Reviewer 1

Overall Evaluation

The manuscript presents a valuable and well-structured reconstruction of snow cover using a hybrid gap-filling framework that combines decision-tree and machine learning approaches. The long-term dataset and the integration of multiple satellite sources represent a significant contribution to snow monitoring and hydrological applications.

The methodology is generally sound, and the results are relevant and promising. However, several methodological aspects require clarification to improve transparency and reproducibility, particularly regarding model configuration, data processing choices, and evaluation procedures. In addition, some figures and descriptions would benefit from clearer explanations to facilitate interpretation.

Overall, the manuscript is of good quality and suitable for publication after minor revisions addressing the points raised below.

Detailed Comments and Suggestions

Comment on Section 2.2 (snowMapper model overview):

The model overview is clear and well structured, and Figure 2 is informative. However, this section remains largely descriptive and would benefit from additional clarification.

Thank you for your comment. We acknowledge your point regarding Section 2.2 of our manuscript being largely descriptive. This decision was made intentionally, in an effort to deliver a brief overview of a relatively complex model, to accommodate the varied interests of a wider audience. This way, readers who are interested in the results of the climatological analysis rather than the model, would be able to get a quick understanding about how the model operates before proceeding to the following sections. On the other hand, readers interested in the technical aspect of snowMapper, will be able to get high detail from the following sections/subsection of the methodology.

Specifically:

- The novelty of snowMapper relative to existing approaches is not clearly articulated. Please highlight the key contributions and what differentiates this framework from previous methods.

  We will articulate more clearly the novelty of snowMapper relative to existing approaches presented in the Introduction.

  The novelty of snowMapper lies within its iteratively solved physics-informed machine learning approach, explained in response to the comment below.

- The term “physics-informed” should be better defined (e.g., are physical constraints explicitly enforced, or are only physically meaningful variables used?).

  We will add more information regarding the term “physics-informed” machine learning.

  Lines 101-110: “Physics-informed machine learning (PIML) integrates data-driven machine learning algorithms with physical constraints embedded within the training process (Karniadakis et al., 2021; Meng et al., 2025). It has been applied across a wide range of domains, including weather and climate modelling (Kashinath et al., 2021), daily snow and ice albedo reconstruction (Ye et al., 2023), snowpack modelling (Maharjan et al., 2025; Zhao et al., 2026), and the simulation of climate-driven runoff shifts in alpine catchments (Zhong et al., 2023). PIML offers a promising framework for reconstructing snow cover time series, as spatiotemporal variability in snow cover reflects underlying physical processes such as accumulation and ablation. Incorporating these processes as constraints can improve model realism and predictive skill (Meng et al., 2025). Achieving this requires a reconstruction workflow that mimics the iterative solution strategies of physically-based numerical models. In this framework, the evolving state of the system provides context for estimating the probability of snow presence or absence at subsequent timesteps.”

- The “assimilation” step appears to rely on direct replacement of observations rather than a formal data assimilation approach; this should be clarified.

  While this is described in section “2.4.4 Assimilation” (lines 272–275), we appreciate your comment. Our approach does not constitute formal data assimilation, as it involves direct replacement of simulated pixel values with available clear-sky satellite observations without statistical weighting. We will revise the manuscript to replace the term “assimilation” with “direct insertion of observations” (or similar wording) to more accurately reflect the method.

  Based on the above, we made the following changes to the manuscript:

  Lines 147-148: “[...] and updated through direct insertion of binary snow cover [...].”

  Line 159: “[...] direct insertion of observations [...].”

  Line 363: “2.4.3 Direct insertion of observations”

  Lines 364-365: “[...] snow cover observations are used to replace the corresponding predicted values”

  Line 383: “[...] but prior to the direct insertion of available Landsat- [...].”

  We have also adapted Figure 2 (Step 3 title on the schematic) to be in line with that change.

- The workflow could be described more explicitly (e.g., step-by-step process or simplified schematic), as Figure 2 is relatively complex and not always easy to follow.

  We will modify section “2.2 snowMapper model overview” in order to link it better with the different steps displayed in the schematic of Figure 2, thus helping the reader to follow.

  We have made the following addition:

  Lines 148-156: “Its modular structure comprises of three main steps (Fig. 2):

  - Step 1: Preprocessing This includes the preparation of (a) satellite imagery (Section 2.3.1), (b) meteorological forcing data (Section 2.3.2), (c) terrain data (Section 2.3.3), and (d) in situ training data (Section 2.3.4), followed by (e) training of the machine learning classifier (Section 2.3.5).
  - Step 2: Snow cover reconstruction. This step consists of (a) model initialisation (Section 2.4.1) and (b) model execution (Section 2.4.2).
  - Step 3: Validation and correction. This involves (a) the extraction of evaluation metrics and a post-processing validation routine (Section 2.5), and (b) the insertion of observed snow cover values (Section 2.4.3).”

  We also moved the subsection titled “Machine learning classifier” to the preprocessing section (“2.3 Data & preprocessing”), in agreement with the layout of Figure 2.

Comment on Section 2.3.1 (Satellite imagery and MODIS processing):

The satellite data processing is generally well described; however, several methodological choices require further justification:

- The use of a fixed NDSI threshold (NDSI > 0.4) is not justified. Since optimal thresholds can vary depending on region, illumination conditions, and land cover, please explain how this value was validated for the Greek mountains. A sensitivity analysis or regional calibration would strengthen the approach.

  While optimized thresholds have been shown to increase accuracy in the classification of binary snow cover from optical satellite sensors, this calibration requires a large amount of *in situ* observations (Notarnicola, 2020; Poussin et al., 2023). Unfortunately, Greece has no such network of snow depth stations, and therefore such an optimization is not yet possible. The use of a steady NDSI > 0.4 threshold was selected due to its ability to still offer high classification accuracy and maintain consistency across time and space. Furthermore, while testing different thresholding schemes falls outside of the scope of this present study, the model does allow users to (a) choose from a list of five threshold-based snow binarization schemes, (b) configure their own thresholds, or (c) add a new custom snow binarization scheme that best fits their research needs.

Notarnicola, C.: Hotspots of snow cover changes in global mountain regions over 2000–2018, Remote Sensing of Environment, 243, 111781, https://doi.org/10.1016/j.rse.2020.111781, 2020.

Poussin, C., Timoner, P., Chatenoux, B., Giuliani, G., and Peduzzi, P.: Improved Landsat-based snow cover mapping accuracy using a spatiotemporal NDSI and generalized linear mixed model, Science of Remote Sensing, 7, 100078, https://doi.org/10.1016/j.srs.2023.100078, 2023.

We have provided further context regarding our decision to use a steady NDSI threshold:

Lines 188-191: "While spatiotemporally optimized NDSI thresholds can improve classification performance (Härer et al., 2018; Poussin et al., 2023), their calibration typically relies on in situ observations, which are unavailable in our study area. We therefore adopt a fixed NDSI threshold, which has been shown to provide high accuracy in binary snow cover classification while maintaining consistent performance across time and space (Gascoin et al., 2015; Koehler et al., 2022; Moazzam et al., 2022)."

- The 50% FSC threshold used to binarize MODIS data also appears empirical. Please clarify whether this threshold was calibrated or evaluated against alternative values.

The FSC threshold used to binarize MODIS was derived from the literature (lines 151-152), where it has been widely used (Notarnicola, 2020; Shen et al., 2025). While we did not perform any further calibration or evaluation, as this would fall outside of the scope of this study, as we mentioned above, the model does allow users to configure their own custom thresholds. Furthermore, in the case of MODIS data, we would like to highlight that it is only used in an auxiliary capacity. This means that MODIS data will only be used in pixels that (a) do not have satellite observations, and (b) do not satisfy the first temperature-based decision-tree gap-filling criterion (lines 259-261).

Notarnicola, C.: Hotspots of snow cover changes in global mountain regions over 2000–2018, Remote Sensing of Environment, 243, 111781, https://doi.org/10.1016/j.rse.2020.111781, 2020.

Shen, Y., Wang, X., Zhu, R., Che, T., and Hao, X.: A Downscaling Algorithm for Snow Cover Extent Over the Tibetan Plateau Based on a Similar Conditional Probability and Otsu's Method, IEEE Transactions on Geoscience and Remote Sensing, 63, 1–14, https://doi.org/10.1109/TGRS.2025.3543433, 2025.

No change was made, based on the explanation provided above.

- MODIS data were resampled from 500 m to 100 m using bicubic interpolation. Please justify this choice, as such resampling does not introduce new spatial information and may lead to smoothing artifacts. Why was this approach preferred over simpler methods (e.g., nearest neighbor) or dedicated downscaling techniques?

  We chose bicubic resampling over nearest neighbour, not despite its smoothing effect, but because of it. We believe that at the 100 m scale, smoothing the hard edges of 500 m MODIS grid cells before extracting binary snow cover values from them, allows for a more realistic representation of snow cover at our final spatial resolution. At the same time, although, as you correctly point out, new spatial information and a more dedicated downscaling technique would offer a more accurate representation, we have seen bicubic resampling used in a similar context with good results in a much less computationally demanding approach (Kollert et al., 2024; lines 152-154). Although a more dedicated MODIS downscaling module might be developed as part of future versions of the snowMapper, we would like to emphasize that MODIS is currently only used at an auxiliary capacity.

  Kollert, A., Mayr, A., Dullinger, S., Hülber, K., Moser, D., Lhermitte, S., Gascoin, S., and Rutzinger, M.: Downscaling MODIS NDSI to Sentinel-2 fractional snow cover by random forest regression, Remote Sensing Letters, 15, 363–372, https://doi.org/10.1080/2150704X.2024.2327084, 2024.

  No change was made, based on the explanation provided above.

- The use of MODIS Terra only is not justified. Combining Terra (MOD10A1) and Aqua (MYD10A1) products is commonly used to reduce cloud contamination and improve temporal coverage. Please explain why Aqua data were not included, particularly given the importance of gap-filling in this study.

  While MODIS Aqua may have provided additional snow cover information, we decided against its integration for three reasons:

  1. Overpass time in our study region is ~13:30. This is inconsistent with MODIS Terra as well as the higher-resolution missions (Landsat, Sentinel-2), which all pass around 10:00, allowing us to calculate daily aggregates of the meteorological conditions around that time, in order to obtain the most up-to-date meteorological information for each pixel (lines 164-166).
  2. In a comparison of MODIS Terra's and Aqua's snow detecting capabilities over the Pyrenees, the latter was found to be less accurate (Gascoin et al., 2015).
  3. As mentioned earlier, MODIS data are used only in an auxiliary capacity, during gap filling, and therefore we believe that the Terra collections alone are able to satisfy the needs of the model at that stage.

  Gascoin, S., Hagolle, O., Huc, M., Jarlan, L., Dejoux, J.-F., Szczypta, C., Marti, R., and Sánchez, R.: A snow cover climatology for the Pyrenees from

MODIS snow products, Hydrology and Earth System Sciences, 19, 2337–2351, https://doi.org/10.5194/hess-19-2337-2015, 2015.

We have provided further context regarding our decision to exclude MODIS Aqua data here:

Lines 213-216: "MODIS Aqua-derived snow cover was not used due to its lower snow detection accuracy relative to MODIS Terra (Gascoin et al., 2015), as well as its later overpass time (~13:30 compared to ~10:00 for MODIS Terra, Sentinel-2, and Landsat). This temporal offset causes inconsistencies in the calibration of the meteorological forcing data described in the following subsection."

Comment on Section 2.3.4 (In situ data):

The training data are derived from stations in the Alps and Pyrenees rather than from Greece. Please justify the transferability of the model to Mediterranean snow conditions, which may differ significantly.

As we point out in section "2.3.4 In situ data" (lines 199-200), no data are available for Greece. However, this is what inspired us to create our 'physics-informed' machine learning algorithm, which simulates snow cover conditions by performing iterations that take into account the continuously cumulating meteorological conditions, including heating, cooling, precipitation, and continuous snow cover days. To ensure that the machine learning algorithm takes into account only physical variables, we have removed all geographical ones (lines 224-225). We believe that the accuracy assessment of the model provides satisfactory evidence that this approach does indeed work, and can be transferable to other regions. Having said that, users are of course still able and welcome to create their own classifier with data from their own study region, or a region with similar climate conditions to theirs, enabling the model to always provide the most accurate simulation (modules for preprocessing station data & training a machine learning classifier are included in our code repository to facilitate this process).

No change was made, based on the explanation provided above.

Comment on Section 2.4.1 (Machine learning classifier):

The Random Forest hyperparameters (e.g., number of trees = 30, minimum leaf size = 1, bag fraction = 0.5) are specified, but their selection is not justified. Please clarify how these values were chosen (e.g., cross-validation, sensitivity analysis, or empirical testing).

In the case of number of trees, we used 30, following an error matrix sensitivity analysis. In the case of minimum leaf size and bag fraction, we used the default settings recommended in Google Earth Engine, due to no information on these appearing in our search in the literature for applications of machine learning methods in snow science.

We have provided more clarity regarding the selection of the machine learning hyperparameters here:

Lines 307-309: "In our case, the classifier was configured with 30 trees following an error matrix sensitivity analysis during trial runs. The minimum leaf population and a bag fraction were set to their default values of 1 and 0.5, respectively."

Comment on Section 2.4.5 (Final output):

The computation of monthly aggregates is not clearly described. Please clarify how daily snow cover is aggregated to monthly values (e.g., mean, maximum, or fraction of snow-covered days). In addition, the method used to convert daily binary snow maps into monthly fractional snow cover (FSC) should be explicitly defined.

In this step, daily binary snow cover values are subject to two consecutive aggregations:

- First, an aggregation by mean is applied in the temporal domain, ultimately describing the fraction of snow-covered days in a given pixel (i.e., temporal FSC at pixel scale), and
- Second, an aggregation by mean is applied in the spatial domain, ultimately resulting in a final value per mountain/study area, per month (i.e. spatio-temporal FSC at study scale).

Indeed, this information was missing, and we thank the reviewer for pointing it out. We will provide further clarification on the updated manuscript.

We have applied the following change:

Lines 374-376: "Once snowMapper has fully reconstructed daily snow cover across the domain and simulation period, we aggregated binary snow cover values to monthly means for each massif. The resulting variable represents monthly spatiotemporal fractional snow cover (FSC) at the massif scale."

Comment on Figure 4:

Figure 4 is not easy to interpret. The definition of "fraction of pixels" is unclear, and it is not specified how these monthly proportions are computed. Please provide additional information in the figure caption. In addition, the machine learning contribution appears relatively constant over time; please clarify how this fraction is computed and whether it varies across years.

By fraction of pixels, we simply refer to the percentage of pixels that each month came from clear-sky satellite observations (Landsat/Sentinel-2), or were gap-filled using decision trees, or machine learning. These metrics are derived from flags during the daily snow cover reconstruction, which are then aggregated on a monthly scale. Therefore, 100% = all pixels of the area, across all days of that given month.

Regarding the contribution of the machine learning gap-filling, the graph correctly shows that more weight is given to decision tree gap filling and consequently less to machine learning after 2000, due to MODIS (decision tree gap filling step No.2 - lines 262-263). This variation across years is described in lines 353-355.

We will provide further clarification, as requested.

We have updated the y-axis of Figure 4, to display values in percentages instead of fractions. We have also edited the caption to provide further clarification.

Lines 491-492: "For each month-year, 100% represents the sum of all grid cells of the study area across all days of that given month.

Comment on Figure 5:

Although Figure 5 describes the temporal aggregation of the metrics, the evaluation methodology is not fully clear. Please clarify what datasets are being compared (e.g., model outputs vs. observations) and whether the evaluation is performed at the pixel level over the study area.

As described in section "2.5 Model evaluation & bias correction", first a pixel-level true positive/false positive/true negative/false negative classification is given to the model data, by comparing pre-assimilation results with any clear-sky observations (Landsat/Sentinel-2). We do not need to use a subsection of those observations for validation, as they are still independent during the evaluation step – replacing simulated values with the observed clear-sky ones comes after that evaluation. Once the model run is complete, during a postprocessing step, all tp/fp/tn/fn values are aggregated in time, and space (over the study area), and used to calculate the monthly accuracy, underestimation, and overestimation metrics using Eq. 11-13 (lines 290-298).

We applied the following changes:

Lines 385-386: "All available observations were used for validation, as they remained independent at this step and were only replaced after evaluation, following the process described in Section 2.4.3."

Lines 391-392: "At the end of the model run, we cumulate these daily grid cell-level evaluations across each month and massif, and calculate monthly performance metrics"

In the figure caption (**Line 399**), it is stated that the evaluation is performed at a study-area level.

Reviewer 2

This preprint presents an analysis of snow cover trends in the Greek mountains, a region where knowledge of the cryosphere remains limited. The manuscript reads very well. The methods and analyses are well illustrated, and the conclusions are clearly outlined. The code developed in this study should be readily transferable to other regions. Overall, I believe this is an original and important contribution.

Main comments

The paper has a double objective: (1) to introduce a new algorithm for snow cover area reconstruction, and (2) to present an original analysis of snow cover in the Greek mountains. The introduction is not focused specifically on snow cover in Greece, which is a good choice since it broadens the potential applications of the SnowMapper algorithm. However, it lacks a review of previous studies on snow cover in Greece, which is important for contextualizing the second part of the paper dedicated to snow cover trends in this region. In particular, the authors could better position their analysis with respect to the previous work of Masloumidis et al. (2025). I am also somewhat surprised that there appear to be no other relevant studies on this topic.

We will discuss further the work done by Masloumidis et al. (2025), as well as any other relevant regional/global-scale studies, in order to better contextualize the climatology part of the manuscript. Unfortunately, the available literature on this topic in Greece is extremely limited (Fayad et al., 2017), highlighting the extent of the current knowledge gap, the relevance of the present study, and the need for more studies like it.

Fayad, A., Gascoin, S., Faour, G., López-Moreno, J. I., Drapeau, L., Page, M. L., and Escadafal, R.: Snow hydrology in Mediterranean mountain regions: A review, Journal of Hydrology, 551, 374–396, https://doi.org/10.1016/j.jhydrol.2017.05.063, 2017.

We've provided additional context on climatologically-relevant literature:

Lines 50-72: "In the Balkan Peninsula, extreme minimum temperatures exhibit a positive latitudinal gradient, and the frequency of winter cold spells declines towards the end of the 1961-2019 period (Tringa et al., 2022). Large negative trends in snowfall have also been identified across the Peninsula's mountain regions between 1979-2018 (Faranda, 2020).

In Greece, analyses of snow cover days between 1991-2020, indicate an initial increase followed by rapid declines of up to -1.5 days yr-1, with reductions becoming more frequent and severe in the latter half of the study period (Masloumidis et al., 2025). Similar patterns are reported for snow depth and snow water equivalent in two lowland catchments in northern Greece, based on the Global Land Data Assimilation System Noah Land Surface Model L4 V2.0, at a 0.25° resolution (Voudouri et al., 2023). According to Masloumidis et al. (2025), the strongest and statistically significant trends are concentrated in winter, particularly in the mountains of northwestern Greece. However, these findings may be constrained by the coarse spatial resolution of the underlying reanalysis products. Although CERRA-Land provides relatively high resolution for such datasets (~5.5 km), its ability to resolve snow processes at finer spatial scales (such as small mountain massifs) or during transitional periods (autumn and spring), remains limited (Monteiro and Morin, 2023). The inclusion of September and October in the definition of autumn snow cover by Masloumidis et al. (2025) may have also obscured early snow season trends, given that snow onset in the Greek mountains typically occurs in November. Consequently, analyses based on higher resolution datasets may reveal stronger and more spatiotemporally heterogenous trends.

Indeed, a MODIS-based snow cover phenology dataset at 500 m resolution for the 2000-2023 period show declines in snow cover duration of up to -4.1 days yr-1 across the Greek mountains (Notarnicola, 2024a), nearly three times the maximum rate reported in courser-resolution analyses (Masloumidis et al., 2025). While these high-resolution data offer valuable insight into sub-kilometre-scale snow cover variability, expanding the temporal coverage to earlier decades would further improve understanding of long-term changes and their relationship to ongoing climate change."

As acknowledged by the authors, a key challenge is the lack of satellite data prior to the 2000s. I believe the results would be strengthened if the authors tested the robustness of their trend analysis with respect to the increasing availability of satellite observations (Bayle et al., 2024). For example, the trend analysis could be repeated without the assimilation of satellite data. Another option (potentially better, but more difficult to implement) would be to limit the annual number of assimilated satellite observations to a constant value. At least, this issue should be discussed.

We acknowledge the need for further deliberation on this issue, which we will do in our discussion section. While testing the sensitivity of trend analysis to the availability of imagery is important when trends are relying purely on imagery, we do not think it would be particularly informative in our case. The reason is that our trend analysis does not depend on the irregularly captured and irregularly spaced satellite imagery, but rather on monthly aggregates derived from reconstructed daily snow cover maps. The main purpose of satellite observations in our model is to 'course-correct' the simulated snow cover and limit the propagation of gap-filling errors, in time. Furthermore, based on our evaluation of snowMapper's performance, it appears mainly to underestimate snow cover, rather than overestimate it, and this has remained fairly consistent over the four decades. In fact, this underestimation is more evident in the first two decades of the reconstruction, which, if anything, would then tend to underestimate our negative snowcover trends over the full period.

We have added the following discussion point:

Lines 576-580: "Although our analysis incorporates satellite-derived snow cover data, these observations are used within snowMapper primarily to constrain the simulated snow cover and limit the temporal propagation of gap-filling errors. Therefore, the reported trends are unlikely to be biased by the increasing availability of satellite observations in recent decades (Bayle et al., 2024). This is because the analysis is based not on irregularly captured satellite imagery, but on monthly aggregates derived from reconstructed daily snow cover maps."

As noted by Reviewer 1, the calculation of the monthly SCA should be clarified: "Daily binary snow cover is converted to monthly fractional snow cover (FSC).". This step involves both temporal and spatial aggregation. I assume that the authors computed the monthly average of the daily snow cover fraction within each mountain polygon. In that case, it is possible to obtain identical FSC values from very different snow conditions. For example, a constant 10% snow cover throughout an entire month would yield the same FSC as a short-lived 100% snow cover lasting only three days (assuming a 30-day month).
10% SCA × 30 days / 30 days per month = 10% / month
100% SCA × 3 days / 30 days per month = 10% / month
If my understanding is correct, this limitation should be acknowledged and discussed.

Your understanding and examples are indeed correct. As you point out, this simplification of a spatiotemporally distributed dataset is a limitation, necessary to obtain a monthly spatiotemporal FSC. However, we believe that the relevance of this with respect to the climatological analysis is limited. While several different combinations of (a) daily snow cover extent and (b) its associated duration within a month, can result in the same monthly spatiotemporal FSC, this metric encapsulates that information. Ultimately, as described in more hydrologically-focused studies, the value of the snowpack is a relationship of "how much" water is stored as snow, and "how long" it stays stored in that frozen form for (Aragon & Hill, 2024). The spatiotemporal FSC informs on both the extent of the snow cover and its duration.

Aragon, C. M. and Hill, D. F.: Changing snow water storage in natural snow reservoirs, Hydrology and Earth System Sciences, 28, 781–800, https://doi.org/10.5194/hess-28-781-2024, 2024.

We have made the following change:

Lines 376-386: "Although this aggregation reduces the dimensionality of the dataset, such that different combinations of daily snow cover extent and duration within a month may produce identical monthly FSC values, the metric implicitly captures both spatial extent and temporal persistence of snow cover. This is consistent with hydrological perspectives, in which the value of the snowpack depends on both the volume of water stored as snow and the duration of its storage in that frozen form (Aragon and Hill, 2024). Accordingly, spatiotemporal FSC provides an integrated measure of snow cover extent and duration, offering a versatile metric for dataset evaluation and analysis."

Minor comments
In their literature review, the authors may also consider the study by Zakeri and Mariethoz (2024).

Thank you for bringing this study to our attention; we will reference it.

We have added the following:

Lines 85-87: "More recently, a climatology-based method was introduced for gap-filling by linking preceding meteorological conditions to snow cover patterns and identifying analogous states through a k-nearest neighbour algorithm (Zakeri and Mariethoz, 2024)."

L169: Is this the full MicroMet algorithm, including the Barnes objective analysis scheme?

The quick answer is no. Apart from this being inherently complicated and computationally expensive within the GEE environment, it is our understanding that the Barnes objective analysis scheme is most useful when the MicroMet algorithm is applied to interpolate station data 'randomly' located in an area, rather than a uniform grid of points extracted from a reanalysis product. In our case, we perform a nearest neighbor resampling of the original grid to the desired spatial resolution, and we then correct the meteorological values using the respective MicroMet equations and lapse rates/correction factors for temperature and precipitation (Liston & Elder, 2006). We will include this information in the manuscript.

Liston, G. E. and Elder, K.: A Meteorological Distribution System for High-Resolution Terrestrial Modeling (MicroMet), Journal of Hydrometeorology, 7, 217–234, https://doi.org/10.1175/JHM486.1, 2006.

We applied the following changes:

Lines 233-235: "We first resampled ERA5-Land data to the target resolution using the nearest-neighbour interpolation. Downscaling was then performed using the 30 m SRTM DEM v3 (Farr et al., 2007) resampled bicubically to 100 m, together with the respective MicroMet temperature and precipitation functions."

L366: typo (parentheses)

Thank you – in those brackets we will add the missing legend item "standard deviations (shading)".

We applied the correction described above (**Line 501**).

L399: I wonder if the significance in the "increase in the frequency of extremely low snow cover instances" is influenced by the temporal autocorrelation in the negative anomalies timeseries (i.e. a negative anomaly in April is more likely if there was a negative anomaly in March). I am not an expert in trend analysis but I suspect that this could inflate the trend significance?

We will explore this further and provide a more complete answer in the revised manuscript.

While 'frequency' in our case refers to the number of months within a hydrological year with extremely high or low snow cover, we have now recalculated the trends in both magnitude and frequency of extremes using a modified Mann-Kendall (MMK) test, ideal for detecting trends in time series data that are influenced by serial correlation. The updated results include a significant decline in both magnitude and frequency of years with extremely high snow cover. At the same time, years of extremely low snow cover present no significant trend in neither magnitude nor frequency. However, we wish to point out that these changes to the results are not linked to the choice of MMK over MK (both deliver comparable p-values overall), but rather to a mistake in the aggregation of extreme values which were previously incorrectly grouped by calendar year instead of hydrological year. This has now been corrected, and the correct p-values are reported on the manuscript. We've made the following changes on the manuscript:

Lines 448-451: "We calculated SCA anomalies relative to the 1984-2025 monthly means and identified extreme high and low snow-cover events based on the monthly 80th and 20th percentile thresholds. Extremes were grouped by snow season, and aggregated as means for magnitude and as counts for frequency. The autocorrelation-sensitive modified Mann-Kendall test was then applied (Hamed and Ramachandra Rao, 1998) to identify trends in both (Li et al., 2025)."

Lines 533-535: "Trend analysis of the monthly SCA anomalies across the study area and time period reveals a significant decline in both the magnitude and frequency of extremely high snow-cover instances, while no significant trend is observed for extremely low snow cover instances (Fig. 8)."

Lines 571-573: "Corresponding patterns are observed in the extremely positive SCA anomalies, which display a marked shift - both in frequency and magnitude - toward milder conditions, consistent with the reported decline in winter cold spell frequency (Tringa et al., 2022) and the overall warming trends (Fig. A8)."

References

Bayle, A., Gascoin, S., Berner, L. T., and Choler, P.: Landsat-based greening trends in alpine ecosystems are inflated by multidecadal increases in summer observations, Ecography, e07394, https://doi.org/10.1111/ecog.07394, 2024.

Masloumidis, I., Dafis, S., Kyros, G., Lagouvardos, K., and Kotroni, V.: Snow Cover and Depth Climatology and Trends in Greece, Climate, 13, https://doi.org/10.3390/cli13020034, 2025.

Zakeri, F. and Mariethoz, G.: Synthesizing long-term satellite imagery consistent with climate data: Application to daily snow cover, Remote Sens. Environ., 300, 113877, https://doi.org/10.1016/j.rse.2023.113877, 2024.

References added

Alexopoulos, K., Willis, I., Pritchard, H., Kyros, G., Kotroni, V., and Lagouvardos, K.: Monthly Fractional Snow Cover aggregates over the Greek mountains reconstructed with snowMapper, Zenodo [data set], https://doi.org/10.5281/zenodo.19335252, 2026.

Aragon, C. M. and Hill, D. F.: Changing snow water storage in natural snow reservoirs, Hydrology and Earth System Sciences, 28, 781–800, https://doi.org/10.5194/hess-28-781-2024, 2024.

Bayle, A., Gascoin, S., Berner, L. T., and Choler, P.: Landsat-based greening trends in alpine ecosystems are inflated by multidecadal increases in summer observations, Ecography, 2024, e07394, https://doi.org/10.1111/ecog.07394, 2024.

Bousbaa, M., Boudhar, A., Kinnard, C., Elyoussfi, H., Karaoui, I., Eljabiri, Y., Bouamri, H., and Chehbouni, A.: An accurate snow cover product for the Moroccan Atlas Mountains: Optimization of the MODIS NDSI index threshold and development of snow fraction estimation models, International Journal of Applied Earth Observation and Geoinformation, 129, 103851, https://doi.org/10.1016/j.jag.2024.103851, 2024.

Faranda, D.: An attempt to explain recent changes in European snowfall extremes, Weather and Climate Dynamics, 1, 445–458, https://doi.org/10.5194/wcd-1-445-2020, 2020.

Hamed, K. H. and Ramachandra Rao, A.: A modified Mann-Kendall trend test for autocorrelated data, Journal of Hydrology, 204, 182–196, https://doi.org/10.1016/S0022-1694(97)00125-X, 1998.

Härer, S., Bernhardt, M., Siebers, M., and Schulz, K.: On the need for a time- and location-dependent estimation of the NDSI threshold value for reducing existing uncertainties in snow cover maps at different scales, The Cryosphere, 12, 1629–1642, https://doi.org/10.5194/tc-12-1629-2018, 2018.

Kashinath, K., Mustafa, M., Albert, A., Wu, J.-L., Jiang, C., Esmaeilzadeh, S., Azizzadenesheli, K., Wang, R., Chattopadhyay, A., Singh, A., Manepalli, A., Chirila, D., Yu, R., Walters, R., White, B., Xiao, H., Tchelepi, H. A., Marcus, P., Anandkumar, A., Hassanzadeh, P., and Prabhat: Physics-informed machine learning: case studies for weather and climate modelling, Philos Trans A Math Phys Eng Sci, 379, 20200093, https://doi.org/10.1098/rsta.2020.0093, 2021.

Maharjan, S., Li, W., Thomas, R., Fazli, S., Ansari, A., Morgan, H., Elgendy, A., Allali, M., and El-Askary, H.: Physics-informed deep learning reveals climate-driven snowpack decline and threatens ecological water availability in a Californian snow-fed catchment, Ecological Informatics, 92, 103526, https://doi.org/10.1016/j.ecoinf.2025.103526, 2025.

Meng, C., Griesemer, S., Cao, D., Seo, S., and Liu, Y.: When physics meets machine learning: a survey of physics-informed machine learning, Mach. Learn. Comput. Sci. Eng, 1, 20, https://doi.org/10.1007/s44379-025-00016-0, 2025.

Poussin, C., Timoner, P., Chatenoux, B., Giuliani, G., and Peduzzi, P.: Improved Landsat-based snow cover mapping accuracy using a spatiotemporal NDSI and generalized linear mixed model, Science of Remote Sensing, 7, 100078, https://doi.org/10.1016/j.srs.2023.100078, 2023.

Tringa, E., Tolika, K., Anagnostopoulou, C., and Kostopoulou, E.: A Climatological and Synoptic Analysis of Winter Cold Spells over the Balkan Peninsula, Atmosphere, 13, https://doi.org/10.3390/atmos13111851, 2022.

Voudouri, K. A., Ntona, M. M., and Kazakis, N.: Snowfall Variation in Eastern Mediterranean Catchments, Remote Sensing, 15, 1596, https://doi.org/10.3390/rs15061596, 2023.

Ye, F., Cheng, Q., Hao, W., Yu, D., Ma, C., Liang, D., and Shen, H.: Reconstructing daily snow and ice albedo series for Greenland by coupling spatiotemporal and physics-informed models, International Journal of Applied Earth Observation and Geoinformation, 124, 103519, https://doi.org/10.1016/j.jag.2023.103519, 2023.

Zakeri, F. and Mariethoz, G.: Synthesizing long-term satellite imagery consistent with climate data: Application to daily snow cover, Remote Sensing of Environment, 300, 113877, https://doi.org/10.1016/j.rse.2023.113877, 2024.

Zhao, W., Fang, J., Yang, T., Lian, X., Winkler, A. J., Sun, F., and Gentine, P.: Observation-Constrained Physical Snow Water Equivalent Simulations Using a Physics-Guided Machine Learning Approach, Water Resources Research, 62, e2025WR041406, https://doi.org/10.1029/2025WR041406, 2026.

Zhong, L., Lei, H., and Gao, B.: Developing a Physics-Informed Deep Learning Model to Simulate Runoff Response to Climate Change in Alpine Catchments, Water Resources Research, 59, e2022WR034118, https://doi.org/10.1029/2022WR034118, 2023.

We would like to thank both reviewers for supporting the improvement of our paper! All your comments are very valuable.