Dear Editor,

We would like to thank the two Reviewers for the reviews and suggestions that helped us improve our manuscript. We kindly request that the editor grant us one more opportunity for peer review. We believe that this work will advance the application of machine learning in snow depth and snow water equivalent (SWE) monitoring with remote sensing.

We have addressed all the points highlighted by the reviewers and we modified the manuscript accordingly. In particular:

(1) Clarification of the scope and title of the review

In response to concerns that the title did not fully reflect the content, we revised the title to "A decadal review (2015–2025) of machine learning models applied to remote sensing-based snow depth/snow water equivalent retrieval: statistical analysis, existing paradigms, and future directions". We also clarified that snow depth and SWE are two key parameters characterizing snow mass, that they are highly correlated, and that both were included in the search criteria; this decision is explained in Section 2. In addition, we clarified the scope of "remote sensing," noting that ground-based, airborne, and spaceborne observations were considered.

(2) Addition of quantitative accuracy metrics

The reviewer asked for quantitative comparisons rather than qualitative statements. In response, we have added accuracy metrics throughout the manuscript, including the Abstract. For example, we now report that compared with purely data-driven machine learning, coupling machine learning with snow physical models (data-model-driven) improves snow depth/SWE retrieval accuracy by approximately 53%; that machine learning can predict microwave brightness temperature with an accuracy of approximately 2–4 K, which is comparable to that of snow radiative transfer model (RTM) models; that machine learning can accelerate RTM simulations by at least two orders of magnitude; that data fusion and bias correction improve gridded snow depth products by approximately 55%; and that multilayer RTMs reduce the bias from −13.4 cm to 2.4 cm and RMSE from 15.9 cm to 10.2 cm.

(3) Reorganization of the structure, and removal of unsupported content

We simplified the review of RTM and traditional snow depth/SWE retrieval algorithms in Introduction, removing material that was unclear or insufficiently connected to snow depth/SWE. We have also rewritten the Abstract and Introduction so that they begin directly with the importance of snow depth/SWE and the challenges in monitoring them, rather than with extreme weather or snow cover extent. In addition, we added an explicit description of the overall structure at the end of the Introduction and reorganized the paper so that Sections 2, 3, and 4 correspond directly to "statistical analysis" "existing paradigms" and "future directions". We also corrected typos and unclear expressions throughout.

(4) Clarification of the five existing paradigms and the five future challenges

We clarified the five existing paradigms in Section 3: coupling of machine learning and snow physical model (snow RTM or process model), substituting RTM with machine learning for predicting microwave signals in assimilation or iteration algorithms, optimizing RTM by providing key inputs or accelerating operational efficiency, improving existing gridded snow depth/SWE products, and downscaling coarse snow depth/SWE products to a fine-scale resolution. Our study concluded that data-model-driven machine learning retrieval algorithms outperform purely data-driven ML models, but this advantage relies on sufficiently accurate physical model simulations. Therefore, Section 4 now clearly presents five interconnected challenges from three

perspectives: physical models, machine learning, and engineering applications. Physically, we need to bridge the disparity between RTM simulations and remote sensing observations (Section 4.1). For machine learning, we need to integrate fundamental physical laws into machine learning structures to enhance interpretability (Section 4.2) and improve the quality of machine learning training samples to strengthen generalization capability (Section 4.5). Regarding water resource management, we need to focus on achieving fine-scale snow depth/SWE monitoring at large scales by integrating multi-source remote sensing data using machine learning (Section 4.3), and on enhancing snow depth/SWE estimates under complex conditions to achieve spatiotemporally continuous snowpack monitoring.

(5) Revision of the bibliometric analysis

To better understand trends and the impact of machine learning models in remote sensing-based snow depth/SWE monitoring, we conducted a statistical analysis of 127 publications. We added the research purpose in each part of Section 2 to improve the logical flow and readability. Meanwhile, we simplified the bibliometric analysis, deleted the top 10 cited articles (Section 2.4), clarified the citation source and search procedure, and explained how this analysis supports the subsequent paradigm synthesis.

In the following, we answer in detail the specific comments made by the reviewers. The author responses are reported in blue right below the reviewers' comments. Line and page numbers indicated by the reviewers are referred to the submitted paper.

Reviewer 1 (anonymous)

Thank you very much for providing the chance to review "Review article: A decadal review (2015–2025) of machine learning models applied for satellite-based snow depth retrieval." The authors present the last decade of machine learning research on snow depth from spaceborne sensors. While I believe this topic is certainly worthwhile, I found the methods to be poorly justified and the topics to drift substantially throughout the paper.

We would like to thank the two anonymous reviewers for their affirmation of the originality of the paper. Both reviewers considered the originality of the paper to be good. Our main motivation for this study was to summarize the progress and existing challenges over the past decade and to better understand how machine learning techniques can serve snow remote sensing. We modified the manuscript accordingly.

Major Comments

Since each comment involves multiple issues, we have addressed them in separate responses.

1-1: I found the writing of this paper to be disorganized at times, with several paragraphs drifting in direction. For example, in the Abstract, the authors begin with extreme weather, then switch to snow depth as an indicator for climate change, and then switch to snow cover extent. A stronger beginning would simply be to discuss the importance of snow and the current uncertainties in measuring.

We fully agree with the reviewer's comments. Therefore, we have rewritten the abstract of the paper as follows:

"Understanding snow depth and snow water equivalent (SWE) distribution is crucial for studying

water processes, weather, and climate. Numerous remote sensing-based approaches have been used to derive spatially continuous snow depth/SWE. However, challenges remain in capturing their changes because of the nonlinear and so-called 'ill-posed' problems in retrieval algorithms. Machine learning (ML) techniques (including deep learning) are beginning to play important roles in advancing snow depth/SWE retrieval, particularly with microwave remote sensing, owing to their strong ability to fit nonlinear, nonexplicit functional relationships between snow depth/SWE and massive amounts of geoscience data. However, a systematic review of ML applications in snow depth/SWE retrieval with remote sensing is notably absent from the literature, and the trajectory for future advancements remains ambiguous. This review comprehensively summarizes the implementation (with an overview of the literature) and current paradigms of snow depth/SWE retrieval using ML techniques and remote sensing observations over the past decade (2015–2025) and proposes future research directions in response to the existing challenges. A statistical analysis of 127 studies revealed that the number of ML-related articles has increased over the past 10 years, from 3 to 33 per year. By first author affiliation, China and the United States lead in terms of contributions, accounting for almost 70% of the papers. We also report that nearly 50% of the studies focus on high-resolution snow depth/SWE retrieval (from a few metres to hundreds of metres) at regional or catchment scales, especially over mountains, which can be attributed to comprehensive ground-based and airborne field campaigns (e.g., SnowEx, NoSREx, and Airborne Snow Observatory (ASO) Lidar) and spaceborne SAR observations (e.g., Sentinel-1). Our decadal review concludes with five existing paradigms: (1) coupling ML with snow physical models (data-model-driven ML), which improves snow depth/SWE retrieval accuracy by 53% compared with purely data-driven ML methods; (2) substituting snow RTM with ML for simulating microwave signals, where ML can simulate signals with an accuracy of 2–4 K; (3) optimizing snow electromagnetic models by improving their operational efficiency, with ML accelerating RTM simulations by at least two orders of magnitude; (4) improving existing gridded snow depth/SWE products, which enhances their accuracy by approximately 55% through data fusion; and (5) downscaling coarse snow depth/SWE products from the conventional tens of kilometres to the hundred-metre level. However, some challenges and unresolved issues still exist. Future efforts in snow remote sensing research should aim to bridge the disparity between forward model simulations and remote sensing observations to facilitate data-model-driven ML; integrate fundamental physical laws into ML structures to enhance the interpretability of ML; achieve fine-scale snow depth/SWE monitoring at large scales through the integration of multisource remote sensing data (PM, SAR, Lidar, etc.); enhance ML-based snow depth/SWE estimates under complex conditions to achieve spatiotemporally continuous snowpack monitoring; and improve the quality of ML training samples to strengthen its generalizabillity. Overall, this paper provides a comprehensive review of the applications of ML techniques in snow depth/SWE remote sensing, focusing on statistical analysis, current paradigms, and potential future research directions. We believe that ML techniques hold significant potential for addressing the challenges associated with the quantitative inversion of snow depth/SWE and deepening our understanding of the spatial variability of snowpack globally."

1-2: Throughout the paper their structure is somewhat unclear. They present a synthesis on snow depth (but also at times SWE).

We sincerely apologize for the confusion caused by the unclear structure.

Since the SWE is the product of snow depth and snow density, snow depth and the SWE are highly correlated. Even in snow remote sensing, snow depth and SWE are often not distinguished from each other; for example, when snow density is assumed to be a fixed value, 0.24 g/cm^3 (Takala et al., 2012; Luojus et al., 2021). According to our 127 remaining publications, 89 articles focused on snow depth, 32 on SWE retrieval, and 7 on snow depth and SWE simultaneously. Most importantly, the machine learning methods they proposed are highly representative, e.g., Xue et al., RSE, 2015; Bair et al., TC, 2018; Santi et al., TGRS, 2022; Cui et al., JH, 2023; Yang et al., RSE, 2025; Kumawat et al., RSE, 2025.

Therefore, we retained the SWE-related references. To clarify this for readers, we have revised the title of the paper: "A decadal review (2015–2025) of machine learning models applied to remote sensing-based snow depth/snow water equivalent retrieval: statistical analysis, existing paradigms, and future directions". In addition, we have added an explanation in the Introduction: (Page 2, Lines 51-52 in the revised manuscript): "The amount of snowfall (snow mass) is typically measured by two different parameters, snow depth and snow water equivalent (SWE). Since SWE is the product of snow depth and snow density, snow depth and SWE are highly correlated."; Section 2 (Page 6, Lines 164-165 in the revised manuscript): "Since snow depth and SWE are both key parameters for characterizing snow mass and are highly correlated, they, together with related variables, were included in the search criteria."

1-3: And then proceed to focus mostly on microwave-focused retrieval algorithms, and then they end with future directions. The ending is the strongest part, but the authors do not report specific accuracy metrics across studies so it is difficult to know which challenges are the most important.

We sincerely apologize for the poor readability caused by the unclear logic. The several challenges we summarized follow the logic below:

(1) According to the literature review, combining the ML technique and physical models has been a mainstream method for retrieving snow depth/SWE (see Section 3.1). Especially for microwave remote sensing, radiative transfer models (RTMs) are key physical models; thus, improving their simulation accuracy is necessary, particularly at the remote sensing scale. Our review revealed that ML is a promising tool for monitoring and imaging the internal snowpack structure, which is expected to improve the simulation accuracy of the RTM and reduce the discrepancy between simulations and satellite observations (Fig. 16).

(2) Current approaches for coupling ML with physical models are relatively limited, which constrains the physical interpretability of ML. Therefore, we summarized existing methods (which have been applied in other fields) for combining ML with mechanism-based models, such as differentiable modelling (DM) and loss function embedding (LFE) methods (Fig. 17).

(3) According to our literature review, most studies focus on coarse spatial resolution, whereas practical applications require higher-resolution snow depth/SWE products. Therefore, how to combine existing remote sensing data with ML methods to achieve high-resolution snow depth monitoring is challenging and represents a future research direction. We further proposed a method of combining multisource remote sensing data (e.g., radiometer, SAR, and LiDAR) and using ML methods to obtain high spatial resolution products (Fig. 18).

(4) Existing studies have begun to focus on snow depth/SWE in complex scenarios (alpine regions, polar regions, and snowmelt seasons), but snow cover is highly heterogeneous and difficult to

monitor using traditional methods. We call for more research teams to pay attention to snow depth/SWE distribution in complex scenarios so that full-spatiotemporal (daily and global) and full-life cycle (freezing and melting periods) snow observations can be obtained (Fig. 19).
(5) Since our professional background is not in computer science or artificial intelligence, I do not have much insight into the machine learning techniques themselves. Our review revealed that most training samples for ML are from stations. Although station observations are relatively accurate, their spatial representativeness and sample balance are difficult to guarantee. Therefore, we recommend combining multisource observations (station, Lidar, SAR, simulations, etc.) to enhance the quality of training samples, including their accuracy, representativeness, and balance (Fig. 20).

The five challenges and future research directions we propose are interconnected and mutually responsive. Challenge 1 targets physical models, Challenge 2 addresses the interpretability of ML, Challenge 3 focuses on spatial resolution, Challenge 4 concerns complex scenarios, and Challenge 5 is oriented towards ML training samples. Frankly speaking, we cannot determine which challenge is the most important; the ideal outcome would be for all of them to be addressed.
The reviewer's question is very interesting, and it prompted us to think more deeply about how to test or demonstrate, within a single framework, the potential for improving snow depth/SWE accuracy in each challenge. However, the present work is a review article and the corresponding results will be presented in our future research articles.

2: Additionally, the authors focus on snow depth for the entirety of the paper, but they include SWE in the search term without providing justification for why/why not this might not be related and whether SWE was later screened out or not. These two properties are different and require different both ML models and algorithms, and it would be good for the authors to justify this methods decision.
We have made the appropriate revisions accordingly. Please refer to the response to "Major Comments 1-2" above.

3-1: Critically, I think the authors need to spend more time digging substantially through the papers to provide quantitative comparisons (not just one is superior to the other, but by how much) and repeat their search to either definitively include SWE or not.
We sincerely thank the reviewer for the constructive comments. We have rewritten the abstract to add quantitative comparisons; please refer to the response to "Major Comments 1-1" above.
We are indeed attempting to compare the strengths and weaknesses of various methods within a unified framework. On the one hand, this requires a relatively long period for systematic evaluation; on the other hand, given that the present work is a review article, the corresponding results will be presented in our future research articles.

3-2: Finally, I think some stronger organization would help the reader to understand the overall flow, and I think the title does not reflect what the paper is about.
We thank the reviewer for the constructive suggestion. We have adjusted the structure of the paper in the following aspects:
(1) We modified the title of the paper to make its content immediately clear to readers: "A

decadal review (2015–2025) of machine learning models applied to remote sensing-based snow depth/snow water equivalent retrieval: statistical analysis, existing paradigms, and future directions".

(2) We have rewritten the Introduction section and added a description of the overall structure of the paper. According to our ten-year review, we compiled statistics on the literature, including the number of publications, main lines of development, spatial resolution, study areas, training sample data sources, and their relationships. Afterwards, we summarized five existing paradigms for improving snow depth/SWE with ML approaches. Finally, we defined the key challenges and proposed possible strategies for advancing research in this promising direction. Sections 2, 3, and 4 correspond precisely to the statistical analysis, existing paradigms, and future directions, as reflected in the title, thus rendering the structure of the paper more coherent and transparent.

(3) We added the overall framework to the last paragraph of the Introduction (Page 5, Lines 154-161 in the revised manuscript): "The remainder of this article is organized as follows. Section 2 provides statistics on the literature published over the past decade, including the number of publications, research scales, key milestones, and citation counts. Section 3 summarizes five existing paradigms for improving snow depth/SWE with ML approaches, namely, the coupling of ML and snow physical models, the substitution of RTM for simulating microwave signals, the optimization of snow physical models, the improvement of gridded snow depth/SWE products, and the downscaling of coarse snow depth/SWE datasets. Section 4 details the challenges faced and proposes solution directions, involving the simulation accuracy of RTM, the interpretability of ML, the enhancement of spatial resolution, spatiotemporally continuous monitoring in complex scenarios, and the quality of training samples. Finally, Section 5 concludes this study."

3-3: The current paper provides an overview of the literature, but it has a strong focus on microwave, that I think this needs to be better reflected.

Owing to the ability of microwave signals to penetrate snowpack, microwave remote sensing is a valid tool for retrieving snow depth/SWE. Among the 127 reviewed papers, 89 employed microwave observation data (e.g., SSM/I, SSMIS, AMSR-E/2, FY-3, HY-2, and Sentinel-1) as one of the ML training variables. In addition, 15 papers did not use microwave remote sensing observations directly; they adopted microwave remote sensing snow depth/SWE products (e.g., GlobSnow, AMSR2, and C-snow) as training variables and thus are closely related to microwave remote sensing. Accordingly, we have simplified the review of microwave radiative transfer models and snow depth/SWE retrieval algorithms in the introduction section (Pages 3–4).

Admittedly, approximately 10 studies directly employed reanalysis data (e.g., ERA5, MERRA2, and SNODAS), meteorological data (e.g., air temperature and wind speed) or auxiliary data (e.g., DOY, DEM, and snow cover) to drive machine learning models. To avoid ambiguity, we have provided an explanation in the introduction section (Page 2, Lines 63-66 in the revised manuscript): "In the literature reviewed in this study, nearly 80% of the studies applied machine learning (ML) models and microwave remote sensing observations to retrieve snow depth/SWE. Only a few studies have directly used reanalysis data (e.g., ERA5 and MERRA2) or auxiliary data (e.g., DOY, DEM, and snow cover) for snow depth/SWE retrieval. Therefore, we briefly review the microwave radiative transfer model (RTM) and snow depth/SWE retrieval algorithms."

Abstract

Line 25: “This review comprehensively summarizes the implementation and progress of snow depth research using microwave remote sensing over the last decade (2015–2025).” The title of this paper is related to ML, why do you say the review is summarizing implementing microwave remote sensing?

We revised this sentence as follows: “This review comprehensively summarizes the implementation (with an overview of the literature) and current paradigms of snow depth/SWE retrieval using ML techniques and remote sensing observations over the past decade (2015–2025) and proposes future research directions in response to the existing challenges”, making it more consistent with the title.

With respect to why microwave remote sensing is frequently mentioned, please refer to the response to “Major Comments 3-3” above.

Line 30: What is the importance of breaking down where the research takes place? How does that help us understand the progress of ML for snow depth? Is there a call for collaboration or ways that data can be shared across geographic boundaries?

We revised it to “We also found that nearly 50% of the studies focused on high-resolution snow depth/SWE retrieval (from a few metres to hundreds of metres) at regional or catchment scales, especially over mountains, which can be attributed to comprehensive ground- and airborne-based field campaigns (e.g., SnowEx, NoSREx, and ASO Lidar, etc.) and spaceborne SAR observations (e.g., Sentinel-1).”

In our view, large-scale field campaigns (e.g., SnowEx, NoSREx, and ASO Lidar) have indeed contributed to a better understanding of snow processes and improved model simulation accuracy. We therefore strongly recommend that China also conduct similar comprehensive field experiments, especially in high mountain regions.

Line 39: Who is “our”? is it the group of researchers on this paper? Researchers in a specific country or the snow / remote sensing community at-large?

We revised it to “Future efforts in snow remote sensing research should aim to bridge the disparity between forward model simulations and remote sensing observations to facilitate data-model-driven ML; integrate fundamental physical laws into ML structures to enhance the interpretability of ML; achieve fine-scale snow depth/SWE monitoring at large scales through the integration of multisource remote sensing data (PM, SAR, Lidar, etc.); enhance ML-based snow depth/SWE estimates under complex conditions to achieve spatiotemporally continuous snowpack monitoring; and improve the quality of ML training samples to strengthen its generalizability”.

Line 39: What is “model-observation mismatch”? Is that another word for error? Can they be more specific from their review of what some of the biggest drivers of model-observation mismatch/ aka error is?

We revised it to “bridge the disparity between forward model simulations and remote sensing observations (model–observation) to facilitate data-model-driven ML”.

Can this abstract say anything about well we are doing using ML for snow depth? I found a lot of this information could be summarized from abstracts and the title/author affiliation from the papers, but I would curious from the author's review if from across their review they could report on how well ML-physical models are performing and include those metrics in the abstract.

We thank the reviewer for the constructive suggestion. We added the metrics and rewrote the abstract: "Understanding snow depth and snow water equivalent (SWE) distribution is crucial for studying water processes, weather, and climate. Numerous remote sensing-based approaches have been used to derive spatially continuous snow depth/SWE. However, challenges remain in capturing their changes because of the nonlinear and so-called 'ill-posed' problems in retrieval algorithms. Machine learning (ML) techniques (including deep learning) are beginning to play important roles in advancing snow depth/SWE retrieval, particularly with microwave remote sensing, owing to their strong ability to fit nonlinear, nonexplicit functional relationships between snow depth/SWE and massive amounts of geoscience data. However, a systematic review of ML applications in snow depth/SWE retrieval with remote sensing is notably absent from the literature, and the trajectory for future advancements remains ambiguous. This review comprehensively summarizes the implementation (with an overview of the literature) and current paradigms of snow depth/SWE retrieval using ML techniques and remote sensing observations over the past decade (2015–2025) and proposes future research directions in response to the existing challenges. A statistical analysis of 127 studies revealed that the number of ML-related articles has increased over the past 10 years, from 3 to 33 per year. By first author affiliation, China and the United States lead in terms of contributions, accounting for almost 70% of the papers. We also report that nearly 50% of the studies focus on high-resolution snow depth/SWE retrieval (from a few metres to hundreds of metres) at regional or catchment scales, especially over mountains, which can be attributed to comprehensive ground-based and airborne field campaigns (e.g., SnowEx, NoSREx, and Airborne Snow Observatory (ASO) Lidar) and spaceborne SAR observations (e.g., Sentinel-1). Our decadal review concludes with five existing paradigms: (1) coupling ML with snow physical models (data-model-driven ML), which improves snow depth/SWE retrieval accuracy by 53% compared with purely data-driven ML methods; (2) substituting snow RTM with ML for simulating microwave signals, where ML can simulate signals with an accuracy of 2–4 K; (3) optimizing snow electromagnetic models by improving their operational efficiency, with ML accelerating RTM simulations by at least two orders of magnitude; (4) improving existing gridded snow depth/SWE products, which enhances their accuracy by approximately 55% through data fusion; and (5) downscaling coarse snow depth/SWE products from the conventional tens of kilometres to the hundred-metre level. However, some challenges and unresolved issues still exist. Future efforts in snow remote sensing research should aim to bridge the disparity between forward model simulations and remote sensing observations to facilitate data-model-driven ML; integrate fundamental physical laws into ML structures to enhance the interpretability of ML; achieve fine-scale snow depth/SWE monitoring at large scales through the integration of multisource remote sensing data (PM, SAR, Lidar, etc.); enhance ML-based snow depth/SWE estimates under complex conditions to achieve spatiotemporally continuous snowpack monitoring; and improve the quality of ML training samples to strengthen its generalizability. Overall, this paper provides a comprehensive review of the applications of ML techniques in snow depth/SWE remote sensing, focusing on statistical analysis, current paradigms, and potential future research directions. We believe that ML techniques hold significant potential

for addressing the challenges associated with the quantitative inversion of snow depth/SWE and deepening our understanding of the spatial variability of snowpack globally".

Introduction

The introduction currently reads as a comprehensive literature for the progress made in microwave remote sensing for snow depth, evolving from various microwave models. They also switch between multiple snow properties, including snow cover, snow depth, and SWE. I believe their title on decadal review of ML for snow depth is misleading. It seems a better title would be to clarify that the bulk of their review focuses on the evolution of spaceborne remote sensing for snow depth, touching on both physical and ML methods. As it reads right now ML isn't introduced until paragraph 5.

We thank the reviewer for the constructive suggestion. We have adjusted the structure of the paper in the following aspects:

(1) We have rewritten the first paragraph to directly emphasize the importance of snow depth and snow water equivalence: "The amount of snowfall (snow mass) is typically measured by two different parameters: snow depth and snow water equivalent (SWE). Since the SWE is the product of snow depth and snow density, snow depth and SWE are highly correlated. The successful estimation of snow depth/SWE can improve the accuracy of snowmelt-runoff prediction and the management of water supply (Tanasienko et al., 2011; Barnett et al., 2005; Sturm et al., 2017; Vano et al., 2020), contribute to predicting potential snow-related disasters, e.g., flooding in spring, snowstorms, and snow avalanches (Gao et al., 2016; Eckert et al.,2024), and enhance our understanding of global water cycle in the climate system (Hock et al., 2019). Remote sensing techniques are the only means to monitor spatiotemporally continuous snow depth/SWE at large scales".

(2) We modified the title of the paper to make its content immediately clear to readers: "A decadal review (2015–2025) of machine learning models applied to remote sensing-based snow depth/snow water equivalent retrieval: statistical analysis, existing paradigms, and future directions".

(3) We have deleted the review of microwave radiative transfer models and simplified the traditional snow depth/SWE retrieval algorithms (Page 3, Lines 67-90 in the revised manuscript).

Line 46: Sensitive to what? Temperature?

Revised.

Line 54: They introduce both snow depth and SWE, but they proceed to focus on snow depth. Can the authors include a sentence for why it is important to look at snow depth only? Why did they only look at snow depth?

We revised and clarified this accordingly; please refer to the response to "Major Comments 2" and "Introduction Comments" above.

Line 81: why is Model-drive method capitalized?

Revised.

Line 140 – “Additionally, owing to snow metamorphism, the compaction and restructuring of snow crystals lead to changes in their size, shape and bonds, greatly affecting their microwave scattering characteristics (Tsang et al., 2022). Thus, temporally and spatially dynamic snow characteristics still limit the use of ML in snow depth retrieval.” This reads to me as better fitting into their actual review (so not their intro) as they are drawing a conclusion from the literature. This is another example, where the paper felt disorganized and I believe the authors could spend some time ensuring that their introduction clearly sets up for why their paper matters before following into methods and analysis.

Here, we emphasize that snow metamorphism alters microscale snow parameters, which complicates microwave scattering and ultimately leads to larger errors in ML-based snow depth retrieval.

We revised it to “Additionally, snow compaction and metamorphism can cause variations in snow grain size, density, shape, and bonds, which in turn affect microwave scattering signals. Thus, the seasonally and spatially dynamic characteristics of snow limit the use of standalone data-driven ML for snow depth/SWE retrieval.”

We have reorganized the paper’s structure and rewritten the introduction so that it more clearly highlights the importance and novelty of our work. Please refer to the response to the Editor and two anonymous Reviewers above.

Line 155: They state a review on ML applications for snow depth is missing, but don’t give a justification for why it’s needed. Please add one.

We revised it to “These studies consistently indicated that coupling RTMs or process models with ML can yield more accurate snow depth/SWE data than can either physical snow models or data-driven ML alone.”

Overview

Line 165: they included snow water equivalent in the search term, but then hardly bring that up at all in the synthesis. Is there a reason why?

The snow depth and SWE (the product of snow depth and snow density) are two key parameters used to characterise snow mass. To avoid ambiguity or confusion for readers, we have revised the title and modified the corresponding parts of the paper accordingly.

Line 167: “…with remote sensing techniques were included in the review, we then manually filtered the search results.” How did you define remote sensing techniques? Just ones from satellites? What about airborne sensors? Again in their title they say their review focuses on just satellites, but it’s unclear in their review if it includes all sensors that would qualify as remote sensing or just spaceborne satellites.

We thank the reviewer for this reminder.

We compiled a database, and all our literature statistics and reviews were based on it. If necessary, we would be happy to share this database. After reexamination of the database, we reported that several studies performed ML-based snow depth/SWE retrieval using airborne or ground-based remote sensing observations. Accordingly, we have revised the paper in the following aspects:

(1) We replaced ‘satellite’ with ‘remote sensing’ in the title and revised the relevant sections of

the manuscript accordingly.
(2) We have clarified the scope of remote sensing (Page 6, Lines 169-172 in the revised manuscript): "Notably, the concept of remote sensing is relatively broad, encompassing ground-based, airborne, and spaceborne sensors; studies using related remote sensing datasets were also included."

2.1 Timeline of ML for snow depth retrieval from 2015–2025
Line 180: "Their studies revealed that the ML-simulated brightness temperature is sensitive to snow depth and could potentially serve as a more efficient and effective model operator within the data assimilation frameworks (Xue et al., 2018; Kwon et al., 2019)". By how much? How sensitive was it?
Revised.
"To summarize the contributions that ML has made to snow depth/SWE monitoring and the paradigms that have emerged, we reviewed the key milestones during the past decade. On the basis of the quantitative statistics (Fig. 1) and a literature review, we divided the ML-based snow depth/SWE retrieval into three stages: ML as an observation operator in assimilation systems, standalone data-driven ML models, and data-model-driven ML models (Fig. 2). Pioneering studies in each stage are regarded as representative work along the milestone timeline. It should be noted, however, that key milestones may be identified differently depending on individual perspectives or disciplinary backgrounds. Initially, ML approaches (e.g., ANN and SVM) were used as alternatives to RTM for simulating microwave brightness temperature, with an RMSE of less than 10 K at all frequencies compared with satellite (e.g., SSM/I, AMSR-E) observations (Xue et al., 2015; Forman et al., 2015, 2017). Subsequently, the AMSR-E brightness temperatures were successfully assimilated into the catchment model using an SVM as the observation operator, and the snow depth retrieval accuracy improved by 31% (Xue et al., 2018; Kwon et al., 2019)."

Line 182: "Subsequently, more communities have focused on developing snow depth retrieval algorithms based on ML techniques, by combining the PM remote sensing data and other auxiliary data (Fig.2), e.g., the SVR (Xiao et al., 185 2018) and the RF (Bair et al., 2018)." Was this successful? Why did they combine PM with auxiliary data?
Revised.
"More recently, given the complexity of assimilation systems, numerous studies have directly developed ML-based snow depth/SWE retrieval algorithms using microwave remote sensing data, such as SVR (Xiao et al., 2018) and RF (Bair et al., 2018)."

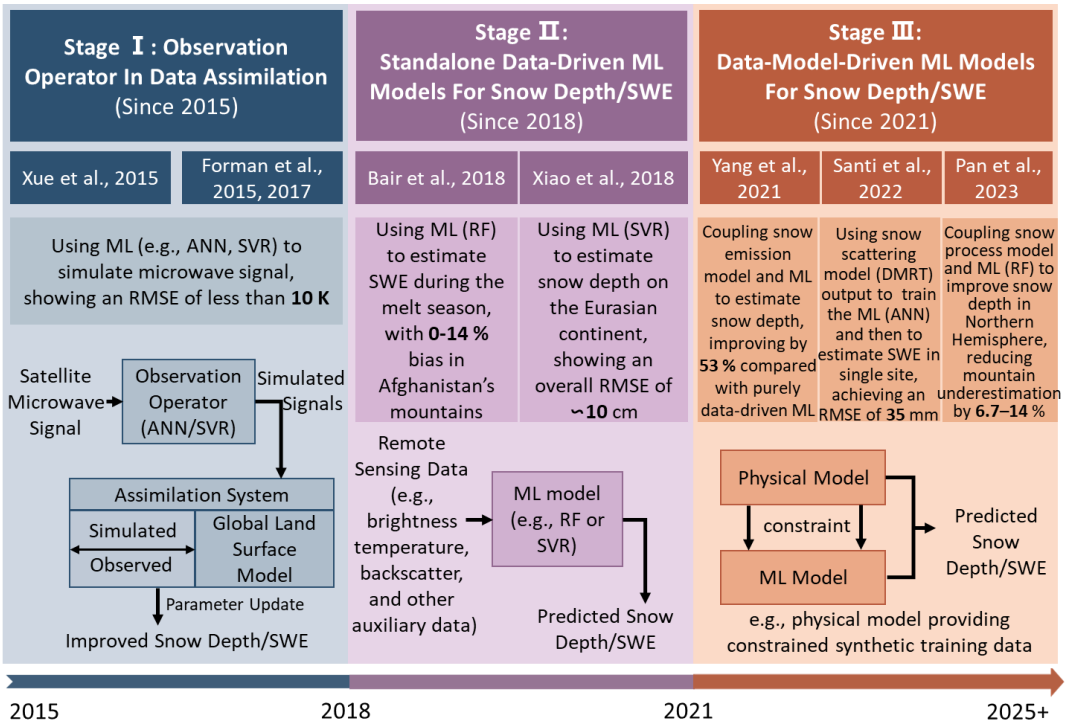
Line 193: "was applied to simulate snow grain size and snow density, and to first assess snow depth as prior information for the RF model; the authors reported that the proposed method significantly reduced the snow depth underestimation in mountain areas (Fig.2)." How much did it reduce?
Revised.
"The results revealed that the proposed method significantly reduced snow depth underestimation by 6.7%–14% in mountainous areas."

I like the concept behind Figure 2, however when the authors discuss it, I would like them to

mention how much that improves our ability to measure snow depth. Are we doing better? Or are these milestones simply following the hype of the field?

We thank the reviewer for this reminder. We added an explanation to Fig. 2: "To summarize the contributions that ML has made to snow depth/SWE monitoring and the paradigms that have emerged, we reviewed the key milestones during the past decade. On the basis of the quantitative statistics (Fig. 1) and a literature review, we divided the ML-based snow depth/SWE retrieval into three stages: ML as an observation operator in assimilation systems, standalone data-driven ML models, and data-model-driven ML models (Fig. 2). The pioneering study in each stage is regarded as the representative work along the milestone timeline."

To demonstrate the representativeness of these studies, we redrew the figures and added some quantitative metrics.



2.2 Main countries and journals of published articles from 2015–2025

Line 198: I am afraid I am lost as to why a geographic distribution of paper authors is helpful. Is that where the studies took place? Is this saying something about data availability? The connection for the research gap and how this fills a research gap is needed here.

In our view, a basic bibliometric analysis is necessary. On the one hand, it supports the subsequent paradigm synthesis and corresponding strategies; on the other hand, it demonstrates the reliability and authenticity of our reference database. To improve the logical flow and readability, we revised it to "To better understand country contributions, a summary of the geographic distribution (according to first author affiliation) of published articles is shown in Fig. 3a. China and the United States are the dominant contributors, together accounting for almost 70% of the papers, followed by Iran, Canada, Italy and Germany, Morocco, India, Switzerland, Norway, and Belgium. To highlight the contributions of the journals and to help readers understand the topics of interest for each journal, the publication distribution from 2015 to 2025 is shown in Fig. 3b."

Line 204: How does knowing what journals publish what help us understand the evolution of the field generally? This is circular, as the papers published in certain journals will depend on the journal focus.

Revised. Please refer to the response above.

2.3 Spatial resolution and source of target predictors in snow depth retrieval

- Line 230-265: This is summary information, and in order to assess the progress and evolution of the field, it is necessary for the authors to dig more into these papers and examine how well each of these papers did for snow. Does the number of publications reflect the utility and where the future direction should go? Given that the authors claim that this is one of the primary goals for this paper, I believe it is necessary for this whole section to be expanded upon and for the authors to compare accuracy metrics from each of these papers.

We fully agree with the reviewer that a quantitative comparison would be valuable. However, such an analysis would require substantial additional work—including data collection, methodology reproduction, normalization, and benchmarking—that goes well beyond the scope of a review article and would be more appropriate for a dedicated research paper.

On the basis of these summaries of spatial resolution and target predictor sources, we identify the existing challenges and future directions, which are discussed in Sections 4.3 and 4.5, respectively.

2.4 Top 10 most cited articles in the past 10 years

- Line 293 – 297: Where was the citation information pulled from? Web of science? This needs to be defined in the methods with the search term. The analyses presented, as it stands now, appear to be somewhat ad hoc, without much justification for why they looked at certain criteria. The authors need to outline this at the beginning and how this relates to their overall goal of understanding progression and future direction.

Deleted.

- Line 297-300: "To improve SWE estimates in a semiarid interior region, Broxton et al. (2019) used an ANN to estimate snow density by combining snow surveys and Lidar measurements. Yang et al. (2020) first explored the potential of a RF approach in snow depth retrieval and demonstrated that its spatial transferability is poor due to the dynamic snow characteristics." This is the first time the authors bring up SWE in their literature (besides mentioning it in the search term). It's unclear how this relates to snow depth, and why it is included.

Additionally, please refer to the response to "Major Comments 2" and "Introduction Comments" above.

- Figure 6. How many citations were these top 10 papers? What do the authors think about these papers besides the fact that they were cited the most? It's difficult to know how to interpret this information and what will be useful for the snow community based off this information.

We deleted this subsection; please refer to the response above.

- Line 313: "ML contributes to accurate snow depth retrieval with remote sensing techniques. Here, we summarize five main directions for improving snow depth estimation." The authors say

they will talk about five main directions but it is unclear what those directions are. The five main directions need to be stated at the beginning.

This has been added (Page 12, Lines 313-318 in the revised manuscript).

"ML contributes to accurate snow depth/SWE retrieval with remote sensing techniques. Here, we summarize five main paradigms for improving snow depth/SWE estimation, including the coupling of ML and snow physical models (snow RTM or process model), substituting RTM with ML for predicting microwave signals in assimilation or iteration algorithms, optimizing RTM by providing key inputs or improving operational efficiency, improving existing gridded snow depth/SWE products by data fusion, bias correction or assembly techniques, and downscaling coarse snow depth/SWE products to a fine-scale resolution."

3.1 Coupling ML technique with snow physical model

- line 316-319: typo for examples. Also what are the physics-based snow models? How much was the error? This is all quite superficial information, and for this literature review it would be useful for the authors to include the current baseline of these methods.

Section 3.1 provides two examples, namely, ML combined with a radiative transfer model (HUT) and with a process-based model (SNTHERM), which correspond to Sections 3.1.1 and 3.1.2, respectively. In these two subsections, the HUT and SNOTHERM models are explained in detail.

We revised the manuscript as follows:

"Previous studies have shown that combining ML and snow physical models (data-model-driven ML) can retrieve snow depth/SWE with lower errors than physics-based or ML models alone can. Here, we use a snow RTM and a snow process model as case studies, which are coupled with ML models to improve snow depth/SWE retrievals." (Page 12, Lines 320-322).

"Pan et al. (2023) proposed a method that uses simulated snow parameters from a snow process model (SNTHERM) to construct an ML retrieval algorithm, achieving an RMSE of 12.4 cm at the global scale (Fig. 8). In this methodology, SNTHERM, driven by the GLDAS meteorological forcing datasets, was run to generate priors for snow depth, profile-averaged snow grain size, and snow density. In particular, combining ML and SNTHERM significantly reduces snow depth underestimation by 6.7%–14% in mountainous areas. These findings demonstrated that priors from a snow process model introduce another group of independent information into the ML algorithm and can greatly improve the accuracy of PM snow depth/SWE products." (Page 13, Lines 341-345).

3.2 Applying RTM for simulating microwave signals

- This is another place where the authors introduce an analysis that is separate from the literature (generating a database) without justification. Please bring this up in the beginning and explain how this fits with your overall view.

Added: "Accurately and efficiently simulating microwave signals (e.g., brightness temperature and backscatter) with an RTM can improve our ability to retrieve snow depth/SWE from spaceborne observations. However, RTMs typically require extensive prior information and are computationally inefficient, which restricts the development of snow depth/SWE retrieval algorithms. Moreover, while previous studies have proven that ML can be used as an observation operator (Xue et al., 2015; Forman et al., 2015, 2017), it has not been validated against ground-based observations."

3.3 Optimizing snow RTM model

3.3.1 Providing key parameters for snow electromagnetic model

- Line 363: I like this section, but I am afraid you've lost this reader for how this relates to the initial research gap that was presented. This paper feels like it is discussing more the evolution of RTM models/ other types of microwave models for snow depth, using ML, and I believe this needs to be better stated in the beginning.

We thank the reviewer for the positive recognition. To improve the logical flow and readability, we added the following content: "Snow volume scattering is typically affected by snow depth, snow grain size, snow density, and layering structure. Thus, snow priors are necessary for driving snow RTM. Previous studies have typically relied on meteorological forcing datasets to drive snow process models (Pan et al., 2023) or have used satellite microwave observations (e.g., brightness temperature) to constrain RTM simulations (Yang et al., 2021) to obtain snow priors. However, the complexity of these methods limits the operational efficiency of the snow depth/SWE retrieval algorithm."

4.1 Bridging the disparity of model–observation mismatch

Line 482: "has been has been" please fix this typo

Deleted.

Line 495: "Previous studies have demonstrated that a multi-layer RTM is superior to a one-layer model (Zhu and Tan et al., 2018; King et al., 2018, 2020; Rutter et al., 2019; Pan et al., 2024)." By how much?

Added: "Previous studies have shown that multilayer RTMs outperform single-layer models in snow depth/SWE retrieval (Zhu and Tan et al., 2018; King et al., 2018, 2020; Rutter et al., 2019). For example, Pan et al. (2024) reported that a dual-layer model reduced the bias from −13.4 to 2.4 cm and the RMSE from 15.9 to 10.2 cm, significantly reducing the underestimation and uncertainty of snow depth."

4.3 Developing high-spatial-resolution snow depth products

- Line 600: Here and throughout, I found that the Figures were dropped in without much description, and I found myself having to jump back and forth between the text and the figure to determine what the authors were trying to say. Can the authors be sure to incorporate what the main takeaway points are from each figure into the text? This will help with readability.

We appreciate the reviewer's constructive suggestions. We have revised Sections 4.4 and 4.5.

Sections 4.4:

"Our decadal review revealed that although existing studies have attempted to retrieve snow depth/SWE in mountainous regions (see Section 2.3), most rely on airborne or ground-based observations and are therefore limited to small areas with discontinuous temporal coverage. We also reported that few studies have addressed snow depth/SWE over polar sea ice or during the snowmelt season. Therefore, we address the challenges associated with snow depth/SWE monitoring in complex scenarios, including alpine regions, polar regions, and snowmelt seasons (Fig. 19). Our main purpose is to call for more research teams to pay attention to snow cover distribution in complex scenarios, thereby achieving spatiotemporally continuous snowpack

monitoring.

The seasonal snowpack in the mountains acts as a vital natural reservoir that regulates runoff downstream in snowfed basins. Estimating snow depth/SWE in mountainous terrain is currently the most important unsolved problem (Dizoer et al., 2016; Fayad et al., 2017). These challenges include sparse in situ measurements, high spatiotemporal heterogeneity of snowpack due to inclement weather and rugged topographical conditions, and complex snow layering structures (Fig. 19). Snow cover in alpine regions is typically thick, likely causing signal saturation for conventional X- and Ku-band SAR (Tsang et al., 2022). C-band SAR, in contrast, has a strong penetration capability and is therefore more suitable for retrieving deep snow (Lievens et al., 2019, 2022). Therefore, combining C-, X-, and Ku-band SAR observations is a primary approach for snow depth/SWE monitoring in alpine regions. For example, NASA is developing a three-frequency (two in the Ku-band and one in the C-band) airborne SAR system, the Snow Water-equivalent Wide Swath Interferometer (SNOWWI) (Siqueira et al., 2024). Airborne testing has been conducted in Grand Mesa, Colorado, and a spaceborne constellation concept led by Capella Space is developing a space-ready SNOWWI for satellite missions (Tarres et al., 2025).

Snow depth/SWE monitoring over polar regions (e.g., Arctic and Antarctic sea ice) is another unsolved research task. The greatest limiting factor is the lack of in situ snow depth/SWE measurements since 2022 (Fig. 19). In particular, some observation missions have ceased the collection and updating of datasets, e.g., airborne campaigns such as the Operation IceBridge (OIB) and IceBird programs and those based on snow buoys that provide ice mass balance (IMB) data. Current spaceborne altimetry sensors (e.g., ICESat/ICESat-2, CryoSat/CryoSat-2, and SARAL) offer a notable advantage for supplementing polar snow depth samples, primarily owing to their high coverage in both polar regions. Another limiting factor is that the underlying surface of polar snow, sea ice, is highly dynamic and variable in its properties, making accurate RTM simulations difficult. Brine and air inclusions, for instance, dominate volume scattering (Nandan et al., 2016; Soriot et al., 2022), whereas snow–ice interface roughness enhances surface scattering (Liu et al., 2025; Wang et al., 2025). We particularly call for enhanced international cooperation to sustain polar sea ice and snow observations and ensure the availability of abundant first-hand data, which is essential for supporting the latest satellite missions and improving polar snow depth/SWE monitoring.

Snowmelt is a major component of streamflow in cold regions and serves as an important supply of freshwater resources. However, we have still not resolved the challenge of snow depth/SWE monitoring during the snowmelt season with remote sensing techniques. Theoretically, monitoring snow depth/SWE under wet snow conditions with microwave remote sensing is difficult because of the high liquid water content of the snowpack. Reconstruction via reverse calculation of the daily snowmelt depth is currently the mainstream approach (Premier et al., 2023). However, the reconstructed snow depth/SWE data are available only retrospectively and do not support forecasts (Bair et al., 2018). During the snowmelt season, accurately determining four snow parameters (maximum snow depth, snowmelt onset date, snow disappearance date and snowmelt rate) is critical for monitoring and understanding the snowpack depletion process. Microwave remote sensing is effective at determining the maximum snow depth/SWE during the snow season. Moreover, the microwave backscattering signal (e.g., in the C-band) is sensitive to wet snow and can be used to determine the snowmelt onset date. Additionally, combining optical and microwave remote sensing methods can reduce the effects of cloud cover, making calculating

the snow disappearance date possible. A snow process model or hydrological model can be used to simulate the snowpack depletion process, and an ML method can be used to spatiotemporally predict the snowmelt rate at a large scale. Thus, combining snow physical models (RTM and snow process models), remote sensing observations and ML techniques is expected to improve snow depth/SWE retrieval across a complete hydrological year in the future."

Sections 4.5:

"The performance of ML models heavily relies on the quality of the training samples. Our decadal review revealed that most existing studies rely on station observations as target predictors to train the ML model (see Section 2.3). Most current efforts focus on bias correction—e.g., linear fitting (Venäläinen et al., 2025) and data fusion (Qiao et al., 2025)—to improve training sample accuracy but often neglect the representativeness and balance of samples (Fig. 20).

To assess the spatial representativeness of the point-scale measurements, some metrics, such as clustering indicators (Galeati et al., 1986), the mean relative difference based on dense sampling data (Herbert et al., 2024, 2025) or data from neighbouring sites (Meromy et al., 2013; Chen et al., 2017), a snow cover day-based score (Pan et al., 2023), and the triple collocation-derived error (Wang and Zheng, 2020; Yang et al., 2025) or correlation coefficient (Peng et al., 2025), were used to characterise the spatial heterogeneity. Recently, the ML technique has been used to study the representativeness of field sampling locations. For example, Liljestrand et al. (2025) applied an unsupervised Gaussian mixture model to identify sampling locations within the Franklin Basin on the basis of topographic features and reported that information from 10 sampling sites can be used to determine snow depth across the basin with high accuracy. Xie et al. (2024) proposed a physics-constrained ML scheme to identify optimal sensor placement in a field, thereby enhancing soil moisture prediction for the entire study area.

In the field of ML, it is typically assumed that the number of samples in each class (e.g., shallow and thick snowpack) is approximately equal. However, the lack of measurements in mountains and for deep snow conditions typically results in an imbalance of training samples. Some scholars have attempted to increase deep snow samples using airborne (Daudt et al., 2023; Dunmire et al., 2024; Feng et al., 2025) and spaceborne (Xiong et al., 2022b; Yang et al., 2024) measurements. Moreover, a simulation database from a snow RTM and a process model can contribute to balancing training samples (Santi et al., 2022a, 2022b). Additionally, data resampling (Carvalho et al., 2025) and category weight adjustment (Zhang and Yu et al., 2022) are typically used to solve sample imbalance issues. Overall, we call on the snow research community to devote more attention to the quality of ML training samples, especially to integrating station measurements, airborne and satellite observations, and model simulations to improve sample representativeness and balance (Fig. 20)."