We are grateful to the referee for their constructive review. Please find below our point by point response to the review. The comment of the referee are shown in blue and our response in black below. Proposed modifications of the manuscript are shown in green with page and line numbering corresponding to the preprint version of the article.

Summary:

This work proposes a methodological improvement to surface energy balance modeling over frozen ice surfaces by merging the benefits of two diverging current approaches to coupling air temperature and ice temperature. The coupling approach appears effective and insightful and is an important contribution to the field. The paper presents two case studies, one over snow and one over a glacier (with highly idealized implementations) as demonstrations of the accuracy. There is a well motivated exploration of the implementation's dependence on time and spatial resolution, which are not only practically important for anyone wishing to implement this method, but also provide the opportunity to discuss numerical stability.

General comments:

It is exciting to see this paper address both snowpack and glacier surface energy balance. It would be good to discuss (briefly) the physical similarities and differences (structure, air content) between the two.
We have added to the manuscript that snowpacks and glacier surfaces can be modeled in a similar framework as they share (i) the same fundamental governing equation (i.e. the energy conservation equation with heat conduction and shortwave absorption as a processes), and (ii) a first order phase change transition, where melt/refreeze occurs with latent heat. These similarities have already been used in the literature to treat snow and glacier ice in a unified framework, for instance in the model COSIPY. However, snowpacks and glacier surfaces present some differences that might complexify this unified treatment, for instance the fact that water does not percolate similarly in snow and glacier ice or that vapor movement plays a role significant role in snow but not in glacier ice.

We revised the manuscript to:

P3 - L60
"As snowpack and glaciers share many similarities and processes, such as heat conduction or the presence of a phase transition when the melt temperature is reached, they can be represented by the same type of equations. These similarities enable simulations mixing snow and glacier ice within a single framework (e.g. Sauter et al., 2020). Hence, for the sake of generality, the equations discussed in the following sections apply to both snow and glacier ice. That being said, snow and glacier ice present some differences, notably concerning liquid water percolation. As addressed later, this might require a differential treatment of glacier ice and snow when implementing the liquid water percolation scheme."

There is a consistent overuse of commas in the setup ',that' (many of which should be 'which' with no comma)
This was also pointed out by the review Richard Essery. This is now corrected.

The manuscript is clearly structured in introducing a new method to approach temperature and melt numerical modeling and then applying that method to two test cases. However, the test cases are very specific and thus convey limited information about the broader

application of the method – these limitations should be discussed, especially as a future goal would likely be to apply this numerical routine to more complicated cases.
This point was also stressed in the review of Richard Essery. We were also aware of this potential limitation when doing this study, and wondered if more realistic cases should be analyzed. We however decided to limit this study to simple idealized cases. Our goal behind this choice was to provide simple cases from which the impact of the numerical implementation can be clearly analyzed.

We also decided not to include comparisons with direct observations. Indeed, it would not be possible to decipher errors due to the numerical implementations (which is the focus of our paper) from errors due to the assumed physics, parametrizations, and forcings (which we do not and cannot not address in this study). Therefore, we think that to meaningfully analyze numerical implementations in terms of cost, accuracy and robustness, the use of simplified test cases is appropriate. We however agree that the test cases should not be too unrealistic if we want their results to be informative of how a numerical scheme might behave in a realistic settings. That it is why we have used realistic forcings and initial conditions.

We revised the manuscript to specify our intention more clearly. We explain that our simple test cases are meant to ease the comparison of the numerical implementations of the surface-internal energy budgets, but that our toy-model should not be viewed as proper a snowpack/glacier model as many important components are lacking.

P12 - L330
"Two simple examples, showcasing the differences between numerical treatments, are presented below. We note that these simulations cannot be considered as fully realistic simulations of a snowpack or glacier surface, as many processes, such as the deposition of atmospheric precipitation or mechanical settling, are lacking. The goal is rather to provide a simplified setting in which the impact of the numerical implementation of the SEB can be analyzed. In the same idea, we do not attempt to compare the simulation results to field observations. Indeed, it would not be possible to decipher errors due numerical discretization (the focus of this paper) from errors due to the assumed physics, parametrizations and atmospheric forcing. Nonetheless, in order for the results to still be informative of how a given numerical implementation might behave in a realistic setting, we use realistic atmospheric forcings, initial conditions, and physical parametrizations. The first simulation is meant to highlight the behavior of the numerical models when simulating the surface energy balance on a snow-free glacier. The second one focuses on the impact of the model implementations on the simulation of the energy budget of a seasonal snowpack, during the melting period."

Lastly, the finding that a coupled surface model can outperform other models at coarser grid sizes is implied here to be more computationally efficient due to the change in mesh size. However, this is not generally true when you are also changing the numerical scheme, so the assertion of computational cost savings which maintaining accuracy (as claimed here) should be backed up by either reports of the time taken for the computations and/or a clear statement that the numerical implementations are computationally identical by construction. This, if true, should also be mentioned in the conclusion as it is an important outcome! This is somewhat related to the discussion of numerical reduction (back to the same order of the original models) that you get from the Schur complement, but they are not discussed together and the data are not shown.
To answer this question we have computed the number of basic operations (addition/substraction and multiplication/division) required to perform the linear algebra

problem solvings of the three presented models, including the use of Schur-complements. The exact number are now presented in the new Appendix A and discussed in the article. We found that in terms of operations the standard skin-layer scheme requires about 40% less operations than the coupled-surface and no-surface schemes (which both require very similar number of operations). The last two schemes are more computationally expensive as they require the extra computation of the Schur-complement that is a bit more costly than the standard inversion used in the standard-skin layer formulation. Therefore, and based on the Figures 5 to 9, the introduction of a coupled degree of freedom at the surface (to transform a Class 1 into the coupled-surface scheme) is an interesting numerical trade-off, as it only marginally increase the numerical cost of the method while allowing coarser meshes. Concerning the standard skin-layer models, the trade-off of transforming into a coupled-surface scheme is not as evident as the numerical cost is multiplied by a bit less than 1.70. It allows the use of large time steps, without numerical instabilities, but at the expense of an increased number of steps.

We propose to discuss in more details the numerical cost of the methods in Section 4.1.1

P11 - L294
"An analysis of the numerical cost (in terms of number of basic operations) of this numerical scheme is given in Appendix A, alongside analyses of the numerical cost of Class 1 and 2 models. It shows that the proposed scheme and the Class 1 models have similar numerical costs, which a bit less than 1.7 times larger than the standard-skin layer."

in the conclusion:
P25 - L551
"Furthermore, a reduction technique, based on the computation of a Schur complement, is presented so that the numerical cost of the proposed framework remains of the same order as that of the standard implementations for the same mesh. In particular, for a given mesh, the numerical cost is similar to that of models not explicitly having a surface and about 1.7 larger than that of the standard-skin layer formulation."

P26 - L563
"Moreover, a tightly-coupled treatment of the SEB allows unconditional stability, while the standard skin-layer formulation can be unstable and displays large spurious oscillations with large time steps and small mesh sizes. Thus, while a bit more numerically costly, the formulation presented in this article can be used to overall reduce the numerical cost of a snowpack/glacier model through the use of larger time steps."

As well as in the new Appendix A:
"We see that whole system of Eqs. (A1) is a tri-diagonal system of dimension (N+1)x (N+1), with N the number of cells. Without a Schur-complement, the computation of $A^{-1}B$ can thus be solved with Thomas algorithm in 10N -1 base operations (addition, subtraction, multiplication, and division) per non-linear iteration (neglecting the time spent assembling the matrices). We also note that A_{diag} is a tri-diagonal matrix, and thus Thomas algorithm also applies. Moreover, we see that A_{up} and A_{low} are almost empty matrices, which simplifies the number of operations necessary to compute $A_{diag}^{-1} A_{up}$ and $A_{low} A_{diag}^{-1} A_{up}$. Specifically, the Schur-complement technique used in this paper can be employed with 7N-9 ($A_{diag}^{-1} A_{up}$, once per time step) + 10N-21 ($A_{diag}^{-1} B_{int}$, once per time step) + 15 (assembly and solving of Schur-complement, once per iteration) + 2N (re-injection to compute T_{int}, once per time step) steps, i.e. a total of 17N-6 + 15n_{it} steps, with n_{it} the number of non-linear iterations. We see, that the advantage of the Schur-complement technique is that the cost of performing non-linear iterations do not increase with the mesh

resolution, yielding a smaller numerically cost than inverting the while system for each non-linear iteration.

One may then wonder how the numerical cost of the scheme proposed in the article compares to the Class 1 and 2 models discussed in the paper. The Class 1 model (once a Schur-complement technique has been employed) as a similar numerical cost as the proposed coupled-surface scheme approach, namely 17N-23 + 15n_{it} steps. For a given mesh, it has one less degree of freedom as the coupled-surface scheme and is thus only marginally cheaper. The Class 2 model is the cheapest of all schemes discussed in the paper. Indeed, once the SEB and the surface temperature have been solved through scalar non-linear iterations, it relies on a single tri-diagonal inversion of dimension NxN, which can be done in 10N-11 steps. The ratio of the numerical cost of the scheme proposed in the article over that of the standard skin-layer is of about 1.7."

Finally, we note that we cannot analyze this numerical cost directly in terms of computation time in our implementations. Indeed, they were implemented using the (interpreted) python language with only some parts using pre-compiled (and thus much faster) libraries. Directly comparing computation time would unfairly favor the schemes using pre-compiled librairies.

Specific comments:

L3-4: "This surface energy budget is the sum of the various surface energy fluxes, that depend on the input meteorological variables and surface temperature, and to which heat conduction towards the interior of the snow/ice and potential melting need to be added." the comma between 'fluxes' and 'that' is incorrect, as are similarly positioned commas throughout, and 'that' should be 'which.'
This is now corrected. The same error is also corrected elsewhere in the text.

L2-4: 'and to which heat conduction towards the interior…" this sentence is unclear to me
We wanted to highlights that the conduction of heat towards the interior of the snowpack/glacier is an important factor that affects the SEB and hence the surface temperature. We clarified this in the revised manuscript:

P1 - L2
"This surface energy budget is the result of various surface energy fluxes, which depend on the input meteorological variables and surface temperature, of heat conduction towards the interior of the snow/ice, and potentially of surface melting if the melt temperature is reached."

L26: once the SEB acronym is introduced, it should be used consistently in the paper
We now systematically use SEB instead of "surface energy budget" once introduced.

L25-30: There is a large focus on the nonlinearity of SEB processes, which is important but not hugely challenging in the modeling field, as many of the nonlinearities are easily solved. It would be good to mention this and discuss sources of nonlinearity in a more mechanistic sense. For example, the "regime change" mentioned is due to thermal energy being used for processes with different reaction coefficients in warming frozen ice vs. phase change. This will help build intuition to support the truncation method discussed later. Perhaps mention another example.

We have reformulated the paragraph to lighten the references to non-linearity and clarified that the regime change between a melting and non-melting surface occurs at the fusion point (and not above as previously stated).
We have also precised how the SEB of melting and non-melting surface differs.

P2 - L29
"Mathematically, the SEB thus appears as a non-linear top boundary condition for snowpacks and glaciers. This non-linearity is even reinforced by the existence of a regime change between a melting and non-melting surface, with different thermodynamical behaviors below and at the melting point. Indeed, once the melting point is reached at the surface, the SEB becomes more akin to a Stefan-problem with a discontinuity in the energy fluxes and can no longer be simply described in terms of surface temperature. This leads to numerical challenges when solving the governing equations."

L42: which domain? The ice domain?
By domain we mean the physical space over which the equations are solved, that is to say in our case the snowpack or the glacier. This is now clearer in the text.

P2 - L40
"On the other hand, some FVM implementations do not define a specific temperature associated with the surface, but rather use the temperature of the top-most numerical layer of the domain (i.e. the top layer of the simulated snowpack/glacier) for solving the SEB (Anderson, 1976, Brun et al., 1989, Jordan, 1991, Vionnet et al., 2012, van Kampenhout et al., 2017)."

L63: specify Fourier's law of heat conduction
This is now specified.

L90-95: specify the sign convention used for fluxes
We now specify the sign convention for the fluxes.

Figure :1: clarify the meaning of the blue/orange colors of dots in the figures. Additional labels within the diagram would improve the clarity of the figure. It is also somewhat redundant to label Class 1 as a), class 2 as b) etc. since they are all in essentially the same panel. Consider just labeling the columns as class 1, class2, this paper.
We now specify that the nodes corresponds to variables to be solved (i.e. the cell temperatures and the surface state) and their position in space. This was also added to the caption. The color are meant to group the variables that are solved simultaneously and will be explained in the caption. We also revised the Figure to change the panels labeling to "Class1", "Class2", and "proposed scheme".

We propose for the new caption:
P6 - Fig 1
"Classification of FVM models with respect to their treatment of the SEB. Class 1: The surface energy and the internal temperatures are solved in a tightly-coupled manner but there is no explicit surface. Class 2: An explicit surface temperature (and surface melting) exists but it is solved in sequential manner with respect to the internal temperatures. Proposed scheme in this article: An explicit surface temperature is considered and is solved in a tightly-coupled manner with the internal temperatures. In the schematic, dots represent the prognostic variables of the schemes (with or without temperature at the surface) while the colors indicate which variables are solved simultaneously."

L115: “We therefore do not treat the finite elements method, which is for instance used in the SNOPACK model.” -> “We therefore exclude implementations of the the finite elements method, such as in the SNOWPACK model.”
Following the review of Michael Lehning, we now discuss the equivalent of our implementation in a FEM setting. We explain that by construction, the FEM posses a surface node which naturally allows one to computed a tightly-coupled SEB with the interior of the snowpack, but that the mix of node-wise (temperatures) and element-wise (energy content, liquid water content) variables in the FEM complexifies its use. The implementation of an equivalent FEM scheme is presented in the new Appendix C (attached at the end of this response) and discussed in the manuscript:

P4 - L112
“Moreover, we focus on numerical schemes based on FVM, as it is the method employed by most models (e.g. Anderson, 1976, Sauter et al., 2020, Westermann et al., 2023). We note that, contrary to the FVM, the use of the finite element method (FEM) naturally incorporates the presence of a surface temperature, which can be used for a fully-coupled treatment of the SEB, as done in SNOWPACK for instance (Bartelt and Lehning, 2002).”

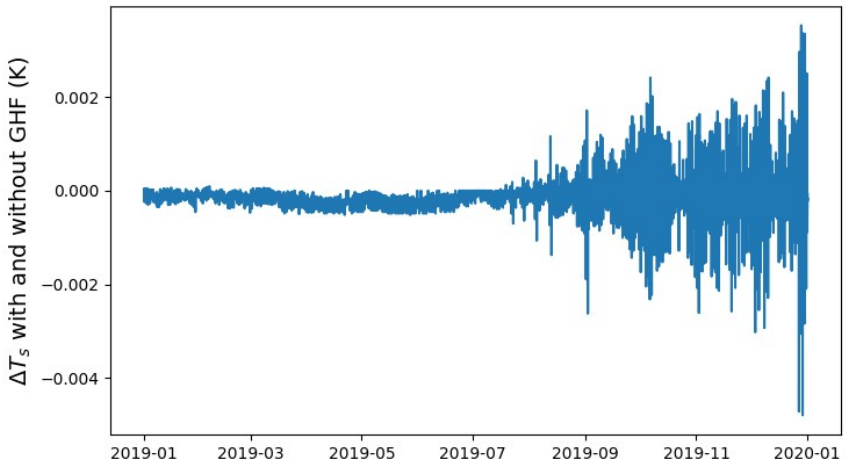
P11 - L292
“Finally, a translation of this numerical strategy (including the fictitious variable and the Schur-complement technique) in a FEM framework is presented in Appendix C.”

P12 - L329
“Finally, note that we do not include the FEM in this comparison. As detailed in Appendix C, a specificity of FEM models is to rely on a temperature field that can be defined element-wise or node-wise. It is thus required to convert back and forth between these two representations. However, the relation between the two is not bijective. This prevents an unambiguous transformation from element-wise to node-wise temperatures, affecting the end-result of our simulations. Because of this problem, the FEM is not further explored in this article, as a direct comparison to the FVM models is not possible.”

L179: omit the word “let’s” or use “we” instead
We modified the manuscript accordingly.

L275: the introduction of new terminology in representing equations 5 and 9 as a block-matrix system with decomposed components (Adiag, etc.) requires additional explanatio of the correspondence between terms in Equations 5 and 9 and their placement in Equation 11.
We now clearly explain how Eqs (5) and (9) can be cast as the block-system of Equation (11). This is done in a new Appendix A, attached at the end of this response.

L350: the bottom boundary no heat flux assumption seems strong to me, or at least appropriate in a limited set of conditions. A citation or further discussion of this would help.
We think that in the glacier test case, the assumption of a no-flux boundary condition at the bottom is appropriate as the temperature is essentially isothermal in this region (as given by the initial conditions derived from a COSIPY run). Moreover, as this boundary condition is far away from the surface (~189m), it would take much more than a year for it to influence the surface where we perform our analysis. To quantify this point we have run a simulation of the glacier test case with a 64.7 mW/m2 geothermal heat flux (GHF; Davies, 2013, Talalay et al., 2020) instead of a no-flux conditions. This difference in surface temperature between the simulation with and without GHF is displayed in the Figure below. It barely exceeds 4mK over the simulation, with a standard deviation of 0.4mK.



We added this number in the text:

P13 - L352

"For instance, we performed a simulation in which a 64.7 mW m^{-2} geothermal heat flux is applied instead (Davies, 2013). The impact on the surface temperature remains below 0.4 mK."

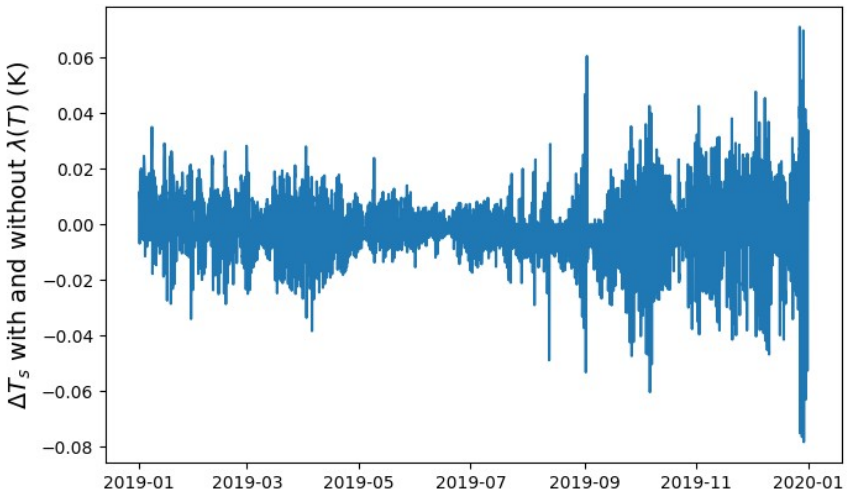
Finally, we also want to note that the goal of our simplified simulation set-up is to provide an easy framework for the comparison of numerical methods. While more realistic boundary conditions could be used, this will not change our conclusion that are confined to behavior of the numerical schemes.

L353-355: these constants are also introduced on L 66 and 72-75, use the same symbols here to connect them.
We now re-use the alredy introduced symbol to refer to the physical variables.

L354: thermal conductivity of ice is temperature sensitive! If making this assumption, please explain why it is warranted in this case (i.e., the temperature ranges reasonably experienced in this case are small enough that there is not meaningful variation?)
Indeed, the thermal conductivity of ice is expected to vary of about 10% over the range of temperature considered in this test case (from 2.5 W/K/m at 240K to about 2.22 W/K/m at 273K[1]).
We however chose not to introduce the temperature dependence of ice in our computation as (i) this is the assumption followed by the other models discussed in the paper (i.e. COSIPY or Crocus) and (ii) this added complexity would not influence the numerical benefit of tight-coupling the surface and internal energy budgets, and the properties we want to study (time step and mesh sensitivity, stability, etc). We also want to add that including a temperature-dependence for the thermal conductivity would render the system of equation globally non-linear (rather than just locally near the surface) and would thus obscure the advantage of variable elimination to speed up the resolution of system of equation where non-linearities only appears locally. We think this last point is important has it is relevant for simplified snowpack/glacier models, which do not necessarily include such temperature dependence, and that are part of larger climate and Earth system models and where speed up of the snowpack/glacier component would be beneficial.

1 https://www.engineeringtoolbox.com/ice-thermal-properties-d_576.html

Furthermore, we have run a simulation of the glacier test case with the temperature dependence on the thermal conductivity. The difference in surface temperature between the simulations with and without this temperature-dependence is visible in the Figure below. It shows that the difference remains below 0.06K, with a standard-deviation of 0.01K.



We now in the revised manuscript that considering the thermal conductivity (and specific thermal capacity) as temperature-independent is a simplifying assumption that is regularly made in models and that allows the internal heat budget equation to be linear (and hence more easily solvable). This assumption could be relaxed, but to the detriment of a more numerically costly system to be solved.

P3 - L82

"Finally, in this article we consider the thermal conductivity λ and capacity c_p not to depend on temperature. The motivation for this is twofold as it (i) corresponds to a simplifying assumption regularly made by snowpack and glacier surface models (e.g. van Pelt et al., 2012, Vionnet et al., 2012, Sauter et al., 2020, Covi et al., 2023) and (ii) it allows keeping the internal heat equation linear"

P11 - L294

"We also note that to apply this technique, the assumption of temperature-independent thermal capacity and conductivity is important, as otherwise the internal heat equation system would not be linear and thus the matrices A_{diag}, A_{up}, and A_{low} not constant."

L385: albedo over what wavelength range? In most of the visible spectrum, this would be a quite low value in clean snow. Further, the longwave emissivity of snow is more density dependent due to the presence of air. It seems reasonable to use 1 for this approach, as the emissivity is still usually quite high

The albedo used in this work refers to the broadband albedo (i.e. integrated over the while solar spectrum). This is now specified in the manuscript.

We chose a simple constant value of 0.6 as the simulation is meant to take place during the melting season, when the snow albedo is at its lowest. We agree that this value is on the lower-hand of snow albedo. Thus, we have changed this value to 0.7 in the article (Section 5.2), increased the duration of the snowpack simulations, re-ran them, and

updated the numbers in the manuscript (notably Figures 4, 7, 8, 11, 12, and 14). The conclusions of the article remain unchanged.

L435: the “lag” of one time step mentioned here is interesting and well explained. The impacts of this on interpreting a snow model output may be sensitive to the time step. If there is a long time step, this would be problematic as it may prevent modeling melt. A short time step may be more resilient to this lag.
We added to the text that this lag become less problematic with short time step.

P16 - L435
“The impact of this lagging problem can be mitigated by the use of small time steps, but with the drawback of numerical cost.”

L440: the observation that numerical instability is leading to differences between class 2 models and other models is interesting and seen clearly in Figure 4. The fact that this is not happening in the glacier model is only vaguely referenced. A direct comparison of the reasons for this – if there is an inherent numerical instability in class 2 models, why is there not an instability in the glacier model? Is all of the oscillation occurring in the meltwater percolation?
We do not think that the overall difference between models visible in Figure 4 can be readily explained by the presence of oscillations in the Class 2 model. For instance, at the beginning of the plot the Coupled surface and Class 2 appear quite in agreement on average, despite the occasional instabilities of Class 2. The two models then diverges (even not considering the presence of instabilities) before re-agreeing later in the graph. It is therefore not straight forward to link the overall agreement/disagreement of the two models with the presence/absence of instabilities, as there are periods with a good agreement despite instabilities, and periods a divergence despite the absence of instabilities.

While less visible, instabilities in the glacier test case are still possible, as for instance seen in Figure 13. As far as we understand, the presence or absence of oscillations is linked to the stiffness of the equations that relates the SEB and the internal temperature, and will depend on the specific thermal conductivity, thermal capacity, cell sizes, and on the time step at play.

To illustrate this point we performed a simple stability analysis of the standard skin-layer scheme (keeping in mind that this kind of stability analysis requires to linearize the system of equations, which departs from the actual scheme). The derivation is available at the end of this response and in the new Appendix E. It shows that the standard skin-layer scheme is not unconditionally stable, and that there is exist a maximum time step size. The presence of instabilities is favored in the case of large thermal conductivities or of a large derivative of the atmospheric fluxes with respect to the surface temperature in the SEB. On the contrary, these instabilities are hindered in the case of large cell sizes or large specific thermal capacity.

This is now discussed in the manuscript:

P21 - L503
“The unstable nature of class 2 models can be shown with a linear stability analysis, provided in Appendix E. Such analysis shows that class 2 models are only conditionally stable, and confirm that instabilities are favored in the case of large time steps and small mesh sizes. We stress that these oscillations can appear even if the time integration of the

internal energy budget relies on the Backward Euler method, known for its robustness against instabilities (Fazio, 2001, Butcher, 2008). Our understanding is that the sequential treatment of the standard skin-layer formulation breaks the implicit nature of the time integration by using "lagged" (in other words, explicit rather than implicit) terms. This, combined with the fact that the surface layer does not possess any thermal inertia and that its temperature can thus vary rapidly in time, permits large temperature swings if the time step is too large or the mesh size too small. On the other hand, it can be shown that the two schemes with a tightly-coupled SEB are unconditionally stable (Appendix E), in agreement with the absence of oscillations in their simulations. Notably, the unconditional stability of the coupled-surface scheme proposed in this article entails that the model does not need an adaptive time step size strategy depending on the mesh size. This ensures that it remains robust, regardless of the time step and mesh size.

P26 - L563
"Moreover, a tightly-coupled treatment of the SEB allows unconditional stability, while the standard skin-layer formulation can be unstable and displays large spurious oscillations with large time steps and small mesh sizes."

Figure 4: it is impossible to see the coupled surface line in panel b – consider adding markers or some other formatting choice which would allow us to see it clearly. Layering the coupled surface model on the front may help if markers are not working favorably.
Indeed, for some reason panel b of the Figure was done using a lighter shade of blue for the coupled-surface line. This was fixed and the Figure should be more readable now.

Figure 6: right panel y axis would benefit from additional labels
We added additional labels in Figures 6.

L490-495: as worded, the phrase "the class 2 model exhibits the largest phase change rate errors for an initial number of cells of 225" is ambiguous – is 225 the worst number of cells for C2 models or is C2 the worst option when working with 225 cells? From the graph, it is the second option, which is potentially less important than discussing the fact that for the other two model options, a larger number of cells generally confers better performance (within the parameter space explored here), but that is not the case for class 2 when moving from 90 to 225. Why might this be?
This deterioration is due to the development of numerical instabilities with small mesh size in the Class 2 model. This is now specified in the text.

P20 - L 490
"Finally, Fig. (10) reveals that in the glacier test case, the phase change rate errors of the class 2 tend to deteriorate with further mesh refinement past a certain point (here for an initial cell number above 90). We interpret this deterioration as a result of the appearance of numerical instabilities that develop with small mesh sizes."

L504: implicit backward Euler method?
There is one backward too much. It is now corrected.

L506: "explicit" ?
We modified the text to:

P21 - L506
"'(in other words, explicit rather than implicit)".

L509: “too”
We corrected the typo.

519: is the Dirichlet approach actually used? If not, it is not relevant to compare it here
The same comment was brought-up by Richard Essery. The potentiality of using the surface temperature as a Dirichlet condition rather than the subsurface conduction flux was made aware to us from reading the publicly available COSIPY code (cosipy/modules/heatEquation.py files, last accessed 08/11/2023) and EBFM codes. However, we stress that these codes use a Forward Euler time stepping, and in this case the using the sub-surface conduction flux or a Dirichlet condition are equivalent.

As mentioned in our response to the Richard Essery’s review, we think it is important to mention and show that using a Dirichlet condition will lead to greatly deteriorated simulations, since the use of a Dirichlet condition actually numerically stabilizes the system (which can be seen with the absence of instabilities in the orange curve of Fig. 14 and can be demonstrated with a stability analysis, provided at the end of this document and in the new Appendix E) and might be used in this attempt. However, this stabilization is at the detriment of accuracy and energy conservation.

We propose to better justify in the manuscript that the use of a Dirichlet condition might be tempting to obtain stability, but that it will produce large errors in response. We also propose to shorten the first part of the Section:

P23 - L511
“As explained in Section 2.2, the heat conduction flux from the surface to the interior of the domain (i.e. G in Equation 3) needs to have the same value in the computation of the SEB and in the computation of the energy budget of the first interior cell. Inconsistencies in G between these two budgets lead to the violation of energy conservation and create an artificial energy source/sink near the surface. Such inconsistencies could be created when implementing the standard skin-layer formulation (class 2 models) due to the sequential treatment of the surface and internal energy budgets. Indeed, after solving the SEB, one can either use the surface temperature or the subsurface heat flux G as a boundary condition for the computation of the internal temperatures. We note that the use the computed surface temperature as a boundary condition leads to an unconditionally stable numerical scheme (Appenddix E). However, using such Dirichlet condition in order to stabilize the standard-skin layer formulation comes at the expense of energy conservation and deteriorates of the simulated results.”

Figure 13: this is a time series of temperature, not a graph of numerical instabilities and should be labeled as such. It seems that the goal is to point out the larger variance in the higher number of cells-driven runs, so I would recommend either adding the time average standard deviation, or converting this plot to a time series of deviation from some sort of rolling mean in order to focus more on the instability-driven variance. Or, add a second column that contains histograms of that variance for each case.
We modified the caption to refer to the Figure as a temperature time series, and we will add the computation of a rolling standard deviation to quantify the instabilities and their presence.

P24 - Fig. 13
“Time series of surface temperatures (in blue, left y-axis) and of their 24hr-running standard deviations (in orange, right y-axis) highlighting the presence of numerical instabilities with the standard skin-layer scheme. The simulations correspond to the glacier

test case with a time step of 2 hr. Each panel corresponds to a level of mesh refinement. The lowest mesh refinement is at the top and displays the smallest level of instabilities, while the highest mesh refinement is at the bottom and displays numerous large instabilities in the first half of the simulation."

L560: the level of accuracy is similar but not identical
We are not sure to fully understand the comment of the referee. We have have modified the manuscript to explain that the tightly-coupled scheme results *overall* in a better accuracy, but not always.

P26 - L559
"Mesh and time step convergence analyses show that combining a coupled treatment of the SEB with the explicit introduction of a surface results in an overall better accuracy when compared to the classical implementations."

L613 "This approach is, for instance, used in the Crocus model" add commas
We added the commas.