

Response to RC1

Leam Howe

October 13, 2026

Manuscript Title: SPUN: Deep learning for continuous snow cover fraction retrieval in marginal environments

Authors: Leam Howe, Richard Essery, and Elliot J. Crowley

Journal: Hydrology and Earth System Sciences (HESS)

We thank the referee, Carlo Marin, for the well thought-out, insightful, and constructive review of our manuscript. Our point-by-point response is provided below with the original reviewer comment is copied in *italic* followed by our response to each point (inset).

Point-by-point response

This manuscript presents an interesting and potentially valuable contribution to optical snow-cover mapping. The development of a high-resolution reference dataset for a challenging marginal snow environment, together with a deep-learning framework specifically targeting continuous Snow Cover Fraction (SCF), is timely and relevant. I particularly appreciate the effort to move beyond conventional clear-sky validation, to investigate difficult illumination and atmospheric conditions, to make the model and data openly available, and to assess transferability beyond the Scottish training domain. The results are promising and suggest that the proposed approach has potential for improving snow mapping in conditions where traditional NDSI-based methods struggle.

However, in its current form, I think several methodological and interpretative aspects require substantial clarification before the main conclusions can be fully supported. I believe the manuscript could become a strong contribution after these issues are addressed, and my comments below are intended to help strengthen the analysis and better delimit the conclusions supported by the experiments. My main concerns relate to:

Figure 1 does not currently demonstrate the observational difficulties that are most relevant to the study i.e., retrieval of snow in marginal environments. It uses simulated spectral albedo to motivate effects on Sentinel-2/NDSI even though Sentinel-2 measures directional TOA reflectance, and the distinction may be especially relevant under the low-sun and potential complex topographic conditions emphasized in this study. Albedo is essentially hemispherically integrated reflectance, whereas the satellite sees directional (radiance) reflectance for a particular sun-surface-sensor geometry. For NDSI, which is generally calculated from B11 and B3 (and not from B2 as shown in the figure), a low illumination does not necessarily behave as an identical multiplier in both bands. Topographic shadow, atmospheric scattering/path length, snow anisotropy/BRDF, and very low signal can alter the spectral relationship altering the expected behavior of NDSI. Given that the authors reported rapid melt and low solar elevation are central characteristics of the Scottish dataset it would be more informative to quantify these effects instead of difference in grain size, for example by evaluating retrieval behavior as a function of ageing, wetting and also of solar zenith angle, topographic illumination and shadow. I like a theoretical background but why not convolve the simulations with your reference Sentinel-2 spectral response functions and actually show the real B3, B11 and NDSI values under different conditions? Furthermore,

the definition and representativeness of the assumed 0.1–1 mm “grain size” should be clarified (is it the optical-equivalent radius, effective radius, physical grain radius, diameter, or a value derived from SSA?)

Author response: In the original manuscript figure 1, B2 was mistakenly highlighted green instead of B3 which is conventionally used for NDSI. This was just an error in plotting this figure, not in the NDSI values quoted in the results. We will correct this.

Our original version of figure 1 was intended to be illustrative of how snow reflectivity is impacted with age (affecting grain size and ‘dirtiness’), but we accept the reviewers comment that these theoretical simulated albedos are not the same as what Sentinel-2 ‘sees’ and does not represent many of the other key difficulties that we are attempting to overcome in this study. We therefore update this figure to instead directly show Sentinel-2 TOA values from our Scottish dataset, as seen below in the updated figure 1. Panel (a) now shows the mean reflectance values from our dataset for each band instead of theoretical albedos, and panel (b) shows the distribution by class of NDSI. As this is an illustrative figure and not part of our analysis we argue that adding a more detailed breakdown, e.g. exploring the distribution of B3 and B11 and by age, solar angle, and topographic shadow, is beyond the scope of this paper. Since the figure now shows information about our dataset, not just theoretical background it will be moved to the results. The mention of grain size in this figure has been removed as it is not relevant to the results.

L27: The manuscript repeatedly refers to Sentinel-2 as having a spatial resolution of 10 m. This is imprecise, as the MSI bands have native resolutions of 10, 20, and 60 m. The paper later clarify that all bands are resampled to 10 m. Please distinguish throughout between native spatial resolution and the 10 m model grid. In particular, upsampling the 20/60 m bands does not add spatial information, which should be acknowledged when describing the effective resolution of the SCF product.

Author response: We will distinguish throughout between native spatial and 10 m model grid, and acknowledge upsampling coarser bands does not add information.

L30: The limitation of MODIS for resolving heterogeneous snow cover in complex mountain terrain is well established, but Gascoin et al. (2024) is a review article. More importantly, the statement that snow-cover variability “typically occurs at length scales of less than 100 m” is rather general. The spatial resolution required to represent snow-cover heterogeneity depends on the spatial scales of the underlying snow patterns and should ideally be discussed in the context of spatial sampling principles (e.g. Shannon–Nyquist), rather than through a single generic threshold such as 100 m. In practice, the ability of a sensor to resolve these patterns also depends on its effective spatial response and on the prevalence of mixed pixels. I therefore suggest qualifying this statement and supporting it with primary studies explicitly investigating the scale dependence of snow-cover heterogeneity and mixed pixels (e.g. Selkowitz et al. [2014], Premier et al. [2021]). Selkowitz et al. [2014], for example, explicitly demonstrate how the occurrence of mixed snow pixels increases with pixel size from meter to MODIS-scale resolutions.

Author response: We agree that our original statement was too generalized and lacked the appropriate theoretical context. Relying on a single 100-meter threshold oversimplified the complex, scale-dependent nature of mountain snowpack, and we appreciate the recommendation to rely on primary literature rather than a review paper. In the revised manuscript, we address this by removing the generic 100 m threshold and rewriting this paragraph to explicitly discuss snow-cover heterogeneity in the context of spatial sampling principles and the physical reality of mixed pixels, and include appropriate reference to relevant primary studies including Selkowitz et al. [2014] and Premier et al. [2021], and studies on the spatial scaling of snow distribution (e.g. Blöschl [1999], Deems et al. [2006]).

L39–40: This sentence is central to the motivation of the study, but the link to Figure 1 is not fully clear (see previous comment). The text identifies patchiness, wet snow, and rapid melt-freeze cycling as defining characteristics of marginal Scottish snowpacks, whereas Figure 1 only explores changes in simulated snow albedo associated with grain size and soot contamination. Patchiness is a sub-pixel mixing problem not represented in figure 1, wetness is not represented, and the relationship between the prescribed grain-size range and melt-freeze metamorphism is not explained. That’s why I suggest

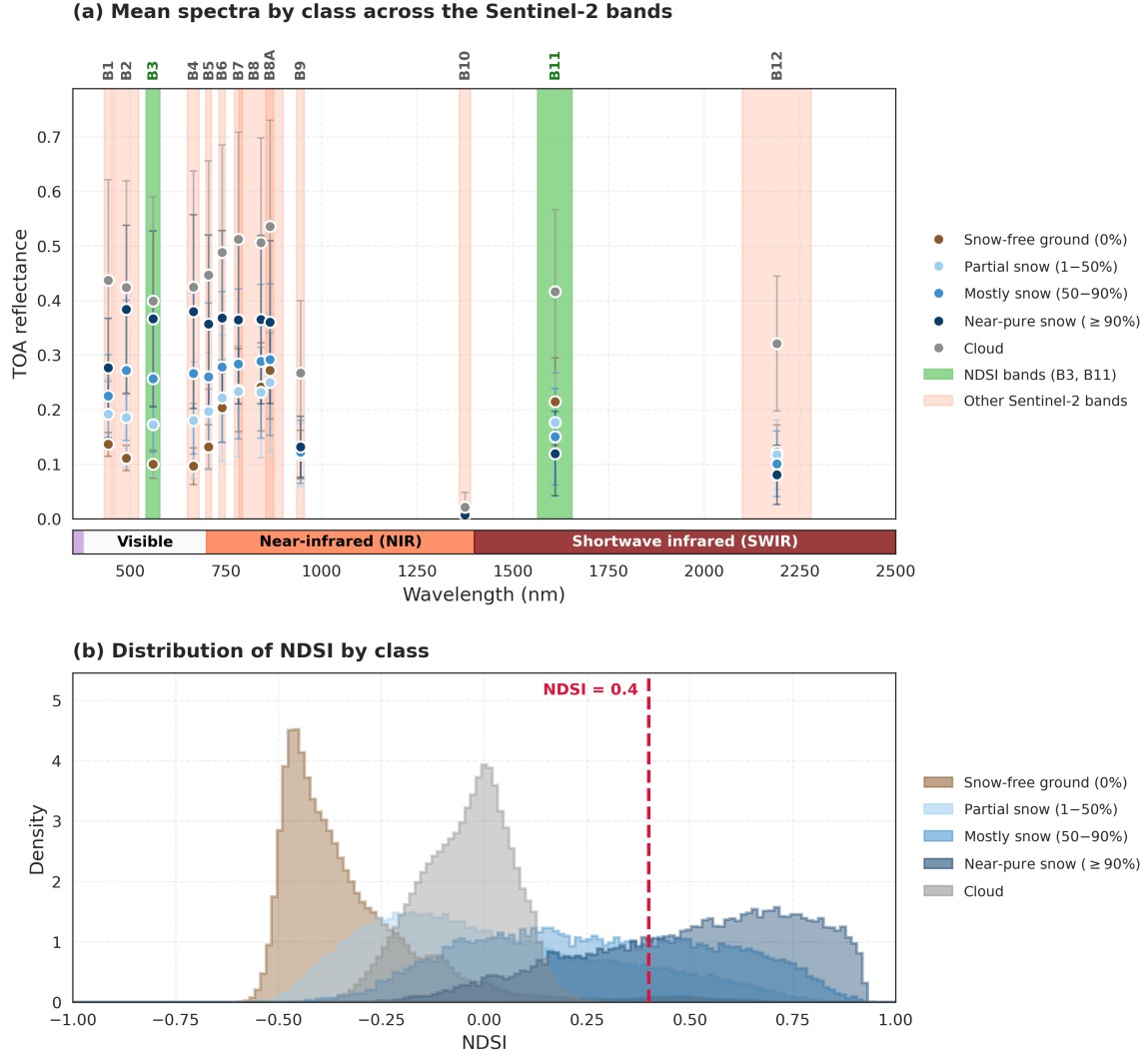


Figure 1: Sentinel-2 Level-1C top-of-atmosphere reflectance by class across all 13 bands (a), and the corresponding NDSI distributions (b), for the Scottish SCF dataset. Markers show class means with ± 1 standard deviation; the two bands used to compute NDSI (B3, B11) are highlighted in green. Of pixels with a reference SCF of 90 % or above, 33 % fall below the conventional 0.4 threshold; recomputing NDSI from Level-2A surface reflectance reduces this to 30 %.

either revising the interpretation of Figure 1 or expanding the analysis so that it better represents the processes emphasized in the text, particularly wet snow/melt-freeze conditions and illumination effects relevant to the Scottish dataset.

Author response: We believe our earlier proposed changes to figure 1 largely address these concerns. We acknowledge that the link between grain size and wetness and melt-freeze metamorphism was not fully explained, these specificities have been removed and are no longer in the figure.

Study region / Figure 2: I found it difficult to navigate between the geographical description of the study region and Figure 2. The text refers to the Isle of Skye, the Nevis range, the Cairngorms, and later Glen Coe, but these locations are not labelled on the map. Please consider adding the main geographic features discussed in the text to Figure 2 and improving the locator inset so that readers unfamiliar with Scotland can immediately understand where the study area lies within Scotland/Great Britain. It could also be useful to indicate, at least approximately, the western versus eastern climatic regions described in Sect. 2.1.

Author response: We agree that Figure 2 could be improved to make more accessible to those unfamiliar with the region. We will label any locations or climatic regions referred to in the text. And update the caption to explicitly state that the inset map is Scotland.

L 96: Please specify how the Sentinel-2 20 m and 60 m bands were resampled to the 10 m model grid (e.g. nearest-neighbour, bilinear, cubic interpolation). This is important for reproducibility and may affect the spectral information and spatial gradients at snow-patch boundaries. It would also be useful to clarify the reference grid/projection used for the resampling.

Author response: We will add details of the resampling method used to the text here: The 20 m and 60 m Sentinel-2 bands were resampled to the 10 m grid using nearest-neighbour interpolation. This was handled dynamically by the Google Earth Engine (GEE) API, which defaults to nearest-neighbour resampling when data is requested at a non-native scale without specifying otherwise.

Regarding the reference grid, the resampling was performed onto the local Universal Transverse Mercator (UTM) projection (WGS84 datum). The specific UTM zone was automatically determined by the pipeline based on the central latitude and longitude of each targeted scene.

L98–101: The justification for selecting Level-1C TOA reflectance should be strengthened. First, TOA reflectance is not inherently comparable across multispectral sensors because sensor spectral responses, calibration, acquisition geometry, and atmospheric conditions still differ. If the intended advantage is consistent availability of Level-1 products and independence from sensor-specific atmospheric-correction processors, this should be stated more precisely. Second, the argument that TOA data preserve detail in topographic shadow is plausible, but TOA reflectance also retains atmospheric and illumination effects; please provide supporting references or an empirical L1C–L2A comparison for representative shaded snow patches. A potentially important additional justification is that standard L2A processing may rely on prior scene classification, including snow/cloud probabilities (as Sen2cor for example), creating a possible dependency on an existing snow-detection algorithm. Please identify the relevant L2A processor and clarify whether avoiding this potential circularity motivated the use of L1C data.

Author response: The justification for the use of Level-1C TOA reflectance will be clarified in the manuscript. We agree that TOA reflectance is not directly comparable across sensors, and we have removed this claim. The stated advantage is now that L1C products are consistently available across the full Sentinel-2 archive and are independent of any atmospheric-correction processor. The reviewer also highlights a strong additional motivation: keeping the snow mapping product independent of specific L2A processing pipelines. For example, the standard processor provided by ESA, Sen2Cor, depends on scene classification algorithms and is known to deteriorate in performance in topographic shadow. Furthermore, the Let-It-Snow product relies on a different Level-2 processing pipeline (the CNES MAJA processor). Therefore, maintaining independence from these pipelines to avoid circularity is a major advantage of our approach. We thank the reviewer

for this astute comment.

L125–133: The use of Cloud Score+ appears reasonable, and the selected cs_cdf threshold of 0.5 lies within the recommended range. However, because CS+ is used to define cloud-contaminated pixels in the reference dataset, the choice has a direct influence on both training and evaluation. I therefore suggest providing more evidence that this configuration is appropriate for the particularly challenging Scottish conditions considered here, ideally through a manual validation on a representative subset and/or a sensitivity analysis to the CS+ threshold. In addition, cs_cdf is a temporally normalized quality/visibility score rather than a strict binary definition of “clear sky”; consequently, the terminology should distinguish truly cloud-free observations from partially affected but potentially usable pixels (e.g. haze, cloud margins, cloud shadow). This distinction seems particularly important given that exploiting such marginal observations is one of the central claims of the study. For example, how would the conditions shown in the first three rows of Fig. D2 be classified according to CS+ and according to the Tier 2a clear-sky criterion?

Author Response: We agree with the reviewer comment that more justification for the 0.5 threshold would be beneficial.

Manual validation was already explored and we will present visual examples of this on the first three tiles of fig. D2. We will investigate the sensitivity of this threshold across the recommended range from 0.5-0.65.

The reviewer is correct that the central aim of this study is to better utilise these “partially affected but potentially usable pixels”, therefore, clearly defining this would be a beneficial addition. We have somewhat implicitly implied in the manuscript that the difference between the LIS cloud mask (MAJA derived) and the CS+ mask ($\text{cs_cdf} < 0.5$) defines this. We will explicitly add this definition and integrate with the rest of the manuscript. We note that this definition introduces a dependence on two different cloud masks to define “partially affected but usable” and a more practical implementation would be the difference between e.g. $\text{cs_cdf} < 0.5$ and $\text{cs} < 0.65$, but exploring this definition further we believe to be beyond the scope of this study.

L146–149: The use of a U-Net with a ResNet50 encoder and continuous output is reasonable for pixel-wise SCF regression, but the architecture used to produce the fractional estimate should be described in greater detail. In particular, please specify the regression head, output normalization/activation, treatment of padded pixels, and how the final 0–100% SCF values are obtained. The combined Dice-regression objective may favor spatial detection and predictions close to 0 or 100%, an effect that the authors themselves note later in the manuscript. Performance stratified by reference SCF, or a predicted-versus-reference SCF analysis, would help demonstrate that the model is retrieving fractional cover rather than primarily performing accurate snow-extent segmentation. Please also clarify how Soft Dice is computed when the reference SCF itself is continuous (or, if it is binarized, specify the threshold). Finally, the manuscript is inconsistent regarding the selected regression loss: the Abstract and Fig. 4 state L1, whereas the hyperparameter optimization identifies Huber as the selected loss. This should be reconciled throughout.

Author response: We agree with the comment and will add a more detailed description of the regression to the manuscript.

We did provide “predicted-versus-reference SCF analysis” in the paper (table 2), but we agree that analysing to what extent the dice-regression loss function is pushing the model values to the extreme and prioritising segmentation over regression would be useful future work and important for operationalising such a method. We will investigate performance stratified by reference SCF but argue that including these results is beyond the scope of the current manuscript which is presenting a novel method, not an operational product. The small size and class minority of snow pixels within our Scottish SCF dataset will make performance stratified analysis of limited value at this stage, but it would be more insightful and reflective of true performance to do this analysis on a large-scale SCF dataset (such as that suggested in the ‘Future Work’ section from community driven collation of VHR imagery). To the best of our knowledge a large representative SCF dataset does not yet exist, but we will add this point to the discussion.

The Soft Dice is computed by binarising the continuous signal and this was one of the parameters that we tuned, the ‘snow decision boundary’ which is quote in the text but, as this comment from the reviewer shows, its was not explained clearly. We will add this clarification.

During iterations of the manuscript we forgot to update L1 to Huber (the best option identified from the parameter tuning), this will be corrected throughout.

L161–178: The overall U-Net/ResNet50 architecture and use of Sentinel-2 pretrained weights appear reasonable, but several aspects of the implementation require clarification for reproducibility. Please distinguish between the backbone architecture and the different pretraining/initialization strategies (ImageNet, DINO, MoCo, DeCUR, SoftCon), and describe how the 13-band Sentinel-2 input is handled by the pretrained encoder, including band ordering, input normalization/scaling, adaptation of the first layer, and whether encoder weights are fully fine-tuned or partially frozen. There also appears to be an inconsistency regarding class weighting: the text states that a per-pixel snow weighting was used, whereas the selected hyperparameters report “no snow weighting.” The multi-objective Optuna procedure should also explain how the final model was selected from the Pareto-optimal solutions. Finally, the optimized snow decision threshold (0.3997) appears inconsistent with the 0.5 threshold used later for evaluation; please clarify which threshold was used at each stage and why. Additional training details such as number of epochs/stopping criterion, data augmentation, learning-rate scheduling and random-seed treatment would also improve reproducibility.

Author response: We thank the reviewer for the constructive feedback. We will revise the manuscript to address these critical implementation details to ensure reproducibility:

- **Architecture & Input:** We will explicitly distinguish the ResNet50 backbone from the pretraining strategies (e.g., DINO, ImageNet). We will also detail the 13-band input handling, covering band ordering, normalization, first-layer adaptation, and confirming that encoder weights were [fully fine-tuned / partially frozen].
- **Class Weighting:** We will correct the text to state that per-pixel snow weighting was *explored* to mitigate class imbalance, but the optimal tuning configuration ultimately applied no class weight.
- **Model Selection:** We will explain our Pareto-front selection process. We prioritised the Dice score for spatial boundary accuracy, choosing the model that maximized Dice without the severe RMSE degradation observed at the extreme tail of the curve.
- **Thresholds:** We will clarify that 0.3997 was used strictly to binarise labels for the training Dice loss—acting as a boundary relaxation to help the model learn snow edges. For all final evaluations and product comparisons, we applied the literature-standard 0.5 threshold.
- **Training Details:** We will add the requested training specifics, including maximum epochs, early-stopping criteria, data augmentations, learning rate, and random seed protocols.

Experimental design (sec 2.4): The independence of the training, validation, and test datasets should be described more explicitly. The manuscript states that the splits “aim to preserve” spatial separation and that an attempt was made to keep different Sentinel-2 retrievals separate, but it is not clear whether complete Sentinel-2 acquisition scenes/dates are exclusively assigned to a single split. This distinction is important because multiple 1 km image chips extracted from the same Sentinel-2 acquisition share atmospheric conditions, illumination geometry, snow state, and radiometric characteristics. If chips from the same acquisition occur in both training and test sets, the reported test performance may therefore not represent a fully independent evaluation. Given that the complete dataset comprises only 37 Sentinel-2 scenes, this information is particularly important for assessing model generalization. In addition, I suggest being more precise about what is demonstrated by the global validation. The Wang et al. dataset provides categorical snow, cloud, background labels and therefore allows evaluation of the generalization of snow extent detection, but not of the continuous SCF regression itself. Consequently, statements about global generalization should clearly distinguish between global transferability

of snow detection and global validation of fractional snow-cover retrieval. Demonstrating the latter would require independent continuous SCF reference data outside Scotland.

Author response: We thank the reviewer for this important point about dataset splits. The splits were designed primarily to preserve spatial separation and a representative distribution of snow conditions across splits. We prioritised spatial separation because spatial autocorrelation between nearby chips is widely recognised as a major source of optimistic bias when evaluating models on spatially structured data (e.g. Roberts et al., 2017; Ploton et al., 2020). With a limited number of snow-covered acquisitions, strict separation by acquisition would have compromised both. One Sentinel-2 product (S2A_MSIL1C_20170508T113321_N0205_R080_T30VUJ_20170508T113346) contributes chips to both training and test sets. These chips are 38–53 km from the nearest training chip from the same acquisition, so no pixels are shared, although we acknowledge they share acquisition conditions. In the revised manuscript, we will describe the split design and its rationale explicitly, report this overlap, and investigate the impact on test results if these chips are removed.

We agree an independent SCF reference dataset outside of Scotland would be better but unfortunately we do not know of a global dataset of such extent, and this is part of what we advocate for in the discussion section of this paper. We will ensure consistent phrasing in the manuscript to reflect that the global generalisability only shows generalisability of SCE retrieval, not SCF.

Table 1/Table 2: I found the distinction between Tier 1 (“Intrinsic”) and Tier 2b (“Reality”) unclear. Both scenarios appear to use all-sky observations, with the principal difference being that Tier 1 evaluates SPUN and the 10 m baselines at 10 m, whereas Tier 2b aggregates/resamples the products to 20 m to enable comparison with the native 20 m LIS product. If this interpretation is correct, I suggest using more descriptive terminology (e.g. “10 m native-resolution evaluation” and “20 m operational comparison”) and stating explicitly whether exactly the same test scenes/pixels and cloud treatment are used in Tier 1 and Tier 2b. Spatial aggregation can naturally reduce MAE/RMSE because positive and negative sub-pixel errors cancel within the larger 20 m pixel; therefore, this improvement should be interpreted primarily as a scale effect rather than improved retrieval performance. More importantly, please clarify how the 10 m products and reference were aggregated to 20 m. For an area-preserving aggregation over an identical set of pixels, mean bias would normally be expected to remain approximately invariant, whereas SPUN snow-only bias changes from +8.36% to +6.10%.

Author response: We agree this is a fair critique of the terminology used for the different evaluation scenarios. We will update to the recommended names: “10 m native-resolution evaluation” and “20 m operational comparison”.

We thank the reviewer for this point on spatial aggregation. We will clarify in our manuscript that any improvement in performance from 10 m to 20 m spatial resolution is likely due to aggregation not improvement in retrieval performance.

The 10 m product and reference were aggregated to the 20 m LIS grid by 2×2 block averaging, ignoring cloud-masked pixels. A 20 m pixel was retained if any of its four 10 m pixels were valid. This masked averaging is a common approach for aggregating gridded data with missing values. We used it to retain partially cloudy blocks, which would otherwise be lost along cloud edges. This aggregation is bias-preserving when all four pixels are valid. In partially cloudy blocks, however, the valid pixels are up-weighted, so pixels near cloud edges carry more weight at 20 m than at 10 m. The 10 m and 20 m evaluations are therefore not over identically weighted pixel sets. This explains the small change in all-class bias (+0.08 to +0.04). The snow-only bias is further affected because the snow-only subset is defined after aggregation and includes snow-free margin pixels at 20 m. We have added these details to the methods and mention this impact in the discussion.

L365–370: I suggest revising the interpretation that bare rock, water bodies and varying illumination “mimic the spectral signature of snow.” These conditions may produce ambiguous values when the spectral information is reduced to an index such as NDSI, but this does not necessarily imply that their full (or multispectral) signatures are similar to snow. This distinction is particularly relevant here

because SPUN uses all 13 Sentinel-2 bands, whereas NDSI uses only two bands. Consequently, the improvement over NDSI cannot necessarily be attributed to the use of spatial context alone; it may also result from the much richer spectral information available to the network. If the authors wish to demonstrate specifically the added value of spatial context, a pixel-wise classifier or regressor using the same spectral inputs would provide a more direct ablation. Comparison with a spectral-unmixing SCF method would also substantially strengthen the evaluation. Otherwise, I suggest moderating this interpretation.

Author response: We agree that “mimic the spectral signature of snow” is inaccurate. These surfaces can produce high NDSI values because of the B3/B11 ratio, not because their full spectral signatures resemble snow. We will revise the text accordingly. We also agree that SPUN’s improvement over NDSI cannot be attributed to spatial context alone. We will ensure our phrasing is consistent and that improvement could be from either, or both, spatial and spectral information.

We agree with the reviewer that this would be a great way to isolate the contribution of spatial context and we will reflect as such in the text. We acknowledge that a pixel-wise ablation and a comparison with a spectral unmixing method would strengthen the analysis, but both are beyond the scope of this study. An unmixing baseline could be built using endmembers from our training data, but these would be specific to Scotland and not directly comparable with SPUN in the global evaluation. We will recommend these comparisons as future work in the Discussion, including comparison with LAMSU once it becomes publicly available.

L402 and conclusion: I suggest reconsidering the term “catastrophic overfitting.” The experiment presented here assesses geographic generalization—i.e. whether a model trained on Scottish data transfers to unseen regions—rather than what is usually referred to as catastrophic overfitting in the machine-learning literature. “Overfitting to the Scottish training domain” or “limited out-of-domain generalization” would be more appropriate. Moreover, the Wang et al. dataset provides categorical snow/cloud/background labels, so this experiment demonstrates geographical transfer of snow detection, but does not establish global generalization of the continuous SCF regression. The corresponding conclusions should therefore be moderated accordingly.

Author response: Agreed. We will reword the term “catastrophic overfitting” to “overfitting to the training domain”, and clarify conclusions to say that we are only addressing generalisation of SCE not SCF.

L414–416: The modular separation between a snow-specific model and an independent cloud-mask algorithm is an interesting and potentially useful design choice. However, I suggest discussing the corresponding propagation of errors through the two-stage pipeline. The performance of the final product depends jointly on both modules: snow pixels incorrectly masked as cloud cannot subsequently be recovered by SPUN, whereas clouds incorrectly retained by the mask must then be correctly rejected by SPUN. These errors are also unlikely to be statistically independent because both modules respond to the same spectral and illumination conditions. Consequently, replacing CloudScore+ with another cloud algorithm would not simply preserve SPUN performance but would require re-evaluation of the complete pipeline. It would be useful to report the performance of CloudScore+ itself on the global reference data and explicitly quantify how its omission and commission errors propagate into the final snow product.

Author response: Agreed that CloudScore+ performance on the global dataset should also be quoted. This will be added.

L455–465: I agree that reproducibility, accessibility and model interpretability are important barriers to wider adoption of ML-based snow products. However, some of the discussion here is quite general and could be more closely connected to the findings of this study. Claims regarding the limited availability of code or training data in the existing literature would benefit from supporting references. I would also distinguish more carefully between explainable AI and physics-informed learning: attribution or perturbation methods can provide insight into the information used by a model, but they do not necessarily demonstrate that the model has learned or represents the underlying physical processes correctly. In this context, it may be particularly useful to discuss the concrete requirements for operationalizing SPUN

identified by this study, including cloud-mask dependency, preprocessing, uncertainty characterization and validation of continuous SCF outside the Scottish training domain.

Author response: Great points. We will clarify throughout that our investigation using perturbation methods provides some interpretation of what information the model is using but is not “explainability”. We are not offering an operational product here, but we are certainly suggesting that these methods could and should be explore more for operational pipelines. But we agree adding further description of the requirements and steps towards implementing this would strengthen the manuscript and we will add this.

References

- Günter Blöschl. Scaling issues in snow hydrology. *Hydrological Processes*, 13(14-15):2149–2175, 1999. ISSN 1099-1085. doi: 10.1002/(SICI)1099-1085(199910)13:14/15<2149::AID-HYP847>3.0.CO;2-8. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291099-1085%28199910%2913%3A14/15%3C2149%3A%3AAID-HYP847%3E3.O.CO%3B2-8>.
- Jeffrey S. Deems, Steven R. Fassnacht, and Kelly J. Elder. Fractal Distribution of Snow Depth from Lidar Data. *Journal of Hydrometeorology*, 7(2):285–297, April 2006. ISSN 1525-7541, 1525-755X. doi: 10.1175/JHM487.1. URL https://journals.ametsoc.org/view/journals/hydr/7/2/jhm487_1.xml. Publisher: American Meteorological Society Section: Journal of Hydrometeorology.
- Valentina Premier, Carlo Marin, Stefan Steger, Claudia Notarnicola, and Lorenzo Bruzzone. A Novel Approach Based on a Hierarchical Multiresolution Analysis of Optical Time Series to Reconstruct the Daily High-Resolution Snow Cover Area. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:9223–9240, 2021. ISSN 2151-1535. doi: 10.1109/JSTARS.2021.3103585. URL <https://ieeexplore.ieee.org/document/9511108>.
- David J. Selkowitz, Richard R. Forster, and Megan K. Caldwell. Prevalence of Pure Versus Mixed Snow Cover Pixels across Spatial Resolutions in Alpine Environments. *Remote Sensing*, 6(12):12478–12508, December 2014. ISSN 2072-4292. doi: 10.3390/rs61212478. URL <https://www.mdpi.com/2072-4292/6/12/12478>. Publisher: Multidisciplinary Digital Publishing Institute.