Response to comments by Anonymous Referee # 1

Before starting the point-by-point response, we would like to sincerely thank the reviewer for the thorough and insightful evaluation of our manuscript. We greatly appreciate the time and effort invested in providing detailed and constructive feedback, which will be very helpful in improving the manuscript should we be invited to submit a revised version.

Please note that, in the responses below, the reviewer's comments are reproduced in red letters, followed by our replies in black text.

Major Comments

1. In the sub-sections in 2.3, it would be nice to have a bit more discussion on how the simplifications would be expected to compare to a fully distributed model run at a higher resolution with more robust physics. I am not asking the authors to run a higher resolution to compare to, but rather would like some discussion on how the simplified equations here compare to more robust physics based solutions to these various fluxes (that may be computationally impractical), and/or to the core model simply run at a high enough resolution to consider these flows (and if this is even possible or if these represent processes that would be important even at higher spatial resolutions). This qualitative discussion would also open up an opportunity to highlight the computational benefit compared to what would be needed to fully represent these processes on a typical grid.

We would be happy to provide some discussion concerning the question of tiling in comparison to high resolution. In general, our main motivation for designing the scheme was that it may provide a perspective on a level beyond even ultra-high-resolution setups, e.g. the patterned ground of the second example or the surface depressions in the first example, often feature typical length-scales of tens of meters, yet may be extremely relevant for the hydrological conditions and the thermophysical state of the surface. Thus, we always thought of it as being complementary to rather than in competition with high model resolutions. However, given the abilities of most current-generation land surface models, we would argue that in many ways our scheme is advantageous even for landscape features that can be resolved explicitly and not only with respect to computational demand. Here, a key issue is that most land surface models do not represent any lateral exchange between grid-cells other than river discharge. This means that, by explicitly resolving certain features that, in reality, form a connected landscape the results may actually become less plausible. If, for example, uplands and slopes of a catchment are located in one grid cell and the lowlands in the next, it is quite likely that the lowlands would be too dry without the (sub-) surface runoff coming from the slopes, possibly resulting in an underestimation of evapotranspiration on the catchment-scale. Having, these features in the same coarse resolution grid cell and instead representing them by laterally connected tiles may therefore provide a more realistic picture of the overall energy and moisture exchange with the atmosphere.

2. 2.2.1: Why are circles chosen as the characteristic shape as opposed to, say, rectangles where the shape can be preserved in the n-nested case? Rectangles also would follow the underlying grids informing the tile setup more naturally. Are there presumed issues with the rectangle format? The authors point out some potential issues with the choice of circles, namely the nested setup. I find the potential errors that result from different assumed shape to be concerning; could this be significant? If there is a valley, for example, would the different assumed shape depending on the connection create long term inaccuracies? Would it result in quicker or slower draining of the ridges to the valley? How big are these issues?

The circular geometry was chosen mainly because it is possibly the least bad option for many landscape features such as the patterned ground targeted in our second example or smaller lakes and ponds. Here, the problem with the higher-level nested setups doesn't necessarily arise from the assumed circular shape, but rather from nesting itself, that is having a patch inside another patch, inside another patch, inside . . . , since the contact length and center-to-center distance at nesting level n depend on the characteristic areas of all tiles on nesting levels 1 to level $n-1$. Obtaining this information is, unfortunately, not all that straightforward, due to certain infrastructural choices we had to make in the model. If the need to run higher-level nested cases arises, it is, therefore, not impossible to adapt the model, but it will require a good amount of work. With the present limitations, the errors for a nesting with more than 3 levels could become very large, because the diffusive fluxes scale linearly with the contact length for two connected patches and are inversely proportional to the centre-to-centre distance, with both of these properties depending directly on the assumed shape of the patches. In a ridge-to-valley example that employs more than three tiles – e.g. ridge, upper slope, lower slope and valley bottom – the drainage would be too slow mainly because the scheme would overestimate the flowpaths and underestimate the interface areas between slope-low, slope-high and ridge.

In a revised manuscript version we would be happy to discuss the above issue in more detail, and also provide an outlook on how the geometry assumptions could be made more flexible to better approximate the fluxes between non-circular patches, e.g. using (tile- and grid cell specific) shape-factors (S_1; S_2) in the key ratios of perimeter (C) and radius (r): $C = r \cdot 2 \cdot pi \cdot S1$ and radius and area (A): $r = (A \cdot pi)^{0.5} \cdot S2$.

3. I am not immediately convinced that the given approach would be appropriate for snow, whose redistribution will likely be driven substantially by wind which is excluded from both the scheme itself and the definition of the connectivity matrix (i.e. if the wind usually flows along the boundary between two tiles, the effective connectivity would be very low compared to a situation where the wind flows directly tile to tile).

Is there any evidence that the redistribution of snow is simply elevation based? How strong is this assumption?

What are the implications when it breaks down?

Is it a fair assumption to have it entirely independent of wind direction or magnitude?

More discussion is needed on this. If, for example, the dominant velocity is going upslope it is doubtful that snow would redistribute downslope. If velocities are very low, snow redistribution will be radically different from high velocities. Are tiles, and connectivity, designed for vegetation and subsurface flow the same tiles that would be chosen if snow redistribution was the primary concern? Considering the complex physics that should be at play for snow redistribution, and the importance of wind velocity both in magnitude and direction, and the difference in processes relative to subsurface lateral flow, I think at the least additional background information is necessary in this section to justify this decision, and significant discussion on the implications for complex terrain, different meteorological conditions etc. The snow redistribution method is not seriously considered in the results either. Overall, the snow part feels relatively poorly considered compared to the other sections on lateral connections for heat and liquid water and relies on very different fundamentals (i.e. atmospheric boundary layer becomes important, wind is important, obstructions such as vegetation become important) compared to heat/liquid water. This difference in fundamentals requires much deeper thought and discussion into snow that is not included here. I would encourage the authors to either be much more thorough in the discussion/determinations/limitations of snow processes or, likely better for a succinct publication, save this for a future publication where more background research, justification for the scheme, details on limitations, results focused on the impacts of this process and perhaps a scheme that considers velocity can be included.

We thank the reviewer for this detailed comment and admit that the snow redistribution scheme neglects many factors that determine the real-world process in a large number of settings. As a result, its applicability is very limited, which we would discuss in more detail in a revised version of the manuscript.

The reason why the scheme neglects many of the real-world factors is that it is exceedingly difficult to account for the wind direction in our modeling approach. And without the latter, it is at least arguable whether including a wind-speed scaling would notably improve the model. Instead we decided to employ an approach that neglects the near-surface winds entirely. The inability to account for the wind direction stems primarily from a structural limitation of ICON's land-surface component. Subgrid tiles coexist statistically within a grid cell but are not spatially arranged explicitly; therefore, the model does not "know" the relative orientation of tiles and cannot diagnose whether wind preferentially connects specific tile pairs or how the connectivity would be for a given wind direction. This would require assumptions about subgrid spatial organization that are not represented in the current framework. Here, it is conceivable to derive such information in the pre-processing and then use direction-specific connectivity matrices in the model. However, while we agree that a more physically complete treatment of wind-driven snow transport is highly desirable, we fear that we do not have the resources to work on this shortcoming in the foreseeable future.

The elevation-based redistribution is therefore a first-order simplification whose applicability is limited to a subset of surface heterogeneity. It implicitly assumes that the patches represented by the tiles are sufficiently interspersed for the snow redistribution to become independent of the wind direction. We therefore only apply the snow redistribution in case of the micro-scale heterogeneity for which a certain randomness in the distribution may be a permissible assumption.

4 Minor Comments

Figure 4 and lines 243-250: I find this a bit unclear. I understand the tile refined grid, but it definitely takes some work and staring at the image. I would consider trying to do this with a more simplified grid example. In particular, it is hard to understand what tile grid levels are interacting with what sibling grid levels.

We will try to reduce the complexity of Fig. 4 without loosing any of the key information.

Figure 4: it appears as though no water is flowing into the bottom level of the sibling native grid, if I am correct in interpreting the color scheme. This implies that the water in the tile would not flow along the bedrock. Does this introduce errors? Particularly in arid locations with a high elevation gradient? In an extreme case, one could even imagine a saturated zone with an unsaturated zone beneath it if horizontal flow is large, and subsurface is dry. Perhaps I am not understanding this correctly; I would appreciate clarification.

This is correct and we hope to clarify the description accordingly. However, we could not find any sign of this leading to large errors, because the vertical transport subsequently redistributes the water within the column, with (downward) percolation being a key driver.

Line 65: Error in this phrase; "Max Planck Institute for Meteorology's- and the AWI Earth System … "

Will be corrected.

145-150: I find this confusing; are the authors saying that the connectivity matrix is then weighted by the area of the tile? Wouldn't this be problematic for, for example, a riparian zone tile which may be long and skinny and therefore have a small area but large contact perimeter?

We apologize if this formulation was somewhat misleading. This optional formulation does not change any of the geometrical assumptions made for a given tile A. It merely determines the degree to which tile A is assumed to be connected to a sibling B instead of sibling C, based on the (relative) fraction of the grid cell that is covered by each of the siblings. It is therefore only relevant for tiles that are connected to more than one sibling, e.g. the riparian zone being similarly connected to a lowland-ridge and a low-land depression tile. In such cases the characteristic connectivity can be weighted by the respective cover fractions to avoid extreme situations, e.g. that 50 % of floodwater from the riparian zone flows onto the lowland-depressions even though these only cover 1% of the lowland area in a specific grid cell, while the lowland-ridge tile occupies 99% of the lowlands.

Lines 580 – 581: A few too many dashes used in place of commas or other separation words which makes it a bit confusing.

Will be corrected.

Does the subgrid scheme you have implemented increase or decrease subgrid variability in heat/moisture fluxes and soil moisture? We see this for temperature, but not for moisture or the surface fluxes (perhaps I missed this). Does this, qualitatively, match the expectations from remote sensing etc.? I ask particularly in reference to land-atmosphere interactions (which, given one way coupling, is not fully considered here). If subgrid variability is increased, as I suspect, does this pose problems for the typical flux homogenization (flux averaging/aggregation) done between the land and atmosphere given increasing work/emphasis on considering subgrid atmospheric affects driven by surface heterogeneity, such as secondary circulations? (Waterman 2025, Arnold 2024, Fowler 2024, Huang 2022) These questions cannot be fully answered with this analysis, and is not expected to be addressed, but has interesting applications in this area facing growing interest.

In general, the diffusive and conductive fluxes reduce the subgrid-scale variability, while the transport of surface- and subsurface runoff predominantly increases the variability in soil moisture, surface temperature and the turbulent fluxes. So the question whether the lateral transport poses problems for the coupling of land-surface and atmosphere depends on which surface features are being resolved in a given setup. In our first topography-based example, the lateral runoff processes dominate the overall response causing a subgrid-scale variability that certainly invites a reconsideration of the concepts that most flux aggregation approaches are based on: At least in cases in which the tiles represent kilometer-scale patches, the assumption that the vertical fluxes within a grid cell have become horizontally homogeneous at the height of the lowest atmospheric model level (often less than 50m) is highly implausible. However, ICON, as most ESMs, is not equipped to represent spatial subgrid-scale variability within the atmosphere so we couldn't say wether this commonly made blending height assumption introduces meaningful biases. We only tested whether the increased spatial variability leads to technical issues, such as the model becoming instable, which fortunately did not appear to be the case.