*K) compared to the other sites and do not indicate a large bias towards too warm radiation based surface temperatures."*

p. 4801, l. 16: at what depth is the lower boundary? Is it below the depth of zero annual amplitude? In this case, the amplitude should be negligibly small. Does this treatment take the geothermal gradient into account at all, i.e. that ground temperatures become warmer at depth? In this case, how can ground temperatures at depths of up to 80m be modeled?? How is the lower boundary condition for the future runs?

Response and changes in manuscript: *The depth of the lower boundary is different for the various sites (due to different maximum borehole depths) but at least 30 meter and therefore always well below the depth of zero annual amplitude. Therefore, the prescribed heat flux at the lower boundary condition is indeed negligible. However, this enables comparatively stable conditions at the lower boundary and accounts for the often isothermal conditions found in Alpine permafrost at this depth. Of course a potential change in ground temperatures at this depth cannot be simulated and are not interpreted within this study. It is only used as lower boundary condition. See also our response to the comment regarding the spin-up time. A corresponding paragraph was added to section 3.3 and the depth of the lower boundary for each site was added in Figure 3.*

p. 4801: I don't understand how Eqs. 1 and 2 are related. There are four fluxes, qh, qv, qh and qin, are they related somehow? In Eq. 2, the units don't match, the first term has the unit K/m, not J/m2d.

Response and changes in manuscript: *We apologise for the mistake in Equation 2, partly because of two different nomenclatures – q_h is the surface heat flux, and q_v and q_w are the vapour and water fluxes (in general in Eq. 1, and at the surface (z=0) in Eq. 2). Infiltration (as mentioned in the original manuscript) would be a different expression for q_w. We corrected this mistake.*

p. 4802, l. 7ff: This treatment seems to be mainly focused on reproducing spatial averages of the ground thermal regime. Is this appropriate for reproducing temperatures close to the surface, which would either "see" snow or now snow, not patchy snow conditions?

Response and changes in manuscript: *The critical snow depth parameter is in our case less focused on the spatial average but more on the surface conditions, i.e. the roughness. It is one of the main tuning parameter concerning snow cover conditions and their temporal variation. In theory it should reflect amongst others the roughness of the surface, e.g. if it consists of large blocks of 1-2m height (e.g. rock glacier Murtèl) 10cm snow cover has a different thermal insulation effect as at Schilthorn, where the surface is covered by sandy material. Corresponding explanatory sentences were added to this part.*

p. 4802, l. 11: What is Eq. 9?

Response and changes in manuscript: *This equation is included in Table 2. We changed the reference to "see Table 2".*

p. 4802, l. 25: I can't believe that this procedure will create anything realistic for the "deep" ground temperatures. At a depth of almost 80 m, like in Stockhorn, the ground temperatures should still be influenced (if not completely determined) by times before 2000, and even 1981. The modeled ground temperatures at depth would then only reflect unphysical steady-state conditions for the applied forcing data. If this is the case, the deep ground temperatures should be completely removed from the optimization routine and analysis, and the limitations on the future simulations resulting hereof should be mentioned.

<u>Response</u>: *This is of course true, but as mentioned above we are not aiming at simulating deep ground temperatures or very long time-scales. We consider these depths only as lower boundary condition for our near-surface simulations, and therefore include them in the calibration procedure to (semi-automatically) produce data-consistent conditions. Tests with changed spin-up procedures or lower boundary conditions did not show significant changes for the results of the time-scales considered in this study (similar results were obtained by Ekici et al. 2015 and Scherler et al. 2013).*

<u>Author changes in revised manuscript:</u> *We added the following sentences to this section : "Model tests with longer spin-up times showed only negligible differences with respect to the procedure described above. However, this approach clearly neglects all long-term effects of past climatic conditions on the ground thermal regime at larger depths. Therefore, simulation results at larger depths should not be interpreted in a climate context."*

p. 4803, l. 26ff: This is unproblematic if it is done for periods when observations are available. It is extremely problematic as a basis for future simulations, as it is done here. It is not clear at all, if the parameter set is also an optimal one for the future forcing, or if it creates a fake model reality. See also major comment.

<u>Response and changes in manuscript</u>: *See our response to the major comment. We changed the respective paragraph as follows: "In addition, a model set-up which is consistent with present day conditions may not be optimal for future climatic conditions. This well-known problem is inherent to most long-term transient simulations with a high number of parameterised and calibrated processes. One possibility to avoid compensation of two or several parameters, which all show unphysical or unrealistic values is to (1) constrain the parameter range to physical plausible values and (2) verify if the obtained calibration values for all parameters contain any outliers, which cannot be explained by site-specific conditions. However, it has to be kept in mind that the aim of the calibration procedure is not to get a physical determination of the parameter values themselves, but to get a model that is thermally most representative of for the ground thermal regime at a given site. Keeping the above constraints in mind, for long-term simulations, where no observations are available, it has to be assumed that the parameters governing the ground thermal regime do not change significantly over the duration of the simulation."*

p. 4804. L. 4: How about the parameters that do have on-site measured values. Is the uncertainty/spatial variability/changes over time taken into account? This could play a major role for parameters to which the model is highly sensitive.

<u>Response and changes in manuscript</u>: *we apologise for the incorrect wording and the misunderstanding – usually, and also in our case, no subsurface data except ground temperature is available for high mountain permafrost site. Even if one would analyse the rock type in detail, e.g. by collecting of rock samples from the surface, the physical properties of the subsurface could differ strongly, due to the strong weathering, fracturing and heterogeneity at most sites. Because of this also porosity distribution with depth is not available for the sites. Similarly, no in-situ data of snow characteristics are available (thermal conductivity, density, etc). We rephrased the sentence by omitting the term "no on-site measured values" which was misleading in the original version of the manuscript.*

p. 4804: In the fitting routine, only ground temperatures are used to determine the model performance. This could be problematic as the freezing point of water is not exceptional in this procedure. It is the same if the model is wrong by 0.2 degree C at +10 degrees or at -0.1degree C. For the latter, the differences in ground ice and thus the energy content of the ground would be substantial, with pronounced impacts on the future simulations. This would in particular influence deeper ground temperatures. The effect could potentially be moderated by using the energy content of the ground (e.g. as in Jafarov et al., 2012, calculated with the freeze curve also assumed in the model) instead of temperature. The authors should at least comment and discuss this limitation. Note that this comment does not refer to using additional measured parameters, such as water contents.

<u>Response and changes in manuscript</u>: *we thank the reviewer for this important comment, which we are fully aware of – as it complicated the calibration routine for us in the first place. This problem was also the reason that we used a joint bias (mean error, ME) and correlation (r2) based statistical approach to evaluate the calibration performance. While minimising the ME ensures that the absolute values are near the observed ones, the correct simulation of the timing of freeze/thaw events can be improved by maximising r2. It is clear that in the case of long freeze/thaw events a good correlation will always be difficult to achieve, but we found a reasonable good match at least for the near-surface layers by optimising the correlation. In a future step, other quantities such as the energy content of the ground as cited by the reviewer, but also the water content (see our Fig. 9) can be used to enhance the calibration. Regarding its influence on deeper temperatures the reviewer is of course right. However, as mentioned above, the temporal change of deeper ground temperatures near the lower boundary is not subject of the present study. A corresponding discussion was added in section 5.1.*

p. 4805, l. 7: see comment above on initialization/deep ground temperatures.

<u>Response and changes in manuscript</u>: *see our response to the respective comment above: the deep temperatures are only used to get the lower boundary right for the different sites. In addition, their influence on the overall calibration procedure is small compared to the near-surface levels (cf Fig. 5). On the other hand, the benefit of including them into the calibration routine is not only to improve the lower boundary, but also to identify possible differences between the sites regarding the sensitivity of model processes/parameters a these model levels. Therefore, we propose to keep these levels within the calibration routine, but omit their discussion and any interpretation about the temporal change of deep temperatures in the manuscript. We modified/added the following sentence to the final part of section 5.1: "When considering the absolute importance (% in Figure 5, left), we notice that deep boreholes (COR, RIT and STO) have low percentages, which is not surprising as the temperatures at those depths vary on much longer time-scales and depend primarily on the structural set-up of the model. As their future evolution is influenced by past climates, which are not included in the present study, simulated temperature changes at large depths will not be discussed within this study. However, their correct representation for present day climate is important as lower boundary condition for shallower levels."*

p. 4808, l. 5: I suggest leaving the deep temperatures out, see above.

Response and changes in manuscript: *see our response above; we included a clarification regarding the deep temperature calibration as proposed, and changed the text of the deep temperatures accordingly.*

p. 4809, l. 14ff: Which criteria is the statement on equifinality based upon?

Response and changes in manuscript: *In his PhD thesis the main author A. Marmy (2015) addressed the problem of equifinality, i.e. the fact that several model set-ups may result in different but equally well-calibrated models, which may consequently lead to different projections into the future. Sensitivity tests showed that in the context of our study (and using the parameter and parameter ranges indicated), this effect seems to play only a minor role, as model set-ups with similar calibration performances lead also to similar simulation results. However, as a thorough discussion of this topic would be beyond the scope of the present paper, we propose to omit this paragraph in the revised version.*

p. 4817, l. 20: So what's the value of the simulations in these cases then? The authors clearly state that the model cannot represent site-specific conditions and the associated ground thermal regime, so simulations of the future ground thermal regime could feature a strong bias. Wouldn't it be better to clearly state that the procedure is inappropriate for such conditions, and that it is not meaningful to conduct future simulations with the scheme in this case?

Response and changes in manuscript: *We thank the reviewer for this comment and apologise for the apparent misunderstanding: With saying that the structure was not manually adapted we meant that we did not hard-code a specific porosity structure or specific additional heat sources/sinks into the model, as was being done in the paper by Scherler et al. (2013). On the contrary, we gave realistic ranges of parameters (especially porosity) and let the model calibration find the optimal values – and by this giving an indication of a physically-based estimate of the parameters which we have to guess anyway (also in Scherler et al. 2013 real observed values were not available and values were obtained by "educated guesses"). If an automatic calibration leads to a mismatch in temperature or in unrealistic parameter combinations at a certain depth, we can deduce in a more scientific way that an additional heat sink would be necessary to get realistic results and, therefore, that a process is really missing. Because of this, we think that the present approach is more objective, than starting with a (potentially wrong) hypothesis already during model set-up. We tried to explain this concept and the difference to the Scherler et al. paper more clearly in the revised version and rephrased the corresponding paragraph. For the specific case of rock glacier Murtel you are right in saying that the original long-term simulation of Scherler et al. 2013 is probably more correct, even though we do not know whether this process will be unchanged during the next hundred years. Because of this we included a comparison of our and the Scherler et al. 2013 simulation results for rock glacier Murtèl in the revised version (Figure 11).*

p. 4813, l. 17ff: It would be nice if the authors could directly provide some statements on how these two points affect the outcome of this study.

Response and changes in manuscript: *We added a corresponding paragraph to the discussion, as mentioned by the reviewer: "The calibration with GLUE depends on several subjective initial assumptions: a) choice of tested parameters and their range:*

this choice has to be made by the modeler prior to the calibration and is a result of previous tests to identify relevant and sensitive parameters and, b) the choice of criteria of acceptance. For the former, we tried to include a representative set of parameters for surface processes (snow, albedo, evaporation), subsurface processes (thermal and hydraulic conductivity) and properties which are characteristic for the specific geomorphological sites (porosity) in order to provide enough degrees of freedom for a satisfactory calibration. In addition we used our prior experience with CoupModel (cf. Engelhardt et al. 2010, Scherler et al. 2010, 2013, 2014, Marmy et al. 2014, Staub et al. 2015) to identify the most sensitive parameters. We tried to fix the allowed parameter range to physically plausible ranges and verified that the obtained values during calibration were not distributed at the limits of these ranges. Regarding the choice of criteria of acceptance, we gave priority to good correlation coefficients near the surface and at intermediate levels while making sure that absolute values (via the RMSE) were acceptable at all depths. Whereas a different set of calibration parameters would not change Here, different simulation results would have been obtained by e.g. giving more weight to intermediate levels, however, due to the uncertainties regarding the influence of past climates at the lower boundary and regarding the exact representation of temperature evolution near the freezing point, the results would be less certain than in the case of a well-calibrated model at the upper boundary."

p. 4815, l. 28: remove "provide again . . . "?

<u>Response and changes in manuscript</u>: *Sorry for this editing error! The missing link was inserted.*

p. 4816, l. 8: How is the partitioning between transpiration and evaporation controlled in COUP? Are there really plants on Stockhorn, and is it realistic to assume that the latent heat flux is strongly controlled by transpiration rather than evaporation? If no, this would be a good example where the effects of incomplete or even flawed model physics, input data and/or other biased model parameters is compensated by tuning the model in an unphysical way. In my opinion, this does not result in a model that can describe reality in a better way (although it can describe the training data sets).

<u>Response and changes in manuscript</u>: *We are thankful to the reviewer for spotting this error. Indeed no vegetation is present at all our modelled sites, and the corresponding vegetation module in CoupModel was switched off for all simulations. The wilting point parameterisation mentioned in our old Figure 9 (and corresponding text) is used in Coup as a soil physical property and is part of the water retention parametrisation (corresponding to a pF of 4.2 or 150 m of water). This applies also in the absence of plants. We apologise for the incorrect usage of the term wilting point. The corresponding sentences were changed to:*

"In a second step, we manually calibrated the soil physical parameter used in the water retention curve to define the minimal residual water, which has also notable influence on the freezing-point depression." and

"Figure 9. comparison of the simulated (red) and measured (black) soil moisture data at 12 cm (left panels) and 60 cm (right panels) at SCH. (a) and (b) are the results for soil moisture of the best thermal calibration while (c) and (d) are the results after a further calibration of the soil physical parameter of the water retention curve, showing that the calibration can be further improved with additional data sets."

p. 4817, l. 23+25: I would interpret these two statements in the way that the results are not meaningful for this case. The optimization procedure yielded unphysical model parameters to compensate for the incomplete model physics, and then the model is run in this configuration with the future forcing, not knowing anything about the effect of the biased configuration under the warmer future conditions.

Response and changes in manuscript: *see our response to the same comment above. The aim of this study was to introduce and evaluate a new calibration approach, and show the advantages and disadvantages for various (and different) mountain permafrost sites. The long-term simulations (and their comparison with the earlier published estimates of Scherler et al. 2013) show the effects of incomplete calibration if specific processes (here convection within the coarse blocky surface layer) are not included. We added a paragraph discussing these effects, and also included the comparison between the Scherler et al. (2013) results and the results of the present study as new Figure 11. We find a thorough discussion of the advantages/disadvantages better than just leaving out the long-term simulations for rock glacier Murtèl. A corresponding paragraph was also added in the Discussion chapter.*

p. 4820, Conclusions: I expect a clear statement from the authors if the considerable efforts involved in this method can bring a performance gain over much simpler methods (e.g. the comparatively primitive approach to simulate future scenarios for borehole temperatures in Etzelmüller et al. 2011, Hipp et al. 2012, or the spatial modeling of Jafarov et al., 2012).

Response and changes in manuscript: *We added a corresponding statement to chapter 8, Conclusion: "The method was generally suitable for large-scale or long-term modelling but is not recommended for site-specific process analysis, if there are existing dominant processes which are not included in the CoupModel formulation. In these cases, manual calibration and parameterization of the missing processes have to be added. In comparison to other, simpler approaches to simulate future scenarios for borehole temperatures (as e.g. in Etzelmüller et al. 2011, Hipp et al. 2012 or, regarding spatial modelling, in Jafarov et al. 2012) the approach of this study focuses more on the site-specific processes understanding, while the long-term simulation results will not necessarily be better than results from simpler approaches as in the above cited studies. But we believe that the considerably higher efforts of our approach are well justified by the knowledge gain regarding the effect of the dominant processes of the different sites. Of course, future work has to be directed into including the already identified missing processes into the model formulation (i.e. convection)."*

II. Anonymous Referee #2

Received and published: 1 February 2016

The manuscript by Marmy et al. is concisely written and dealing with an important question in permafrost modeling, trying to bridge the gap between site specific calibration and its practical adaptation to spatially distributed modeling on larger scales. In my opinion it is a valuable contribution to permafrost research using a new and time saving approach to calibrate complex models at a set of sites in the Swiss Alps and is as such suited to be published in this journal after some minor revisions.

II.1 General comments

The English should be improved and the manuscript should be shortened. The site descriptions are quite extensive and could in part be moved to a table instead. Also the abstract should be shortened.

Response and changes in manuscript*: Thank you very much for your constructive comments! In general we improved the English, shortened the abstract and tried to shortened the paper wherever possible. We shortened the site description but did not include an additional table, as we feel that it would not shorten the paper with respect to space, and would not improve the readability, due to the many specific characteristics of the sites. As response to reviewer 1, certain paragraphs especially regarding the discussion of the calibration procedure had to be slightly enlarged.*

- The distinction between 1D, 2D and 3D models is not consistent, as the cited 2D models partly refer to empirical statistical models whereas the 1D and 3D approaches refer to process based numerical models.

Response and changes in manuscript (see also reviewer 1): *that is correct, thanks for this clarifying comment (see also reviewer 1). We rewrote the entire paragraph and included new references*

- It is stated that wind speed has virtually no influence on the ground thermal regime. I would doubt this, so either a reference should be given or this simplification has to be addressed in the discussion.

Response and changes in manuscript (see also reviewer 1): *We apologise for the misunderstanding! The reviewers are correct in pointing out that wind speed may have a pronounced effect, what we wanted to say is that the wind speed scenarios of the different GCM-RCM chains are very similar so that the choice of the specific chain does not play a major role in this case. The sentence: "As the wind plays only a minor role in the long-term trend of soil, it was satisfactory to use the median series of the seven other chains" was accordingly changed to: "As the wind speed scenarios of all available GCM-RCM chains are very similar, we consider it acceptable to use the median of these as a substitute for the seven chains with missing wind speed scenarios".*

- The improved calibration using soil moisture data (Figure 9) is very interesting – it would be of great interest for the reader to see how the future projections for this specific site would change by applying the new calibration.

<u>Response and changes in manuscript:</u> We supplied a new figure (Fig. 10) and explanations in the text, which shows the changes (~0.3K colder and later degradation) for Schilthorn by applying the new calibration.

II.2 Technical corrections

P4788
L1: Permafrost also exists in other regions than the European Alps. Be more precise.

<u>Response and changes in manuscript</u>: *changed to "Permafrost is a widespread phenomenon in mountainous regions of the world such as the European Alps".*

L5: ..which allow for . . .

<u>Response and changes in manuscript</u>: *changed accordingly*

L15: Is the calibration method "automated" or "semi-automated"?

<u>Response and changes in manuscript</u>: "semi-automated" → *changed accordingly*

L21: . . . by the end of . . .
L26: climate input data

<u>Response and changes in manuscript</u>: *changed accordingly*

P4789
L1-2: Sentence is unclear and should be rewritten.

<u>Response and changes in manuscript</u>: *changed to: "Permafrost is the thermal state of a soil or rock subsurface with a temperature that remains below 0°C for two or more consecutive years"*

L12: . . . has had notable effects on permafrost that are apparent..

<u>Response and changes in manuscript</u>: *changed accordingly*

P4791
L15: .. multiple sites . . .
L20: .. interpolating in between by the help of . . .

<u>Response and changes in manuscript</u>: *changed accordingly*

P4792
L14: ..massif site ..

<u>Response and changes in manuscript</u>: *changed accordingly*

L16: The expression "non-vegetated lithology" sounds unfamiliar to me.

Response and changes in manuscript: *changed to "The lithology of this non-vegetated site is dominated by…"*

P4797
L23: . . . complete on-site meteorological . . .

Response and changes in manuscript: *changed accordingly*

P4798
L7: ..resolution for the . . .
L9: . . . that some . . .

Response and changes in manuscript: *changed accordingly*

P4800
L13-15: See "General comments"

Response and changes in manuscript: *see response to the general comments*

L17: For calibration, we used . . .
L19: ..errors are possible due to . . .

Response and changes in manuscript: *changed accordingly*

P4803
L6: ..them based on/using . . .
L27: ..for the ground thermal . . .

Response and changes in manuscript: *changed accordingly*

P4804
L4-6: Give references to why these parameters are important in reality. It would also be interesting to read what kind of preliminary analysis is mentioned.

Response and changes in manuscript: *the references were added. With "preliminary analysis" we meant the previous CoupModel permafrost studies by our research group. These references were also added and the sentence was changed accordingly..*

L13: ..tested as well as their range . . .

Response and changes in manuscript: *changed accordingly*

P4805
L15: "As expected": Why was this expected? Give a reference or a reasonable for this expectation.

Response and changes in manuscript: *the respective references were added.*

L24: ..thick snow cover...

Response and changes in manuscript: *changed accordingly*

L24-27: Sentence is unclear.

Response and changes in manuscript: *If the snow cover for a specific field site is very thick (e.g. maximum snow cover thickness of >2m) the ground is decoupled from atmosphere during most part of the year (often October/November-July). During this time, small inaccuracies in the snow cover parameterisation will not influence the model results as they do not affect the ground thermal regime. The site is therefore less sensitive. On the contrary, if a site has generally only a thin snow cover (e.g. maximum snow cover <1m), small uncertainties in the snow cover parameterisation will directly affect the simulated ground thermal regime, so the site is more sensitive. From a modellers point of view these sites are more easy to calibrate, as the snow parameters present an easy possibility for an automatic calibration. As we see that this might be misleading in the sentence we omitted the sentence in brackets (“*and therefore more difficult to calibrate”*) in the revised version.*

P4807
L9: How broad is the range of thermal conductivities needed?

Response and changes in manuscript: *we added the respective values to the text (between 0.3 and 2.5 W/m*K).*

L24: ..should be.. Or should it read "has to be"?

Response and changes in manuscript: *see the changes to this sentence shown in the comment below.*

L24-26: This is somehow contradictory to the findings that evaporation is not important for the ground thermal regime at most sites, as it seems to be a tuning parameter for one site. Can you explain why?

Response and changes in manuscript: In the beginning *we were not completely sure of the reasons, because the obtained value for the parameter which was used for evaporation tuning was not very different for RIT as for the other sites (in fact the value for RIT is in the middle of the spread of all sites). Meanwhile we looked more in detail into all parameters from the calibration (see also response to reviewer 1) and found the probable reason for this behaviour. We therefore included the following paragraph into the discussion of the calibration results:*

“When analyzing the specific values obtained for the different calibration parameters at all sites, it was noted that even though the parameter related to evaporation (water tension Ψ_{eg}, cf. Table 2) did not show specifically high or low values for RIT, the parameterized values for T_{snow} (minimum temperature at which precipitation falls only as snow) and ΔS_{crit} (critical snow depth, at which the whole surface is considered to be covered by snow) are very low (T_{snow} = -4.86°C) and very high (ΔS_{crit} = 1.9m), respectively. Whereas the former leads to comparatively large precipitation input as rain, the latter leads to an almost never completely snow-covered surface. In addition, the wet soil albedo for RIT is calibrated with the lowest value of all sites (α_{wet} = 7.0) whereas its dry albedo is comparatively high (α_{dry} = 34.6). In total, this parameter combination enables additional energy input by liquid water into the subsurface, which of course also explains the high sensitivity to evaporation. Even though this parameter combination may lead to an unrealistic process representation in the model, it is still in good accordance with observations, as the effect of 3-d advective water flow from the melting snow cover has been observed in the borehole temperatures (see also Luethi et al. 2016), which is probably the reason for this specific calibration outcome. Of course, the real 3D-process of melt water infiltration is not included in the model."

P4808
L6: boreholes
L17: ..which is because..

Response and changes in manuscript: *changed accordingly*

P4811
L3:. . . slightly positive. . .
L11:. . . set of..

Response and changes in manuscript: *changed accordingly*

P4812
L4:. . . artifact is due to. . .
L5:. . . propagated when run..
L6:. . . by 2080..

Response and changes in manuscript: *changed accordingly*

P4813
L9-15: Does the model have structural errors (L14) or is it realistic (L10)? Be more consistent.

Response and changes in manuscript: The paragraph was changed as follows: "The GLUE calibration method is not meant to determine the physical values of a parameter. The model is physically-based regarding its underlying equations, but has to rely on parameterisations for many of the complex processes in the subsurface and at the soil-snow-atmosphere boundary. The values for all model parameters at all depths cannot be known exactly, especially as almost no direct measurements of these properties are available., The GLUE method gives the ability of finding the value which gives the best fit with the observations within the number of tested runs. But as the system is complex, with sometimes highly uncertain initial and boundary conditions, non-linear processes and model structural simplifications make an optimum calibration impossible (Beven, 2002). It is therefore more meaningful to analyze the residuals and the sensitivity to parameters than the values of the parameter themselves."

P4814
L8: What is meant by "semi-infinite"?

Response and changes in manuscript: *changed to "extremely large number…"*

L14: parameters

Response and changes in manuscript: *changed accordingly*

P4815
L28-29: What is meant by "provide again the P3-link"?

<u>Response and changes in manuscript</u>: *Sorry for this editing error! The missing link was inserted.*

L29: .. as a validation . . .

<u>Response and changes in manuscript</u>: *changed accordingly*

P4816
L9: Give a proper definition of the wilting point.

<u>Response and changes in manuscript (see also reviewer 1)</u>: *See our answer to the respective comment of reviewer 1 – we apologise for the error. Indeed no vegetation is present at all our modelled sites, and the corresponding vegetation module in CoupModel was switched off for all simulations. The wilting point parameterisation mentioned in our old Figure 9 (and corresponding text) is used in Coup as a soil physical property and is part of the water retention parametrisation (corresponding to a pF of 4.2 or 150 m of water). This applies also in the absence of plants. We apologise for the incorrect usage of the term wilting point. The corresponding sentences were changed to:*

*"*In a second step, we manually calibrated the soil physical parameter used in the water retention curve to define the minimal residual water, which has also notable influence on the freezing-point depression." *and*

"Figure 9. comparison of the simulated (red) and measured (black) soil moisture data at 12 cm (left panels) and 60 cm (right panels) at SCH. (a) and (b) are the results for soil moisture of the best thermal calibration while (c) and (d) are the results after a further calibration of the soil physical parameter of the water retention curve, showing that the calibration can be further improved with additional data sets."

L13: ..further improved . . .

<u>Response and changes in manuscript</u>: *changed accordingly*

P4819
L24: . . . analyze the . . .
L27: . . . can be drawn . . .

<u>Response and changes in manuscript</u>: *changed accordingly*

P4820
L4: ..precisely whereas . . .

<u>Response and changes in manuscript</u>: *changed accordingly*

L25-26: Is this sentence complete?

<u>Response and changes in manuscript</u>: *changed to:* “The degradation is primarily driven by the change in air temperature during the snow-free period and the change in snow cover duration.”

P4821
L1: . . . decrease from . .

<u>Response and changes in manuscript</u>: *changed to:* “to decrease by values between 20 % and 37 %”

Semi-automated calibration method for modelling of mountain permafrost evolution in Switzerland

A. Marmy[1], J. Rajczak[2], R. Delaloye[1], C. Hilbich[1], M. Hoelzle[1], S. Kotlarski[2], C. Lambiel[3], J. Noetzli[4], M. Phillips[5], N. Salzmann[1], B. Staub[1] and C. Hauck[1].

[1]{Department of Geosciences, University of Fribourg, Switzerland}

[2] {Institute for Atmospheric and Climate Science, ETH Zurich, Switzerland}

[3] {Institute of Geography, University of Lausanne, Switzerland}

[4] {Department of Geography. University of Zurich, Switzerland}

[5] { WSL, Swiss Federal Institute for Snow and Avalanche Research, Switzerland}

Correspondence to: A. Marmy (antoine.marmy@gmail.com)

Abstract

Permafrost is a widespread phenomenon in mountainous regions of the world such as the European Alps. Many important topics such as the future evolution of permafrost related to climate change and the detection of permafrost related to potential natural hazards sites are of major concern to our society. Numerical permafrost models are the only tools which ~~facilitate~~ allow for the projection of the future evolution of permafrost. Due to the complexity of the processes involved and the heterogeneity of Alpine terrain, models must be carefully calibrated and results should be compared with observations at the site (borehole) scale. However, for large-scale applications, a ~~large number of local point data are necessary to obtain a broad overview of the thermal evolution of mountain permafrost over a larger area, such as the Swiss Alps, and the~~ site-specific model calibration ~~of each~~ for a multitude of grid points would be very time-consuming. To face this issue, this ~~paper~~ study presents a semi-automated calibration method using the Generalized Likelihood Uncertainty Estimation (GLUE) as implemented in a 1-d soil model (CoupModel) and applies it to six permafrost sites in the Swiss Alps. ~~prior to long-term permafrost evolution simulations.~~ We show that this semi-automated calibration method is able to accurately reproduce the main thermal condition characteristics with some limitations at sites with unique conditions such as 3-d air or water circulation, which have to be calibrated manually. The calibration obtained was

used for Global and Regional Climate Model (GCM-RCM) ~~RCM~~ based long-term ~~simulations~~ climate projections under the A1B climate scenario (EU-ENSEMBLES project) specifically downscaled at each borehole site. The projection shows general permafrost degradation with thawing at 10 m, even partially reaching 20 m depth~~s~~ ~~until~~ by the end of the century, but with different timing among the sites and with partly considerable uncertainties due to the spread of the applied climatic forcing. ~~The degradation is more rapid at bedrock sites whereas ice-rich sites with a blocky surface cover showed a reduced sensitivity to climate change. The snow cover duration is expected to be reduced drastically (between -20 % to -37 %) impacting the ground thermal regime. However, the uncertainty range of permafrost projections is large, resulting mainly from the broad range of input climate data from the different GCM-RCM chains of the ENSEMBLES data set.~~

1. Introduction

Permafrost is the thermal state of ~~the~~ a ~~subsurface~~ soil or rock ~~that~~ subsurface with a temperature that remains below 0°C for two or more consecutive years (Harris et al., 2009). It occurs in the Arctic (~~Langer et al., 2013~~Romanovsky et al. 2010) and Antarctic ice-free regions (Vieira et al., 2010) as well as in mid-latitude mountain ranges such as in the European Alps (Boeckli et al., 2012), the Andes (Trombotto, 2000) and the Himalayan range (Weiming et al., 2012). In the last few decades, in the context of global warming, interest for permafrost has increased for various reasons such as greenhouse gas releases (e.g. Anthony et al., 2012), engineering and construction issues (e.g. Lepage and Doré, 2010, Bommer et al., 2010), water management issues (e.g. Quinton et al., 2011) and slope stability concerns (McColl, 2012). In mountain environments, the increase in air temperatures observed in the last decades (Mountain Research Initiative EDW, 2015) has had ~~a~~ notable effects on permafrost that are ~~visible~~apparent: i) in the borehole data series by higher surface and subsurface ground temperatures and significantly deeper active layers (e.g. PERMOS, 2013) ii) in geophysical data with a decrease of the electrical resistivities (Hilbich et al., 2008, 2011, PERMOS, 2013) and of seismic velocities (Hilbich, 2010) indicating a reduction of ice-content and iii) in the increased activity of permafrost creep (Kääb and Kneisel 2006; Barboux et al., 2013) and increased velocities of instable rock glaciers (Kääb et al., 2007; Gärtner-Roer, 2012).

Therefore, increasing effort has recently been put into permafrost modelling across different temporal and spatial scales. The conceptual and spatial range of modelling approaches include: i) physically-based process and/or energy balance models which focus either on 3-d ~~models~~ applications ~~that can be used to~~

by simulating~~e~~ a limited number of processes such as heat conduction, latent heat and the effect of topography (e.g. Noetzli and Gruber, 2009; Noetzli et al., 2007) or 1-d simulations to analyse a large number of complex subsurface processes with a potentially high number of feedback mechanisms (e.g. Westermann et al. 2015, 2016; Langer et al., 2013; Hipp et al., 2012; Scherler et al., 2010; Luetschg et al., 2008), and ii) empirical-statistical ~~2-d~~ distribution models (e.g. Etzelmüller et al., 2006; Hartikainen et al., 2010; Boeckli et al., 2012; Sattler et al. 2016) which are often based on rock glacier inventories or other permafrost evidences (Cremonese et al. 2011). ~~that are applicable for permafrost distribution or zones of potential ground instability and iii)1-d models (e.g. Hipp et al., 2012; Westermann et al., 2011; Scherler et al., 2010; Luetschg et al., 2008) that are used to simulate many complex processes with a high number of feedback mechanisms~~. Recently, new model approaches were developed which are able to simulate hydrological processes in 3-d, while keeping most thermal processes in 1d (Endrizzi et al., 2014). On hemispheric and global scales, ~~3-d~~ spatially distributed 1-d models (also called 2.5d models) and land surface schemes are used to assess permafrost evolution. Here, ground temperatures are only calculated along 1-d soil columns, but on a large regional or hemispheric grid (e.g. Jafarov et al. 2012, Zhang et al. 2012, Westermann et al. 2013, 2016, Ekici et al., 2014, 2015, Chadburn et al., 2015) without lateral interaction.

The 3-d and 2-d approaches can be related more easily to geophysical or remote sensing methods, especially in Arctic lowlands where methane release is a major issue (Anisimov, 2007). In mountain environments, 1-dimensional modelling is widely used due to the spatial heterogeneity of surface and subsurface composition, topography, morphological landforms and microclimatic processes. Moreover, 1-d approaches are easier to relate to borehole temperature time series that are common in alpine permafrost research and are usually the only validation or calibration data available. However, the final goal of most permafrost modelling studies, especially in the Arctic (e.g. Ekici et al., 2015), is the representation of permafrost and permafrost processes in a distributed model. Whereas this is common in the A~~a~~rctic, this is still at a beginning stage in alpine environments due to many limiting factors, including the scarcity of input data, heterogeneity of surface, subsurface and microclimatical conditions. Fiddes et al., (2015) proposed a scheme that is leading in the direction of combining physically-based land surface models (LSMs) and gridded climate data to efficiently simulate air temperature and near-surface ground temperature but does not include borehole data validation.

Site-specific calibration is an important prerequisite for successful permafrost modelling with complex models. However the process of calibration often faces the scarcity of measured input parameters such as porosity, ice and water content, or thermal and hydraulic conductivities. All modelling approaches trying to ~~model~~ simulate real conditions should use a ~~certain~~ specific procedure (Westermann et al., 2013), which can also ~~and some integrate more~~ include empirical calibration methods by manual tuning (Gruber and Hoelzle, 2001; Hipp et al., 2012; Scherler et al., 2013). With recent improvements in computing capacity, the use of automated procedures of inverse modelling approaches using Monte-Carlo chains has become increasingly attractive (Jansson 2012, Heerema et al., 2013), but so far this approach has not been tested in permafrost research.

The final goal of most permafrost modelling studies is their application to long-term climate impact simulations. Previous studies of combined climate-permafrost simulations with explicit subsurface simulations for the Alps are rare and were focused only on 1 or 2 sites (e.g. Engelhardt, et al., 2010; Scherler et al., 2013) because of the limitations in the availability of ground temperature data and/or on-site meteorological data for calibration/validation purposes. Atmospheric forcing data for permafrost models can be derived from ~~Global~~ General Circulation and/or Regional Climate Models (GCMs, RCMs). Especially for alpine terrain, RCMs offer an added value with respect to coarse resolution GCMs (e.g. Kendon et al., 2010; Torma et al., 2015) and are now widely used in scientific research, especially in the impact modelling community (e.g. Bosshard et al., 2014).

In this study, we present a semi-automated procedure for calibrating a soil model to a large number of points at ~~different~~ multiple permafrost sites. The calibration procedure attempts to understand site-specific differences ~~between the sites~~ as well as to quantify the sensitivity of the soil model to the tested parameters. ~~This method is a first step towards the realization of a well-calibrated distributed model, possibly done in the future by applying the semi-automated calibration method to all sites where surface or subsurface calibration data are available, and interpolating between with the help of additional data (e.g. Digital Elevation Models).~~ The procedure has been applied to six test sites in the Swiss Alps: Stockhorn, Schilthorn, Muot da Barba Peider, Lapires, Murtèl-Corvatsch and Ritigraben. After calibration, the model set-up was used for long-term simulations driven by downscaled climate model data until the end of the 21st century, and an analysis of the evolution of the ground thermal regime and the snow cover is presented. The present work has two main objectives: i) show the benefits and limitations of a semi-automated calibration procedure for detailed soil process modelling in permafrost terrain, use this procedure to identify differences and similarities among the test sites and to assess the sensitivity of the soil model to certain parameters, and ii) develop ~~potential~~ scenarios of the possible evolution of mountain permafrost in Switzerland.

2. Study sites

In the framework of the SNF-funded project "The Evolution of Mountain Permafrost of Switzerland" (TEMPS) (Hauck et al., 2013) and the Swiss permafrost monitoring network ~~(~~PERMOS~~)~~ (PERMOS, 2013), based on the collaboration of five research institutions, the necessary data sets for calibration and validation purposes were available for six different sites in the Swiss Alps. These sites cover a broad geographical range within Switzerland and represent a variety of landforms including rock slopes/plateaus, talus slopes and rock glaciers. The choice of the following sites was mainly driven by the availability of long-term time series of borehole temperatures and meteorological observations.

2.1 Schilthorn

The Schilthorn~~ (SCH)~~ massif site (SCH~~Bernese Alps, Switzerland~~) is situated at 2970 m above sea level (asl) in the north-central part of the Swiss Alps.~~,~~ The lithology of this ~~and its~~ non-vegetated ~~lithology~~ site is dominated by deeply weathered dark limestone schists forming a surface layer of mainly sandy and gravelly debris up to several meters thickness over presumably strongly jointed bedrock. Within the framework of the European PACE project (Harris et al., 2003), the site was chosen for long-term permafrost observation and consequently integrated into the Swiss permafrost monitoring network PERMOS as one of its reference sites (PERMOS, 2013). The monitoring station at 2910 m asl is located on a small plateau on the north-facing slope and comprises a meteorological station (short and longwave radiation, air temperature, humidity, snow height, wind speed and direction) and three boreholes (14 m vertical, 100 m vertical and 100 m inclined) with continuous ground temperature measurements since 1999 (Vonder Mühll et al., 2000; Hoelzle and Gruber, 2008; Noetzli et al., 2008; Harris et al., 2009; PERMOS, 2013). Borehole data indicate permafrost of at least 100 m thickness, which is characterized by ice-poor conditions close to the melting point. Maximum active-layer depths recorded since the start of measurements in 1999 were generally around 4-5 m until the year 2008 but increasing to 6-7 m since 2009. During the superposition of very warm winter 2002/2003 with the summer heat wave 2003 (Schär et al., 2004) the active-layer depth increased exceptionally to 8.6 m, reflecting the potential for degradation of permafrost at this site (Hilbich et al., 2008).

The monitoring station is complemented by soil moisture measurements since 2007 and geophysical (mainly geoelectrical) monitoring since 1999 (Hauck 2002, Hilbich et al., 2011, Pellet et al. 2016). The snow cover at Schilthorn can reach maximum depths of about 2-3 m and usually lasts from October through to June/July. ~~Due to its long time series and fairly homogeneous subsurface characteristics, the Schilthorn data set was already used several times for long-term simulations but also in model sensitivity studies (Noetzli et al., 2008, Engelhardt et al., 2010, Scherler et al., 2010, Marmy et al., 2013, Scherler et al., 2013, Ekici et al., 2015).~~

2.2 Murtèl-Corvatsch rock glacier

The rock glacier Murtèl-Corvatsch (COR) is situated in the Upper Engadine, Eastern Swiss Alps, and ranges from 2750 to 2600 m asl, facing north-northwest. ~~Its surface relief is pronounced with ogive-like structures with elevation differences between the furrows and ridges of up to 8 m.~~ The surface consists of large blocks of up to several meters high which are composed by granodiorite and metamorphosed basalt ~~of the Corvatsch nappe as well as muscovite and calcite marble of the Chastelets series~~ (Schneider et al., 2013). Below this coarse blocky surface layer of approximately 3–3.5 m in thickness, a massive ice core (up to 90 %, Haeberli, 1990; Haeberli et al., 1998; Vonder Mühll and Haeberli, 1990) is present down to 28 m, with a frozen blocky layer below reaching from 28 to 50 m, probably adjacent to the bedrock (Arenson et al., 2002). ~~Borehole temperature data are available from 1987, which is the longest permafrost time series of the European Alps (PERMOS, 2013).~~

The main monitoring station is situated on a flat ridge at 2670 m asl and comprises a meteorological station (short- and long-wave radiation, air temperature, surface temperature, humidity, snow height, wind speed and direction) established in 1997 (Mittaz et al., 2000; Hoelzle et al., 2002, Hoelzle and Gruber, 2008) and two boreholes drilled in 1987 and 2000 (PERMOS, 2013), which show significant small scale heterogeneities in the rock glacier (Vonder Mühll et al., 2001; Arenson et al., 2010).. Permafrost temperatures are around -2°C at 10 m depth ~~and -1.4°C at 25 m depth~~ and the active layer has a thickness of 3.2 m on average. ~~The comparison of the stratigraphy of the drill core of the 1987 borehole with the stratigraphies of the other core of the borehole located within a distance of 30 m shows significant small scale heterogeneities in the rock glacier (Vonder Mühll et al., 2001; Arenson et al., 2010). Permafrost conditions in the vicinity of the rock glacier are heterogeneous. Areas with fine grained subsurface material, vegetated soil and solid rock usually do not show any permafrost conditions, whereas coarse grained surface conditions often have permafrost conditions (Schneider et al., 2012, 2013, Scherler et al., 2014).~~ Annual precipitation at the site is about 900 mm (982 mm St Moritz 1951–1980; 856 mm Piz Corvatsch 1984–1997) with ~~.~~ t~~T~~ypical ~~maximum~~ snow cover thickness ~~is~~ of ~~between~~ 1 ~~m and~~ - 2 m. Mean annual air temperature (MAAT) is -1.7 °C for the observation period of March 1997 to March 2008 (Scherler et al., 2014). Geophysical monitoring (mainly ERT) has been conducted since 2005 (Hilbich et al., 2009).

2.3 Lapires

The Lapires (LAP) talus slope is located on the western slope of Val de Nendaz in Valais (46°06'N, 7°17'E) in the Western Swiss Alps, ranging from 2350 m asl to 2700 m asl with a NNE orientation. Its surface consists of Gneiss schists and the talus shows a thickness of more than 40 m at the locations of the boreholes described below. Snow avalanches and minor rock falls with variable frequencies from one year to another affect the slope (Delaloye 2004, Delaloye & Lambiel 2005, Lambiel 2006). The Lapires talus slope shows an active layer of about 4–5.5 m thickness situated on top of an ice-rich (30-60 %) permafrost layer of around 15 m thickness with temperatures very close to the melting point (Scapozza et al., 2015, Staub et al., 2015).

The monitoring station consists of a meteorological station (air temperature and shortwave radiation since 1998, wind speed and direction and snow depths since 2009) installed in 1998 and three further boreholes installed in 2008 along a longitudinal profile (Scapozza et al., 2014). ~~The site's Mean Annual Air Temperature (MAAT)~~ MAAT was +0.5°C at 2500 m asl.

Compared to the strong microtopography of Murtèl rock glacier, the Lapires talus slope is comparatively homogeneous regarding slope and microtopography. The permafrost distribution within the talus slope is discontinuous (mainly related to heterogeneous substrate dominated by fine-grained material in the western part and coarse-blocky material in the eastern part) and linked to a complex system of internal air circulation also called the "chimney effect" (Delaloye and Lambiel, 2005). This air circulation ~~is most effective when the temperature (and therefore density) gradient between the air and the voids in the porous substrate is significant, and it is~~ responsible for ground cooling at the bottom of the talus slope, where cold air is suck up in winter. These 2-dimensional (or potentially 3-dimensional) processes cannot be explicitly simulated with the COUP-model ~~(see also Staub et al., 2015)~~, however, their effect on the thermal regime ~~they~~ has~~ve~~ been indirectly confirmed by specific 1-d distributed COUP simulations ~~process modelling studies~~ at this site (~~Wicky~~ Staub et al. 2015).

2.4 Ritigraben

The active rock glacier Ritigraben (RIT) is located in the area Grächen-Seetalhorn (46°11N, 7°51E), Valais, western Swiss Alps, and covers an area between elevations of 2260 m to 2800 m asl. Block sizes at the surface range from 0.5 up to several cubic meters. Active layer depth is almost constant at 4 m.

A 30 m borehole was drilled in 2002 in the lower part of the rock glacier at an altitude of 2615 m asl~~ and was equipped with thermistors~~, which is gradually being sheared off from the base upwards ~~.~~ d~~D~~ue to the movement of the rock glacier. ~~, the borehole is gradually being sheared off from the base upwards, and as~~ As a result temperature is currently only measured to a depth of 13 m. Borehole temperatures indicate the formation of a seasonal talik between 11 and 13 m depth, which appears to be directly linked to snow melt water and rainfall infiltration (Zenklusen Mutter and Phillips, 2012). The effect of these processes on the thermal regime has recently been analysed by explicit process modelling using the model Snowpack (~~Luethi and Phillips~~Luethi et al., 2016~~5~~).

The monitoring station is complemented by an automated weather station (net radiation, air temperature and relative humidity, surface temperature, snow depth, precipitation and wind speed and direction) installed in 2002 (Herz et al., 2003).

2.5 Muot da Barba Peider

The Muot da Barba Peider (MBP) talus slope is located near the top of the NW-oriented flank of the Muot da Barba Peider ridge at 2960 m asl above the village of Pontresina, Upper Engadine, Eastern Swiss Alps. The slope is 38° steep and is covered with coarse ~~-~~blocks (Zenklusen et al., 2010). The bedrock consists of gneiss from the upper Austroalpine nappe. Two adjacent (50 m apart) 18 m deep boreholes were drilled in 1996.~~ and are 18 m deep.~~

The drilling stratigraphy shows ground ice occurrences inside the talus, which reach a depth of about 4 m, with frozen bedrock below (Rist et al., 2006). Active layer depth varies between 1 and 2 m (Zenklusen et al., 2010). Due to the presence of experimental snow avalanche defence structures near borehole 1, the snow cover persists longer there in spring/summer and thus influences the ground thermal regime (Phillips, 2006).

An~~u~~ automatic weather station was installed in 2003 ~~and since then,~~ showing MAAT ~~is~~ of -3°C. R~~r~~egional values for mean annual p~~P~~recipitation ~~is not measured on-site but mean annual regional values~~ are around

1500 mm at this elevation (Zenklusen and Phillips, 2012). Maximum snow depths have ranged between ~~50 and 300 cm~~ 0.5 and 3m since 2003. ~~The bedrock consists of gneiss from the upper Austroalpine nappe.~~

2.6 Stockhorn

The study site of the Stockhorn (STO) plateau is situated on an East-West oriented mountain crest around 3410 m asl, to the west of the Stockhorn summit (3532 m asl) above Zermatt ~~, Valais~~ (45°59'N 7°49'E), Western Swiss Alps. The lithology consists of Albit-Muskowit schists and the surface is characterized by patterned ground that has developed in a thin debris cover. Significant amounts of ground ice could be observed in large ice-filled cracks during construction works of a new ski lift in summer 2007 (Hilbich 2009). Two boreholes only 30 m apart were drilled in 2000 as part of the PACE project (Harris et al., 2003). The recorded borehole temperatures show that the Stockhorn plateau is strongly affected by 3D topography effects (Gruber et al., 2004), because the 100 m deep borehole close to the north face exhibits significant colder temperatures than the 17 m deep borehole located close to the southern edge of the plateau. ~~The monitoring station is complemented by a~~ A meteorological station (short- and long-wave radiation, air temperature, humidity, snow height, wind speed and direction) ~~that~~ was installed in 2002. A soil moisture station was added in 2014.

The MAAT at this site is -6.4°C for 2002-2012 and the annual precipitation ~~at~~ is around 1500 mm (Gruber et al., 2004, based on King, 1990, Begert et al., 2003). This site is characterized by low precipitation and high solar radiation (mean short-wave incoming radiation from 2002-2013: 209.3W m^{-2}) due to particular conditions created by surrounding mountain ranges exceeding 4000 m asl (Gruber et al., 2004).

3. Data and model

One of the main challenges in the modelling of permafrost evolution is the general lack of long (> 15 years) and ~~gapless~~ complete on-site ~~multivariate~~ meteorological data necessary as input for the calibration of the soil model. Similarly, data from GCM/RCM derived climate scenarios have to be downscaled and bias corrected for specific on-site conditions, which is non-trivial due to the high altitudes of most permafrost stations and the above mentioned short length of on-site meteorological data. In this section we will explain the downscaling approach used, and introduce the available borehole data-sets used for calibration of the soil model. Finally, the physical basis of the COUP soil model as well as its major parameterizations will be explained.

3.1 Climate scenarios: Statistical downscaling and bias correction.

Site-specific climate scenarios have been developed for eight meteorological variables at daily resolution ~~and~~ for the period from 1951 to 2099 (Rajczak et al. 2016). The scenarios are based on an ensemble of 14 regional climate model (RCM) projections from the EU ENSEMBLES project (van der Linden and Mitchell, 2009). It should be noted that~~,~~ some variables have fewer GCM-RCM chains available: 7 for mean wind-speed and maximum wind-gusts and 13 for global radiation. Only the 13 chains with global radiation were used in the present study.

The ensemble accounts for a comprehensive range of model uncertainty and is forced by the IPCC SRES A1B emission scenario (Nakicenovic and Swart, 2000). Due to their limited spatial resolution, site-specific features are typically not resolved by climate models and even on resolved scales, models are subject to biases (e.g., Kotlarski et al., 2014). Statistical downscaling (SD) and bias correction (BC) techniques serve to attain representative conditions for the site scale and to remove model biases. SD/BC applications derive an empirical relationship between observations and model output. The established relationships are in turn used to translate long-term climate simulations to the site scale. Calibrating SD/BC techniques, however, requires long-term observations (e.g. 30 years and more), a prerequisite not met by the monitoring sites of the present study.

To obtain robust and reliable climate scenarios at the six considered sites, a newly implemented SD/BC method was used that specifically targets locations that lack long-term data. A detailed description and comprehensive validation of the approach is given by Rajczak et al.~~,~~ (2016~~5~~). It is designed as a two-step procedure sketched in Figure 1. In the first step, climate model simulations are downscaled to match long-term observational measurements at a most representative site (MRS) within a surrounding measurement network (e.g. MeteoSwiss weather stations). In the second step, the downscaled and bias-corrected time series from the MRS are spatially transferred to the site of interest (e.g. a permafrost monitoring site). Both steps rely on the quantile mapping (QM) method, a well-established statistical downscaling and bias correction technique (e.g. Themessl, 2011). The concept behind QM is to correct the distribution of a given predictor (e.g. climate model output) in such a way that it matches the distribution of a predictand (e.g. observations of the same variable at a monitoring site). Values outside the range of calibrated values are treated using the correction for the 1st (99th) quantile. Within this study, the spatial transfer is performed from an objectively selected MRS within the MeteoSwiss monitoring network. Consequently, Rajczak et al., (2016~~5~~) show that the MRS is, in many cases, not the closest station but rather one at a similar altitude.

3.1.1 Reconstruction of meteorological observations

The two-step procedure (Figure 1) additionally facilitates the reconstruction of data at the monitoring site for non-measured periods. The concept behind reconstructing data is to spatially transfer (Figure 1, step 2) observed values from an MRS to the target site. In the framework of the present study, data were reconstructed for some periods between 1981 and 2013. Note, that reconstruction is constrained by the availability of data at the MRS. An extensive validation of the reconstruction performance is given by Rajczak et al., (2016~~5~~).

3.1.2 Climate Scenarios: Projections of 2 m-temperature

Based on the developed site-scale scenarios, Figure 2 provides the projected evolution of mean annual air temperature (MAAT) at 2 m above ground for the six considered sites in the period between 1961 and 2099. The projections assume an A1B emission scenario and include model uncertainty (i.e. range of estimates). While MAAT is predominantly negative in ~~the~~ present-day climate, all six sites are subject to a significant increase in temperature and ~~virtually~~ the majority of ~~all~~ climate models ~~agree~~ indicate at four of the six sites ~~on positive mean temperatures by the end of the 21st century~~positive mean annual temperatures by the end of the 21st century.

For each site, the reconstructed meteorological data used consists of daily series' for the period between 1981-2013 for five variables: mean air temperature, precipitation sum, mean wind speed, mean relative humidity and global radiation. For the site MBP, the global radiation series could not be reconstructed because of a lack of validation data and could therefore ~~. Therefore, this variable was not used~~ not be used as forcing variable in the calibration for this site. Global radiation for MBP has therefore been estimated ~~and has been estimated by the model~~ by CoupModel based on potential global radiation (depending on latitude and declination) and atmospheric turbidity (Jansson 2012). ~~and air temperature.~~ Independent comparison between measured, reconstructed and CoupModel estimated global radiation values for COR showed an overestimation of global radiation by the CoupModel leading to near-surface maximum temperature biases of up to 10°C in summer (cf. supplementary material). However, the calibration technique applied (see sections 4 and 5) would compensate potential biases in the temperature simulations by adjusting related parameters in the model, e.g. snow cover parameters or the albedo. Corresponding uncertainties arising from a potential compensation in the MBP results will be further discussed below.

Despite the good quality of the reconstruction, some short gaps could not be avoided. These gaps have been filled by artificial random selection of data from other years at the same date. This method is satisfactory as the gaps are short and infrequent.

For seven of the chains, ~~the~~ wind speed scenarios were not available. As the wind speed scenarios of all available GCM-RCM chains are very similar, we consider it acceptable to use the median of these scenarios as a substitute for the seven chains with missing wind speed scenarios~~As the wind plays only a minor role in the long-term trend of soil, it was satisfactory to use the median series of the seven other chains~~.

3.2 Borehole data

For calibration~~ data~~, we used series of ~~measured~~ borehole temperature data for each ~~borehole~~ site with a minimum length of 10 years (Table 1). Borehole data is often considered as “ground truth data” but potential measurement errors ~~cannot be excluded~~ are possible due to several reasons (such as sensor or logger drift, logger failure and infiltration of water inside the borehole casing, to name a few). Further, an unequal repartition of ~~longer~~ data gaps may introduce a bias in the calibration. The gaps within the borehole temperature series have not been filled in order to avoid the introduction of inconsistency and additional errors in the data used for calibration. ~~The p~~Periods with gaps are consequently ignored in the calibration process. ~~In addition, one has to be aware that a level of uncertainty is introduced by calibrating the model only against temperature measurements and not against other sets of independent data such as snow depth, water content and/or apparent electrical resistivity.~~

3.3 COUP model description & experimental set-up

The model used for this study is the CoupModel, a 1-dimensional numerical model coupling ~~the~~ soil, snow and atmospheric~~e~~ processes (Jansson and Karlberg, 2004; Jansson 2012). This model has already shown ~~good aptitude~~ that it is well suited to simulate mountain permafrost processes at Schilthorn

(Engelhardt et al., 2010, Scherler et al., 2010; Scherler et al., 2013; Marmy et al., 2013) and ~~at~~ Murtèl rock glacier (Scherler et al., 2013; Scherler et al., 2014). It also includes an optional procedure for semi-automatic calibration based on statistical indicators (see section 4).

The model couples the water and heat transfer of the soil using the general heat flow equation:

$$\frac{\delta(CT)}{\delta t} - L_f \rho \frac{\delta \Theta_i}{\delta t} = \frac{\delta}{\delta z}\left(k \frac{\delta T}{dz}\right) - C_w T \frac{\delta q_w}{\delta z} - L_v \frac{\delta q_v}{\delta z} \quad (1)$$

where C (J K^{-1}) is the heat capacity of soil, C_w (J K^{-1}) is the heat capacity of water, T *(z,t)* (K) is the soil temperature, L_f and L_v (J kg^{-1}) are the latent heat of freezing and vapor, ρ (kg m^{-3}) is the density, Θ_i *(z,t)* is the volumetric ice content, k (W m^{-1} K^{-1}) is the thermal conductivity, t is the time, z is the depth and q_w *(z,t)* and q_v *(z,t)* (kg m^{-2} s^{-1}) are the water and vapor fluxes.

The lower boundary condition is derived from the sine variation of the temperature at the soil surface and a damping factor ~~for the~~ with depth. The maximum model depth is different for the various sites -due to the varying maximum depth of the available boreholes, but it is at least 30 meters for all sites and well below the depth of zero annual amplitude (see Figure 3). The prescribed heat flux at the lower boundary condition is therefore negligible. This enables comparatively stable conditions at the lower boundary, and accounts for the often isothermal conditions found in Alpine permafrost at this depth (Scherler et al. 2013, PERMOS 2013). However, long-term variability of permafrost conditions at the lower boundary cannot be simulated with this approach. The hydraulic boundary condition is given by gravity-driven percolation i~~I~~f the lowest compartment is unsaturated.~~, a percolation is calculated by gravitational forces~~.

The u~~U~~pper boundary condition~~s~~ ~~are~~ is calculated ~~with a~~ using the complete energy balance at the soil surface (or snow surface, if present). ~~, or at the snow surface, if snow cover is present.~~ The convective heat inflow of water is given by precipitation and snow melt multiplied by the surface temperature and the heat capacity of liquid water (C_w):

$$q_h(0) = \frac{T_s - T_1}{\Delta_z/2} + C_w(T_a - \Delta T_{Pa}) q_{w\cancel{in}}(0) + L_v q_{\cancel{w}v}(0) \quad (2)$$

where $q_h(0)$ (J m^{-2}d^{-1}) is the soil surface heat flow, T_s is the soil surface temperature, T_1 is the temperature in the uppermost soil layer, ΔT_{Pa} is a parameter representing the temperature difference between ~~the~~ air and ~~the~~ precipitation, $q_{v\cancel{in}}$ (0) ~~is the water infiltration rate, Δ_z is the depth change,~~ and q_w (0) ~~is~~ are the ~~water~~

vapour and ~~vapour~~ water ~~flow~~ fluxes at the surface and L_v is the latent heat of vapour. For periods with snow cover, the upper boundary condition is calculated assuming a steady state heat flow between the soil and a homogeneous snow pack using the thermal conductivity of snow. Temporally changing insulation conditions of the snow cover can be simulated ~~Patchy snow cover conditions can be parameterized with~~ by a critical snow height that corresponds to the snow height that completely covers the soil. It depends mainly on the surface roughness and reflects the fact that 50 cm of snow induces different insulation properties for a surface consisting of 1-2 m high boulders (e.g. for a rock glacier, COR) compared to a rather homogenous surface covered by sandy soil (e.g. at SCH, cf. also the discussion in Staub and Delaloye 2016). ~~and is one of the most influencing parameters used later in the calibration procedure.~~ The fraction of bare soil is then calculated by a ratio between 0~~cm~~ and this threshold (see ~~Equation 9~~Table 2) ~~.~~ ~~The fraction of snow free ground is then~~ and further used to estimate the average soil surface temperature and ~~the~~ surface albedo. This critical snow height is one of the parameters having the largest influence in our calibration procedure.

Snow is simulated by ~~CoupModel~~ partitioning ~~of the~~ precipitation into rain and snow depending on temperature threshold parameters. The snow cover is assumed to be horizontally and vertically homogenous. ~~The s~~Snow melt is estimated as part of the heat balance of the snow pack, including net radiation, sensible and latent heat flux to the atmosphere, heat flux in precipitation, snow temperature change and heat flux to the soil. Further important processes in COUP are listed in Table 2 together with the respective equations.

The soil structure consists of 18 to 25 compartments (depending on the site)~~,~~ with increasing thickness with depth, ranging from 0.1 m in the upper layers to 4 m in the lower layers (Figure 3). ~~The i~~Initial conditions are estimated by the model ~~by~~ using the first values of the meteorological data series. To avoid imprecise initial conditions, the model is run from 1981 onwards, although ~~the~~ observational time series usually begin ~~only~~ around the year 2000. No additional spin up is needed as the model usually reaches ~~the~~ stable conditions (i.e. not influenced by initial conditions) after 10 to 15 years. Model tests with longer spin-up times showed only negligible differences with respect to the procedure described above. However, this approach clearly neglects all long-term effects of past climatic conditions on the ground thermal regime at larger depths. Therefore, simulation results at larger depths should not be interpreted in a climate context.

4. Calibration procedure: GLUE

With the recent increase in computing power, the automation of the calibration of soil models, also called inverse modelling, has been used increasingly (e.g. Finsterle et al., 2012; Cui et al., 2011; Boeckli et al., 2012; Tonkin et Doherty, 2009). This method can handle complex systems with a large number of free parameters and calibrate them ~~against~~ using on-site measured data. Among the many statistical methods available, the Generalized Likelihood Uncertainty Estimation (GLUE), developed by Beven and Binley (1992), is implemented in the COUP model (Jansson, 2012) and has been used in the present study. GLUE assesses the equivalence of a large number of different parameter set-ups stochastically selected among a given set of parameter value ranges. It is based on the premise that any model set-up is, to a certain extent, in error with reality (Morton 1993). Assigning a likelihood to any model set-up will allow the selection of the most correct one within the number of ~~the~~ tested sets of model parameters. The probability of getting a result with reasonable likelihood increases with the number of simulations, especially ~~in~~ for a complex system with a large number of parameters. ~~Moreover, e~~Expert knowledge of the system is required to (a) select the parameters to test and (b) to define their ranges in order to minimize the error sources resulting from physically intercorrelated parameters, autocorrelation, insensitive parameters and heteroscedasticity (sub-populations that have different variabilities from others which invalidate~~ing~~ statistical tests) in the residuals (Beven and Binley, 1992). However, a large number of simulations with different sets of parameters may also raise the equifinality problem:~~;~~ several model set-ups can lead to an acceptable calibration (Beven and Freer, 2001) which may lead to an uncertainty into the prediction. For example, two model set-ups giving the same likelihood during the calibration process could lead to different results when used for long-term simulations.

~~The~~ In addition, a model set-up which is consistent with present day conditions may not be optimal for future climatic conditions. This well-known problem is inherent to most long-term transient simulations with a high number of parameterised and calibrated processes. One possibility to avoid compensation of two or several parameters showing unphysical or unrealistic values is to (i) constrain the parameter range to physical plausible values and (ii) verify if the obtained calibration values for all parameters contain any outliers, which cannot be explained by site-specific conditions. However, it has to be noted that the aim of the calibration procedure is not ~~to get a~~ the ~~physical~~ determination of the parameter values (e.g. physical properties) themselves, but to get a model that is thermally most representative ~~of~~ for the ground thermal regime at a given site. Keeping the above constraints in mind, for long-term simulations, where no observations are available, it has to be assumed that the parameters governing the ground thermal regime do not change significantly over the duration of the simulation. ~~Therefore, the value range of the tested parameters may be beyond the values found in literature but this is necessary to report and compensate some imprecision and the incompleteness in the model structure as well as the under-determination of other parameters.~~

~~In the present study, w~~We selected 14 parameters ~~that have no on-site measured value and~~ that have either shown a large influence on modelled temperature variations in ~~a preliminary analysis~~ previous studies and/or are known to be important in reality (cf. Lütschg et al. 2008, Schneider et al. 2012, Scherler et al. 2013, 2014, Gubler et al. 2013, Marmy et al. 2013). The 14 parameters (listed in ~~the~~ Table 2) were tested for each site in a first iteration of 50’000 simulations. Each of the simulations was run with stochastically ~~different~~ selected parameter values~~, selected stochastically,~~ creating thus 50’000 different model set-ups. The most sensitive model parameters were then identified for each site based on their relative importance on ~~the~~ calibration performance~~s of calibration~~ (Figure 4 and 5). Those four to six sensitive parameters were then used in a second iteration of 20’000 simulations to refine the calibration. It is important to note that the sensitive parameters may differ from site to site depending on site-specific characteristics, although ~~the~~ initial parameters ~~tested and their~~ and their ranges~~s~~ were equivalent for all ~~the six~~ sites~~, the sensitive parameters may differ from site to site depending on site-specific characteristics~~ (see next section). From the 20’000 simulations of the second GLUE calibration iteration, an optimal ~~five~~ model set-up~~s~~ for each site ~~were~~ was then selected based on statistical performance indicators (r2 and the mean error, ME) for ground temperature at several depths. The calibration procedure is summarized in Figure 4. In addition to useful information about site-specific processes and their representation in the model (see next section), t~~T~~he calibration obtained by this method led to the selection of a model set-up~~, the closest to the optimum,~~ for the long-term simulations forced by the GCM/RCM data.

5. Calibration results

5.1 Relative importance metrics

The GLUE method was used to test a large number of parameters at each site and to statistically assess their relative importance in the model. The relative importance of each parameter in the model is calculated based on the standardized covariance matrix of the tested parameters and related model performances using the LGM (Lindemann, Gold and Merenda) method (Lindeman et al., 1980) that averages the sequential sums of squares over all orderings of regressors. We group the parameters into six categories: 1) Snow parameters ($\mathbf{T_{rain}}$, $\mathbf{T_{snow}}$, $\boldsymbol{\rho_{snowmin}}$, $\mathbf{S_k}$, $\mathbf{Melt_{rad}}$, $\mathbf{Melt_{temp}}$, $\mathbf{\Delta S_{crit}}$), 2) Albedo parameters ($\boldsymbol{\alpha_{dry}}$, $\boldsymbol{\alpha_{wet}}$), 3) Hydraulic conductivity ($\mathbf{k_w}$, $\mathbf{g_m}$), 4) Porosity ($\mathbf{\Phi}$), 5) Thermal conductivity ($\mathbf{K_{soil}}$) and 6) Evaporation ($\boldsymbol{\psi_{eg}}$) and evaluate the influence of each parameter group on the statistical performance indicators r2 and ME at three different depths (near-surface, around 10 m, and the maximal depth of each borehole). The r2 accounts for variance whereas ME accounts for absolute errors. This joint analysis of correlation and mean error is needed, as small temperature biases near the freezing point may result in large errors when latent heat processes are not adequately represented. While minimising the ME ensures that the absolute values are near the observed ones, the correct simulation of the timing of freeze/thaw events can be improved by maximising the correlation coefficient. It is clear that in the case of long lasting freeze/thaw events a good correlation will always be difficult to achieve, but a reasonable good match was achieved at least for the near-surface layers by optimising the correlation. In a future step, other quantities such as the energy content of the ground (Jafarov et al. 2012) could be used for calibration, but also variables directly related to the amount of water present (see section 7) can be used to enhance the calibration.

The results of the calibration are shown in Figure 5 (left). Hereby, the relative importance of the six groups of parameters are shown for the three different depths as well as the absolute importance of the varying parameters on the simulations results (in %). A large relative importance identifies a parameter or process as being dominant with respect to the other parameter groups, however, it can still have a low overall importance on the simulation results, if the absolute importance is low.

As ~~expected~~already noted in many previous studies (e.g. Lütschg et al. 2008, Gubler et al. 2013, Scherler et al. 2013, Atchley et al. 2016), Figure 5 shows that the snow parameters have the greatest importance on the calibration performance for all sites. This importance is obviously pronounced at the surface, as the snow conditions ~~are~~ represent a large part of the upper boundary condition by influencing the ground surface temperature during the snow-covered period. The variation of ~~the~~ r2 at the surface is explained by snow with a relative importance usually above 50 %, ranging from 34 % at RIT up to 72 % at COR and 90 % at LAP. ~~These~~ The differences between the sites ~~are~~ can be explained by different snow conditions: there is a mean of about 280 days with snow cover per year at RIT whereas COR only has about 200 days of snow cover days per year, ~~meaning that the more snow there is~~indicating ~~, the smaller its~~ that the relative importance of snow ~~in the~~ for model calibration decreases with increasing snow cover. Hence a site with long and ~~high~~ thick snow cover is less sensitive to variations in the snow parameters ~~(and therefore more difficult to calibrate)~~, as the snow persists anyway during a long period, than sites with less snow and a faster transition between snow covered and snow-free ground. At~~T~~ LAP, ~~the~~ snow cover conditions are ~~unique because of~~ additionally influenced by the presence of ski tracks and frequent occurrence of avalanches (~~see~~ cf. Staub et al., 2015).

-In comparison with ~~the~~ r2, ~~the~~ ME is less influenced by snow parameters, as snow cover is ~~less~~ more important for temporal temperature variability (i.e. by accurately reproducing the transition between snow-covered and snow-free ground) than ~~in regards to the~~ for absolute ~~surface~~ temperatures values. ~~than the temporal variation of temperature, i.e. by accurately reproducing the transition between snow-covered ground decoupled from atmosphere and snow-free ground~~. Interestingly, the relative influence of snow on ground temperatures is still large ~~even~~ at greater depths:, ~~the~~ snow explains 12 % of the r2 and 10 % of the ME at MBP at 17.5 m; 65 % of the r2 and 43 % of the ME at RIT at 30 m, 19 % of the r2 and 54 % of the ME at SCH at 13.7 m; 8 % of the r2 and 8 % of the ME at STO at 98.3 m. At 57.95 m at COR, the snow shows a very limited influence as it explains only 0.1 % of the r2 and 0 % of the ME at this depth. This is probably related to the thick model layer with high porosity (~~62 % / 43 %,~~ cf Figure 3), where massive ice is permanently present ~~due to negative temperatures. Such a thick ice layer~~ which decouples the lowest layers from processes ~~of~~ at the upper boundary.

The albedo parameters have a significant influence on the calibration results at all sites, ~~especially regarding temperature amplitudes at the surface~~ with relative importance for the ME ranging from 17 % at SCH to 74 % at LAP, reflecting the calibration of the surface temperature amplitudes. The r2

(~~variability~~reflecting the inter-seasonal variation) is less or not influenced by the albedo. ~~Indeed, the albedo effect only occurs during snow-free period, changing the amount of heat kept by the uppermost layer therefore influencing the absolute value of temperature but not really the inter-seasonal variation.~~ In some cases at greater depths (r2 and ME at 10 m at COR, 8 m at LAP), albedo appears to have a high relative influence, sometimes higher than ~~its importance~~ at the surface. This is most likely not related to realistic physical processes~~ in reality~~: ~~those~~ intermediate depths, which are located between the well calibrated upper and lower boundary conditions, are difficult to calibrate with any of the parameters tested (see the low percentages of absolute importance in Fig. 5), ~~which are located between the well calibrated upper and lower boundary conditions~~, ~~are difficult to calibrate with any of the parameters tested (see the low percentages of absolute importance).~~ Therefore, those values are interpreted as statistical artefacts.

~~The evaporation only has a very limited importance, with values between 0 and 10 % for all the sites. As an exception, RIT is sensitive to changes in evaporation as it influences both the r2 and the ME values at the surface and at greater depth (up to 44 %).~~

The sum of the influence of snow~~ parameters~~, albedo ~~parameters~~ and evaporation parameters ranges from 58 (SCH) to 100 % (RIT) near the surface, from 26 % (COR) to 97 % (RIT) at medium depth and from 7 (STO) to 96 % (RIT) at larger depth for ~~the~~ r2. This highlights the major role played by the upper boundary condition in the calibration. LAP and COR are exceptions as the importance of the upper boundary parameters is high at the surface (90 % for r2 and 97 % for ME at LAP and 78 % for r2 and 86 % of the ME at COR) but negligible at ~~19.6 m (LAP) and 57.95 m (COR)~~ larger depths, where ~~. At this depth at LAP, the~~ variation of ~~the~~ r2 and ~~the~~ ME is mainly due to variation of the thermal conductivity. The model needs to broadly tune the thermal conductivities (between 0.3 and 2.5 W/m*K) of certain layers (10-15 m) to correct the temperature ~~at this depth~~ where a missing process or an incorrect soil structure parameterization need to be corrected. LAP and COR are ~~2~~ two ice-rich sites (as seen e.g. in the geophysical results by Hilbich 2009, Hilbich et al., 2009), with large blocks at the surface and high ~~modelled~~ porosity.~~:~~ ~~t~~The combination of these ~~three~~ effects decouples ~~more~~ the intermediate ~~depths~~ layers from the upper boundary conditions to a larger extent than at MBP and RIT, which are also ~~talus slopes / rock glaciers~~sites with coarse-grained material, but with a smaller ~~modelled~~ estimated porosity by the model (Figure 3).

Not surprisingly, t~~T~~he thermal conductivity plays a large role at depth where ~~its~~ the relative importance (ME) ranges from 34 % at SCH ~~(ME, 13.6 m)~~ to 89 % at STO ~~(ME, 98.3)~~ and ~~, as mentioned before,~~ even 100 % at LAP ~~(r2 and ME at 19.6 m)~~ (an exception is ~~.~~ COR ~~shows~~ with ~~a relatively small sensitivity of changes in the thermal conductivity (~~11 %). ~~of the ME at 57.95 m).~~ At MBP, thermal conductivity plays a large role~~,~~ even at the surface (67 % of the r2). As mentioned above, a potential radiation bias could be present in the input data of MBP due to the absence of on-site measured global radiation. A compensation of a potential bias would be expected either in the near-surface thermal conductivities or in the albedo values. Although the critical snow height parameter ΔS_{crit} for MBP is very low, indicating a potential model compensation as it affects the albedo calculation, the albedo values themselves were calibrated with average values (α_{dry} = 24.1%, α_{wet} = 19.1%), which rather points to the absence of a large radiation bias. Similarly, the calibrated thermal conductivity values for the near-surface layer are about average (around 2-4 W/m*K) compared to the other sites and do not indicate a large bias towards too warm radiation based surface temperatures.

~~Only~~ Among all sites, only RIT is insensitive to changes in the thermal conductivity (3 % of ME at 30 m depth). ~~At this site, opposed to the others~~On the other hand, evaporation ~~at the soil surface~~ has a strong influence on calibration~~,~~ even at larger depths (44%), which is in strong contrast to all other sites, where this parameter shows only little influence (between 0 and 10 %). When analyzing the specific values obtained for the different calibration parameters, the parameter related to evaporation (water tension Ψ_{eg}, cf. Table 2) did not show specifically high or low values for RIT, but the parameterized values for T_{snow} (minimum temperature at which precipitation falls only as snow) and ΔS_{crit} (critical snow depth, at which the whole surface is considered to be covered by snow) were very low (T_{snow} = -4.86°C) and high (ΔS_{crit} = 1.9 m), respectively. Whereas the former leads to comparatively large precipitation input as rain, the latter leads to an almost never completely snow-covered surface. In addition, the wet soil albedo for RIT is calibrated with the lowest value of all sites (α_{wet} = 7.0) whereas its dry albedo is comparatively high (α_{dry} = 34.6). In total, this parameter combination enables additional energy input by liquid water into the subsurface, which of course also explains the high sensitivity to evaporation. Even though this parameter combination may lead to an unrealistic process representation in the model, it is still in good accordance with observations, as ~~This is most likely due to~~ at RIT the effect of 3-d advective water flow from the melting snow cover has been observed in borehole temperatures (Zenklusen Mutter ~~Luethi~~ and Phillips, 201~~5~~2, Luethi ~~and~~et al. 2016), which explains this specific calibration outcome. Of course, the real 3D-process of melt water infiltration ~~. Therefore, the evaporation parameter should to be adjusted during the calibration process in order to compensate the effect of lateral water circulation~~, ~~a process that is not~~ cannot be explicitly included in ~~the~~ our model.

~~The calibration of the p~~Porosity and ~~the~~ hydraulic conductivity of different horizons show little or no influence on calibration performance. For ~~the~~ porosity~~,~~ this is not surprising as the parameter ranges ~~tested are~~ are narrow to keep porosity close to reality. The only site showing sensitivity of changes in porosity is MBP (20 % of importance for the r2 at 10 m and 19 % at 17.5 m), ~~. This site is mostly~~ which is specifically sensitive to changes of the porosity of the 2nd soil layer (1.6 to 3.6 m depth). This ~~could~~ points ~~be due~~ to an imprecise initial soil structure set-up that the model needed~~s~~ to correct, in this case the thickness of the surface blocky layer with high porosity~~.~~

When considering the absolute importance (% in Figure 5, left), we notice that deep boreholes (COR, RIT and STO) have low percentages, which is not surprising as the temperatures at those depths ~~are less prone to variance~~ vary on much longer time-scales and depend ~~more~~ primarily on the structural set-up of the model. As their future evolution is influenced by past climates, which are not included in the present study, simulated temperature changes at large depths will not be discussed within this study. However, their correct representation for present day climate is important as lower boundary condition for shallower levels. ~~For all sites~~Contrary to these deep levels, the surface at all sites shows the highest sensitivity to the tested parameters mainly due, as explained above, to the high importance of snow and albedo parameters. ~~At RIT, the r2 at the surface is not easily calibrated with the current set-up, because the r2 is relatively constant for all tested set-ups: as the temperature is very close to the surface (0.1 m), the meteorological forcing is the main driving factor for the r2 at this depth. However, even if near-surface temperature variation depends mainly on the meteorological conditions, the ME can be tuned as some parameters (e.g. albedo, evaporation) are able to modify the energy balance at the surface. At several places (LAP, MBP) the absolute importance is low around 10 m and this is because this intermediate depth is harder to calibrate than the upper and the lower boundary conditions.~~

After the LGM analysis, the most sensitive parameters for each site were identified to be used in the second iteration of the GLUE calibration procedure (cf. Figure 4) to refine the calibration. ~~focusing on sensitive parameters.~~ The parameters listed in Figure 5 (right), are the four to six most important parameters in the variation of statistical indicators; their relative importance for the variation of ~~the~~ r2 and

~~the~~ ME at three different depths is represented by the pie charts. One parameter that shows high sensitivity is ~~the~~ ΔS_{crit} ~~(threshold snow height for the snow to be considered as fully covered)~~ which allows the model to correct for the imprecise snow conditions and systematic biases in the building of the snow cover. The biases regarding the disappearance of the snow cover in early summer are corrected by the parameter $Melt_{rad}$ (coefficient for the importance of global radiation in the melt function of the snow). The thermal conductivity (k_{soil}) is important to adjust temperatures at middle and lower depth (COR, LAP, SCH and STO) but also at the surface (MBP). It can also be seen that snow parameters (blue colors in the pie diagrams) have stronger influence at the two bedrock sites (SCH, STO) compared ~~with~~ to talus slopes and rock glaciers (COR, MBP, RIT, LAP), where other processes such as advection, convection and latent heat processes (due to the higher ice content) play a major role at depth.

5.2 Ground temperatures

To identify the most accurate runs among the 20’000 runs of the second iteration, we apply a selection based on two balanced criteria: i) selecting the runs with the highest r2 (= seasonal and interannual variability) in layers close to the surface and ii) reducing as much as possible the ME (= model temperature bias, leading to a globally too warm or too cold model) at greater depth. This option has been preferred over a globally best r2 or ME averaged over all depths because ~~this option~~ the latter would put the weight ~~consider~~ equally to all depths, whereas the surface is more important (and more accurate) regarding decadal changes.

~~To address the potential problem of equifinality during long-term simulations (section 6), we select an ensemble of five model set-ups giving similarly good calibration performances and analyzed whether they differ significantly regarding long-term simulation results and calibrated values. This analysis showed that we can exclude that the equifinality plays a major role in our study.~~

~~The Figures 6a, 6b and 6c~~ Figure 6 shows the performance of the calibration at each site ~~, comparing the measured and the simulated temperatures at~~ for three different depths, ~~also~~ indicating the obtained value of ~~the~~ r2 and ME. It has to be pointed out that a low r2 or high ME value does not mean that a better result at a certain depth cannot be obtained by GLUE, because ~~this~~ the selection process is a compromise between r2 and ME at several depths. Most calibration runs produce either well calibrated temperatures near the surface or at greater depths, but not both for the same set of calibration parameter.

~~The c~~Calibration at the surface is very good at LAP and STO ~~with~~ (r2 ~~higher than~~ > 0.8), indicating a ~~meaning that~~ good representation of the upper boundary condition, especially ~~the~~ regarding snow timing and duration. ~~could be well represented.~~ At the four other sites, ~~the~~ r2 at the surface ranges between 0.65 (COR) and 0.77 (MBP). The comparatively low values at COR are not surprising due to the presence of very coarse blocks (>2 m) at the surface inducing additional processes in the active layer that influence the near-surface sensors in the borehole (~~cf.~~ Scherler et al., 2014). The general variation and absolute values of near-surface ground temperature is satisfactory. Some systematic mismatches ~~notwithstanding~~ exist, such as insufficient cooling during winter at SCH and LAP, or excessive cooling in winter at MBP. At MBP, this is compensated by an equally high excessive warming during summer. At STO, the general behavior of the near-surface temperature is accurately reproduced by the calibration, but ~~both the winter temperatures and the summer temperatures appear as smoothed~~ with a reduced amplitude (warmer in winter and cooler in summer). At COR, there is an insufficient warming in summer, leading to a negative bias at the surface.

Temperatures at or around 3 m are the most challenging to calibrate as the influences of the upper and the lower boundary conditions have to be balanced. Moreover, this depth is ~~often~~ usually within the ~~seasonally thawing~~ active layer and a small error in temperature (and/or soil water content) will lead to a mismatch in active layer thickness (e.g. at STO). Without putting a specific focus on matching the active layer thickness, the transition between ~~the~~ frozen and unfrozen conditions is difficult to reproduce, especially given that subsurface structure and composition is generally unknown. The selection process showed that the selection of the best r2 at this depth led to the introduction of a strong positive bias in the absolute value (leading to disappearance of permafrost) and to poor calibration results at lower depth. ~~This is why we put special emphasis on the~~ A reduction of the ME at this depth led to a better representation of the permafrost conditions at all sites, but ~~rather than having a high r2 in order to be close to the mean annual ground temperatures. As~~ as a consequence~~,~~ the seasonal variations at this depth could not always be reproduced.

~~They are~~ Seasonal variations at this depth are only reproduced correctly ~~only~~ at MBP (low ME and high r2) and, to a certain extent, at COR and SCH (cf. Figure 6). At SCH a warm bias is introduced in the model at 3 m depth because of an ~~insufficient cooling during winter~~. ~~This~~ The warm bias at SCH ~~is~~ can be explained by insufficient cooling during winter at the surface ~~the bias introduced at the upper boundary~~ which propagates to larger depths. At COR, the warm bias at the surface ~~in~~ is not reproduced at 3.55 m due to the permafrost conditions at this depth in the model. At RIT and STO, the model shows a constant temperature at the freezing point, leading ~~therefore~~ to a large positive bias (1.74 K at RIT and 1.32 K at STO). At LAP, the model also shows temperatures at the freezing point at 3.6 m, and it is able to reproduce some seasonal variations only at the end of the calibration period. The bias at LAP is slightly positive (0.25 K).

~~The c~~Calibration of the lowermost layer is always satisfactory even though the model shows a ~~slight~~ small positive bias at STO (0.56 K), RIT (0.29 K) (probably originating from the propagation of the warm bias at 3 m) and MBP (0.28 K), and a negative bias at SCH (-0.25 K).

Even if the calibration resulting from the GLUE procedure is not always satisfactory, it represents the optimal set-up for the given initial model for each site under the constraints of this semi-automated calibration approach presented in this study.

6. Long-term simulations

One of the goals of any calibration is to get a suitable ~~setoff~~ set of model parameters to be used in further analysis. ~~In~~ The TEMPS project ~~,~~ had the overall goal ~~is the~~ to investigat~~ion~~e ~~of~~ the present and possible long-term evolution of mountain permafrost in Switzerland. Hence, ~~we~~ the calibrated ~~the~~ model set-ups for each ~~of the six~~ sites ~~as explained in the previous section to simulate the possible thermal evolution of permafrost until the end of the century. For this, the model was~~ were forced with ~~the~~ downscaled and bias corrected climate model output data from 13 GCM/RCM chains as explained in section 3.1~~2~~. The corresponding changes of the two main meteorological driving variables air temperature (see Figure 2) and precipitation are summarized in ~~the~~ Table 3.

Figure 7 shows the simulated evolution of ground temperature at 10 m and 20 m, both as mean of 13 scenario simulations for each site as well as the corresponding ensemble range. The chosen depths show permanently frozen conditions during the observation period (cf. ~~also~~ PERMOS 2013) but are subject to thaw in a climate warming perspective. Because the calibration procedure identified implausible combinations of parameter values for RIT (due to 3D advective processes as described by Luethi et al.

2016), which may lead to erroneous projections for the future, no long-term projections are shown for this site.

At all sites, the 10 m layer ~~(Figure 7a, 7b and 7c, top)~~ is projected to be unfrozen by the end of the century, but there is a considerable difference regarding the timing between the sites. Moreover, there is uncertainty among the 13 different GCM/RCM chains (grey area in Figure 7). The 10 m layer is projected to become unfrozen between the decades 2060 and 2090 at COR, 2030 and 2060 at LAP, 2020 and 2030 at SCH and 2010 to 2060 at STO. ~~At RIT, it is projected to have seasonal positive temperatures at 10 m already from model year 2010 onwards, although the active layer is presently around 3 to 4 m deep in reality. This artefact comes from the positive bias introduced during calibration which is reproduced and enhanced when run with the climate model data.~~ At MBP, the 10 m layer is projected to be unfrozen ~~from around~~ by 2080 for certain chains but remains frozen until the end of the century for other chains.

Once its ice has permanently melted, the 10 m layer is subject to ~~undergo~~ significant seasonal variations (see COR, RIT and STO). SCH is not as much affected by the seasonal variations although the layer is projected to be unfrozen early in the century because of a smaller decrease in snow cover duration in comparison with other sites. In addition, its permafrost degradation is less pronounced than projected in Scherler et al.~~,~~ (2013). This is most probably due to the cold bias introduced during the calibration and to a slightly higher porosity value at depth (7 % as opposed to 5 % in Scherler et al., 2013), leading to higher ice content and therefore a slower degradation. Note as well that the air temperature warming at SCH is the lowest (+3.36K, see Table 3) compared to other sites.

At LAP, COR and MBP, the soil is projected to ~~still be~~ remain frozen at 20 m ~~at~~ until the end of the century ~~(Figures 7a, 7b and 7c, bottom)~~. At ~~COR,~~ SCH and STO, some chains project a thawing, occurring around ~~2069 at COR,~~ 2080-2090 ~~at SCH and 2085 at STO~~, while other chains project negative temperatures at 20 m until the end of the century.

As ~~already~~ mentioned above, the snow cover duration is one key element for the evolution of the ground thermal regime ~~is the snow cover duration~~. Its evolution in the future is expected to be mostly influenced by changes in air temperature: the changes in the annual sum of precipitation are highly uncertain and do not generally exceed ±5 % in the GCM/RCM output (see Table 3, ~~) (~~but with high variability among the chains), while the simulated mean change in snow cover duration ranges from -20 % (SCH) to -37 % (LAP). Figure 8 shows the relationship between the air temperature ~~warming~~ increase and the decrease in snow cover duration. For all sites, the correlation is linear and the trend of snow cover duration decrease~~s~~ per degree of warming ranges from -5.98 d/K (COR) to -8.76 d/K (LAP). This decrease represents a shortening of the snow cover duration ~~shortening~~ of 48 days (COR) to 88 days (LAP) until the end of the century. ~~The differences in the trend are due to site specific model set-up characteristics as the slope, the snow parameters (e.g. snow density) or differences in other meteorological variable (e.g. global radiation).~~ The range of the different GCM/RCM chains is broad, confirming the high uncertainty and the general difficulty in ~~predicting~~ projecting the evolution of precipitation.

7. Discussion

7.1 Approach

The GLUE calibration method is not meant to determine the physical value~~s~~ of a parameter. The model is physically-based regarding its underlying equations, but has to rely on parameterisations for many of the complex processes in the subsurface and at the soil-snow-atmosphere boundary. ~~realistic in the equations, but the~~ The values for all model parameters at all depths cannot be known exactly, especially as almost no direct measurements of these properties are available. ~~and~~ The ~~the~~ GLUE method enables to find ~~gives the ability of finding~~ the value which gives the best fit with ~~the~~ observations within the number of tested runs. But as the system is complex, with sometimes highly uncertain initial and boundary conditions, non-linear processes and simplifications of the model structure~~al~~ ~~errors~~ make an optimum calibration impossible (Beven, 2002). It is therefore more ~~relevant~~ meaningful to analyze the residuals and the sensitivity to parameters than the values of the parameter themselves.

The calibration with GLUE depends on several subjective initial assumptions: a) choice of tested parameters and their range: this choice has to be made by the modeler prior to the calibration and is a result of previous tests to identify relevant and sensitive parameters and, b) the choice of criteria of acceptance. For the former, we tried to include a representative set of parameters for surface processes (snow, albedo, evaporation), subsurface processes (thermal and hydraulic conductivity) and properties which are characteristic for the specific geomorphological sites (porosity) in order to provide enough degrees of freedom for a satisfactory calibration. In addition we used our prior experience with CoupModel (cf. Engelhardt et al. 2010, Scherler et al. 2010, 2013, 2014, Marmy et al. 2014, Staub et al.

2015) to identify the most sensitive parameters. We tried to fix the allowed parameter range to physically plausible ranges and verified that the obtained values during calibration were not distributed at the limits of these ranges. Regarding the choice of criteria of acceptance, we gave priority to good correlation coefficients near the surface and at intermediate levels while making sure that mean errors were acceptable at all depths. Here, different simulation results would have been obtained by e.g. giving more weight to intermediate levels, however, due to the uncertainties regarding the influence of past climates at the lower boundary and regarding the exact representation of temperature evolution near the freezing point, the results would be less certain than in the case of a well-calibrated model at the upper boundary. ~~The results of the calibration would have been different with different assumptions. Moreover, the~~ Finally, uncertainties of the calibration add themselves to the uncertainties of ~~the~~ observation and ~~to the uncertainties of the~~ climate models when considering the long-term simulations.

7.2 Calibration

One challenge of the calibration with GLUE is that there are many parameters to calibrate which are often underdetermined with respect to the available data. Therefore, the optimum is sometimes poorly-defined especially for sites that include processes like 2-d air circulation which is not taken into account in the present model formulation. According to Beven (2002), an increased physical realism of the model structure does not aid in obtaining a better calibration. The perfect model would include an ~~semi-infinite~~ extremely large number of parameters and be unique to each site, and this is of course unrealistic.

In comparison with other permafrost modelling studies (e.g. Scherler et al., 2013; Westermann et al., 2013; Fiddes et al., 2015), the calibration method reaches a satisfactory calibration level for most of the sites. The obtained biases in the calibration may originate from several phenomena (which are very likely linked): a) neglecting a sensitive model parameter in the calibration process~~the set of tested parameters missed one or several calibration parameter(s) that would have been sensitive regarding the calibration performances~~, b) too narrow parameter ranges which do not allow to reach the global optimum~~the ranges of the parameter tested were not large enough and did not allow the model to sufficiently use extreme values for certain parameters to compensate for a physical process that is not included in the model~~, c) insufficient ~~the~~ number of runs ~~was not sufficient~~ to find the optimum for each site, d) errors regarding the initial model structure (soil type, horizons, …), e) biases introduced in the reconstruction of the input meteorological data or f) errors or imprecision in temperature measurements.

In addition, ~~Regarding point b), we particularly think about~~ several potentially relevant processes such as convective flow of air in the coarse blocky layer or 2-d air or water circulation ~~which is~~ are not included explicitly in the COUP model. In a previous study this was solved by artificially creating a heat source/~~ or~~ sink to reproduce convection within the coarse blocky layer of rock glacier Murtèl (Scherler et al., 2013). A similar parameterization for advective water flow within the SNOWPACK model has been ~~proposed~~ published by ~~Luethi and Phillips~~Luethi et al. (2016~~5~~) for Ritigraben.

~~In the present study, we focused on finding a semi-automated calibration method to simulate the long-term evolution at many sites. In this context, it would be unclear how these processes (and the depths where they are active), and therefore the specifications of heat source/sinks, would evolve over time in a degrading permafrost.~~

Other processes that were not taken into account in the model concern the snow redistribution by avalanches or by wind that often takes place in high mountain environments (Hoelzle et al., 2001; Lehning et al., 2008; Mott et al., 2010, Gisnås et al. 2016). ~~Snow has been shown to have a strong influence on the ground thermal regime both in field measurements and in the soil models because it is a boundary decoupling the atmosphere and the soil (e.g. Ling and Zhang, 2003; Zhang, 2005; Hoelzle and Gruber, 2008). In this study~~ However, we could quantify the influence of several snow parameters. Snow has an especially strong influence at sites with shorter snow cover duration: there it is the most important parameter for the variations at the surface, but it also has a strong influence at deeper layers. The sites with a long-lasting snow cover (RIT and MBP) showed a reduced sensitivity to snow parameters as the snow is present most of the time and the transition between snow-covered and snow-free conditions is less difficult to simulate. ~~This is in accordance with the analysis of Zhang (2005) who showed a strong influence of snow on permafrost, especially in warmer regions. The influence of snow has already been pointed out by many authors.~~ In general, the definition of the upper boundary conditions (snow, albedo, evaporation) appears to be a crucial issue as they influence the performance of the calibration of the whole soil column.

Facing the scarcity of measured data, it is difficult to check whether the calibration obtained by the semi-automated procedure is robust for outputs other than temperature. ~~However, there are p~~Possibilities exist to validate the calibration with electrical resistivity data (related to water/ice content) or direct soil moisture data but a thorough analysis of the quality of the present calibration or a calibration improvement by including these data in the calibration routine ~~this is~~ would be beyond the scope of this paper, especially as ~~these~~ data do not exist for all ~~modelled~~ sites. ~~Efforts are currently~~ First tests have ~~being~~ been made in this direction ~~on the site level~~at STO, with promising results of a joint calibration using temperature and electrical resistivity data ~~at STO~~ (Python, 2015). Efforts are also currently being made towards the installation of a soil moisture network in mountain environments (SNF project SOMOMOUNT, http://p3.snf.ch/project-143325~~provide again the P3-link~~). ~~These~~ Soil moisture and geophysical monitoring data could then serve as additional validation of the thermal calibration (Pellet et al. 2016) as shown for the example of SCH (Figure 9~~9~~). Figs. 9~~9~~A and B show the soil moisture output of the model set-up giving the best fit with observed temperatures in comparison with on-site measured data ~~(Figure 9 A and B)~~ that stem from soil moisture sensors adjacent to the borehole (see Hilbich et al., 2011). Although some biases are present, like the absolute value of the maximal peak in early summer (about 10 % mismatch at 12 cm), the absolute minimum during winter (about 7 % mismatch at 12 cm), or the stable summer maximum at 60 cm, the general behavior is well reproduced: the mean values and the timing of freezing-thawing is satisfying. In a second step, we manually calibrated the soil physical parameter ~~wilting point (soil water content which a plant dies,~~ used in the water retention curve to define the minimal residual water, which has also ~~). The wilting point is part of the water retention function and has~~ a notable influence on the freezing-point depression. By this, the agreement with measured soil moisture was substantially improved (Figure 9~~9~~, C and D) showing that model calibration can easily be improved ~~further~~ if additional data sets are available. Figure 10 shows the resulting temperature difference at 10 m depth in the long-term simulations between the improved and the reference run indicating colder temperatures (~0.3K) and later permafrost degradation at 10 m depth compared to the reference run.

7.3 RCM-based simulations

Given the various sources of uncertainty mentioned above and the choice of only one emission scenario (A1B) in the climate simulations, the results of the long-term simulation should not be considered as a prediction but rather as a projection of the range of the possible ~~future~~ evolution of permafrost in the Swiss Alps under a given emission scenario. Our long-term simulations showed that the permafrost evolution is ~~more~~ strongly influenced by the specific regional climate scenario applied (i.e., ~~b~~ the specific

GCM/RCM chain) ~~than~~ but also by differently calibrated CoupModel set-ups. Climate scenario uncertainty appears to be the dominant component of uncertainty in this study.

A similar climate impact study has been carried out by Scherler et al.~~,~~ (2013), but with a different calibration procedure of the CoupModel and a different RCM downscaling technique for SCH and COR. In comparison to their results for SCH, the timing of permafrost degradation at 10 m around 2020-2030 and the moment when the entire seasonal thaw layer cannot refreeze anymore in winter is modelled similarly, but the consecutive warming after the start of degradation is smaller in the present study. Similarly, the 20 m layer shows a rapid degradation in Scherler et al.~~,~~ (2013) whereas it remains below the freezing point for most of the GCM/RCM chains in the present study. The discrepancies are mainly explained by a slightly different soil structure, which was part of the calibration approach in the present study. At COR, the results of Scherler et al.~~,~~ (2013) show slow warming at 10 m and at 20 m. In the present study, the warming is also slow but once the 10 m-layer is thawed, the warming propagates faster to deeper layers than in the results of Scherler et al., (2013), see Figure 11. This difference is not surprising as Scherler et al.~~,~~ (2013) manually introduced a site-specific seasonal heat sink/source to compensate for the effect of air convection in the coarse blocky surface layer. By this, permafrost was conserved longer in the model than in a model set-up without parameterized convection. In addition, higher ice contents within the rock glacier ice core were simulated in Scherler et al., (2013) than in the present study (85 % versus 62 %, cf. Fig. 3), which decelerates warming as well. On the contrary, the calibrated porosity values near the surface ~~of the present study~~ are higher in the present study (49 %) than ~~in~~ the manually calibrated values of the previous study (10 %). Porosity values in heterogeneous rock glaciers are of course always highly uncertain, but it has to be noted that the best results of the GLUE procedure were not obtained with the highest porosities for the deeper layers: during the selection process, the consideration of the r2 tended towards high porosities, but the best performances were obtained with lower porosities when considering the ME (cf. Figure 3).

In contrast to Scherler et al., (2013), the cooling effect of convection in the coarse blocky surface layer was not hard-coded by an explicit source/sink term, ~~structure of the model in the present study was not manually adapted to site-specific conditions, because the goal was to develop a calibration method that finds the model set-up with the best fit to observations in a semi-automated procedure. Therefore, the cooling of the ground by convection in the coarse blocky surface layer was only~~ but rather represented indirectly through automatic adaption of site-specific subsurface parameters during calibration, e.g. a comparatively high albedo (~25%), low critical snow depth parameter and specifically a larger porosity (see above). Nevertheless, the absence of an explicit convection parameterisation for coarse blocky subsurfaces is still the major shortcoming of the CoupModel regarding mountain permafrost applications (cf. also Staub et al. 2015) and leads~~ing~~ to a probable overestimation of the warming at this site (Figure 11). However, it is not yet clear how the cooling by convection would evolve in a context of climate change and permafrost degradation, which is why an explicit treatment of this process would be favourable compared to a static, hard-coded energy source/sink approach used in Scherler et al. (2013).

At all six sites, significant permafrost degradation is projected, driven mostly by the projected increase in air temperature during snow-free periods and the prolongation of these periods due to snow cover decrease~~its prolongation in time due to snow cover decrease~~. This is in good agreement with earlier sensitivity studies using the same model (Marmy et al., 2013) and similar studies from other regions (Etzelmüller et al., 2011; Hipp et al., 2012). In general, the sites with blocky material and higher porosity (COR, LAP, MBP) show a lower sensitivity to climate change whereas the bedrock sites (SCH and STO) tend to have a more rapid degradation. At most places, a high porosity is coupled with higher interstitial ice contents, hence requiring more energy to melt the ice and warm the ground. ~~As an exception, the simulated degradation at RIT is relatively fast despite a rather high porosity. This originates from two factors: i) this site is projected to have one of the highest decreases in snow cover duration with -20.57 % to -49.42 %, representing a period of -140 to -74 days per year and ii) this site had the greatest warm bias at depth introduced during calibration which was responsible for the overestimation of degradation in the simulation results. The warm bias in the model was created because cooling by convection is not taken into account in the model (see COR) and because at this site a seasonal intra-permafrost talik has formed in the past years due to the infiltrating melt water from the snow cover (Luethi and Phillips, 2015). The calibration is forced to create unrealistic parameter settings as the model tries to reproduce those seasonal warm anomalies. At MBP, the degradation is projected to be slow compared to other sites. This reduced degradation does not originate from the bias in the calibration (cf. Fig. 6a), but may stem from the combination of various explanatory factors: i) the mean air temperature warming is average (below +4 K), ii) this reduced air temperature increase, coupled with the mean precipitation sum that is projected to increase for most of the GCM/RCM chains leads to a smaller decrease of the snow cover (mean: -22.81 %, i.e -48 days), therefore maintaining an efficient insulation layer decoupling the ground thermal regime from the atmosphere, iii) one of the highest porosity of all the sites, and therefore a high interstitial ice~~

~~content or iv) one of the coldest initial ground thermal regimes, that could not be reproduced by the calibration.~~

Changes simulated in the snow cover duration are mostly influenced by the increasing air temperature and much less by change in mean annual precipitation sum. This is in agreement with both Wang et al., (2014) who stated that the increase in atmospheric freezing level is responsible for most cryospheric changes ~~in cryosphere~~ in the future and with Steger et al. (2013) who found that Alpine snow cover changes in the ENSEMBLES GCM/RCM chains are mostly driven by temperature increases. Our CoupModel simulations showed a decrease of snow cover duration of about -20 % to -37 %, which is in the same order of magnitude than the results by Bavay et al., (2009) who projected a mean reduction of snow cover duration of ca. 30-35 % for two alpine catchments (run under the B2 and A2 scenarios), and by Schmucki et al., (2014) who projected a decrease of snow cover of 32-35 % for high-elevation sites. These numbers are furthermore consistent with Steger et al. (2013) who analyzed Alpine snow cover changes in the ENSEMBLES climate models themselves. During the next 10-20 years this reduction of snow cover may have an opposite effect to ground warming in summer: a decrease of the snow cover in fall and early winter can lead to a cooling of the ground, because the cool winter temperature can better penetrate the ground with no or reduced snow cover. However, sensitivity studies for a whole range of air temperature and precipitation changes suggest that until the end of the century the effect of warming ~~is dominant~~ will dominate over ~~this~~ the potential cooling effect in late autumn/early winter (Marmy et al., 2013). In spring and late summer, the decrease of snow cover has always had a warming feedback because the snow is no longer present to isolate the ground from the positive summer temperatures.

The results of the long-term simulations have to be considered with caution as uncertainty may arise at several steps of the model chains: errors in the measurements used for calibration, structural errors of the model, choice of parameters and choice of their tested ranges, biases introduced during the calibration, emission scenario uncertainty or GCM/RCM chains uncertainty.

8. Conclusion

The present paper tested a semi-automated method for a soil/permafrost model calibration, in order to be able to use ~~the method to calibrate~~ it for a potentially large number of sites (e.g. in a distributed model).

Other goals were to analyze ~~of~~ the sensitivity of the model results to certain parameters, to identify site-specific processes which play a major role for the thermal regime at the individual permafrost sites, and to use the calibrated model set-ups for long-term RCM-based simulations of the permafrost evolution.

The following conclusions can be ~~derived~~ drawn from the study:

- The method of semi-automated calibration using the Generalized Likelihood Uncertainty estimation (GLUE) showed an efficient ability to reproduce permafrost conditions at several permafrost sites in the Swiss Alps: the upper boundary conditions were simulated precisely ~~and~~ whereas the absolute errors in the deepest layers were ~~below~~ within a satisfactory error ~~level~~range. The r2 at the surface ranged from 0.72 to 0.84 and the mean error at depth was usually smaller than 0.5 K except at STO and RIT
- Some site-specific characteristics, such as vertical or 2-d circulation of air (convection) or lateral flows of air and water could not be reproduced by the ~~method~~approach, hence leading to warm biases at depth.
- ~~The method was generally suitable for large-scale or long-term modelling but is not recommended for site-specific process analysis, if there are existing dominant processes which are not included in the CoupModel formulation. In these cases, manual calibration and parameterization of the missing processes have to be added.~~
- The calibration of upper boundary parameters, especially parameters related to snow cover, ~~were~~ ~~was~~once more showed ~~shown~~ to have a large influence on the calibration performance~~s~~, also on deeper ground layers. Therefore, efforts to obtain a precise upper boundary calibration must be undertaken, especially by increasing the length and the quality of surface measurements (GST, radiation, snow cover, soil moisture etc).
- The long-term simulations have shown a degradation trend at all sites, with an increasing active layer depth to at least 10 m at all sites until the end of the century, and even to 20 m at SCH and STO. However, strong uncertainty exists among the different GCM/RCM.
- The degradation is primarily driven by the change in air temperature during the snow-free period and the ~~changing~~ change in snow cover duration.

- The snow cover duration is projected to decrease by values ~~of~~ between -20~~.03~~ % ~~to~~ and -37~~.06~~ % and is mainly driven by the change in air temperature.

In general, the calibration method can be suitable for large-scale or long-term modelling but it is not recommended for site-specific process analysis, if there are existing dominant processes which are not included in the CoupModel formulation. In these cases, manual calibration and parameterization of the missing processes have to be added. In comparison to other, simpler approaches to simulate future scenarios for borehole temperatures (as e.g. in Etzelmüller et al. 2011, Hipp et al. 2012 or, regarding spatial modelling, in Jafarov et al. 2012) the approach of this study focuses more on the site-specific processes understanding, while the long-term simulation results will not necessarily be better than results from simpler approaches as in the above cited studies. But we believe that the considerably higher efforts of our approach are well justified by the knowledge gained regarding the effect of the dominant processes at the different sites. Of course, future work has to be directed into including the already identified missing processes into the model formulation (i.e. convection).

-

We believe that the method presented here can be used as a starting point for large-scale modelling of the permafrost distribution in the Alps provided that an increased number of sites with high quality data series of observed ground temperature become available. A distributed model could be derived from the numerous calibrated sites by interpolation, in combination with digital elevation models, remote sensing data, GST measurements and subsurface data from geophysical surveys. ~~As the model performance proved to be strongly influenced by the upper boundary condition, such a distributed model could also be generated from an interpolation between several GST measurements.~~

Acknowledgements

We would like to acknowledge the Swiss National Science Foundation for the funding of the TEMPS project (project no. CRSII2 136279) as well as the Swiss PERMOS network for the data provided. A special thank you to Prof. Per-Erik Jansson from Kungliga Tekniska Högskolan of Stockholm for the technical and scientific

support with the CoupModel. The critical and constructive comments of two reviewers very much helped to improve the manuscript.

References

Anisimov, O. A.: Potential feedback of thawing permafrost to the global climate system through methane emission. Environmental Research Letters, 2(4), 045016, 2007.

Anthony, K. M. W., Anthony, P., Grosse, G., & Chanton, J.: Geologic methane seeps along boundaries of Arctic permafrost thaw and melting glaciers. Nature Geoscience, 5(6), 419-426, 2012.

Arenson, L., Hoelzle, M., & Springman, S.: Borehole deformation measurements and internal structure of some rock glaciers in Switzerland. Permafrost and Periglacial Processes, 13(2), 117-135, 2002.

Arenson, L. U., Hauck, C., Hilbich, C., Seward, L., Yamamoto, Y., & Springman, S. M.: Sub-surface heterogeneities in the Murtèl-Corvatsch rock glacier, Switzerland. In Proceedings of the Sixth Canadian Permafrost Conference, Calgary, Alta (pp. 12-16), 2010.

Atchley, A. L., E. T. Coon, S. L. Painter, D. R. Harp, and C. J. Wilson: Influences and interactions of inundation, peat, and snow on active layer thickness, Geophys. Res. Lett., 43, doi:10.1002/ 2016GL068550, 2016.

Barboux, C., Delaloye, R., Lambiel, C., Strozzi, T., Collet, C., & Raetzo, H: Surveying the activity of permafrost landforms in the Valais Alps with InSAR. GRAF, C.(Red.) Mattertal–ein Tal in Bewegung. Publikation zur Jahrestagung der Schweizerischen Geomorphologischen Gesellschaft, 29, 7-19, 2013.

Bavay, M., Lehning, M., Jonas, T., & Löwe, H.: Simulations of future snow cover and discharge in Alpine headwater catchments. Hydrological Processes, 23(1), 95-108, 2009.

Begert, M.: Homogenisierung von Klimamessreihen der Schweiz und Bestimmung der Normwerte 1961-1990: Schlussbericht des Projekts NORM90. MeteoSchweiz, 2003.

Beven, K., & Binley, A.: The future of distributed models: model calibration and uncertainty prediction. Hydrological processes, 6(3), 279-298, 1992.

Beven, K., & Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. Journal of hydrology, 249(1), 11-29, 2001.

Beven, K.: Towards a coherent philosophy for modelling the environment. Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 458(2026), 2465-2484, 2002.

Boeckli, L., Brenning, A., Gruber, S., and Noetzli, J.: A statistical approach to modelling permafrost distribution in the European Alps or similar mountain ranges. The Cryosphere, 6(1), 125-140, 2012.

Bosshard, T., Kotlarski, S., Zappa, M., and Schär, C.: Hydrological Climate-Impact Projections for the Rhine River: GCM–RCM Uncertainty and Separate Temperature and Precipitation Effects. Journal of Hydrometeorology, 15(2), 697-713, 2014.

Bommer, C., Phillips, M., & Arenson, L. U.: Practical recommendations for planning, constructing and maintaining infrastructure in mountain permafrost. Permafrost and Periglacial Processes, 21(1), 97-104, 2010.

Chadburn, S., Burke, E., Essery, R., Boike, J., Langer, M., Heikenfeld, M., Cox, P. and Friedlingstein, P.: An improved representation of physical permafrost dynamics in the JULES land surface model. Geoscientific Model Development Discussions, 8(1), 715-759, 2015.

Cremonese, E., Gruber, S., Phillips, M., Pogliotti, P., Böckli, L., Noetzli, J., Suter, C., Bodin, X., Crepaz, A., Kellerer-Pirklbauer, A. and Lang, K. Brief Communication:" An inventory of permafrost evidence for the European Alps". The Cryosphere, 5(3), pp.651-657, 2011

Cui, T., Fox, C., & O'Sullivan, M. J.: Bayesian calibration of a large-scale geothermal reservoir model by a new adaptive delayed acceptance Metropolis Hastings algorithm. Water Resources Research, 47(10), 2011.

Delaloye, R.: Contribution à l'étude du pergélisol de montagne en zone marginale (Doctoral dissertation, Université de Fribourg), 2004.

Delaloye, R. and Lambiel, C.: Evidence of winter ascending air circulation throughout talus slopes and rock glaciers situated in the lower belt of alpine discontinuous permafrost (Swiss Alps). Norsk Geografisk Tidsskrift-Norwegian Journal of Geography, 59(2), 194-203, 2005.

Ekici, A., Beer, C., Hagemann, S. and Hauck, C.: Simulating high-latitude permafrost regions by the JSBACH terrestrial ecosystem model. Geoscientific Model Development, 7, 631-647, 2014.

Ekici, A., Chadburn, S., Chaudhary, N., Hajdu, L. H., Marmy, A., Peng, S., Boike, J., Burke, E., Friend, A. D., Hauck, C., Krinner, G., Langer, M., Miller, P. A. and Beer, C.: Site-level model intercomparison of high latitude and high altitude soil thermal dynamics in tundra and barren landscapes, The Cryosphere, accepted, 2015.

Endrizzi, S., Gruber, S., Dall'Amico, M. and Rigon, R.: GEOtop 2.0: simulating the combined energy and water balance at and below the land surface accounting for soil freezing, snow cover and terrain effects. Geosci. Model Dev., 7, 2831–2857, 2014.

Engelhardt, M., Hauck, C., and Salzmann, N.: Influence of atmospheric forcing parameters on modelled mountain permafrost evolution. Meteorologische Zeitschrift, 19(5), 491-500, 2010.

Etzelmüller, B., Heggem, E. S., Sharkhuu, N., Frauenfelder, R., Kääb, A. and Goulden, C: Mountain permafrost distribution modelling using a multi-criteria approach in the Hövsgöl area, northern Mongolia. Permafrost and Periglacial Processes, 17(2), 91-104, 2006.

Etzelmüller, B., Schuler, T. V., Isaksen, K., Christiansen, H. H., Farbrot, H. and Benestad, R.: Modeling the temperature evolution of Svalbard permafrost during the 20th and 21st century. The Cryosphere, 5(1), 67-79, 2011.

Fiddes, J., Endrizzi, S. and Gruber, S.: Large-area land surface simulations in heterogeneous terrain driven by global data sets: application to mountain permafrost. The Cryosphere, 9(1), 411-426, 2015.

Finsterle S., Sonnenthal E.L. and Spycher N.: Advances in subsurface modeling using the TOUGH suite of simulators. Computers & Geosciences vol.65 p.2-12, 2012.

Gärtner-Roer, I.: Sediment transfer rates of two active rockglaciers in the Swiss Alps. Geomorphology 167-168 (0), 45-50, 2012.

Gisnås, K., Westermann, S., Schuler, T. V., Melvold, K., and Etzelmüller, B.: Small-scale variation of snow in a regional permafrost model, The Cryosphere, 10, 1201-1215, doi:10.5194/tc-10-1201-2016, 2016.

Gruber, S. and Hoelzle, M.: Statistical modelling of mountain permafrost distribution: local calibration and incorporation of remotely sensed data. Permafrost and Periglacial Processes, 12(1), 69-77.

Gruber, S., King, L., Kohl, T., Herz, T., Haeberli, W. and Hoelzle, M.: Interpretation of geothermal profiles perturbed by topography: The Alpine permafrost boreholes at Stockhorn Plateau, Switzerland. Permafrost and Periglacial Processes, 15(4), 349-357, 2004.

Gubler, S., Endrizzi, S., Gruber, S., & Purves, R. S.: Sensitivities and uncertainties of modeled ground temperatures in mountain environments. Geoscientific Model Development, 6(4), 1319-1336, 2013.

Haeberli, W.: Pilot analysis of permafrost cores from the active rock glacier Murtèl I, Piz Corvatsch, Eastern Swiss Alps. Versuchsanstalt für Wasserbau, Hydrologie und Glaziologie ETH Zürich, 1990.

Haeberli, W., M. Hoelzle, A. Kääb, F. Keller, D. Vonder Mühll and S. Wagner: Ten years after drilling through the permafrost of the active rock glacier Murtèl, Eastern Swiss Alps: answered questions and new perspectives. 7th International Conference on Permafrost. Proceedings, Yellowknife, Canada, Centre d'Etudes Nordiques, Université Laval, 1998.

Harris, C., Mühll, D. V., Isaksen, K., Haeberli, W., Sollid, J. L., King, L., Holmlund, P., Dramis, F., Guglielmin, M. and Palacios, D.: Warming permafrost in European mountains. Global and Planetary Change, 39(3), 215-225, 2003.

Harris, C., Arenson, L. U., Christiansen, H. H., Etzelmüller, B., Frauenfelder, R., Gruber, S., Haeberli, W., Hauck, C., Hölzle, M., Humlum, O., Isaksen, K., Kääb, A., Kern-Lütschg, M., Lehning, M., Matsuoka, N., Murton, J.B., Nötzli, J., Phillips, M., Ross, N., Seppälä, M., Springman, S. M. and Vonder Mühll, D.: Permafrost and climate in Europe: Monitoring and modelling thermal, geomorphological and geotechnical responses. Earth-Science Reviews, 92(3), 117-171, 2009.

Hartikainen, J., Kouhia, R. and Wallroth, T.: Permafrost simulations at Forsmark using a numerical 2D thermo-hydro-chemical model. Svensk Kärnbränslehantering AB, Swedish Nuclear Fuel and Waste Management Company, 2010.

Hauck, C.: Frozen ground monitoring using DC resistivity tomography. Geophysical research letters, 29(21), 12-1, 2002.

Hauck, C., Delaloye, R., Roer, I. H., Hilbich, C., Hoelzle, M., Kenner, R., Kotlarski, S., Lambiel, C., Marmy, A., Müller, J., Noetzli, J., Phillips, M., Rajczak, J., Salzmann, N., Schaepman~~n~~, M. E., Schär, C., Staub, B. and Völksch, I.: The Evolution of Mountain Permafrost in Switzerland. In AGU Fall Meeting Abstracts (Vol. 1, p. 04), 2013.

Heerema, K., Booij, M. J., Huting, R., Warmink, J. J., van Beek, E. and Jigjsuren, O.: Modelling impacts of climate change on the hydrology of a Mongolian catchment using an appropriate permafrost conceptualization. In EGU General Assembly Conference Abstracts (Vol. 15, p. 4525), 2013.

Herz, T., King, L. and Gubler, H.: Thermal regime of coarse debris layers in the Ritigraben catchment, Matter valley, Swiss Alps. In Eighth International Conference on Permafrost, Zürich, Extended Abstracts, 2013.

Hilbich, C., Hauck, C., Hoelzle, M., Scherler, M., Schudel, L., Völksch, I., Vonder Mühll, D. and Mäusbacher, R.: Monitoring mountain permafrost evolution using electrical resistivity tomography: A 7-year study of seasonal, annual, and long-term variations at Schilthorn, Swiss Alps. Journal of Geophysical Research: Earth Surface (2003–2012), 113(F1), 2008.

Hilbich, C.: Geophysical monitoring systems to assess and quantify ground ice evolution in mountain permafrost. Doctoral dissertation, Jena, Univ., Diss., 2009.

Hilbich, C., L. Marescot, C. Hauck, M. H. Loke and R. Mäusbacher: Applicability of Electrical Resistivity Tomography Monitoring to Coarse Blocky and Ice-rich Permafrost Landforms. Permafrost and Periglacial Processes 20(3): 269-284, 2009.

Hilbich, C.: Time-lapse refraction seismic tomography for the detection of ground ice degradation. The Cryosphere, 4, 243-259, 2010.

Hilbich, C., Fuss, C. and Hauck, C.: Automated Time-lapse ERT for Improved Process Analysis and Monitoring of Frozen Ground. Permafrost and Periglacial Processes, 22(4), 306-319, 2011.

Hipp, T., Etzelmüller, B., Farbrot, H., Schuler, T. V. and Westermann, S.: Modelling borehole temperatures in Southern Norway–insights into permafrost dynamics during the 20th and 21st century. The Cryosphere, 6(3), 553-571, 2012.

Hoelzle, M., Mittaz, C., Etzelmüller, B. and Haeberli, W.: Surface energy fluxes and distribution models of permafrost in European mountain areas: an overview of current developments. Permafrost and Periglacial Processes, 12(1), 53-68, 2001.

Hoelzle, M., Mühll, D. V. and Haeberli, W.: Thirty years of permafrost research in the Corvatsch-Furtschellas area, Eastern Swiss Alps: A review. Norsk Geografisk Tidsskrift-Norwegian Journal of Geography, 56(2), 137-145, 2002.

Hoelzle, M. and Gruber, S.: Borehole and ground surface temperatures and their relationship to meteorological conditions in the Swiss Alps. In Proceedings Ninth International Conference on Permafrost, (pp. 723-728), 2008.

Jafarov, E. E., Marchenko, S. S., and Romanovsky, V. E.: Numerical modeling of permafrost dynamics in Alaska using a high spatial resolution dataset, The Cryosphere, 6, 613–624, doi:10.5194/tc-6-613-2012, 2012.

Jansson, P.-E., and Karlberg, L.: Coupled heat and mass transfer model for soil-plant-atmosphere systems. Royal Institute of Technology, Dept of Civil and Environmental Engineering, Stockholm, 2004.

Jansson, P.-E.: CoupModel: Model use, calibration and validation, Transactions of the ASABE, 55(4), 1335–1344, 2012.

Kääb, A. and Kneisel, C.: Permafrost creep within a recently deglaciated glacier forefield: Muragl, Swiss Alps. Permafrost and Periglacial Processes, 17(1), 79-85, 2006.

Kääb, A., Frauenfelder, R. and Roer, I.: On the response of rockglacier creep to surface temperature increase. Global and Planetary Change, 56(1), 172-187, 2007.

Kotlarski, S., K. Keuler, O.B. Christensen, A. Colette, M. Déqué, A. Gobiet, K. Goergen, D. Jacob, D. Lüthi, E. van Meijgaard, G. Nikulin, C. Schär, C. Teichmann, R. Vautard, K. Warrach-Sagi and V. Wulfmeyer: Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble. Geoscientific Model Development, 7, 1297-1333, doi: 10.5194/gmd-7-1297-2014

Kääb, A., Frauenfelder, R., & Roer, I.: On the response of rockglacier creep to surface temperature increase. Global and Planetary Change, 56(1), 172-187, 2007.

Kendon, E. J., Jones, R. G., Kjellström, E. and Murphy, J. M.: Using and designing GCM-RCM ensemble regional climate projections. Journal of Climate, 23(24), 6485-6503, 2010.

King, L.: Soil and rock temperatures in discontinuous permafrost: Gornergrat and Unterrothorn, Wallis, Swiss Alps. Permafrost and Periglacial Processes, 1(2), 177-188, 1990.

Lambiel, C.: Le pergélisol dans les terrains sédimentaires à forte déclivité: distribution, régime thermique et instabilités. UNIL-Faculté des géosciences et de l'environnement-Institut de géographie, 2006.

Langer, M., Westermann, S., Heikenfeld, M., Dorn, W. and Boike, J.: Satellite-based modeling of permafrost temperatures in a tundra lowland landscape. Remote Sensing of Environment, 135, 12-24, 2013.

Lehning, M., Löwe, H., Ryser, M. and Raderschall, N.: Inhomogeneous precipitation distribution and snow transport in steep terrain. Water Resources Research, 44(7), 2008.

Lepage, J. M. and Doré, G.: Experimentation of mitigation techniques to reduce the effects of permafrost degradation on transportation infrastructures at Beaver Creek experimental road site. Proc. 63rd Canadian GeotechnicalmConference, Calgary, Alberta, 2010.

Lindeman, R. H., Merenda, P. F. and Gold, R. Z.: Introduction to bivariate and multivariate analysis. Glenview, IL: Scott, Foresman, 1980.

Ling, F. and Zhang, T.: Impact of the timing and duration of seasonal snow cover on the active layer and permafrost in the Alaskan Arctic. Permafrost and Periglacial Processes, 14(2), 141-150, 2003.

Luethi R., Phillips M, Lehning M: Estimating non-conductive heat flow leading to intra-permafrost talik formation at Ritigraben rock glacier (Western Swiss Alps). Permafrost and Periglacial Processes, in press, 2016.

~~Luethi, R. and Phillips, M.: Talik formation in a rock glacier, Ritigraben, Grächen VS. Poster at TEMPS final symposium, Sion, Switzerland, 2015.~~

Luetschg, M., Lehning, M. and Haeberli, W.: A sensitivity study of factors influencing warm/thin permafrost in the Swiss Alps. Journal of Glaciology, 54(187), 696-704, 2008.

Marmy, A., Salzmann, N., Scherler, M. and Hauck, C.: Permafrost model sensitivity to seasonal climatic changes and extreme events in mountainous regions. Environmental Research Letters, 8(3), 035048, 2013.

McColl, S. T.: Paraglacial rock-slope stability. Geomorphology, 153, 1-16, 2012.

Mittaz, C., Hoelzle, M. and Haeberli, W.: First results and interpretation of energy-flux measurements over Alpine permafrost. Annals of Glaciology, 31(1), 275-280, 2000.

Morton, A.: Mathematical models: questions of trustworthiness. Bri~~s~~tish Journal for the Philosophy of Science. 44 (4): 659-674, 1993.

Mott, R., Schirmer, M., Bavay, M., Grünewald, T. and Lehning, M.: Understanding snow-transport processes shaping the mountain snow-cover. The Cryosphere, 4(4), 545-559, 2010.

Mountain Research Initiative EDW Working Group: Elevation-dependent warming in mountain regions of the world. Nature Climate Change, 5(5), 424-430, 2015.

Nakicenovic, N. and Swart, R.: IPCC—Special Report on Emission Scenarios, Summary for Policy Makers, Cambridge University Press, 2000.

Noetzli, J., Gruber, S., Kohl, T., Salzmann, N. and Haeberli, W.: Three-dimensional distribution and evolution of permafrost temperatures in idealized high-mountain topography. Journal of Geophysical Research: Earth Surface (2003–2012), 112(F2), 2007.

Noetzli, J., Hilbich, C., Hauck, C., Hoelzle, M. and Gruber, S.: Comparison of simulated 2D temperature profiles with time-lapse electrical resistivity data at the Schilthorn crest, Switzerland. In Ninth International Conference on Permafrost, University of Alaska Fairbanks (Vol. 29, pp. 1293-1298), 2008.

Noetzli, J. and Gruber, S.: Transient thermal effects in Alpine permafrost. The Cryosphere, 3(1), 85-99, 2009.

Pellet C., Hilbich C., Marmy A. and Hauck C.: Soil Moisture Data for the Validation of Permafrost Models Using Direct and Indirect Measurement Approaches at Three Alpine Sites. Front. Earth Sci. 3:91. doi: 10.3389/feart.2015.00091, 2016.

PERMOS: Permafrost in Switzerland 2008/2009 and 2009/2010. Noetzli J. (ed), Glaciological Report Permafrost No. 10/11 of the Cryospheric Commisson of the Swiss Academy of Sciences, 80pp, 2013.

Phillips, M.: Avalanche defence strategies and monitoring of two sites in mountain permafrost terrain, Pontresina, Eastern Swiss Alps. Natural hazards, 39(3), 353-379, 2006.

Python, S.: Technical improvement of the 4-phase model to better assess the ice, wáter and air content estimation in permafrost substrates. Master thesis, Department of Geosciences, University of Fribourg, Switzerland, 2015.

Quinton, W. L., Hayashi, M. and Chasmer, L. E.: Permafrost-thaw-induced land-cover change in the Canadian subarctic: implications for water resources. Hydrological Processes, 25(1), 152-158, 2011.

Rajczak, J., S. Kotlarski, N. Salzmann and C. Schär: Robust climate scenarios for sites with sparse observations: a two-step bias correction approach, International Journal of Climatology,36, 1226-1243, 2016.

~~Rajczak, J., S. Kotlarski, N. Salzmann and C. Schär: Robust climate scenarios for sites with sparse observations: A two-step bias correction approach, International Journal of Climatology, accepted, 2015.~~

Rist, A., Phillips, M. and Haeberli, W.: Influence of snow meltwater infiltration on active layer movement in steep alpine scree slopes within the discontinuous mountain permafrost. In Y. Lai, W. Ma, & S. Zhao (Eds.), Proceedings, Asian Conference on Permafrost (pp. 10-11), 2006.

Sattler, K., Anderson, B., Mackintosh, A., Norton, K. and de Róiste, M.: Estimating permafrost distribution in the maritime Southern Alps, New Zealand, based on climatic conditions at rock glacier sites. Frontiers in Earth Science, 4, http://dx.doi.org/10.3389/feart.2016.00004, 2016.

~~Scapozza, C., Lambiel, C., Bozzini, C., Mari, S. and Conedera, M.: Assessing the rock glacier kinematics on three different timescales: a case study from the southern Swiss Alps. Earth Surface Processes and Landforms, 39(15), 2056-2069, 2014.~~

Scapozza, C., Baron, L. and Lambiel, C.: Borehole logging in Alpine periglacial talus slopes (Valais, Swiss Alps). Permafrost and Periglacial Processes, 26(1), 67-83, 2015.

Schär, C., Vidale, P. L., Lüthi, D., Frei, C., Häberli, C., Liniger, M. A. and Appenzeller, C.: The role of increasing temperature variability in European summer heatwaves. Nature, 427(6972), 332-336, 2004.

Scherler, M., Hauck, C., Hoelzle, M., Stähli, M., & Völksch, I.: Meltwater infiltration into the frozen active layer at an alpine permafrost site. Permafrost and Periglacial Processes, 21(4), 325-334, 2010.

Scherler, M., Hauck, C., Hoelzle, M. and Salzmann, N.: Modeled sensitivity of two alpine permafrost sites to RCM-based climate scenarios. Journal of Geophysical Research: Earth Surface, 118(2), 780-794, 2013.

Scherler, M., Schneider, S., Hoelzle, M. and Hauck, C.: A two-sided approach to estimate heat transfer processes within the active layer of the Murtèl–Corvatsch rock glacier. Earth Surface Dynamics, 2(1), 141-154, 2014.

Schmucki, E., Marty, C., Fierz, C. and Lehning, M.: Simulations of 21st century snow response to climate change in Switzerland from a set of RCMs. International Journal of Climatology, 2014.

Schneider, S., Hoelzle, M. and Hauck, C.: Influence of surface and subsurface heterogeneity on observed borehole temperatures at a mountain permafrost site in the Upper Engadine, Swiss Alps. The Cryosphere, 6(2), 517-531, 2012.

Schneider, S., Daengeli, S., Hauck, C. and Hoelzle, M.: A spatial and temporal analysis of different periglacial materials by using geoelectrical, seismic and borehole temperature data at Murtèl–Corvatsch, Upper Engadin, Swiss Alps. Geogr. Helv, 68, 265-280, 2013.

Staub, B., Marmy, A., Hauck, C., Hilbich, C. and Delaloye, R.: Ground temperature variations in a talus slope influenced by permafrost: a comparison of field observations and model simulations. Geogr. Helv, 70, 45-62, 2015.

Staub, B, and Delaloye, R.: Using near-surface ground temperature data to derive snow insulation and melt indices for mountain permafrost applications. Permafrost and Periglac. Process., doi: 10.1002/ppp.1890, 2016

Steger, C., Kotlarski, S., Jonas, T. and Schär, C.: Alpine snow cover in a changing climate: a regional climate model perspective. Climate Dynamics, 41, 735-754, 2013.

Themessl, M.J., Gobiet A. and Leuprecht, A.: Empirical-statistical downscaling and error correction of daily precipitation from regional climate models, International Journal of Climatology, 31(10), 1530-1544, doi:10.1002/joc.2168, 2011.

Tonkin, M. and Doherty, J.: Calibration-constrained Monte Carlo analysis of highly parameterized models using subspace techniques. Water Resources Research, 45(12), 2009.

Torma, C., Giorgi, F. and Coppola, E.: Added value of regional climate modeling over areas characterized by complex terrain—Precipitation over the Alps. Journal of Geophysical Research: Atmospheres, 2015.

Trombotto, D.: Survey of cryogenic processes, periglacial forms and permafrost conditions in South America. Revista do Instituto Geológico, 21(1-2), 33-55, 2000.

van der Linden, P. and Mitchell, J. F. B.: ENSEMBLES: Climate Change and its Impacts: Summary of research and results from the ENSEMBLES project, Met Office Hadley Centre, 2009.

Vieira, G., Bockheim, J., Guglielmin, M., Balks, M., Abramov, A. A., Boelhouwers, J., Cannone, N., Ganzert, L., Gilichinsky, A. A., Goryachkin, S., Lopez-Martinez, J., Meiklejohn, I., Raffi, R., Ramo, M., Schaefer, C., Serrano, E., Simas, F., Sletten, R. and Wagner, D.: Thermal state of permafrost and active-layer monitoring in the Antarctic: Advances during the international polar year 2007–2009. Permafrost and Periglacial Processes, 21(2), 182-197, 2010.

Vonder Mühll, D. and Haeberli W.: Thermal characteristics of the permafrost within an active rock glacier (Murtèl/Corvatsch, Grisons, Swiss Alps). Journal of Glaciology 36: 151-158, 1990.

Vonder Mühll, D., and Haeberli, W.: Thermal characteristics of the permafrost within an active rock glacier (Murtèl/Corvatsch, Grisons, Swiss Alps). Journal of Glaciology, 36(123), 151-158, 1990.

Vonder Mühll, D. S., Hauck, C. and Lehmann, F.: Verification of geophysical models in Alpine permafrost using borehole information. Annals of Glaciology, 31(1), 300-306, 2000.

Vonder Mühll, D. S., Hauck, C., Gubler, H., McDonald, R. and Russill, N.: New geophysical methods of investigating the nature and distribution of mountain permafrost with special reference to radiometry techniques. Permafrost and Periglacial Processes, 12(1), 27-38, 2001.

Wang, S., M. Zhang, N. C. Pepin, Z. Li, M. Sun, X. Huang and Q. Wang: Recent changes in freezing level heights in High Asia and their impact on glacier changes. Journal of Geophysical Research 119: 1753-1765, 2014.

Weiming, C., Shangmin, Z., Chenghu, Z. and Xi, C.: Simulation of the Decadal Permafrost Distribution on the Qinghai-Tibet Plateau (China) over the Past 50 Years. Permafrost and Periglacial Processes, 23(4), 292-300, 2012.

~~Westermann, S., Boike, J., Langer, M., Schuler, T. V. and Etzelmüller, B.: Modeling the impact of wintertime rain events on the thermal regime of permafrost. The Cryosphere, 5, 1697-1736, 2011.~~

Westermann, S., Schuler, T. V., Gisnås, K. and Etzelmüller, B.: Transient thermal modeling of permafrost conditions in Southern Norway. The Cryosphere, 7, 719-739, 2013.

Westermann, S., Elberling, B., Højlund Pedersen, S., Stendel, M., Hansen, B., and Liston, G.: Future permafrost conditions along environmental gradients in Zackenberg, Greenland, The Cryosphere, 9, 719–735, doi:10.5194/tc-9-719-2015, 2015.

Westermann, S., Langer, M., Boike, J., Heikenfeld, M., Peter, M., Etzelmüller, B., & Krinner, G.: Simulating the thermal regime and thaw processes of ice-rich permafrost ground with the land-surface model CryoGrid 3. Geoscientific Model Development, 9(2), 523-546, 2016.

Wicky, J.: Numerische Modellierung von konvektiver Wärmeübertragung durch Luftflüsse in permafrostreichen Blockhalden. Master thesis, Department of Geosciences, University of Fribourg, Switzerland, 2015.

Zenklusen Mutter, E., Blanchet, J. and Phillips, M.: Analysis of ground temperature trends in Alpine permafrost using generalized least squares. Journal of Geophysical Research: Earth Surface (2003–2012), 115(F4), 2010.

Zenklusen Mutter, E. and Phillips, M.: Active layer characteristics at ten borehole sites in alpine permafrost terrain, Switzerland. Permafrost and Periglacial Processes, 23(2), 138-151, 2012.

Zhang, T.: Influence of the seasonal snow cover on the ground thermal regime: An overview. Reviews of Geophysics, 43(4), 2005.

Zhang, Y., Li, J., Wang, X., Chen, W., Sladen, W., Dyke, L., Dredge, L., Poitevin, J., McLennan, D., Stewart, H., Kowalchuk, S., Wu, W., Kershaw, P., and Brook, R. K.: Modelling and mapping permafrost at high spatial resolution in Wapusk National Park, Hudson Bay Lowlands, Can. J. Earth Sci., 49, 925–937, 2012.

Table 1 – Maximal depth, number of temperature sensors and series length of the boreholes used for calibration.

	Maximal depth (m)	**Number of sensors**	**Series length**
COR	**57.95**	**53**	**07/1987 - 02/2013**
LAP	**19.6**	**19**	**10/1999 - 12/2012**
MBP	**17.5**	**10**	**10/1996 - 06/2011**
RIT	**25**	**10**	**03/2002 - 09/2012**
SCH	**13.7**	**17**	**11/1998 - 07/2013**
STO	**98.3**	**25**	**10/2002 - 06/2013**

Table 2 – List of parameters used in the GLUE calibration method and their corresponding equations.

Parameter	Description	Range tested	Equation(s) related
T_{rain}	threshold temperature in the partition of precipitation into rain and snow. Above this value, precipitation falls only in liquid form.	0.1 to 4 (°C)	$Q_P = \begin{cases} \min\left(1, (1 - f_{liqmax}) + f_{liqmax}\dfrac{T_a - T_{rain}}{T_{snow} - T_{rain}}\right), & T_a \le T_{rain} \\ 0, & T_a > T_{rain} \end{cases}$
T_{snow}	threshold temperature in the partition of precipitation into rain and snow. Under this value, precipitation falls only in solid form.	-5 to 0 (°C)	
$\rho_{snowmin}$	density of new snow. Used in the function determining the density of the whole snow pack (new and old snow)	50 to 200 (kg/m^3)	$\rho_{snow} = \dfrac{\rho_{snowmin}}{119.17\, f_{liqmax}}\left(67.92 + 51.25\, e^{\frac{T_a}{2.59}}\right)$
S_k	coefficient used in calculation of the thermal conductivity of snow	10^{-7} to 10^{-5}	$k_{snow} = S_k \rho_{snow}^2$
$Melt_{rad}$	coefficient used to tune the importance of the global radiation on the empirical snow melt function	0 to $3x10^{-6}$	$M_R = Melt_{rad}(1 + s_1(1 - e^{-s_2 2}))$ $Melt_{temp} = \begin{cases} Melt_{temp} & T_a \ge 0 \\ \dfrac{Melt_{temp}}{\Delta z_{snow} m_f} & T_a < 0 \end{cases}$
$Melt_{temp}$	coefficient used to tune the importance of air temperatureon the empirical snow melt function	0.5 to 4	$M = Melt_{temp}T_a + Melt_{rad}R_{is} + \dfrac{f_{qh}q_h(0)}{L_f}$
ΔS_{crit}	threshold snow height parameter for the soil to be considered as completely covered by snow. It is used to calculate the fraction of bare soil during patchy snow conditions by weighting the sum of temperature below the snow and the temperature of bare soil	0.1 to 2 (m)	$\mathrm{f}_{bare} = \begin{cases} \dfrac{\Delta z_{snow}}{\Delta S_{crit}} & \Delta z_{snow} < \Delta S_{crit} \\ 0 & \Delta z_{snow} \ge \Delta z_{cov} \end{cases}$
α_{dry}, α_{wet}	Albedo of dry/wet soil. This parameter is used to define the albedo function of the soil to calculate the net radiation	10 to 40 (%)	$R_{snet} = R_{is}(1 - \alpha)$
ψ_{eg}	factor to account for differences between water tension in the middle of top layer and actual vapour pressure at soil surface in the calculation of the energy balance at the soil surface	0 to 3	$L_v = \dfrac{\rho_a c_p}{\gamma}\dfrac{(e_{surf} - e_a)}{r_{as}}$ $e_{surf} = e_s(T_s)e^{\left(\frac{-\Psi_1 M_{water} g e_{corr}}{R(T_s + T_{abszero})}\right)}$ $e_{corr} = 10^{(-\delta_{surf}\Psi_{eg})}$

k_w	saturated hydraulic conductivity. This parameter is also used in the calculation of the unsaturated hydraulic conductivity	100 to 10^5 (mm/d)	$k_{tot} = k_w S_e^{\left(n+2+\frac{2}{\lambda}\right)}$
g_m	empirical parameter used in the water retention function , in the effective saturation particularly	0.1 to 2	$S_e = \frac{1}{(1+(\alpha\Psi)^{g_n})^{g_m}}$
Φ	porosity, used in the water content calculation	site-specific (%)	$\theta = S_e\left(\Phi - \theta_y\right) + \theta_r$
K_{soil}	multiplicative scaling coefficient for the thermal conductivity applicable for each soil layer. This value is multiplied with the thermal conductivity calculated from the Kerten's equation for unfrozen and frozen soils.	-0.5 to 0.5	$k_{unfrozen} = h_1 + h_2\theta$ $k_{frozen} = b_1 10^{b_2\rho_s} + b_3\left(\frac{\theta}{\rho_s}\right)10^{b_4\rho_s}$

Q_p: thermal quality of precipitation (fraction of solid) (-)

T_a : air temperature (°C)

F_{liqmax} : maximal liquid water content fraction in precipitation (default=0.5) (-)

ρ_{snow}: density of snow (kg/m^3)

k_{snow}: thermal conductivity of snow (W/m/°C)

M_R: melting of snow due to solar radiation (kg/J)

s_1, s_2: empirical parameters (-)

m_f: coefficient to take the refreezing into account

Δz_{snow}: snow depth (m)

M: total snow melt (mm/day)

R_{is}: global radiation (MJ/day)

F_{qh}: scaling coefficient (-)

L_f: latent heat of freezing (J/kg)

F_{bare}: fraction of bare soil

R_{snet}: is the short-wave radiation (W/m^2)

Ris: global radiation (W/m^2)

α: albedo (-)

ρ_a: density of air (kg/m^3)

c_p: heat capacity of air (1.004 J/g/K)

γ: psychometer constant (66 Pa/K)

r_{as}: aerodynamic resistance (s/m)

e_{surf}: vapor pressure at the soil surface (mm water)

e_a: vapor pressure in air (mm water)

e_s : vapor pressure at saturation (mm water)

T_s: soil surface temperature (°C)

Ψ_1: water tension in the uppermost layer (N/m)

M_{water}: molar mass of water (18.016 g/mol)

g: gravity constant (9.81 m/s_2)

R: gas constant (8.31 J/K/mol)

$T_{abszero}$: -273.15°C

δ_{surf}: mass balance of water calculated at the surface (mm water)

$ktot$: total unsaturated hydraulic conductivity (mm/day)

k_w: saturated hydraulic conductivity (mm/day)

S_e: effective saturation (%)

n and λ: empirical parameters (-)

a, g_n and g_m: empirical parameters (-)

Ψ: water tension (N/m)

θ: water content (%)

θ_r: residual water content (%)

θ_y: threshold parameter for water tension (%)

$k_{unfrozen}$: thermal conductivity of unfrozen mineral soil (W/m/°C)

h_1, h_2: empirical constants (-)

k_{frozen}: thermal conductivity of frozen mineral soils (W/m/°C)

b_1, b_2, b_3, b_4: empirical parameters (-)

ρ_s: dry bulk soil density (kg/m^3)

Table 3 – Summary of changes projected for two different decades (2040-2049 and 2090-2099): mean of 13 GCM/RCM chains for change in mean air temperature, in mean precipitation sum and in simulated snow cover duration (number of days per year with snow $\geq$0.1 m) compared to the 2000-2010 decade.

	Δ Air T (K)		Δ Prec (%)		Δ Days of snow (%)	
	2040-2049	2090-2099	2040-2049	2090-2099	2040-2049	2090-2099
Corvatsch	+1.58	+3.97	+8.4	+4.4	-10.84	-23.75
	+1.05 / +2.11	+3.14 / +5.38	+0.31/+18.18	-10.26 / +16.43	-18.77 / +6.74	-45.42 / -4.91
Lapires	+1.67	+4.23	+0.63	-0.82	-16.73	-37.06
	+0.99 / +2.15	+3.04 / +5.83	-6.54 / +9.28	-22.87 - 8.24	-23.67 / -8.00	-57.28 / -27.88
Muot da Barba Peider	+1.58	+3.95	+10.24	+6.74	-8.63	-22.81
	+1.05 / +2.10	+3.13 / 5.33	+0.64 / 21.48	-8.79 / +18.27	-11.94 / -4.13	-37.11 / -8.03
Ritigraben	+1.62	+4.10	+2.14	+4.89	-14.76	-32.31
	+0.97 / 2.08	+2.95 / +5.65	-4.98 / +11.64	-20.25 / +14.68	-20.32 / -5.93	-49.42 / -20.57
Schilthorn	+1.40	+3.36	-1.20	-2.72	-9.21	-20.03
	+0.92 / +1.91	+2.30 / 4.35	-8.50 / +6.51	-16.78 / +11.33	-12.24 / -4.27	-35.73 / -12.09
Stockhorn	+1.55	+3.86	+2.08	+4.68	-10.23	-24.59
	+0.96 / +2.00	+2.84 / +5.31	-5.61 / +11.33	-20.01 / +14.69	-14.18 / -5.35	-39.53 / -16.33

Figure 1: Schematics of the two-step procedure used for the generation of climate scenarios at the six monitoring sites. Figure adapted from Rajczak et al., (2016~~5~~).

Figure 2: Site-scale climate scenarios of mean annual air temperature at 2 m above ground (MAAT) for the six considered permafrost monitoring sites. The results are based on the developed scenarios using the two-step procedure (Figure 1) and are based on 14 ENSEMBLES regional climate models assuming an A1B greenhouse gas emission scenario.

Figure 3 – Description of the model layers as defined in the model (green) and of the simulated subsurface structure for each site. The depths of the horizons were estimated by experts, based on data from boreholes and geophysical surveys whereas the porosity Φ is defined by the GLUE calibration based on the ranges estimated by the experts (given below the GLUE estimated porosity values). The maximum depth for each site (lower boundary, LB) is given below each column.

Figure 4 – Calibration procedure using the GLUE method in the following steps: a) 1st iteration, stochastically testing 14 different parameters in 50’000 runs b) selection of the most sensitive parameters for each site using the LGM method c) Refinement of the calibration with a second iteration of 20’000 runs focusing on the four to six sensitive parameters (may be different for each site) d) selection of acceptable model set-ups among the 20’000 simulations based on statistical performance indicators (r2 and the mean error, ME) for ground temperature at several depths. Among those four to six set-ups, the median (regarding the evolution of active layer thickness) is eventually used for long-term simulations

Figure 5 – Left panel: LGM relative importance of six groups of parameters (snow, albedo, hydraulic conductivity, saturation, thermal conductivity and evaporation) on the r2 (left) and the ME (right) at three different depths. The percentage indicates the total LGM absolute importance. Right panel: LGM relative importance of the most sensitive parameters that were selected for the second step of the calibration procedure.

Figure 6a – Comparison of simulated (black) and measured (red) temperature during the calibration period at six sites at three different depths: one close to the surface, on around 3 m and one close to the lower boundary of the model for Corvatsch and Lapires.

Figure 6b – Comparison of simulated (black) and measured (red) temperature during the calibration period at six sites at three different depths: one close to the surface, on around 3 m and one close to the lower boundary of the model for Muot da Barba Peider and Ritigraben.

Figure 6c – Comparison of simulated (black) and measured (red) temperature during the calibration period at six sites at three different depths: one close to the surface, on around 3 m and one close to the lower boundary of the model for Schilthorn and Stockhorn.

Figure 7 – Long-term evolution of ground temperatures at 10 m and 20 m as simulated with the COUP for The different sites. The black lines represent the median scenario and the grey zone the range of the 13 GCM/RCM chains.

Figure 8 – relationship between the decreasing snow duration and the increase of air temperature for the decades 2040-2049 (dots, representing the 10-year means Δ for each GCM/RCM chain) and 2090-2099 (triangles, representing the 10-year means Δ for each GCM/RCM chain), in comparison with the decade 2000-2010. The trend is variable between the sites (from -5.29 %/K to –8.76 %/K), but all sites shows a linear correlation between Δ air temperature and reduction of days with snow.

Figure 9. comparison of the simulated (red) and measured (black) soil moisture data at 12 cm (left panels) and 60 cm (right panels) at SCH. (a) and (b) are the results for soil moisture of the best thermal calibration while (c) and (d) are the results after a further calibration of the soil physical parameter of the water retention curve, showing that the calibration can be further improved with additional data sets.

Figure 10: Difference in simulated 10 m temperature for the long-term simulation between the reference run for SCH (Figure 7) and the improved calibration of Fig. 9 (c,d).

Figure 11 – comparison of the long-term simulation results for rock glacier Murtèl-Corvatsch at 10 m depth for the present study and the results obtained by of Scherler et al. (2013) with the same model, but a different calibration (see text for details).