Università di Genova

PhD PROGRAM
SECURITY
RISK
VULNERABILITY

PhD Curriculum in Risk, Climate Change and Sustainable Development

Department of Informatics, Bioengineering, Robotics and Systems Engineering

University of Genova, Via Opera Pia, 13, 16145 Genova, Italy

# Unlocking the potential of artificial intelligence in hydrology

A Deep learning framework for snow data assimilation in S3M

15 April 2026

**Tutor:**
**Prof. Dr. Luca Ferraris**

**Supervisors:**
**Dr. Francesco Avanzi**
Dr. Simone Gabellani

**External Reviewers:**
**Prof. Gabrielle DeLannoy**
**Dr. Christian Massari**

**PhD Candidate:**
**Giulia Blandini, XXXVIII Cycle**

## Acknowledgements

I would like to thank the reviewers, Dr. Christian Massari and Professor Gabrielle DeLannoy for their time and insightful comments, which have significantly improved the quality of this manuscript.

I have never been good at acknowledging my achievements and my successes, but today it truly feels like one. Three years of research, conferences, schools, people, papers, good and bad experiences have all been part of an amazing and exciting journey that I know has just begun.

I would have never made it without all the people around me, and that is why my deepest thanks go to them.

Questi tre anni di dottorato non sarebbero stati gli stessi senza la guida fondamentale del mio supervisore Francesco Avanzi. Le parole non sono sufficienti a spiegare la stima che provo nei suoi confronti, come ricercatore e, prima di tutto, come persona. Grazie per avermi insegnato e guidato. Grazie per aver creduto in me e nella mia ricerca anche quando io non riuscivo ad andare oltre le difficoltà. Grazie per essere stato un insegnante appassionato e un fratello maggiore affettuoso.

Grazie a Lorenzo Campo per avermi insegnato tutto quello che so sulla data assimilation. Grazie per i nostri giovedì di studio, per i consigli e per la cura. Grazie per avermi sempre chiesto come stessi prima di "spostarci in uno spazio gaussiano".

Grazie a Simone Gabellani per aver creduto in me e per essere sempre stato dalla mia parte, scientificamente e umanamente.

Grazie ai colleghi e alle colleghe del CIMA, ai dottorandi e alle dottorande. Quello che facciamo forse non salverà vite, ma la mia ogni giorno è più bella perché la condivido con voi.

Thanks to Kristoffer Aalstad, for the guidance throughout my last two years of work. Thank you for being an amazing host supervisor, for the precious lessons, and for always having such kind and encouraging words about my research. You and all the people at GeoHyd have made my Norwegian experience unforgettable, and I will always cherish those memories. To Marco, Julie, and Hannah, to Mikko, Adele and Mina.

Ai miei amici di Torino. Anche se in città diverse, siete sempre la mia famiglia per scelta.

Alla mia famiglia, per avermi supportato dal primo giorno fuori casa, lontano dalla mia Sicilia. Grazie per aver creduto in me, per essere la mia ancora e il mio porto sicuro. Grazie mamma per tutti i biglietti nascosti nelle valigie ogni volta che vado via. Grazie papà perché mi hai insegnato ad essere forte e ad accettare le fragilità. Grazie Claus per essere la mia migliore amica, per non lasciarmi mai. Grazie perché mi avete insegnato ad amare, ad amarmi e a credere in me stessa.

Grazie Lollo, per scegliermi ogni giorno, per credere in me e per tutto il supporto durante il mio percorso. Per quello che non si vede ma si sente. Per l'amore e la nostra vita insieme.

# Unlocking the potential of artificial intelligence in hydrology

A Deep learning framework for snow data assimilation in S3M

## Abstract

Time and computational constraints often limit the retrieval of reliable snow model estimates, particularly in large-domain operational contexts. Artificial intelligence–based modelling approaches can improve the representation of snow storage and dynamics, while reducing uncertainties in snow-dominated regions and enhancing assessments of water availability, ecosystem functioning, and climate feedbacks, with significant societal and environmental implications. Despite this potential, the effective integration of artificial intelligence–based methods into operational snow hydrological modelling frameworks remains an open research challenge.

In this context, the main hypothesis of this doctoral research is that artificial intelligence–based techniques can be effectively embedded within snow hydrological models to enhance predictive accuracy while reducing computational effort and execution time. This integration leverages the potential of artificial intelligence to address key challenges in operational snow modelling, namely: (i) quality control and quality assurance of in situ snow depth observations, and (ii) the assimilation of snow-related variables.

The proposed Random forest–based quality-control algorithm provides a fully automatic, computationally efficient, and scalable approach for quality assurance of snow depth observations. By incorporating expert domain knowledge, it offers an effective alternative to manual screening and substantially reduces the time and resources required for data validation. It achieved F1 scores above 90% for snow versus grass or bare-ground detection, even outside the training domain. Despite less consistent error classification performance, the algorithm's limited sensitivity to variations in snow-season climatology highlights its applicability across heterogeneous environmental conditions.

The Ensemble Kalman filter emulator based on a Long Short-Term Memory network achieved snow depth and snow water equivalent estimates comparable to the traditional ensemble approach while reducing computational time by up to 70% by using ensemble simulations only during training. The framework demonstrated strong spatial transferability, with only a 20% decrease in performance outside the training domain, and highlights the potential for a fast, scalable, and spatially distributed deep data assimilation framework, with future developments envisaging the propagation of pointwise corrections across the domain using Gaussian Process interpolation.

In my view, while the computational efficiency of artificial intelligence-based emulators offers clear advantages for cryosphere and hydrosphere modelling, it should not come at the expense of process transparency and interpretability. I believe that ensuring explainability is essential in operational and decision-support contexts, where trust, accountability, and physical consistency are critical, and that scientists have a key role in linking efficient artificial intelligence outputs with interpretable, physically meaningful results.

# Contents

# List of Acronyms

**AI** Artificial Intelligence. 1–3, 12, 28, 30, 32, 53, 73, 86–88, 90, 98, 103

**ANN** Artificial Neural Network. 100, 103

**DA** Data Assimilation. x, xi, 23–27, 30, 32, 53, 54, 56–58, 60, 62, 64, 66, 67, 69–73, 81–85, 88–90, 92–96, 104

**DL** Deep Learning. 3, 29, 52, 54, 57, 61, 70–72, 82, 88, 89, 100, 103, 104

**EnKF** Ensemble Kalman Filter. 53, 54, 57, 58, 61, 64, 66–68, 70–73, 82, 85, 88–90, 92–94

**FSC** Fractional Snow Cover. 18

**ISPRA** Istituto Superiore per la Protezione e la Ricerca Ambientale. 41

**KGE** Kling–Gupta Efficiency. x, xi, 66–68, 75–78

**LSTM** Long Short-Term Memory. x–xiv, 52, 54, 60–73, 75–79, 82, 85, 88–90, 103, 109, 110

**LWC** Liquid Water Content. 17, 62

**MAE** Mean Absolute Error. 102

**ML** Machine Learning. 3, 20, 28, 29, 33, 34, 36, 37, 39, 71, 98, 100

**MSE** Mean Squared Error. 102

**PF** Particle Filter. 96

**RMSE** Root Mean Square Error. x–xii, xiv, 21, 40, 61–63, 66–69, 72, 75–79, 81, 109

**RNN** Recurrent Neural Network. 60

**SC** Snow Cover. 94

**SCA** Snow Cover Area. 18, 26

**SWE** Snow Water Equivalent. x–xiv, 6–8, 17–19, 26, 36, 54, 56–58, 60, 62–64, 66–70, 73, 75–81, 85, 88, 89, 94–96, 99, 104, 108–111

# List of Figures

# List of Tables

# Chapter 1

# Motivation

Artificial Intelligence (AI) has emerged as a key methodological advancement across multiple disciplines, including the environmental and water sciences. Although in hydrology the development and adoption of AI-based solutions faced some resistance largely due to the complexity of the hydrological system (Shen, 2018), nowadays the interest is growing. Recent progress has demonstrated substantial improvements in predictive capabilities enabled by Artificial Intelligence (AI) models, particularly by Deep learning algorithms (DL), marking a paradigm shift in how hydrological processes are analysed, simulated, and forecasted (Abdulameer *et al.*, 2025).

AI presents significant opportunities to address several persistent challenges in hydrology, including:

- **Interdisciplinary and human–system complexity:** the interaction between humans and water resources is difficult and time-consuming to represent accurately in models (Shen, 2018).
- **Data deluge and data discoverability:** the increasing availability and heterogeneity of environmental datasets pose challenges for efficient data integration and interpretation.
- **Unrecognized or under-represented linkages:** missing or poorly represented processes can introduce systematic model errors.
- **Model scaling issues and equifinality:** hydrological predictions often vary with scale, and multiple parameter combinations may yield similar results (Beven & Freer, 2001).
- **Regionalization and computational feasibility:** the need for spatially explicit models and parameters is constrained by computational costs, particularly in the context of rapid, large-scale assessments required under changing climatic conditions.

AI-based approaches provide a promising pathway to address many of the limitations of traditional hydrological modelling (Yaseen, 2023). They offer computationally efficient, data-driven alternatives that can complement or enhance physically based models.

However, domain expertise remains essential to the successful application of these methods: formulating meaningful scientific questions, selecting appropriate inputs, designing suitable model architectures, and interpreting outputs all require a strong grounding in hydrological knowledge (Shen, 2018). Recent studies further demonstrate that explicitly integrating expert knowledge within AI frameworks can substantially improve model performance, robustness, and interpretability (Karpatne *et al.*, 2017).

In this context, it is worth nothing that a large portion of the planet's freshwater (roughly 70%) is stored as snow and ice within the cryosphere. The cryosphere includes all frozen components of the Earth system found on or beneath the land and ocean surfaces, such as snow cover, glaciers, ice sheets and shelves, icebergs, sea ice, lake and river ice, permafrost, seasonally frozen ground, and all forms of solid precipitation (Beniston *et al.*, 2018). Despite decades of scientific progress, accurately representing snow processes remains difficult due to complex multiscale interactions, non-linear dynamics, and strong spatial–temporal variability (Bormann *et al.*, 2013). In light of this, physically based models face additional challenges related to computational cost, parameter uncertainty, and the assimilation of heterogeneous observational data, limitations that become more pronounced when moving to larger domains or attempting higher-resolution simulations (Pagano *et al.*, 2014).

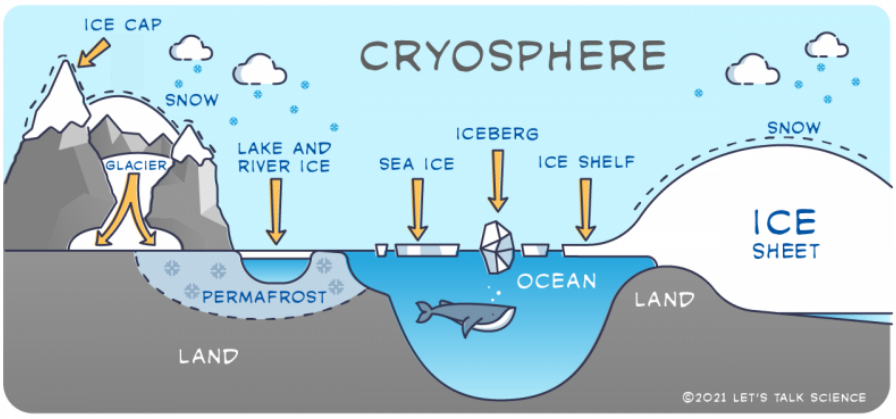


**Figure 1.1:** *Infographic of the components of the cryosphere. Source: (talk science, 2021). Note: The share of seasonal snow shown here may appear disproportionately small and it is intended conceptually, not quantitatively.*

Given the potential of AI to alleviate both computational and epistemic constraints, and considering the pronounced impacts of climate change on the cryosphere, particularly

on snow processes in Mediterranean regions (Matiu *et al.*, 2021; Bozzoli *et al.*, 2024), this research focuses in particular on snow hydrology. Snow is a key component of the climate system and global freshwater resources, yet snow-dominated regions are especially sensitive to uncertainties in estimating snow storage and dynamics (Singh *et al.*, 2024). Such uncertainties can propagate throughout the hydrological cycle, affecting water availability, ecosystem functioning, and climate feedbacks, with significant societal and environmental implications (Dunmire *et al.*, 2024).

Building on these considerations, **the central hypothesis of this thesis is that AI-based methods can be effectively integrated into hydrological, and specifically snow hydrological, modelling frameworks to improve predictive performance while reducing computational costs and processing time**. Moreover, such integration may uncover previously unrecognized non-linear relationships that extend beyond current computational and methodological constraints, offering opportunities to gain deeper epistemic insights into snow processes (Nearing *et al.*, 2021). In this way, improvements in scientific understanding can progress independently of, and potentially ahead of, existing technical limitations.

The remainder of this document is structured as follows:

- **Chapter 2 – Background:** presents the scientific background and motivation of the study, including an overview of snow hydrology, the role of data assimilation in snow modelling, and the emerging potential of AI-enhanced approaches. It ends with a section on my research questions.
- **Chapter 3 – Machine learning-based algorithm for QA/QC:** describes the application of Machine Learning (ML) techniques, specifically Random forest, for quality assurance and quality control of snow and meteorological datasets.
- **Chapter 4 – Deep learning-based data assimilation in 1D:** introduces a one-dimensional framework integrating Deep Learning (DL) algorithms to emulate and enhance traditional data assimilation schemes.
- **Chapter 4 – Deep learning-based data assimilation in 2D:** describes a methodology that will be implemented in future steps of this work to extend the previous approach to a spatially distributed (2D) domain, proposing an innovative way to achieve scalability and enhance the applicability of the deep data assimilation framework.
- **General Discussion and Conclusion:** Discusses the main findings of the research, providing a comparative analysis of the proposed approaches and discussing their implications for AI–hydrology integration. Then it summarizes the key contributions of the thesis and outlines future research directions for the integration of AI and data assimilation within hydrological modelling.

# Chapter 2

# Background

## 2.1 Snow hydrology

According to World Meteorological Organization (Canton, 2021), about 70 % of Earth' fresh water exists in solid form and it is part of the Cryosphere. Around 10% of Earth's land area is covered by glaciers or ice sheets (see Figure 2.1). The cryosphere is tightly linked to the climate system, primarily through global exchanges of water, energy, and carbon (Krissek & St. John, 2025). Indeed, cryospheric components play a crucial role in regulating the Earth's climate by modulating surface albedo, latent heat exchange, and freshwater inputs. Among cryospheric components, snow exerts a particularly strong influence on both the surface energy balance (Zhang, 2005) and freshwater availability (Viviroli *et al.*, 2007).

Because snow has a much higher albedo than bare ground, it produces a strong cooling effect by reducing the absorption of incoming shortwave solar radiation. Surface reflectance of incoming solar radiation, quantified as the ratio of reflected to incident radiation and commonly referred to as albedo, is therefore a key factor in the surface energy balance (Thackeray & Fletcher, 2016). Shortwave electromagnetic radiation corresponds to the main input of solar energy (Rohli & Li, 2021). The high value of snow and ice albedo (about 80–90%), represent an important first-order feedback between the cryosphere and the atmosphere. The enhanced reflectivity of snow-covered surfaces reduces net radiation at the surface, weakening warming effects (Chen *et al.*, 2025). This cooling can increase the likelihood of widespread snow cover, creating a positive feedback loop. Conversely, warming of the Earth's atmosphere can reduce snow and ice cover (Thackeray *et al.*, 2019), lowering the planet's albedo and accelerating melting. However, a warmer atmosphere can also hold more moisture, which in cold regions may lead to increased snowfall, partially offsetting snow loss under certain conditions. These interactions among the cryosphere, atmosphere, and hydrosphere, along with many others, are illustrated in Fig. 2.2.

The seasonal presence of snow governs surface–atmosphere energy exchanges from local

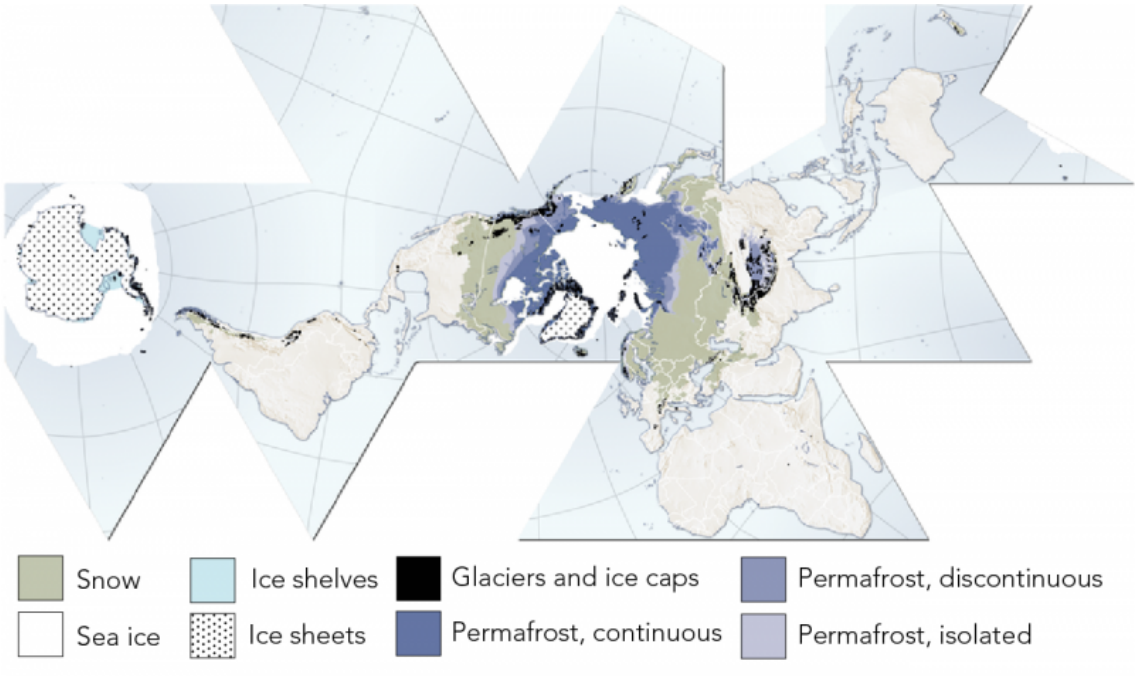


**Figure 2.1:** *Global distribution of the cryosphere. Source: talk science (2021). The sand-coloured areas indicate non-cryospheric regions that are not represented in the legend.*

to regional scales. Owing to its low thermal conductivity, snow insulates the ground surface, dampening soil temperature variability and stabilizing subsurface thermal regimes (Zhang, 2005). The insulating effect of snow also has major implications for the hydrological cycle. In non-permafrost regions,where the ground is only seasonally frozen, the insulating effect of the snowpack limits freezing to the shallow subsurface, allowing deep water drainage to remain uninterrupted and preserved (van Huissteden, 2020).

Lastly, yet of equal significance, the dynamics of snow and snowpack play a critical role in shaping seasonal water resources, exerting a strong influence not only on the timing and magnitude of streamflow, but also on the availability and quality of both surface and groundwater (Dettinger, 2014). In regions characterized by Mediterranean, summer-dry climates, the accumulation of snow during the winter months, followed by its gradual melt in spring and early summer, provides a vital water supply that sustains human societies, agricultural activities, hydropower productions and natural ecosystems precisely during periods of highest water demand, which often coincide with a seasonal decline in precipitation (Avanzi *et al.*, 2023). Similarly, in colder regions (Bravo *et al.*, 2008), the snowpack represents the dominant contributor to streamflow, particularly during the spring and summer months when meltwater release drives river discharge (Bales *et al.*, 2006). In catchments dominated by snow, the long-standing adage that today's snow is tomorrow's

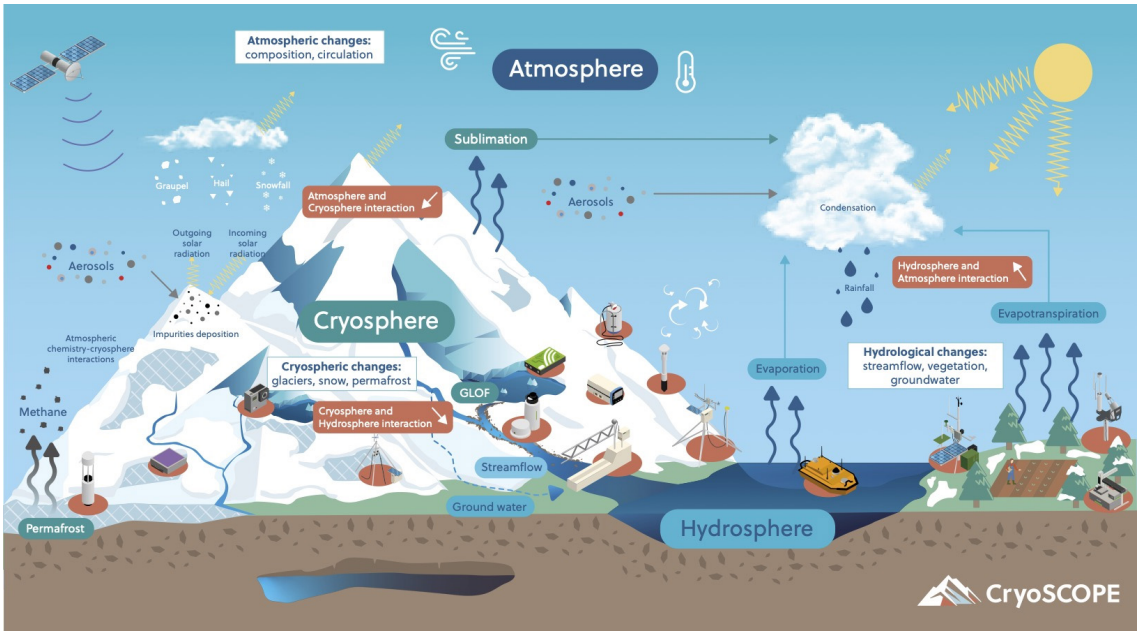


**Figure 2.2:** *Schematic representation of interactions between the atmosphere, cryosphere, and hydrosphere. Source: CryoSCOPE (2026)*

water captures a fundamental principle of hydrology (Avanzi *et al.*, 2025). This concept can be traced back to the pioneering work of James E. Church (Sturm, 2015), widely recognized as the founder of modern snow hydrology, who was among the first scientists in the Western world to rigorously observe, quantify, and emphasize the crucial link between snow accumulation in the catchment and downstream water availability. His work laid the foundation for understanding how seasonal snowpack dynamics underpin freshwater resources and continue to inform water management and climate impact studies today.

From a hydrological perspective, measurements of snow depth and Snow Water Equivalent (SWE ), the amount of water stored in a snowpack (Taheri & Mohammadian, 2022), provide critical information for water management, hydrological forecasting, and emergency preparedness (Hartman *et al.*, 1995). Indeed, these measurements are crucial for real-time applications such as irrigation, agriculture, water supply management, flood prevention, hydropower generation, and overall water resource planning (Hartman *et al.*, 1995).

In recent decades, the already complex hydrological cycle has been increasingly disrupted by the intensification of extreme events such as heavy precipitation, flooding, drought, and wildfires, which are direct consequences of climate change (Ficklin *et al.*, 2022). Altered streamflow regimes (Change *et al.*, 2014), shifts in precipitation patterns (Wang *et al.*, 2022b), and the coexistence of more frequent and intense floods with declining baseflows provide clear evidence of this transformation (Juras *et al.*, 2021).

Recent studies show that reductions in streamflow are often directly linked to declines in snowfall, exacerbated by rising air temperatures (Berghuijs *et al.*, 2014). In particular, snow droughts, periods of below-average snow accumulation, can severely impact the hydrological cycle (Harpold *et al.*, 2017), frequently triggering hydrological droughts downstream (Toreti *et al.*, 2022a). Snow accumulation and melt under a warming climate are governed by competing effects: higher temperatures reduce snowfall and accelerate melt, yet increased atmospheric moisture under warmer conditions can locally enhance snowfall rates (Räisänen, 2008). Climate-driven shifts in circulation patterns further influence snowfall distribution, with mid-latitude snowpacks being particularly sensitive (Fletcher *et al.*, 2009).

Reductions in SWE accelerate the depletion of already stressed groundwater systems (Liu *et al.*, 2021), and across both cold and dry-summer regions declining snowfall and snow cover impact not only water supplies but also ecosystem stability, vegetation and wildlife dynamics, transportation networks, cultural practices, and recreational activities (Bair *et al.*, 2018a). Reduced meltwater availability alters freshwater biodiversity, while changes in soil moisture and thermal conditions, including permafrost thaw, can lead to vegetation degradation in some regions (Walvoord & Kurylyk, 2016). Conversely, increased streamflow in certain Arctic areas may locally enhance vegetation productivity (Skovsholt *et al.*, 2020). Glacier retreat, decreasing snow cover, and permafrost degradation are driving shifts in species distributions, forcing many plants and animals to migrate to higher altitudes (Rew *et al.*, 2020). Microbial processes are also affected, with implications for nitrogen cycling, organic matter decomposition, and broader biogeochemical feedbacks (Bradley *et al.*, 2017). Moreover, changes in snow cover can significantly modify the Earth's surface energy balance and influence large-scale atmospheric circulation in the Northern Hemisphere (Dunmire *et al.*, 2026a). Finally, the retreat of cryospheric elements contributes to the increasing frequency, magnitude, and changing spatial distribution of natural hazards in mountain and polar regions (Beniston *et al.*, 2018).

Within the context of a rapidly changing climate, understanding snowpack dynamics has become increasingly urgent, particularly because nearly one-sixth of the world's population relies on snow-derived water resources (Sturm *et al.*, 2017). Ultimately, changes in the cryosphere represent a multi-dimensional driver that affects climate dynamics, hydrology, ecosystems, geomorphology, natural hazards, and socio-economic systems (Rasul & Molden, 2019). These interconnected impacts underscore the need for improved monitoring and modelling to support sustainable development in regions influenced by snow and ice (Siirila-Woodburn *et al.*, 2021).

In addition to climate-related challenges, the strong spatial and temporal variability of snow cover further complicates the accurate modelling of snowpack dynamics. This variability, combined with the nonlinearity of the precipitation–runoff relationship (Latron *et al.*, 2009), makes it difficult for models to reliably capture snow accumulation, melt processes, and their hydrological impacts.

To address the need for reliable snow variable estimates, snowpack models, when cou-

pled with observational data, provide essential tools for scientists, engineers, and decision-makers engaged in hydrological forecasting and operational water management. Such integrated approaches enable more reliable predictions of snowmelt dynamics, ensuring that water resources are managed sustainably under changing climatic conditions (Kazama *et al.*, 2007).

### 2.1.1 Snow variability

Snow properties such as depth, density, water content, SWE and duration, are characterized by high spatial and temporal variability (Clark *et al.*, 2011), which makes their representation in models inherently challenging (Blöschl, 1999).

Snow cover variability is primarily controlled by meteorological forcing (Bales *et al.*, 2006). Temperature is a primary driver, controlling the precipitation partitioning into rain or snow, influencing the rate of snowmelt and the character of the snowpack (e.g., dry and powdery versus wet and heavy snow, Cline (1997)). The amount, intensity, and type (snow vs. rain) of precipitation events directly affect snow accumulation (Li *et al.*, 2019). Wind can break snow crystals into smaller fragments, increasing density, and causes snow drifting, which leads to large spatial variations in snow depth across a landscape, although its effects on snow are difficult to quantify (Gascoin *et al.*, 2013).

In mountain regions, this complexity is amplified by strong heterogeneities in topography, microclimate, and vegetation. The shape of the land (e.g., elevation, slope aspect) and the presence of vegetation (e.g., forest canopy cover or rocks) significantly influence how much snow accumulates in a given location and how quickly it melts. For example, forest canopies regulate both snow accumulation and ablation (Essery *et al.*, 2009), and vegetation itself is dynamic, responding to climatic changes that feedback into snowmelt and runoff regimes (Strasser *et al.*, 2024).

The reflectivity of the snow surface affects how much solar energy is absorbed,thus how fast snow melts. Fresh, clean snow has high albedo, while older or dust-covered snow has lower albedo and melts faster (Adhikary *et al.*, 2002). On the climatological point of view, large-scale atmospheric-oceanic patterns, such as the El Niño-Southern Oscillation (ENSO) and the Arctic Oscillation (AO, Douville *et al.* (2017)), can influence regional snowfall patterns and contribute to year-to-year variability.

Additionally, climate change is altering historical patterns of snow variability (Wieder *et al.*, 2022). As global temperatures rise, more winter precipitation is falling as rain instead of snow, and snowmelt is occurring earlier in the spring (López-Moreno *et al.*, 2021). This generally leads to an overall decrease in average SWE and the duration of snow cover in many regions, while potentially increasing the relative variability of snow conditions from one winter to the next.

Snow variability has far-reaching implications for both natural systems and human society (Siirila-Woodburn *et al.*, 2021). Changes in the timing, duration, and amount of snow alter soil moisture availability, growing seasons, and habitat conditions for cold-

adapted species, and they influence key eco-hydrological processes such as soil insulation, nutrient cycling, and stream temperature regulation (Ernakovich *et al.*, 2014). At the catchment scale, the spatial heterogeneity of snow accumulation and melt governs the magnitude and timing of runoff, making mountain snowpacks a critical natural reservoir (DeBeer & Pomeroy, 2017). Increased variability in how snow accumulates and melts from year to year complicates water resource management, affecting reservoir operations, agricultural planning, hydropower production, and urban water supply (Soomro *et al.*, 2024).

Earlier and more irregular melt events can lead to both reduced late-season water availability and higher risks of early-season flooding (Frei *et al.*, 2000) or rain-on-snow–induced hazards (Beniston & Stoffel, 2016). Snow-dependent economies, including winter tourism and transportation, are likewise sensitive to shifts in snow persistence and spatial distribution (Leal Filho *et al.*, 2024). These combined pressures highlight the growing need for modelling frameworks capable of capturing fine-scale spatial and temporal snowpack dynamics.

### 2.1.2 Snow modelling

A model can be defined as a simplified representation of a real-world system that captures its essential processes and key interactions. Hydrological modelling provides hydrologists and environmental engineers with a framework to interpret and predict the complexity of the water cycle, aiming to deliver reliable forecasts for sustainable water resource management (Savenije, 2009).

Cryospheric models integrate principles of hydrology, thermodynamics, and rheology (Bitz *et al.*, 2012), with snowpack modelling playing a central role. Snow models seek to reproduce the processes through which snow accumulates, metamorphoses, and melts, thereby regulating the seasonal storage and release of water in snow-dominated regions (Voordendag *et al.*, 2021). The primary goal of snow modelling is to provide a process-based understanding of snow dynamics, allowing for improved quantification of their effects on streamflow timing, magnitude, and downstream water availability (Maurer *et al.*, 2021a). Although cryospheric science, and snowpack modelling in particular, dates back to the early 18th century (Qin *et al.*, 2021), research in this field continues to evolve as these challenges remain central to hydrological and climate studies.

A variety of models, differing in complexity and in the representation of snow processes, have been developed to simulate the accumulation and ablation of seasonal snowpacks (Voordendag *et al.*, 2021). These range from simple empirical formulations to fully process-based, physically explicit models.

Empirical models, such as temperature-index or degree-day models (e.g., Hock (2003); Ismail *et al.* (2023)), rely on statistical relationships between air temperature and snowmelt. Surface energy fluxes are parameterized as a function of air temperature, without explicitly resolving internal snowpack processes. Empirical models are computationally efficient and

widely used. Examples include the Snowmelt Runoff Model (SRM, Martinec (1975)), the one-layer temperature-index snow model (HyS, De Michele *et al.* (2013)), and the Snow Accumulation and Ablation Model (SNOW-17 in Figure 2.3, Anderson (1973)). They estimate seasonal snow cover evolution by combining air temperature, precipitation, and snowmelt (Ohmura, 2001). Despite their heavy parameterization, empirical models remain effective in operational hydrology and applied research due to their ability to reproduce observed melt dynamics with minimal input data (Sahu *et al.*, 2023). However, their empirical nature limits transferability outside the calibration domain, and predictive skill may decline under climate change scenarios (e.g., Li *et al.* (2012); Réveillet *et al.* (2016); Duethmann *et al.* (2020)).

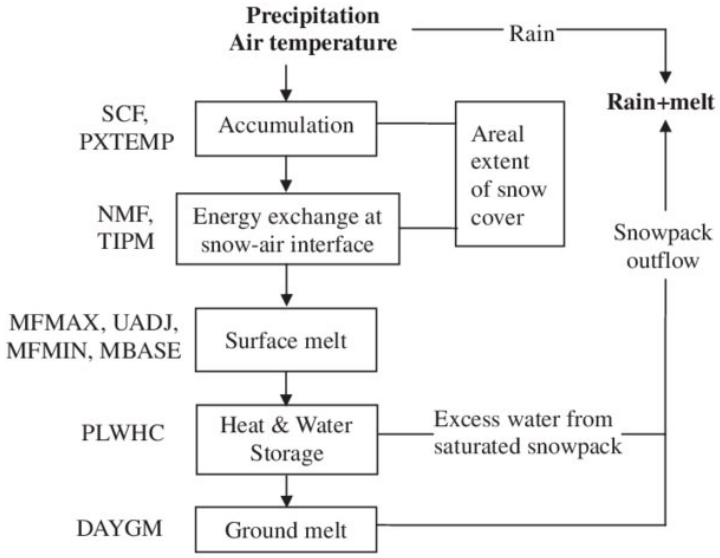


**Figure 2.3:** *SNOW-17 model workflow and parameters. The boxes designate model processes. Model inputs and output are highlighted in bold, source He* et al. *(2011).*

Hybrid models were developed to overcome the limitations of purely empirical approaches (Follum *et al.*, 2015). These models incorporate simplified representations of the surface energy balance to capture key drivers of spatial and temporal variability in snowmelt while retaining the computational efficiency of degree-day formulations (Pellicciotti *et al.*, 2005). Solar radiation is included as an explanatory variable through melt parametrizations (Hock, 1999), enabling these models to reproduce diurnal melt cycles and spatial heterogeneity in snowmelt driven by slope, aspect, and shading. Since potential solar radiation can be computed from solar geometry and topography, hybrid models do not necessarily require additional meteorological inputs beyond air temperature. One example is the Snow Multidata Mapping and Modelling (S3M) model (Avanzi *et al.*, 2022a); it simulates snow mass balance and melt using a hybrid temperature-index and radiation-driven

formulation. S3M accounts for settling, albedo evolution, and liquid water outflow (Avanzi *et al.*, 2022b), balancing process representation with computational efficiency (see Figure 2.4). It supports operational systems that provide real-time, spatially explicit snow cover estimates (Avanzi *et al.*, 2023).

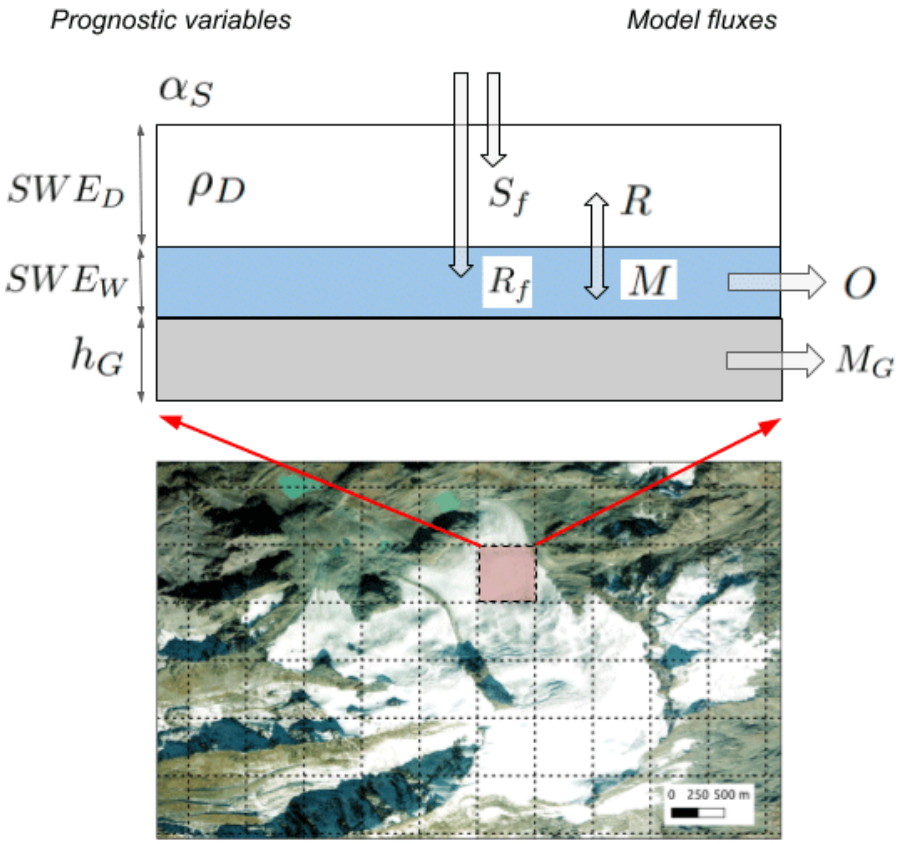


**Figure 2.4:** *Prognostic variables and main mass fluxes in the S3M model: $\alpha_S$ = snow albedo (–); $S_f$ and $R_f$ = snowfall and rainfall rates (mm $\Delta t^{-1}$); $\mathrm{SWE}_D$ = dry-SWE (mm); $\rho_D$ = dry-snow density (kg $m^{-3}$); $\mathrm{SWE}_W$ = wet-SWE (mm); $R$, $M$, and $O$ = refreezing, melt, and outflow (mm $\Delta t^{-1}$); $h_G$ = glacier thickness (m); $M_G$ = ice melt (mm $\Delta t^{-1}$). The schematic represents the distributed snow–glacier model applied independently to each grid cell, where mass and energy fluxes are solved using ordinary differential equations integrated with a forward Euler scheme. All inputs, states, and outputs are spatially distributed as raster fields (e.g., 240 m resolution) with a typical 1-hour time step. The background image shows the Rutor Glacier in northwestern Italy (ESRI Satellite) Source: Avanzi* et al. *(2022a).*

While empirical and hybrid models are attractive for their simplicity and modest data

requirements, their reliance on calibrated statistical relationships limits generalizability and transferability. Physically based models overcome this by explicitly solving the surface energy balance and representing energy and mass fluxes at the snow–atmosphere and snow–ground interfaces (Born *et al.*, 2019). They provide a comprehensive description of snow accumulation, melt, sublimation, and internal processes such as compaction, albedo evolution, temperature gradients, refreezing, and liquid water retention (Ismail-Zadeh *et al.*, 2023).

Physically based, or energy and mass balance, models vary in their representation of snowpack stratigraphy. Single-layer models (Essery *et al.*, 1999) describe bulk snowpack properties such as albedo, density, and thermal conductivity in a depth-integrated form, and remain operationally useful due to their efficiency (e.g., ECMWF Integrated Forecast System snow scheme (IFS), Dutra *et al.* (2010)). Multilayer models explicitly resolve snow stratigraphy and microstructural evolution, ranging from a few layers in intermediate-complexity schemes to 100 layers in highly detailed models. These can resolve processes such as compaction, liquid water percolation, refreezing, and metamorphism. Examples include CROCUS (Vionnet *et al.*, 2012), SNOWPACK (Lehning *et al.*, 2002a), the snow metamorphism and albedo process (SMAP) model (Niwano *et al.*, 2012), SNOBAL (Marks *et al.*, 1998), the Factorial Snow Model (FSM, Essery (2015b)), and GEOtop (Zanotti *et al.*, 2004; Endrizzi *et al.*, 2014). Some models, such as SNOWPACK (see Figure 2.5), adopt a Lagrangian multilayer framework to track snow microstructure evolution explicitly. Some other energy balance models, such as FSM, also differ in their representation of surface fluxes incorporating radiative, turbulent, advective, and conductive heat fluxes using lumped or multi-physics formulations. Another example of this category is the Utah Energy Balance model (UEB, Essery (2015a)). The UEB snow model is an energy-balance–based snowmelt model that represents the snowpack using a lumped, yet physically consistent, formulation that tracks both the water and energy budgets. The model is driven by meteorological inputs including air temperature, precipitation, wind speed, humidity, and shortwave and long wave radiation, provided at a temporal resolution of one or six hours. UEB computes the main components of the surface energy balance,radiative, sensible, latent, and advective heat fluxes, using physically based parameterization ( see Figure 2.6).

Distributed applications in mountainous terrain require consideration of spatial heterogeneity in meteorological forcing and topography, including shading, slope- and aspect-dependent radiation, and downscaling or interpolation of coarse-resolution data to fine model grids (e.g., MeteoIO, Bavay & Egger (2014a)). While physically based models are data- and computation-intensive, they provide the most robust framework for representing fine-scale snow dynamics and assessing climate change impacts.

Recently, data-driven models based on AI have gained attention for their ability to learn complex nonlinear relationships between meteorological drivers and snowpack dynamics without explicit physical formulations (Sit *et al.*, 2020a). By leveraging historical observations, ground measurements, and remote sensing products, these models have demonstrated strong predictive capabilities, particularly for snow depth, and can overcome

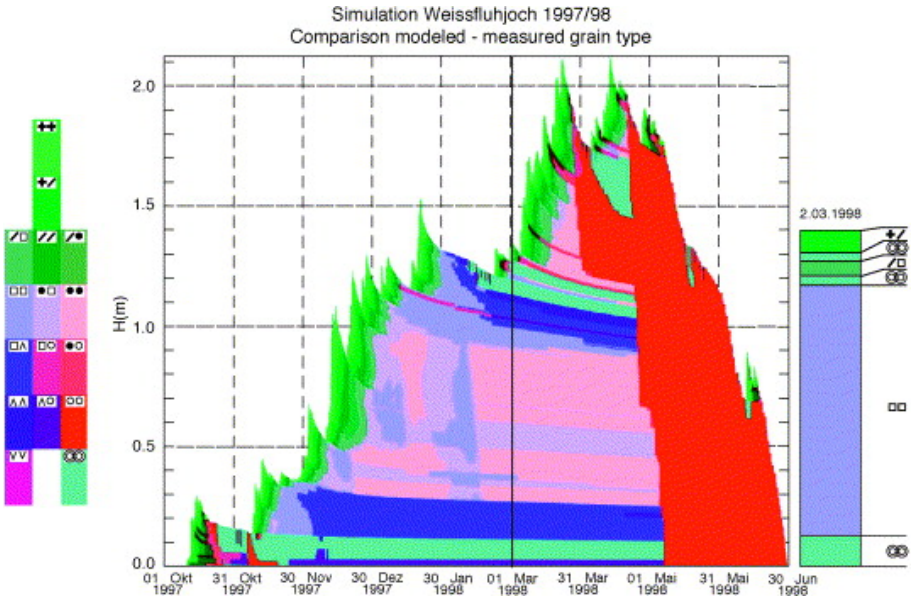


**Figure 2.5:** *Output from the Swiss SNOWPACK model (Lehning* et al.*, 2002c) showing a modelled vertical snowpack structure profile from early October (no snow) to June. Colours and symbols represent different snow grain types.*

some structural limitations of traditional physically based models (Dai *et al.*, 2018; Acharki *et al.*, 2025; Chen *et al.*, 2021b; Bousbaa *et al.*, 2024; Elyoussfi *et al.*, 2025a; LeCun *et al.*, 2015a; Murphy, 2023).

In practice, the choice of model depends on the application, data availability, and spatial scale. Cryospheric models are typically spatially distributed and multi-scale, reflecting the influence of topography and vegetation (Avanzi *et al.*, 2020a). Parsimonious models remain essential in ungauged or data-scarce regions (Avanzi *et al.*, 2021a), while computational efficiency is key to capturing long-term cryospheric processes (Pagano *et al.*, 2014).

The longstanding question of whether a "best" snow model exists remains unsolved. Comparative studies (Essery *et al.*, 2013) show that model performance depends strongly on local conditions, including topography, meteorology, and vegetation. Increased complexity does not necessarily improve accuracy (Magnusson *et al.*, 2015); empirical parameterization can perform as well as or better than more detailed schemes (Beven, 2012). Snow modelling thus seeks adequacy rather than universality: selecting the most appropriate model for a given context, balancing realism, parsimony, and computational cost (Magnusson *et al.*, 2015), trying to answer a central question in snow modelling on how much snow is accumulated across the landscape at a given time (Sturm *et al.*, 2017; Dozier *et al.*, 2016). Calibration improves predictive skill but also highlights the tension between parametric and structural uncertainty, with high-parameter models at risk of overfitting and limited

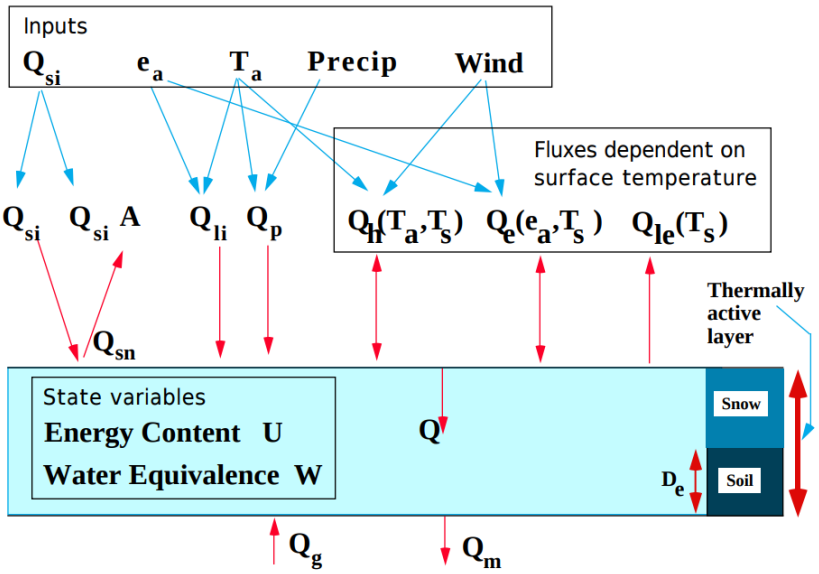


**Figure 2.6:** *Snow model physics and parameterization from Tarboton* et al. *(1996) (UEB model). Meteorological inputs ($Q_{si}$, $T_a$, $e_a$, precipitation, wind) drive energy and water fluxes in the thermally active layer. Fluxes include temperature-independent ($Q_{si}^A$, $Q_{li}$, $Q_p$) and temperature-dependent components ($Q_h(T_a, T_s)$, $Q_e(e_a, T_s)$, $Q_{le}(T_s)$), which modify the state variables—total energy $U$ and water equivalence $W$. Exchanges between snow and soil ($Q'$), snowmelt ($Q_m$), and ground heat flux ($Q_g$) dynamically couple the system, controlling snowmelt, sublimation, and heat transfer.*

transferability (Girotto *et al.*, 2020).

### 2.1.3 Modelling with uncertainty

Despite advances in process representation, models used in snow hydrology remain affected by significant uncertainties (Beven, 2012). Structural uncertainty arises from limitations in the accuracy and reliability of the equations and their numerical discretization used to represent physical processes on a computer. Parametric uncertainty reflects imperfect knowledge of model parameters (Girotto *et al.*, 2020), while forcing uncertainty is particularly pronounced in mountainous terrain; Indeed, both the phase and amount of precipitation are difficult to constrain, temperature distributions are often poorly monitored, and the lack of wind information prevents an accurate representation of snow redistribution processes (Tyson *et al.*, 2023). Moreover, the intermittent nature of snow in space and time, along with stratification and slow processes such as snow metamorphism or wind transport, further complicate the reconstruction of snow dynamics (Piazzi *et al.*, 2019).

Calibration is traditionally employed to mitigate these uncertainties by tuning parameters so as to reduce the mismatch between simulations and observations (Gupta *et al.*, 2006). However, this approach is limited. Different parameter sets may lead to equally good results (equifinality, Beven (2006)), and structural model deficiencies cannot be resolved by parameter adjustment alone. More advanced frameworks, such as the Generalized Likelihood Uncertainty Estimation (GLUE, Beven & Binley (1992)), explicitly consider parameter uncertainty, yet they still provide static corrections and cannot account for the dynamic propagation of state errors during simulations. Thus, calibration is an essential first step, but it cannot fully address the evolving uncertainties encountered in predictive modelling (Acero Triana *et al.*, 2019).

If models are uncertain, so too are the observations used to constrain them. All snow-related observations are subject to measurement errors, representation errors, and retrieval uncertainties (Beven, 2012). Ground-based measurements provide valuable point-scale information, but they suffer from instrumental noise and are heavily influenced by wind redistribution, local topography, and vegetation, which complicates their extrapolation to the scale of model grid cells (Malek *et al.*, 2017a). Their representativeness may further degrade under changing climate and land cover conditions (Cowherd *et al.*, 2024b).

Remote sensing products offer spatially distributed information, but they also carry significant limitations: retrieval algorithms introduce uncertainty, coarse spatial resolution hampers local applications (Van Leeuwen, 2015), and cloud cover or forest canopy may obscure snow signals (Aalstad *et al.*, 2020; Gascoin *et al.*, 2024). Similarly, meteorological forcings derived from reanalyses are affected by large uncertainties. Precipitation is particularly problematic due to gauge undercatch, sparse high-elevation observations, and difficulties in phase partitioning, which are critical during rain-on-snow events (Beniston *et al.*, 2018; Juras *et al.*, 2021). Temperature is generally more reliable, yet high-elevation data remain scarce. Radiation, the dominant energy input for snowmelt, is strongly modulated by topography and vegetation and is often poorly captured in coarse-resolution reanalyses. For instance, ERA5 (Hersbach *et al.*, 2020) and ERA5-Land (Muñoz-Sabater *et al.*, 2021) (31 km and 9 km resolution) are known to overestimate snow accumulation and delay melt timing, partly due to their simplified single-layer snow scheme, wrongly estimating temperature or precipitation. Thus the derived products are often hampered by these propagated errors (Kouki *et al.*, 2023).

Additionally, when dealing with remotely-sensed data it is noteworthy also to consider that they are indirect retrievals of the variables of interest (e.g. electromagnetic radiation is used to indirectly estimate land surface physical properties) and thus significantly affected by uncertainty (De Lannoy *et al.*, 2022).

### 2.1.4 Snow data

Within hydrology-focused cryospheric studies, the availability of accurate snow data is as important as snow modelling itself (Lievens *et al.*, 2019). Changes in the snowpack influence ecological and geomorphological processes, as well as communities and socio-economic systems (Revuelto *et al.*, 2025). Reliable observations are therefore indispensable for advancing the understanding of snow dynamics and assessing their impacts on hydrological processes (Viviroli *et al.*, 2007). Moreover, high-quality datasets provide essential benchmarks for evaluating model performance (Franz *et al.*, 2008). Snow data could be categorized into three main groups: short-, medium-, and long-range measurements (see Figure 2.7).


**Figure 2.7:** *Categories of snow data: ground based measurements, remote sensing, satellite observations and areal ones (Reinking* et al.*, 2022).*

In-situ measurements provide valuable information for capturing snow heterogeneity at fine spatial scales (Pirazzini *et al.*, 2018) and at short ranges. These observations are typically acquired at high frequency, enabling accurate reconstruction of the temporal evolution of the snowpack, particularly when recorded by Automatic Weather Stations (AWS, Largeron *et al.* (2020)). Compared to remote sensing techniques, in-situ measurements are generally less susceptible to environmental interferences such as cloud cover. The use of wind shields and the application of undercatch corrections further reduce uncertainty in solid precipitation measurements (Revuelto *et al.*, 2025).

Among measured variables, snow depth (or height) is the most widely available and can be monitored accurately and continuously in situ. Snow density, which exhibits lower spa-

tial variability than snow depth (López-Moreno *et al.*, 2013), requires fewer measurements but is subject to substantial uncertainties when derived from manual in-situ observations. Direct measurements of SWE are rare and uncertain, similar to snow density. However, SWE can be estimated by multiplying snow depth and density (Sturm *et al.*, 2010).

However, in-situ measurements represent only point-scale information and cannot capture the full spatial variability of snow, which is strongly influenced by climate, terrain heterogeneity, and the density of the observational network (Dong, 2018; Grünewald & Lehning, 2014). Furthermore, snow-dominated regions are often remote, high-elevation, and steep, making access difficult and limiting data availability (Fayad *et al.*, 2017).

To obtain basin-scale snow cover observations, conventional non-invasive remote sensing techniques have been widely used. These methods help address the undersampling inherent to point-scale measurements, but they are constrained by relatively coarse spatial and temporal resolution and by detection limitations when monitoring snow cover (Domine *et al.*, 2022).

A major breakthrough in snow science was the introduction of Light Detection and Ranging (LiDAR, Deems *et al.* (2013)). Numerous studies have used terrestrial and airborne LiDAR to provide valuable information on snow depth distribution (Painter *et al.*, 2016). Terrestrial laser scanning allows spatially continuous mapping of snow depth over hundreds of meters to kilometre scales, with centimetre-level accuracy, by differencing snow-free and snow-covered surfaces (Revuelto *et al.*, 2014). However, these methods are expensive and typically provide only seasonal snapshots (Deems *et al.*, 2013). Their continuous use is also limited by viewing geometry, long-term platform stability, and logistical challenges associated with heavy instrumentation.

In certain regions, airborne sensors mounted on airplanes (e.g. the Airborne Snow Observatory (ASO) initiative, Painter *et al.* (2016)) have enabled detailed snowpack mapping with spatial coverage approaching that of satellite platforms, however their cost remain the major limitation factor.

In recent years, the decreasing cost and increasing reliability of Unmanned Aerial Vehicles (UAVs) have driven rapid growth in UAV-based snow observations (Revuelto *et al.*, 2014). This development has opened new opportunities for characterizing snowpack variables. LiDAR has also recently been deployed on UAV platforms (Marsh *et al.*, 2023). Airborne and terrestrial LiDAR provide centimetre-level snow depth accuracy at regional (medium-range) scales (Sheridan, 2019).

The wide variety of sensors that UAVs can now carry enables direct observation of SWE and liquid water content Liquid Water Content (LWC) using Ground Penetrating Radar (GPR) (Valence *et al.*, 2022). However, UAV-based acquisitions have inherent uncertainties, even under optimal conditions. For example, airborne GPR tends to overestimate LWC and underestimate snow density (Dadrass Javan *et al.*, 2024). Multispectral and hyperspectral cameras mounted on UAVs can detect snow surface impurities, algae, snow grain size, and albedo (Skiles *et al.*, 2023). Yet, the influence of snow impurities on surface darkening remains insufficiently constrained, introducing uncertainties in large-scale

retrievals (Dumont *et al.*, 2014).

UAVs offer many advantages, including flexible acquisition over large domains, rapid data collection, and simplified experimental setups; snow depth can be reliably measured even without ground control points (Eberhard *et al.*, 2020). Nonetheless, recursive visits are necessary to monitor temporal evolution, and UAV operations are subject to legal restrictions on flight time, distance, and altitude, which vary by country.

Satellite observations overcome many of the limitations of in-situ measurements and medium-range sensors (Gascoin *et al.*, 2024) by enabling the monitoring of snow properties over large and often inaccessible areas, particularly in high mountains and ungauged basins (DeWalle & Rango, 2008). The most common snow-related variables retrieved from satellite data include the snow cover area (Snow Cover Area (SCA)), the presence or absence of snow in a pixel, and the fractional snow cover (Fractional Snow Cover (FSC)), which represents the fraction of a pixel covered by snow (Rittger *et al.*, 2013). The identification of snow-covered surfaces from space relies on the interaction between snow and electromagnetic radiation across different frequencies (Dong, 2018). Snow can be detected using microwave, infrared, and visible remote sensing observations, largely due to its high albedo in the visible spectrum (Dozier *et al.*, 2009, 2016). However, visible and near-infrared imagery require daylight and are hindered by cloud cover. Since clouds often exhibit reflectance similar to snow, cloud-removal algorithms are needed (Gascoin *et al.*, 2015). Near-infrared bands improve the discrimination between snow and clouds (Dumont & Gascoin, 2016), yet forest canopies, shading, and surface heterogeneity (e.g., glaciers) remain sources of uncertainty (Webster & Jonas, 2018; DeBeer & Pomeroy, 2017). Microwave sensors play a key role in large-scale snow monitoring. Passive microwave radiometers detect snow through reduced natural microwave emission caused by scattering from snow grains, enabling both snow detection and SWE retrieval (Rouzegari *et al.*, 2025). These sensors operate independently of daylight and most cloud cover but are limited by coarse spatial resolution (up to 25 km), sensitivity to wet snow and vegetation, and saturation in deep snow, making them less effective in complex terrain.

Nowadays, several multispectral satellite missions provide operational visible and infrared data for snow monitoring. NASA's Moderate Resolution Imaging Spectroradiometer (MODIS), on board the Terra and Aqua satellites, was the first mission to provide daily global coverage at 250 m resolution (Justice *et al.*, 1998). MODIS measures radiation across 36 spectral bands and delivers SCA, FSC, and daily snow albedo products at 500 m resolution, with near-daily coverage across mid- to high-latitudes (Justice *et al.*, 2002). These products rely on the Normalized Difference Snow Index (NDSI) (Wang *et al.*, 2022a), which exploits snow's high reflectance in the visible and low reflectance in the shortwave-infrared, combined with additional spectral and thermal filters to reduce classification errors.

The Visible Infrared Imaging Radiometer Suite (VIIRS), aboard the joint NASA/NOAA Suomi National Polar-orbiting Partnership (Suomi NPP) and NOAA platforms (Murphy *et al.*, 2006), continues and extends the observational record established by MODIS and NOAA's Advanced Very High Resolution Radiometer (AVHRR). VIIRS collects visible and

**Figure 2.8:** *Aqua (EOS PM-1) (left) and Terra (EOS AM-1)(right) NASA satellite; Images taken from Wikipedia (Wikipedia contributors, 2004).*

infrared imagery of the Earth's land, atmosphere, cryosphere, and oceans. Its snow cover and sea ice algorithms were specifically designed to ensure compatibility with MODIS, enabling long-term continuity and facilitating climate data record development across the operational lifetimes of MODIS, VIIRS, and future missions (Rittger *et al.*, 2021a).

The Copernicus Sentinel missions (Jutz & Milagro-Perez, 2020) from the the European Union's Earth observation programme, complement these datasets with a combination of radar and optical observations.

Active microwave systems provide weather-independent observations with relatively high spatial resolution, enabling applications such as snow cover mapping and wet snow detection (Demil *et al.*, 2025). They have also shown potential for snow depth retrieval under specific conditions, particularly for dry snowpacks where radar backscatter is sensitive to snow volume scattering, although performance may decrease in areas with dense vegetation (Lievens *et al.*, 2019).

Sentinel-1, the first mission in the series, consists of two sun-synchronous satellites equipped with C-band Synthetic Aperture Radar (SAR) that provides day- and night-time, weather-independent imaging at spatial resolutions as fine as 5 m and swath widths up to 400 km (Potin *et al.*, 2012). Its dual-polarization capability and short revisit time support long-term archives suitable for snow monitoring. Sentinel-1 C-band SAR backscatter is sensitive to snowpack properties, particularly liquid water content, allowing the detection of snow accumulation and melt, while higher frequencies have potential for SWE retrieval (Torres *et al.*, 2012). During melt, the backscatter coefficient typically decreases as water increases dielectric loss and then rises again as meltwater drains and the snow structure

evolves. Empirical and semi-empirical algorithms have been applied for estimating snow depth. Validation studies in the Alps show strong correlation with in-situ data (r approximately 0.89), although errors of 20–30% may occur for deep snow or heterogeneous terrain. Recent advances in snow depth estimation using Sentinel-1 SAR data have progressed from temporal change-detection approaches, such as those implemented in the C-SNOW project (Lievens *et al.*, 2022), toward ML–based methods, that integrate SAR observations with ancillary datasets such as topography, snow cover, and forest information. These ML-based approaches enable estimation of snow depth at finer spatial resolutions (up to 100 m) across the European Alps (Dunmire *et al.*, 2024), improving retrieval performance compared to traditional change-detection techniques.

Finally, Sentinel-1's frequent revisit cycle (6–12 days), high geolocation accuracy, and systematic acquisition strategy make it a key contributor to operational snow monitoring.

**Figure 2.9:** *Sentinel-1 Satellite (ESA, 2024).*

Sentinel-2 complements Sentinel-1 with high-resolution (10–20 m) multispectral optical imagery suitable for detailed snow cover mapping (Spoto *et al.*, 2012). Its Level-2A products enable monitoring of SCA, albedo, and fractional snow distribution, capturing fine-scale spatial heterogeneity in alpine environments. Operational datasets such as the Theia Snow product (Gascoin *et al.*, 2019) combine NDSI with elevation to generate 20 m snow/no-snow maps that generally agree well with observations, though underestimation may occur in shaded or forested regions and cloud cover remains a limiting factor.

Sentinel-3 provides medium-resolution (∼300 m) optical and thermal observations with high temporal frequency (Donlon *et al.*, 2012). Its Ocean and Land Colour Instrument (OLCI) acquires radiance across 21 spectral bands, enabling retrieval of snow grain size, broadband and spectral albedo, and specific surface area. The combined use of Sentinel-

**Figure 2.10:** *Sentinel-2 Satellite (Copernicus, 2025a).*

2 and Sentinel-3 enables multi-scale snow monitoring, where Sentinel-2 captures high-resolution spatial patterns and Sentinel-3 ensures frequent, wide-area coverage.

**Figure 2.11:** *Sentinel-3 Satellite (Copernicus, 2025b).*

Spaceborne lidar missions add an additional dimension by providing surface elevation measurements along satellite ground tracks, enabling snow-depth estimation through differencing between snow-on and snow-off elevations (Deschamps-Berger *et al.*, 2020). The ICESat mission (2003–2010) (Zwally, 2002), equipped with the Geoscience Laser Altimeter System (GLAS) lidar, sampled the surface every 170 m within a 70 m footprint and achieved snow-depth retrievals with an RMSE of about 1 m (Treichler & Kääb, 2017).

However, these were constrained to gentle slopes ($< 10°$) and were temporally sparse, making them suitable mainly for seasonal or elevation-band averages. ICESat-2, launched in 2018, represents a substantial improvement (Zwally, 2002). Its Advanced Topographic Laser Altimeter System (ATLAS) uses photon-counting lidar to generate dense along-track observations with about 0.70 m photon spacing and about 11 m footprints, and geolocation accuracy of 3–4 m (Neumann *et al.*, 2019; Besso *et al.*, 2024). Although the satellite's orbit prioritizes mid-latitude vegetation studies, complicating simple snow-on/snow-off differencing, higher-level products such as ATL06 (20 m land ice elevation) and ATL08 (100 m land surface and canopy height) have been adapted for snow-depth retrieval (Enderlin *et al.*, 2022). ATL08 performs well only for shallow snow ($< 0.35$ m) on very gentle slopes ($< 1.5°$), whereas ATL06 has demonstrated promising results even in mountainous terrain and glacierized areas (Enderlin *et al.*, 2022). Combined with the increasing availability of high-precision snow-free DEMs, ICESat-2 provides a valuable complement to other satellite-based snow observation systems (Deschamps-Berger *et al.*, 2023b).

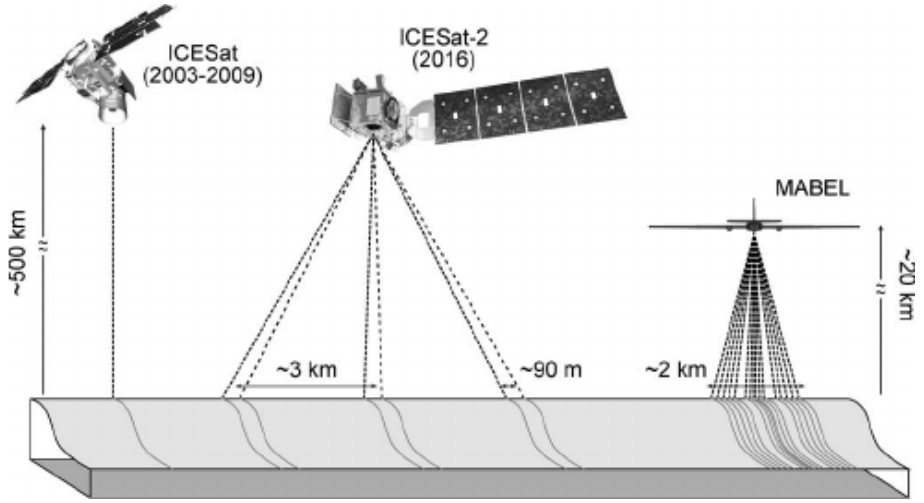


**Figure 2.12:** *ICESat-mission and MABEL, Multiple Altimeter Beam Experimental Lidar(a high-altitude airborne laser altimeter de-signed as a simulator for ICESat-2, Brunt* et al. *(2013)).*

Lastly, it is worth nothing that the scientific community recognizes that there is no clear evidence of substantial advantages in using one sensor over another for snow monitoring. Consequently, the fusion of multiple data sources, including ground-based measurements, remote sensing observations, and derived products, is widely considered the most reliable approach for obtaining accurate snow estimates (De Gregorio *et al.*, 2019).

## 2.2 Data assimilation in snow hydrology

Given the uncertainties inherent in both models and observations, neither alone is sufficient for reliable snow prediction (Girotto *et al.*, 2023). Data assimilation (DA) provides a statistical framework to optimally merge the two while explicitly accounting for their uncertainties (Evensen *et al.*, 2022). Unlike calibration, which is static and offline, DA operates dynamically and online: it updates model states (and sometimes parameters) as new data become available, thereby constraining trajectories to remain consistent with both physical laws and observations (Carrassi *et al.*, 2018). In this way, calibration and DA should be seen as complementary. Calibration defines plausible parameter ranges and ensures structural realism, while DA continuously reduces forecast errors and corrects evolving state trajectories, yielding physically consistent and statistically optimal snowpack estimates (Magnusson *et al.*, 2017). Lastly, according to Metref *et al.* (2023b), assimilating snow data not only improves the representation of snowpack dynamics but also enhances broader hydrological modelling, including seasonal streamflow supply prediction in mountainous basins.

In this context, DA plays a key role in reducing the errors and uncertainties associated with estimates of seasonal snow analysis and prediction (Ismail-Zadeh *et al.*, 2023). By optimally combining two complementary sources of information, the model and the observations, DA seeks to produce a statistically optimal and dynamically consistent estimate of the evolving system state by weighting their contributions according to their respective uncertainties (Carrassi *et al.*, 2018).

### 2.2.1 A mathematical framework for data assimilation

The state of a hydrological model represents the configuration of the system at a given timestep $t_k$. In the context of snow hydrology, the problem of state estimation arises when we seek to characterize the evolving condition of such a dynamical, discrete system (Ismail-Zadeh *et al.*, 2023).

The true state of the system is typically unknown, but it can be approximated through the process known as *analysis*. Formally, the analysis step can be framed as an optimization problem, where the objective is to minimize, on average, the discrepancy between the estimated state and the (unknown) observational truth (Evensen, 2003). To achieve this, DA combines two complementary sources of information: the model dynamics and the observation process (Carrassi *et al.*, 2018).

The discrete model can be described as :

$$\boldsymbol{x}_{k+1} = \mathbf{M}(\boldsymbol{x}_k, \lambda, \boldsymbol{u}_k) + \boldsymbol{\eta}_{k+1}. \tag{2.1}$$

where:

- $\boldsymbol{x}_k \in \mathbb{R}^n$ is the model state at time $k$,

- $\lambda \in \mathbb{R}^p$ is the model parameter vector,
- $\mathbf{M} : \mathbb{R}^n \to \mathbb{R}^n$ are the model equations (the hydrological system),
- $\boldsymbol{\eta}_{k+1}$ is the model error at time $k+1$, represented as an additive stochastic term.
- $\boldsymbol{u}_k$ is the external forcing vector at time $k$.

The model provides an *a priori*, or background state, which is an estimate of the true system state. Since model error accounts for uncertainties related to both the model structure and the forcing data, it is generally state-dependent and unknown. For this reason, it is statistically represented as a stochastic perturbation.

The noisy observations available at discrete times are described as

$$\boldsymbol{y}_k = \mathbf{H}(\boldsymbol{x}_k) + \boldsymbol{\epsilon}_k, \tag{2.2}$$

where:

- $\boldsymbol{y}_k \in \mathbb{R}^m$ is the observation vector at time $k$,
- $\mathbf{H} : \mathbb{R}^n \to \mathbb{R}^m$ is the observation operator mapping model space to observation space,
- $\boldsymbol{\epsilon}_k$ is the observation error at time $k$, modelled as an additive stochastic term.

Two main DA approaches can be distinguished:

- **Non-sequential assimilation**: all available observations within a predefined assimilation window are used simultaneously to estimate the system state. This approach may include observations that are temporally ahead of the current analysis time, provided they fall within the assimilation window, implicitly assuming that all observations over the window are already available. Non-sequential methods are typically formulated within a variational framework (e.g., 3D-Var and 4D-Var) or using smoothing approaches. In both cases, the system state (now the entire trajectory) is optimized to achieve the best overall fit to both the model dynamics and the observations (Montiforte *et al.*, 2024; Carrassi *et al.*, 2018).
- **Sequential assimilation**: observations are assimilated as soon as they become available, and only past and current observations can be used to update the model state. This class of methods, which includes filtering approaches such as the Kalman Filter and its variants, propagates the state estimate forward in time and is therefore particularly well suited for real-time and operational applications (Piazzi *et al.*, 2021).

Sequential DA, also known as filtering, has gained significant attention because it provides a flexible framework for explicitly accounting for different sources of uncertainty (Gong *et al.*, 2023a). A key advantage is its ability to process data sequentially as it becomes available, under the assumption that observations can only influence future model

states (Piazzi *et al.*, 2018a). Additionally, sequential DA schemes satisfy the technical requirements for operational use; indeed, in sequential assimilation, whenever observations are available, the analysis step updates the first-guess model state by incorporating the observation vector (Piazzi *et al.*, 2021). If every degree of freedom were observed, this vector would coincide with the state vector. In practice, however, the number of observations is typically much smaller than the number of state variables. To bridge this gap, the observation operator $\mathbf{H}$ is introduced, mapping the model state space to the observation space. This operator generates model equivalents of the observations, thereby enabling their assimilation.

The analysis problem can be mathematically formalized within a Bayesian framework. Bayesian probability allows uncertainties to be represented by probability density functions (PDFs) describing error statistics (Evensen *et al.*, 2022). The goal is then to estimate the posterior distribution of the unknown process $\boldsymbol{x}$, conditioned on observations $\boldsymbol{y}$, expressed as:

$$p(\boldsymbol{x}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})}{p(\boldsymbol{y})}. \tag{2.3}$$

This approach does not attempt to invert the observation process (i.e., reconstructing the model from the data), but instead updates the prior distribution using the newly available information (Evensen *et al.*, 2022).

For the application of this Bayesian approach, some assumptions are typically made:

- The dynamical model is a first-order Markov process, i.e., the state $\boldsymbol{x}_{k+1}$ depends only on $\boldsymbol{x}_k$.
- Measurements are independent in time and mutually independent, with uncorrelated errors.

Given these assumptions, information about the system state can be recursively updated from observations using Bayes' law. Suppose the posterior PDF at time $k$, $p(\boldsymbol{x}_k|\boldsymbol{y}_k)$, is known. Using the system model, the prior PDF of the current state can be obtained through the Chapman–Kolmogorov equation:

$$p(\boldsymbol{x}_{k+1}|\boldsymbol{y}_k) = \int p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k)\, p(\boldsymbol{x}_k|\boldsymbol{y}_k)\, d\boldsymbol{x}_k. \tag{2.4}$$

In the **prediction step**, the forecast PDF is obtained from the transition model. When a measurement $\boldsymbol{y}_k$ becomes available, it is used to update the forecast via Bayes' rule:

$$p(\boldsymbol{x}_{k+1}|\boldsymbol{y}_{k+1}) \propto p(\boldsymbol{y}_{k+1}|\boldsymbol{x}_{k+1})\, p(\boldsymbol{x}_{k+1}|\boldsymbol{y}_k). \tag{2.5}$$

In the **update step**, the prior PDF is combined with the likelihood of the new observation to obtain the posterior distribution of the state. This recursive formulation constitutes the Bayesian solution to the state estimation problem (Carrassi *et al.*, 2018). The goal of

the analysis is to find the maximum of the conditional PDF $p(\boldsymbol{x}_k|\boldsymbol{y}_k)$, given the observations.

In high-dimensional systems, however, evaluating Bayes' law directly is infeasible due to the multidimensional integration of the forecast density. To overcome this, common approaches target either the mean (minimum squared-error estimator) or the mode (maximum a posteriori estimator). Further assuming Gaussian PDFs yields computationally efficient solutions, as the state distribution can then be fully described by its mean and covariance matrix (Evensen *et al.*, 2022).

Historically, DA in snow hydrology has often relied on relatively simple heuristic approaches. Despite their limitations, these approaches have nonetheless led to noticeable model improvements (Ismail-Zadeh *et al.*, 2023). Commonly used approaches include direct insertion and rule-based adjustments. A representative example is the nudging, or Newtonian relaxation scheme, employed in the DA component of S3M (Avanzi *et al.*, 2022a). In this method, the model state is gradually adjusted towards the observations by relaxing the residual between the simulated and observed variables (Boni *et al.*, 2010). Similarly, Viallon-Galinier *et al.* (2020) reduced model errors through direct insertion of observed snow stratigraphy into a multilayer snow model.However, while these empirical approaches are computationally efficient and can improve model performance, they rely on strong assumptions such as that the model is the sole source of error and that observations are either perfect or can be incorporated deterministically. These assumptions often limit robustness and can even degrade performance, highlighting the need for methods with a rigorous probabilistic framework. This has motivated the development of advanced ensemble- and particle-based DA methods capable of approximating forecast and posterior probability distributions in high-dimensional systems. Indeed, with advances in computational capacity and the increasing availability of satellite products (Houser *et al.*, 2012; Aalstad *et al.*, 2018; Deschamps-Berger *et al.*, 2023a; Lievens *et al.*, 2022; Mazzolini *et al.*, 2024), DA in snow hydrology is progressively converging towards statistically grounded techniques (Largeron *et al.*, 2020; Girotto *et al.*, 2020; Alonso-González *et al.*, 2022).

A review by Helmert *et al.* (2018) highlighted the most widely assimilated categories of data: snow-covered area (SCA) (Margulis *et al.*, 2016), snow depth (Girotto *et al.*, 2024) and SWE (Magnusson *et al.*, 2014). Among state variables, snow depth and SWE remain the most frequently assimilated. At the same time, the scientific community is increasingly investigating the potential of assimilating thermal infrared data and active/passive microwave radar products, which can be transformed into snow-related state variables (Alonso-González *et al.*, 2023a; Cluzet *et al.*, 2024; Largeron *et al.*, 2020).

Building on Bayesian and sequential frameworks, recent developments have focused on statistical assimilation schemes, notably the Ensemble Kalman Filter (EnKF) and the Particle Filter (PF). A detailed description of these methods is provided in the appendix. Smoothers are not discussed here, since the focus is on sequential DA in near-real-time scenarios.

## 2.3 A new modelling approach for snow hydrology: AI enhanced solutions

The aforementioned ensemble-based (Monte Carlo) DA techniques, while powerful, face specific challenges in operational snow hydrology, where rapid turnaround is required for timely forecasts and monitoring updates. In these settings, running large ensembles of model predictions can substantially increase computational demands, particularly across the extensive and complex domains typical of mountain hydrology (Evensen *et al.*, 2022). Beyond computational considerations, ensemble methods are sensitive to model and observation uncertainties and may struggle to fully capture the non-linear and spatially heterogeneous nature of snow processes (Piazzi *et al.*, 2018b; Odry *et al.*, 2022). Their effective use is further complicated by the growing volume of remote sensing observations and by the scarcity of dedicated, well-funded snow-focused remote sensing missions suitable for large-scale operational applications (Gascoin *et al.*, 2024).

These factors can limit the implementation of ensemble-based techniques in real-time applications (Sabzipour *et al.*, 2023), highlighting the need for efficient algorithms and robust computing infrastructure to deliver accurate and timely results. In this context, is worth mentioning the work of Oberrauch *et al.* (2024) which addresses the implementation of a particle-based DA scheme for large-scale, fully distributed, near real-time snow modelling, effectively balancing computational feasibility with operational efficiency.

Moreover, the increasing availability of large and diverse datasets—from in-situ measurements to high-resolution satellite products—introduces additional challenges, requiring complex preprocessing, quality control, and harmonization procedures (Mhangara & Mapurisa, 2019). These steps are computationally intensive and can become major bottlenecks in both research and operational environments. Reducing these computational requirements would enable shorter turnaround times and/or increased model complexity without additional computational burden.

Nevertheless, accurately assimilating snowpack states remains challenging for several reasons. First, the common assumption of unbiased, normally distributed model-state errors is often violated for cumulative variables such as snow depth (Dunmire *et al.*, 2026a). Second, the number of state variables can vary in space and time, particularly when multiple snow layers are considered, and many relevant state variables remain unresolved in the model (Bakketun *et al.*, 2026). Third, the inherent nonlinearity and spatial heterogeneity of snow processes complicate the representation of uncertainties, limiting the effectiveness of traditional ensemble-based approaches under certain conditions (Terzago *et al.*, 2020). Finally, the sparse spatial coverage of snow depth observations relative to model grid scales challenges the propagation of information within DA frameworks, necessitating sophisticated methods to distribute observational information across unobserved areas (Alonso-González *et al.*, 2023b).

These limitations underscore the need for new modelling strategies that combine the

physical interpretability of traditional snow models with the flexibility and computational efficiency required to exploit large, heterogeneous datasets.

In this context, AI offers a powerful complementary framework (Gacu *et al.*, 2025); the term AI refers to a broad family of technologies that enable machines to mimic human cognitive functions such as problem-solving, reasoning, learning, and generalization from past experience (Hunt, 2014). Since its introduction into Earth sciences in the late 20th century (Sun *et al.*, 2022), AI has attracted growing interest for its ability to match, and in some cases outperform, traditional physics-based models across a wide range of applications (Nema & Nagashree, 2024). On the other hand, it is important to note that such performance gains are typically observed in data-rich contexts; in data-scarce domains, such as groundwater studies, AI models may struggle to provide reliable predictions, requiring scientists to combine AI with traditional physically-based approaches or develop novel hybrid methodologies (Borzì, 2025).

AI has prompted a paradigm shift in Earth system modelling, moving from conventional process-based approaches, rooted in physical understanding, to data-driven approaches (Maity *et al.*, 2024). In this paradigm, data are placed at the core of the modelling process: models are trained directly on observations to uncover patterns and relationships without explicit reliance on predefined physical laws.

In the broader context of catchment hydrology, the potential of AI is well established (De la Fuente *et al.*, 2024; Wang *et al.*, 2022b; Di Nunno *et al.*, 2023; Chu *et al.*, 2021), with early applications dating back to the late 1990s and early 2000s (Buch *et al.*, 1993; Smith & Eli, 1995; Toth *et al.*, 2000). However, progress in snow hydrology has been comparatively slower, largely due to the limited availability and sparse distribution of observational data. This gap is now narrowing, thanks to recent advancements in satellite remote sensing, large-scale data repositories, and ML techniques (Liu *et al.*, 2020a; Chen *et al.*, 2021a; Dunmire *et al.*, 2024; Moghaddasi *et al.*, 2025). These developments are paving the way for new AI-based modelling approaches capable of efficiently managing vast datasets, assimilating diverse sources of information, and improving the representation of snow processes across scales.

Looking forward, there is broad consensus in the hydrological community that AI will play a transformative role in shaping the future of the discipline. Indeed, Shen *et al.* (2021) emphasize that ML applications should extend beyond niche use cases to encompass uncertainty quantification, process understanding, and integration with physical models. They highlight the need for close collaboration between hydrologists and data scientists to address key challenges of interpretability, scalability, and robustness. Pursuing these directions would not only improve predictive accuracy but also foster open and reproducible science, promote interdisciplinary collaboration, and generate actionable insights for water-related challenges under climate and societal change (Shen, 2018; van Hateren *et al.*, 2023).

Despite these advances, important challenges remain. AI-based approaches are often criticized for their "black-box" nature, which hinders interpretability and limits trust in operational settings. Moreover, purely data-driven models can struggle under non-stationary

conditions, such as those imposed by climate change, or when extrapolating beyond the conditions observed during training (Naghibi *et al.*, 2017; Reichstein *et al.*, 2019a). To overcome these limitations, recent research has increasingly turned toward hybrid modelling frameworks that integrate physical knowledge with ML. These approaches not only enhance predictive skill and interpretability but also provide a promising pathway toward next-generation snow hydrological models capable of efficient data assimilation, large-scale automation, and physically consistent learning. A description of machine learning (ML) and deep learning (DL) concepts is also provided in the appendix.

## 2.4 Research questions

In recent years, hydrological sciences have witnessed a rapid expansion of AI applications. Snow hydrology, in particular, is benefiting from the increasing availability of high-resolution snow observations and the growing interest in data-driven approaches (Steele *et al.*, 2024). Despite this momentum, the literature still lacks well-documented case studies, especially those targeting operational settings where robustness, computational cost, and transferability are essential (Slater *et al.*, 2025) .

This gap motivates the investigation of how AI can be effectively integrated with traditional physics-based models to enhance snow-related hydrological predictions (Elyoussfi *et al.*, 2025b). Rather than replacing physically based models, current research increasingly points toward hybrid approaches in which AI complements physics by improving computational efficiency, leveraging large datasets, and supporting operational decision-making (Reichstein *et al.*, 2019b).

Building upon this perspective, the overarching goal of this research is to evaluate whether and how AI can strengthen operational snow hydrology workflows. Specifically, the work aims to:

- Assess whether AI can reduce computational and human-resource costs, particularly in workflows that rely on large, high-dimensional meteorological and snow datasets.
- Evaluate the ability of AI models to handle iterative, non-generalized, and site-specific tasks, ensuring that their performance is robust, interpretable, and transferable across different hydrological domains.
- Identify the benefits and limitations of coupling AI-based emulators with physics-based models, with a focus on their potential to accelerate operational applications such as snow monitoring and forecasting.

To address these objectives, the research focuses on two core components that are fundamental for operational snow hydrology: data preprocessing and data assimilation. These topics do not aim to be exhaustive; rather, they serve as key benchmarks illustrating how physically based modelling can be reinforced by AI without sacrificing physical consistency.

The following chapters present three case studies that exemplify this integrated perspective:

- A machine-learning-based quality control (QA/QC) procedure for snow and meteorological datasets.
- A neural-network-based DA emulator in a one-dimensional (1D) framework.
- An extension of the DA emulator to a fully two-dimensional (2D) domain, bridging physics-based modelling with AI tools for operational snow hydrology.

Collectively, these case studies highlight the potential for enhanced AI-based solutions within snow hydrological modelling frameworks and contribute to advancing AI-physics hybridization in the field.

# Chapter 3

# Machine learning-based algorithm for QA/QC

When working with AI, the performance of a model strongly depends on the quality of the input data (Mohammed *et al.*, 2025). A substantial portion of the time required to develop such algorithms is therefore devoted to data processing and quality control. This consideration applies not only to AI-based systems but also to hydrological and snow-hydrological models in general (Pagano *et al.*, 2014).

In the context of snow modelling, noise sources in high-resolution snow-depth data severely limit their automatic use, both in DA and model evaluation, ultimately affecting water management, hydrological forecasting, and emergency preparedness. In particular, snow-depth measurements obtained from ultrasonic sensors are often affected by snow–grass ambiguity and random noise. At the same time, the growing volume of available data makes non-replicable, time-consuming, and error-prone visual screening procedures increasingly impractical.

AI itself, however, can serve as a powerful tool to accelerate and enhance data quality improvement processes (Yasin & Khorsheed, 2025). In this chapter, I present the results of a quality assessment and quality-checking procedure developed using a Random forest algorithm. The outcomes of this work were published in *The Cryosphere* (2023) in the paper titled "A Random forest approach to quality-checking automatic snow-depth sensor measurements" (Blandini *et al.*, 2023). Here, these results are reorganized and contextualized to align with the main thesis presented in this manuscript.

## 3.1 General context

Environmental technologies have made it possible to easily gain new information, even in real time, with an increasing quantity of data made available from remote sensing, and more sophisticated ground sensors. However, high resolution data of snow come with a variety

of noise sources that make Quality Assurance and Quality Control (QA/QC) indispensable to use such data in snowpack modelling (Avanzi *et al.*, 2014; Bavay & Egger, 2014b). A recurring case in this context are snow-depth data, with two frequent noise categories: (1) snow vs. grass ambiguity, due to snow depth ultrasonic sensors detecting not only snow cover but also plant/grass growth in spring and summer (Vitasse *et al.*, 2017), and (2) random errors (e.g., spikes, anomalous data points that protrude above or below an interpolated surface).

Traditionally, in the field of snow cover and snow depth monitoring, QA/QC procedures have been carried out by visual inspection, heavily depending on subjective expert knowledge (Robinson, 1989). While expert-knowledge QA/QC is arguably the most reliable approach to data processing, these practices are not easily reproducible or transferable, and highly time-consuming (Fiebrich *et al.*, 2010). In this context, QA/QC with regard to grass detection is often based on static climatological or minimum-snow-depth thresholds, while random errors are generally detected based on maximum-snow-depth thresholds or criteria based on signal variance (Avanzi *et al.*, 2014). An exception in this regard is the approach implemented by the Swiss MeteoIO algorithm for grass detection, which however requires information on surface snow temperature, ground surface temperature and radiation (Bavay & Egger, 2014b).

In view of this knowledge gap, Jones *et al.* (2018) highlight the burden of subjectivity that may affect overall data quality and comparability, stressing how even expert scientists are not immune to mistakes, especially if performing recurrent unguided quality checking procedures. As explained by Schmidt *et al.* (2018), automatic environment data quality control literature is still fragmented, with heterogeneous applications. It is clear then the necessity for a quality checking procedure that ought to be defined through common and iterable guidelines to guarantee repeatability and consistency (Jones *et al.*, 2018).

Considering the ever growing volume of data and the limitations arising from traditional QA/QC procedures, here I follow intuitions from Schmidt *et al.* (2018) and propose the use of ML to automatically quality check high-resolution snow-depth sensor data from ultrasonic sensors. The choice of ML was driven by its efficiency in dealing with big data sets and as a valid reinforcement of traditional analytic tools (Ferreira *et al.*, 2019). Moreover, ML techniques may also be able to handle different data formats more easily that traditional statistical tools, while they better deal with combination of features that are a-priori unknown to the developer (Zhong *et al.*, 2021). The development of this QA/QC algorithm was guided by three main research questions: (i) what is the accuracy of a Random forest classifier algorithm in automatically performing QA/QC of near-surface snow depth observations? (ii) is the approach transferable to untested regions and, if so, what is the potential drop in performance? (iii) how do meteorological conditions influence model performance and the Random-forest decision process?

## 3.2 Data

To develop, test and validate the ML-based QA/QC algorithm, two different datasets were used: a dataset with 18 years of already classified snow-depth data at 43 locations from Aosta Valley, which was used as intensive study domain to develop the algorithm, and 3 years of data from 27 snow-depth sensors across the rest of Italy, which were used to test the generalization and transferability of the algorithm in time and space.

### 3.2.1 Training and test dataset

Aosta Valley is located in the northwest Italian Alps (Figure 3.1). The region includes some of the highest peaks in the Alps (Mont Blanc - 4808 m asl, Monte Rosa - 4634 m asl, Mount Cervino - 4478 m asl, and Gran Paradiso - 4061 m asl). While some of these peaks, such as Mont Blanc, are inner-Alpine, others, such as Monte Rosa and Gran Paradiso, overlook the Pianura Padana and thus are more exposed to maritime conditions (Sturm & Liston, 2021b). This generates marked precipitation-regime discrepancies, hence climatic differences and different snow regimes, across the region (Avanzi *et al.*, 2021b). Despite being a comparatively small mountainous region, the climate variability and the abundance of already classified snow measurements were the reasons that led us to the choice of this area as training domain.

The Aosta-valley data-set consists of hourly snow depth measurements from 43 ultrasonic sensors (precision in the order of a few cm, Figure 3.1, (Ryan *et al.*, 2008)), provided by the regional Functional Center of the National Civil Protection Department. The period of record goes from August 2003 to September 2021, thus covering a variety of snow seasons across 18 years of data (Avanzi *et al.*, 2023). The elevation range of these sensors goes from 545 m to 2842 m asl, with an average elevation of 2007 m asl that is representative of average elevations across the Italian Alps where the bulk of sensors are located (Avanzi *et al.*, 2021b).

Each data record in this dataset was subject to visual screening by expert hydrologic forecasters during periodical QA/QC manual data processing, with the goal of discriminating random and systematic errors from actual snow depth measurements. This manual processing follows well established practice in the field, including crosschecking with concurrent weather (e.g., air temperature, precipitation, relative humidity) and nearby sensors (Avanzi *et al.*, 2014, 2020b). As a result, each datapoint came with a quality code (Table 3.1): data with code 0 or 1 are valid snow cover data; codes 2 or 4 are for missing data reconstructed from trends or aggregated from different time resolutions; codes 8 and 16 are grass or bare ground; code 32 denotes reconstructed grass data; codes 64 to 256 denote a variety of flags for random and instrumental errors; codes 1024-1032 refer to data classified as invalid after a preliminary procedure based on fixed thresholds (introduced in 2018). While the dataset includes some reconstructed data, these are only 0.03% of the whole dataset, which means they do not affect the analyses.

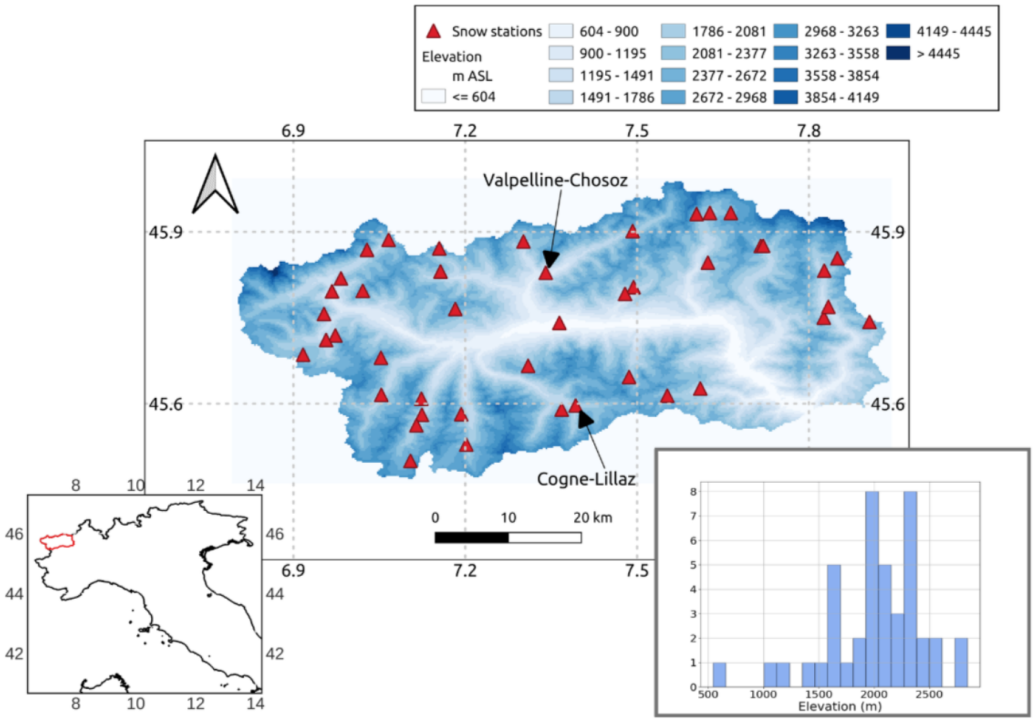


**Figure 3.1:** *Considered snow-depth sensor data across Aosta Valley (see the bottom-left corner for the location of this study region in Italy). The two snow-depth sensors of Valpelline-Chosoz and Cogne-Lillaz were used in section 4.3. The histogram in the bottom-right corner of this figure reports the frequency distribution of the elevation of the Aosta-Valley sensors.*

In this work, the number of classes was reduced to 3 by aggregation: correct snow depth, identified with code "0", grass or bare ground, identified with code "1", and random errors, identified with code "2".

### 3.2.2 Validation dataset

The validation dataset across the rest of Italy comprises hourly data from 27 ultrasonic depth sensors, randomly chosen among the ∼300 Italian automatic snow-depth sensors available outside Aosta Valley. These 27 snow-depth sensors were chosen based on a geographical-diversity criterion to guarantee heterogeneity, especially with regard to the Aosta Valley data (Figure 3.2). This second dataset include data from 3 years: 2018, 2020 and 2022, which were chosen due to their significantly different accumulation patterns (deep snowpacks in 2018, somewhat average snowpacks in 2020, and extraordinarily low

| Code | Data type | % of total | Code | Data type |
|---|---|---|---|---|
| 0 or 1 | Valid snow data | 46.43% | | |
| 2 | Qualitatively (aggregated) valid snow data | < 0.01% | 0 | Snow data |
| 4 | Reconstructed missing snow data | 0.03% | | |
| 8 or 16 | Grass/bare ground data | 49.20 % | 1 | Grass/bare ground data |
| 32 | Reconstructed missing grass/bare ground data | 0.01% | | |
| 64–72 | Random error, invalid data | 3.95% | | |
| 128 | Calibration error | 0.02% | 2 | Errors |
| 256 | Maintenance error | 0.03% | | |
| 1024–1032 | Rejected data based on climatological thresholds | 0.34% | | |

**Table 3.1:** *Snow depth data classification system developed by the Functional Center of Aosta Valley.*

snowpacks in 2022, see Avanzi *et al.* (2023)). No prior processing was available for these data, thus I proceeded with a manual classification to assign codes as in Table 3.1. The procedure included visual screening, checks on seasonality to detect snow vs. grass, and a comparison with measurements from nearby sensors (Avanzi *et al.*, 2014, 2020b).

Italy (approximately 301 $\times$ 103 km$^2$) is a topographically and climatically complex region. Its main mountain chains, the Alps and the Apennines, are among the highest peaks in Europe. Partially snow-dominated regions like the Po river basin or the central Apennines have high socio-economical relevance (Group, 2021). The Italian climate presents a considerable variability from north to south. According to the Köppen-Geiger climate classification (Beck *et al.*, 2018), in the Alps the climate is humid and continental. Central Italy, alongside the Apennines chain, is characterized by a warm, temperate, Mediterranean climate with dry, warm summers and cool, wet winters. In Southern Italy, where the climate is still a warm temperate, Mediterranean climate, winters are mild, with higher humidity and higher temperature during summer. Concerning snow-cover distribution, accumulation across the Alps is generally higher and more persistent than across the Apennines, where it is spatially more limited and more variable from one season to the others (Avanzi *et al.*, 2023). Rivers draining from the snow-dominated Alps and a handful of basins draining from central Apennines host the vast majority of snow water resources across the Italian territory. In particular, the Alpine water basins host nearly 87% of Italian snow. The central Apennines, accumulate about 5% of the national mean winter SWE , leaving the remaining 8% – 9% scattered across the remaining basins over the territory. Intraseasonal melt, expected in a Mediterranean region, is a common feature in sites where cold-alpine and maritime snow types coexist like the Apennines (Avanzi *et al.*, 2023).

## 3.3 Methodology

Among all ML approaches, I chose Random forest due to its benchmarking nature as well as its simplicity of use (Tyralis *et al.*, 2019), as proven by an increasing number of studies

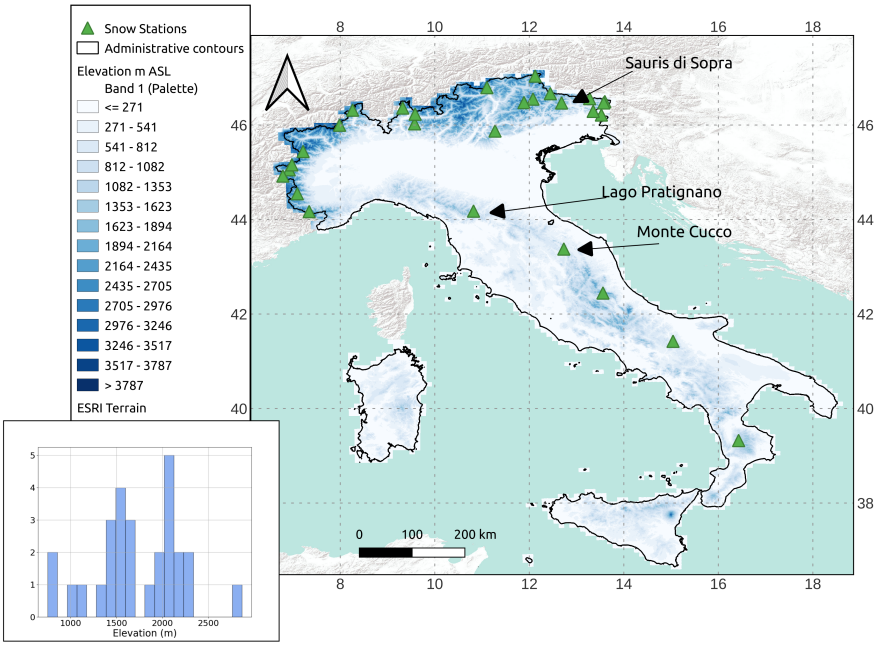


**Figure 3.2:** *Considered snow-depth sensor data across the rest of Italy. Bottom left: frequency distribution of the elevation of these sensors. Three black arrows indicate the location of three snow-depth sensor used in section 3.4.5.*

proving the effectiveness of Random forest as classifier or regressor algorithm. For instance, Desai & Ouarda (2021) developed a flood frequency analysis based on Random forest, which proved to be equally reliable but more efficient than more complex models; Park *et al.* (2020) developed a Random forest classifier for sea ice, using Sentinel-1 data; Random forest proved to be efficient in big data environments (Liu, 2014); recently, Ponziani *et al.* (2023) proved the efficiency of Random forest over other ML algorithms, developing a predictive model for debris flows that could be experimentally implemented in the existing early warning system of the Aosta valley. In the context of snow data, Meloche *et al.* (2022) proved the ability of a Random forest algorithm to predict snow depth distribution from topographic parameters with a root mean square error of 8 cm (23%) in western Nunavut, Canada. In particular, the algorithm object of the present study is a Random forest classifier, an ensemble classifier based on bootstrap aggregation and random features selection.

### 3.3.1 Random forest algorithm

A Random forest is an ensemble of decorrelated decision trees that are let growing and voting for the most popular class (Breiman, 2001). The growth of each tree in the ensemble

is governed by randomness, proved to be a performance enhancer. Randomness is given by two randomization principles: bagging and random feature selection. According to the bagging principle, a large number of relatively uncorrelated trees, each built using a split sample of $n$ dimensions retrieved from the entire training dataset of size $m$, operate as a committee; this ensemble is proven to outperform any of the individual constituent trees. Therefore, the class definition, made by averaging the scores of each tree, is mildly affected by the weight of misclassification done by less performant trees. Furthermore, instead of splitting a node searching the most important feature (i.e., predictor), a Random forest uses the best one among a random subset of features, performing random feature selection, thus increasing the performances. Randomness injection minimizes correlation across trees and reduces variance and overfitting, increasing stability (Breiman, 2001). The algorithm was implemented using Scikit-learn Version 0.20.1, a Python software programming platform, using the class RandomforestClassifier.

### 3.3.2 Training strategies

To train the Random forest, the Aosta Valley classified dataset was used. Based on data frequency (Figure 3.3), this is a typical imbalanced dataset where class distribution is skewed or biased towards one or few classes in the training dataset (Kuhn *et al.*, 2013).

In this framework, data belong either to majority or minority classes. The majority classes are the classes with a larger number of observations, while the minority classes are those with comparatively few observations. In this case, the number of data classified as random errors (code 2) is significantly lower than the number of data from category 0 (snow height) or category 1 (grass/bare ground). Thus, classes 0 and 1 were defined as majority class, while class 2 was defined as minority class.

Class imbalance can severely affect the classification performance (Ganganwar, 2012; Ramyachitra & Manikandan, 2014), and therefore requires a pre-processing strategy. To this end, acknowledging the work of Ponziani *et al.* (2023) in which no clear evidence of outperformance of any such strategy was shown, an oversampling of the minority class was performed by selecting examples to be duplicated and then added to the training dataset; to this end it was used the class RandomOversampler from the package Imbalanced-learn version 0.8.1. To decrease the computational effort that may have stem from this oversampling procedure (Branco *et al.*, 2016), a representative sample of $1.3\times10^6$ measurements was taken from the entire dataset prior oversampling of the minority class for Random forest training. This sample was proven to be representative of the entire dataset distribution (approximately $5.5\times10^6$ datapoints), by performing a two-sample Kolmogorov-Smirnov test, with a significance level equal to 0.05.

After the oversampling procedure, a sample of $1.9\times10^6$ over-sampled measurements (including both the majority and the oversampled minority classes) was used to train the Random forest. From the remaining, not oversampled dataset, an independent test sample of $4.8\times10^5$ measurements was randomly selected. As a result, a train-test split share of

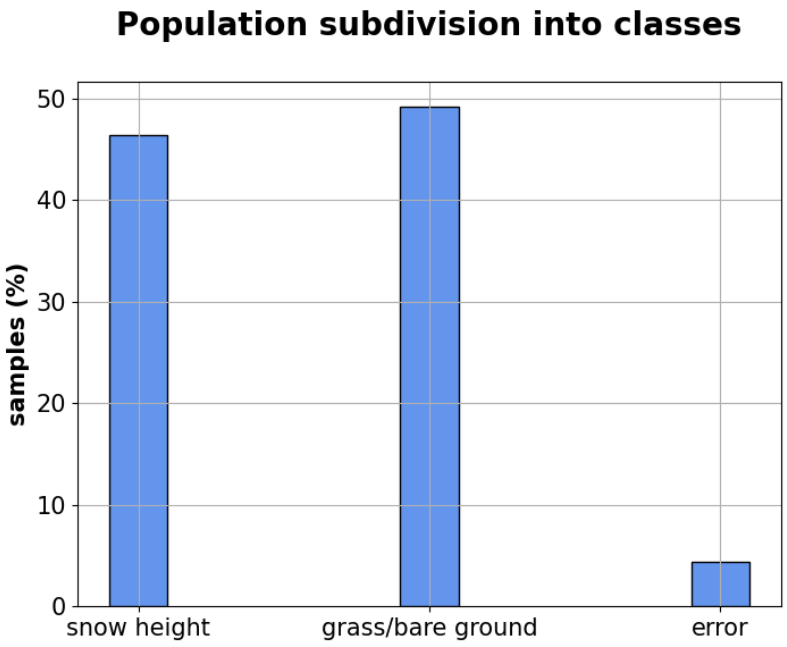


**Figure 3.3:** *Aosta-Valley data subdivision into classes.*

80% training 20 % test was used, in agreement with current standards in ML problems (Harvey & Sotardi, 2018).

When dealing with an imbalance classification, standard evaluation criteria focusing on the most frequent classes may lead to misleading conclusions, because they are insensitive to skewed domains (Branco *et al.*, 2016). For example, accuracy, which is defined as the number of correct predictions over total number of predictions and is a frequently used metrics for classification problems, underestimates the importance of the least represented classes when compared with the majority classes, as it does not take into account data distribution. Adequate metrics need to be used not only for model validation, but also for model selection, given that accuracy scores may ignore the difference between types of misclassification errors, as they seek to minimize the overall error. A good metric for imbalance classification must consider overall data distribution, giving at least same importance to misclassification in both majority and minority classes.

In this paper, it was thus used the F-measure (Van Rijsbergen, 1979), that is, the harmonic mean of precision (measure of exactness), defined as the number of true positives divided by the total number of positive predictions, and recall (measure of completeness), defined as the percentage of data samples that a ML model correctly identifies as belonging to a class of interest out of the total samples for that class.

The harmonic mean is the reciprocal of the arithmetic mean and tends to mitigate the

impact of large outliers, while aggravating the impact of small ones since it tends strongly toward the least represented elements.

$F_\beta$ (the so-called F-measure) is defined as:

$$F_\beta = \frac{(1+\beta)^2 \times \text{recall} \times \text{precision}}{\beta^2 \times \text{recall} + \text{precision}} \tag{3.1}$$

It was set $\beta = 1$ to give equal importance to precision and recall.

As a reference, $F_\beta$ scores in the range of 0.8–0.9 can be considered optimal. Scores between 0.6 and 0.8 indicate good performance, whereas values below 0.5 are generally considered poor (Van Rijsbergen, 1979).

The metrics of precision and recall were used to characterize the performance of the Random forest for each class separately. Then macro-averages of both measures were computed to characterize the multi-class performance. A macro average is the arithmetic mean computed giving equal weight to all classes, and is used to evaluate the overall performance of the classifier.

The performances of the trained Random forest algorithm were tested on the 20% test dataset, using the model in prediction and comparing model's classification with that by the expert forecasters. Validation was also performed by applying the final algorithm on the 3 years of data from the rest of Italy (Section 2.2).

A collection of meteorological, topographic, and temporal variables, known to influence snow accumulation and melt, were chosen as candidate predictors (features) of the Random forest, thus mimicking the decision process made by experts when assigning a classification code. These features include snow-depth values themselves, elevation, aspect, concurrent air temperature, incoming shortwave radiation, total precipitation, wind speed, relative humidity, and the day of the year. Feature values were extracted for each datapoint in both the Aosta Valley and the rest-of-Italy samples, using available geographic information and weather maps operationally developed by CIMA Research Foundation (see Avanzi *et al.* (2021b) for Aosta Valley and Avanzi *et al.* (2023) for other Italian data). The datasets consist of spatially distributed meteorological fields derived from in situ observations, remote sensing products, and statistical spatialization, provided at hourly resolution. Operational checks are performed to detect and fill gaps caused by station malfunctions or transmission failures. Precipitation fields are generated using GRISO, a modified conditional merging method that combines gauge observations (approximately 1 station per 100 $km^2$) with radar data (Avanzi *et al.*, 2023). GRISO estimates the spatial covariance for each gauge and hour, dynamically adapting to radar information. The resulting precipitation fields, with a spatial resolution of 1 $km^2$, have a median root mean square error RMSE below 1 mm for 70 benchmark heavy precipitation events in Italy (2011–2014). Orographic corrections are applied using lapse rates derived from rain gauges and over 11,000 snow course measurements in the Aosta Valley. Air temperature, solar radiation, and relative humidity are derived from station observations and spatialized at approximately 1 km (temperature)

and 500 m (radiation and humidity). Temperature fields are generated via hourly linear regressions with elevation within meteorologically homogeneous regions, applied to a 200 m DEM originally derived from 20 m Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA) data. Static geographical data, such as glacier extent, are incorporated from version 6.0 of the Randolph Glacier Inventory.

A feature importance analysis was also performed. Importance was calculated using the attribute "feature importance" of the class Randomforestclassifier in sklearn.ensemble (Pedregosa *et al.*, 2011). The ranking is driven by each feature contribution to a decrease in impurity over trees.

A set of hyper-parameters were optimized through a combination of automatic, random searching and further manual tuning to reduce overfitting, yet ensuring good F1 score and reliable training time on Aosta Valley dataset. The parameters that were tuned in this work were the number of estimator (namely, the number of trees in the forest), the maximum depth (namely, the maximum number of levels in each decision tree), the minimum sample leaf (namely, minimum number of data points placed in a node before the node is split), and the minimum sample split (namely, minimum number of data point needed to split an internal node). Others default hyper-parameters were not modified.

In addition to the general training strategy above, a Random forest algorithm was also trained using the Valle D'Aosta dataset separately for each year, with 80% of the data used in training and then an out-of-bag validation with the remaining 20% of the same year data. The aim of this further test was to investigate the possible correlation between the performance of the classification by the Random forest algorithm and annual weather characteristics. For each year, the F1 score on the test sample was analyzed against annual mean values of features used for the classification, computing correlation factors.

Finally,the classification results was mapped as a function of feature values to shed light on the decision process taken by the Random forest in classifying snow vs. grass/ground vs. random errors and how they relate with the original classification by operational forecasters.

## 3.4 Results

### 3.4.1 Training and test performances: Aosta Valley

The macro-averaged F1 score of classification for the Aosta Valley dataset during testing phase was 0.96, with a precision value of 0.97 and a recall of 0.95 (Figure 3.4 panel a). In details, the Random forest scored 0.99 in both precision and recall for the classification of snow data. In the classification of grass/bare ground, recall was maximum (1), with precision of 0.99. Lower values were obtained in the classification of random errors, with recall of 0.86 and precision of 0.93, resulting in a F1 score of 0.89. Most of snow depth data and grass/bare ground data were correctly classified (45.94% /46.45 % and 49% / 49.19%), while a comparatively large sample of error data that was miss-classified as snow

(0.50% / 4.43% (Figure 3.4 panel b). Overall, the model resulted in being equally precise and robust in snow and grass/bare ground classification, while the precision and recall of the random-errors class were lower compared to the other two classes (F1 score for snow and grass/bare-ground classes of 0.99 and F1 score for random error classes of 0.89). As a whole, the model tested in Aosta Valley proved to be slightly more precise than robust (precision 0.97 , recall 0.95).

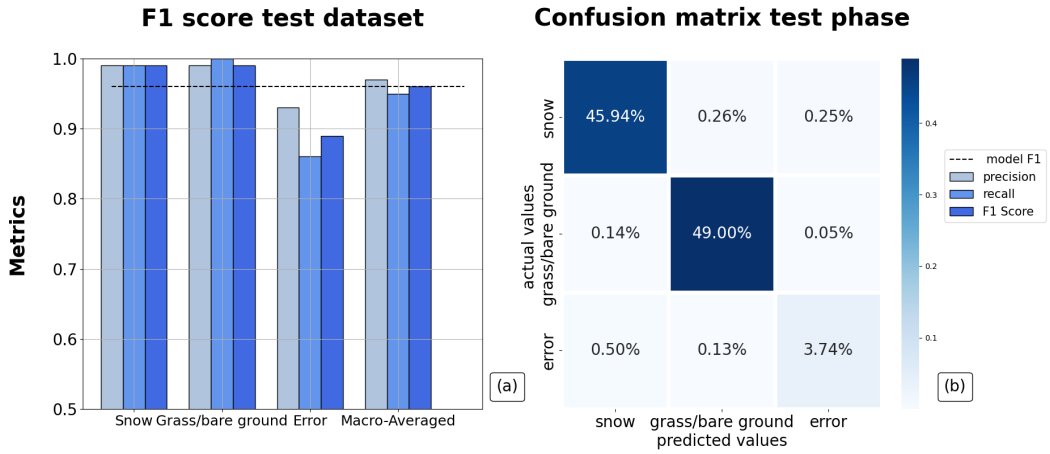


**Figure 3.4:** *Left: model performances in prediction mode for the test dataset in Aosta Valley. Each set of columns reports the values of precision, recall, and F1 score for the three classes, while the last group on the right shows the macro-averaged values referred to the Random forest performances as a whole. The black dashed line is a reference for the macro averaged F1 score of the Random forest. Right: confusion matrix.*

In order to identify recurring patterns in snow cover and grass cover classification during the hydrological year, results of the Random forest classifier for all data and hydrological years were visually screened. Figure 3.5 reports examples for two snow-depth sensor locations (October 2016 to September 2017), which were randomly selected from the entire pool of 43 snow-depth sensors along 18 years of Aosta Valley domain. Note that the samples used for the Random forest training were removed. The Random forest showed an expected tendency of to missclassify snow as grass/bare ground during transitional periods at the beginning and at the end of the snow season (Figure 3.5 pan A2), especially when snow-cover and grass height are comparable (Figure 3.5 pan B2 and B3). Moreover, The Random forest sometimes misinterprets settling during snow period.

### 3.4.2 Model configuration

The best set of parameters for the development of the Random forest resulted in a number of estimator equals to 500, a maximum depth of 40, a minimum sample leaf equals to 1, a

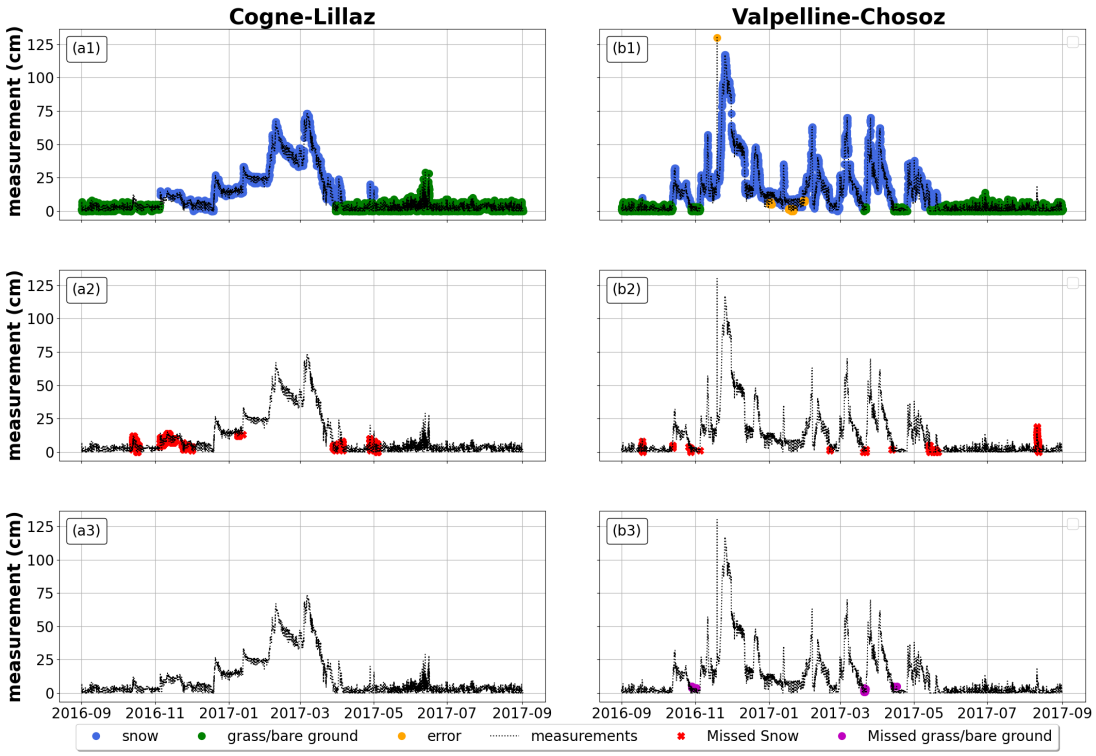


**Figure 3.5:** *Application of Random forest on 2 Aosta Valley snow-depth sensors locations from October 2016 to September 2017. The first row displays the samples of snow height, grass/bare ground and error correctly classified by the model. In blue correctly classified snow sample, in green correctly classified grass sample, in orange correctly classified error. The second row shows miss-classified snow height in red and the third row reports miss-classified grass/bare ground samples in purple. Data are referred to an hydrological year.*

minimum sample split equals to 2. The choice of the best set of features was initially driven by the F1 macro averaged obtained on the test set (Table 3.2, features combination sets from T1 to T7); then, training time was also considered as a discriminant (+10 minutes for T6 compared to T7). Hence, the set of feature selected as the best consisted in: the snow depth record measured by the snow-depth sensor, elevation, aspect, concurrent air temperature, incoming shortwave radiation, cumulative precipitation, relative humidity, and the day of the year to capture seasonality (Table 3.2, set T7). Regarding elevation and aspect, previous studies have shown that geographic location and elevation indeed contribute to improving machine-learning model performance (Bair *et al.*, 2018a).

Feature importance (Figure 3.6) suggested that measured snow-depth itself (regardless of whether is represents actual snow depth, grass, bare ground, or random errors) was

| Features | T1 | T2 | T3 | T4 | T5 | T6 | T7 |
|---|---|---|---|---|---|---|---|
| Snow height | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Aspect | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Elevation | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Air temperature | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Radiation | ✓ | | ✓ | ✓ | ✓ | | ✓ |
| Relative humidity | ✓ | ✓ | | ✓ | ✓ | | ✓ |
| Cumulative precipitation | ✓ | ✓ | ✓ | | ✓ | | ✓ |
| Day of the year | | | | | | ✓ | ✓ |
| Wind velocity | ✓ | | | | | | |
| **F1 score** | 0.84 | 0.87 | 0.85 | 0.86 | 0.93 | 0.95 | 0.96 |

**Table 3.2:** *F1 scores for a variety of tests used to identify the best feature combination for the Random forest Algorithm. T7 was then selected as best option in terms of features.*

the most important feature in the Random forest, followed by the day of the year, air temperature, and aspect. Radiation, relative humidity, and elevation scored similarly, while total precipitation was the least important feature. Feature importance results followed a somewhat intuitive ranking, similar to human perception. For example, the model gave high importance to snow depth, likely to replicate the concept of a "plausible range" of snow depth as opposed to grass, bare ground, or random errors. Seasonality (expressed as day of the year and air temperature) was the second most influencing factor, likely to mimic the concept of a "plausible" period for snow on the ground. Aspect and elevation were less influential, which is likely because of the comparatively small size of the Aosta Valley study region.

It is important to acknowledge that correlation among features and multi-collinearity are problematic for feature importance and interpretation in a Random forest. Features importance may spuriously decrease for features that are correlated with those selected as the most important (Strobl *et al.*, 2007). On the other hand, Hastie *et al.* (2009) point out that the predictive skill of the algorithm is relatively robust to correlations thanks to de-correlation factors involved in bootstrapping. Indeed, even low-importance features may drive the decision process of the algorithm (Avanzi *et al.*, 2019). In this case, it was chosen to use all the features after verifying the lack of correlations across features below -0.5 or above +0.5.

### 3.4.3 F1 correlation with annual climate

To assess potential relationships between the algorithm performance and the characteristics of the input data, we analysed the correlation between the annual F1 score and the mean annual values of each investigated feature.

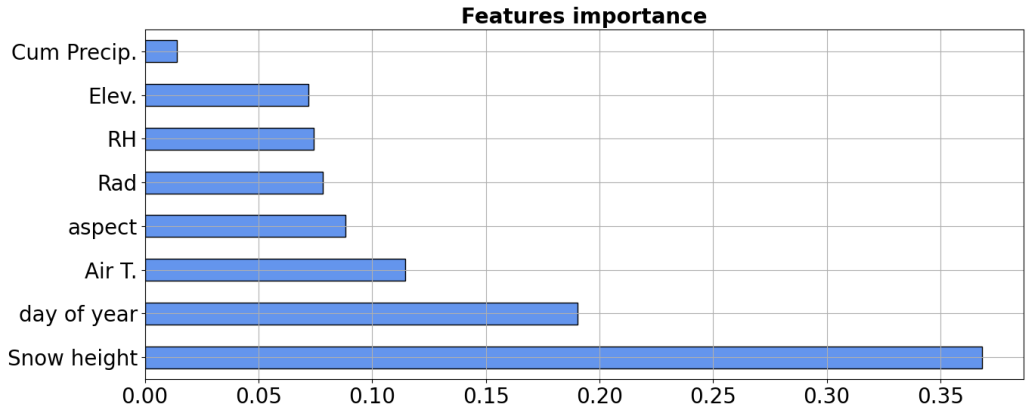


**Figure 3.6:** *Feature importance for the Random forest classification procedure in Aosta Valley. The dimensionless values, along the x-axis, sum up to 1; the higher the value the more important the feature is in the definition of the class. In particular: Cum, Precip. = cumulative precipitation, Elev = elevation, Rh = relative humidity, Rad = radiation, air T = air temperature*

Results showed low or negligible correlation coefficients between the annual mean feature values and the annual F1 score (Figure 3.7), after removing the training data points.

All correlation coefficients were statistically tested, and no significant correlations were found (p-values ranging between 0.21 and 0.40 for all features).

### 3.4.4 Mapping the decision process

Analysis of the Random forest decision process highlighted consistency with the classification procedure by expert forecasters, as well as agreement with the expected decision process behind the human made classification, despite a general underestimation of the number of random error samples (Figure 3.8).

The frequency of data classified as snow decreased with increasing temperature, as expected and in agreement with the original expert classification (Figure 3.8 pan. a1). Simultaneously, the frequency of data classified as grass/bare ground increased with temperature (Figure 3.8 pan. a2), again as expected due to progressive melt and snow disappearance as temperature increases. Regarding random errors, the Random forest underestimated their frequency up to 10°C, while automatic and human-made classifications were more comparable in frequency above that temperature threshold (Figure 3.8 pan. A3).

Considering the day of the year, most snow classifications occurred at the beginning and at the end of the calendar year (thus, in winter); this proved to be consistent between the Random forest and the human classification (Figure 3.8 pan. B1), with then a shift

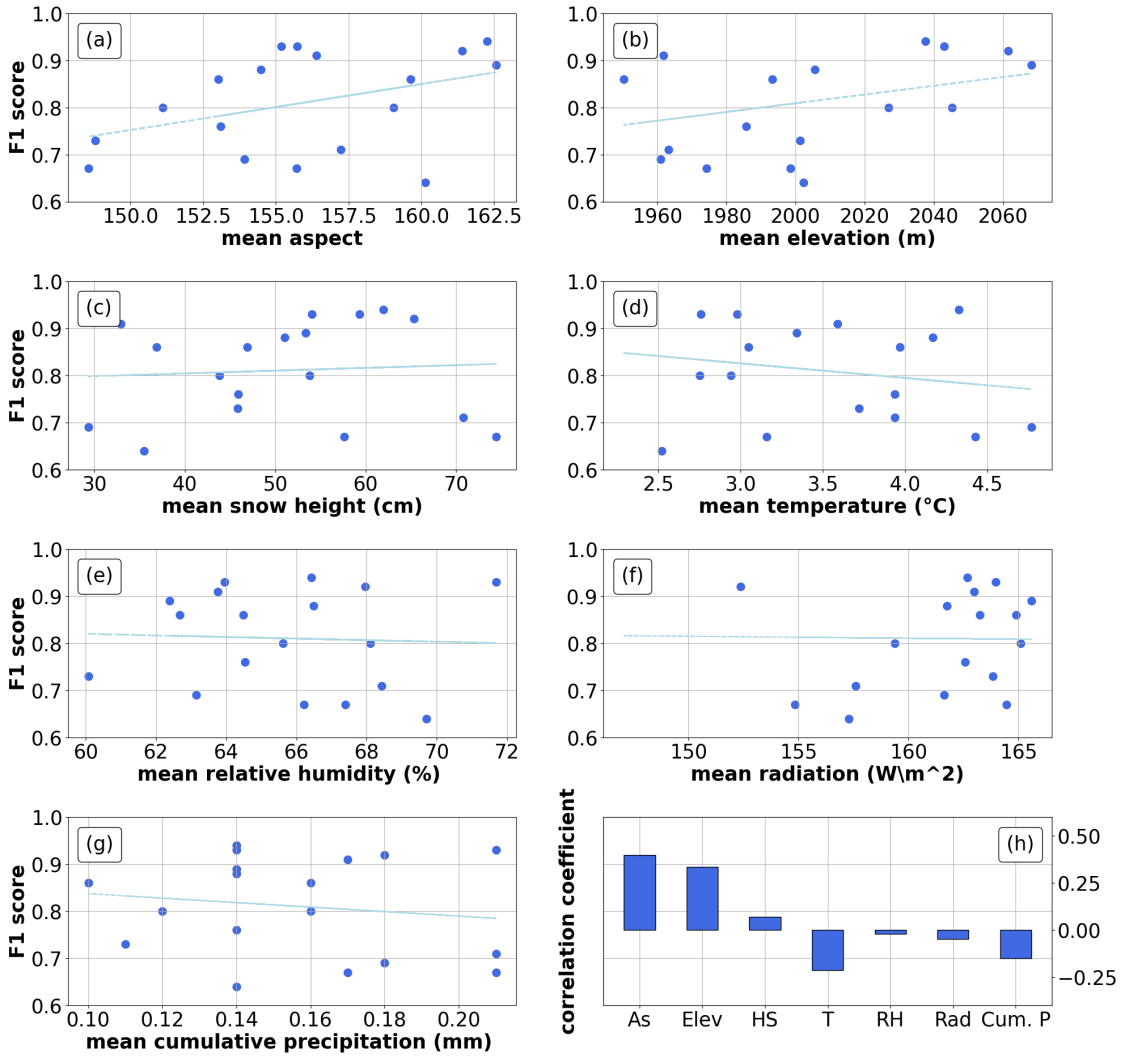


**Figure 3.7:** *Annual F1 score correlated with mean annual feature values. The y-axis reports the F1 score macro averaged for each year, while the x-axis shows the values of annual mean for each feature.The blue straight line indicates a linear regression. The last plot indicates the correlation coefficient between single features and F1 score.*

towards the grass/bare ground class in summer (Figure 3.8 pan. B2). Overall, the random error samples were underestimated throughout the year, especially in the first 150 days of the year (Figure 3.8 pan. B3).

The number of data classified as snow progressively increased with elevation (Figure 3.8 pan. c1), while the number of classification as grass/bare ground samples decreased with elevation (Figure 3.8 pan. c2), consistently between the Random forest and the original dataset. The frequency of random-error classifications generally matched the human classification, except for an underestimation around 2500 m (1% of miss-classified samples) (Figure 3.8 pan. c3).

When looking at aspect, both the automatic and human-made snow vs. ground-soil classification related to local climate. For example, they both classified more snow than grass across south slopes (between 50° and 251°), where precipitation is generally more abundant due to seasonal circulation from the Gulf of Genoa (Figure 3.8 pan.d1 vs. d2, see Rudari *et al.*, 2005; Brunetti *et al.*, 2009). On the other hand, grass classifications increased on north-facing slopes (from 250° to 351°), likely because these areas are exposed to naturally more humid conditions. Overall, the model underestimated the frequency of random-error classification along all aspects.

As for the other, less important features, they generally showed a negligible influence on the decision process. The only clear exception was relative humidity, since it was noted a progressive decrease of snow-classifications as relative humidity increases (Figure 3.8 pan. e1), coupled with an increase of grass/bare ground classification (Figure 3.8 pan. e2).

Finally, considering measured snow depth (by far the most important feature)), the model correctly classified all values above 400 cm as random errors, correctly matching the human classification (Figure 3.8 pan. h3). This is due to an instrumental limit given by the height of the sensor from the ground in this study region. Given that snow depth is the most important feature in driving the classification problem, a perfect match between model and human classifications was traceable (Figure 3.8 pan. h1 and h2).

### 3.4.5 Validation on the rest-of-Italy sample

The application of the Random forest on the 27 ultrasonic snow-depth sensors from the rest of Italy showed a surprising robustness in the classification of snow depth and grass/bare ground, with F1 score values between 0.93 and 0.96 across the three years. The performances of the Random forest on the classification of both snow samples and grass/bare-ground sample proved to be comparable to the ones already noted in Aosta Valley; a severe reduction of performance was registered in the detection of random errors, for which the F1 score was below 0.05 in every year. This is potentially due to the fact that the classification of this dataset by different expert from the one of the Aosta Valley may have introduced an inevitably different subjectivity; this is particularly impactful for random errors due to their smaller frequency in the sample (error sample frequency : 0.36 % in 2018, 0.92% in 2020, 0.52% in 2022).

Results of model application for two years at the exemplary station of Pratignano (Figure 3.10 panels a2 and b2) suggested a better performance for the model in case of higher snow depth. In other words, the model better distinguished snow from grass or bare

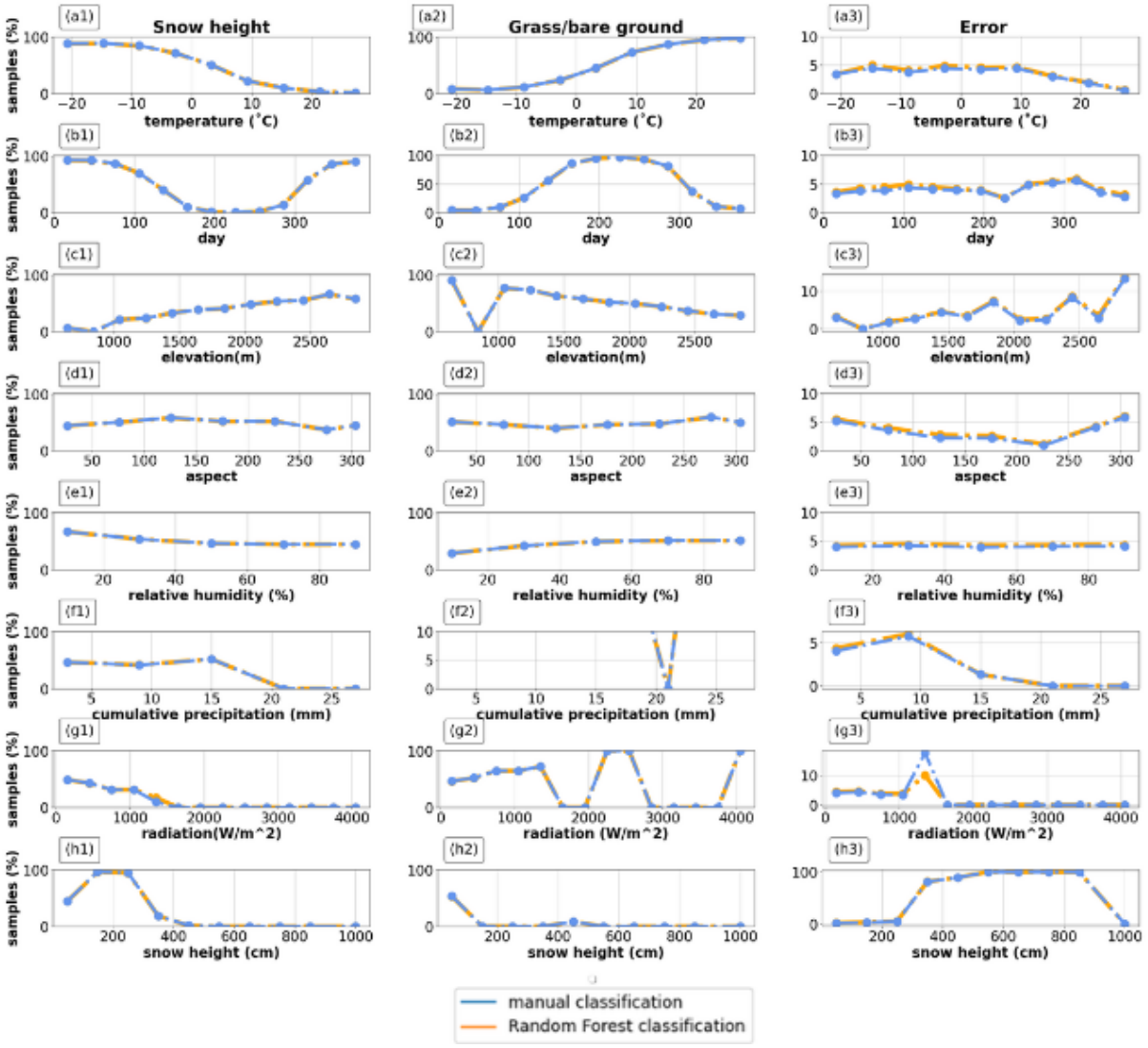


**Figure 3.8:** *Classification results as a function of features values: left are for snow classification, center are for grass/ground classification, right are for random-error classifications. Orange is the human made classification, blue is the classification performed by the Random forest. The x-axis reports feature values, while the y-axis reports the percentage of classification on the total. The plots refer to the test sample in Aosta valley, being it representative of the entire residual dataset. Data are normalized over the total sample size.*

ground when their heights were less commensurable, hence the slightly better performance in a year with higher snow depth (F1 score in 2018 : 0.95 for snow and 0.96 for grass/bare-ground; F1 score in 2022 : 0.93 for snow and 0.94 for grass/ bare-ground). This example also showed a recurring tendency to confounding snow and grass at the beginning and at the end of the season, as already noted in Aosta Valley. Considering grass classification (Figure 3.10 pan. a3 and b3), it was also found a tendency to missclassify snow and grass during periods of intraseasonal melt. Two other examples of application of the Random forest on exemplary sites can be found in the Appendix . A snow-depth sensor located in the Apennines (Figure E.1) and one located in Northeast Italy (Figure E.2) are displayed here to better portrait snow regimes and Random forest performances across Italy.

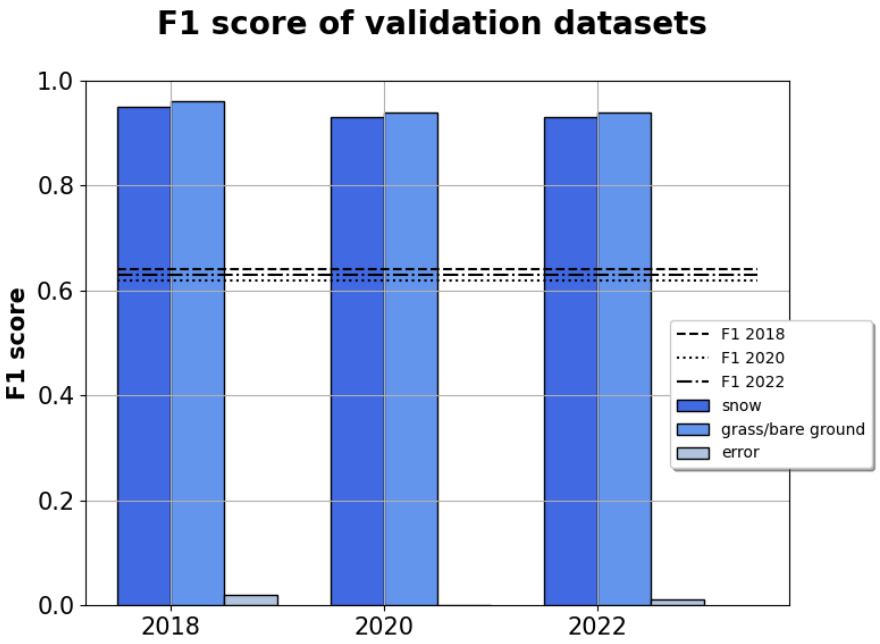


**Figure 3.9:** *Classification performance on the 27 stations across the rest of Italy. The columns grouped along x-axis are the F1 score for snow, grass/bare ground and random-error classes, respectively, subdivided by year. The y-axis reports the dimensionless values of each scoring metrics. The straight lines are the F1 score macro averaged for each year.*

## 3.5 Discussion of results

Due to the central role that snow plays in the global water cycle (Flanner *et al.*, 2011; Beniston *et al.*, 2018), snow measurements have proven to be essential in the development of

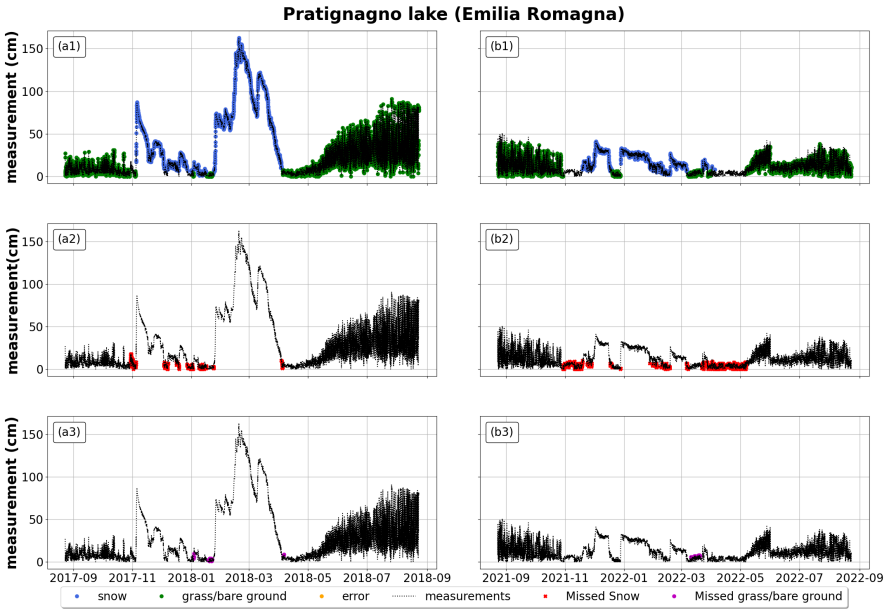


**Figure 3.10:** *Example of application of the Random forest on an Italian station (Lago Pratignano, Emilia Romagna). Left: October 2017 to September 2018; Right: October 2021 to September 2022. First row reports correct classification of snow, grass/bare ground, and random errors (blue for snow depth, green for grass/ground, orange for random errors); second row reports miss-classified snow depth in red; the third row reports miss-classified grass/bare ground (in purple), All plots also report measured snow depth in black (whether it represents actual snow depth, grass/ground, or random errors).*

trustworthy numerical-prediction models and snowpack models (Horton & Haegeli, 2022). In this framework, high-resolution measurements not only include meaningful information, for example related to snowfall intensity and amount (Lehning *et al.*, 2002b,a) or snowmelt patterns (Malek *et al.*, 2017b; Zhang *et al.*, 2017), but also embed a variety of noise sources that hampers their use in operations unless intensive QA/QC is performed (Avanzi *et al.*, 2014). The overarching hypothesis of this paper was that a Random forest classifier could replace expert manual checking and automatically process snow-depth high-resolution measurements from ultrasonic snow-depth sensors and thus add new value to these data for hydrologic practice and research. The main findings of this paper in this regard are three.

First, the proposed Random forest classifier was able to correctly replicate expert-made snow vs. grass/ground classifications, with F1 scores over 90% for the training/test case study of Aosta Valley. These results show that the human assessment based on expert knowledge is largely replicable (see Figure 3.8), at least for what concerns the classification of snow and grass/ground samples. While intuitively simple in nature, this differentiation

is instead complex to automatize due to non-linearities across climate, snow regimes, vegetation patterns, and topography. Meanwhile, differentiating grass/ground from snow bears significant implications with regard to snow-depth assimilation in snowpack models (Lehning *et al.*, 2002a), satellite-data validation using ground-based data (Parajka & Blöschl, 2006; Da Ronco *et al.*, 2020), and a variety of ecological analyses related to snow (Sanders-DeMott *et al.*, 2018). In this regard, the proposed Random forest is a pathway towards minimizing this noise and such accelerating the use of snow depth data in science and technology by opening the way for a fast, objective, and replicable QA/QC of snow-depth data that could complement existing practice (Avanzi *et al.*, 2014; Bavay & Egger, 2014b). Regarding speed, Table 3.3 shows that applying the Random forest of one season worth of data takes about 8 seconds, as opposed to an estimate of hours for visual screening.

| **Phase** | **Execution time** | **Sample size** |
|---|---|---|
| Training | 00:16:29 | $1.9\times10^6$ |
| Testing phase | 00:02:35 | $4.8\times10^5$ |
| Single year validation | 0:00:08 | $2.3\times10^5$ |
| Visual screening | hours/days | $2.3\times10^5$ |

**Table 3.3:** *Execution time per sample size, expressed as hh:mm:sec.*

Second, the algorithm proved to be equally robust and reliable in an independent application across the rest of Italy, at least for what concerns the snow vs. grass/base ground classification (F1 scores above 90% for this larger domain). This outcome could be due to the Random forest including all features of the Sturm & Liston (2021b) snow classification, such as air temperature and precipitation, or proxies thereof (elevation for wind speed). At the same time, the vast majority of Italian sites falls between the maritime and the montane-forest snow-climatology classes, with only a small portion of tundra snow at very high, inner-Alpine elevations (Sturm & Liston, 2021b). In other words, the testing sample might be quite homogeneous with regard to snow climatology, and testing over other regions would still be helpful.

Third, the Random forest algorithm showed a little to none sensitivity to snow-season climatology (Figure 3.7), including temperature or mean snow depth. This result may point to the Random forest being robust to different climatic regimes, including recent dry and warm snow droughts (Hatchett & McEvoy, 2018; Toreti *et al.*, 2022b; Koehler *et al.*, 2022) and future climate change (Beniston *et al.*, 2018). However, long-term climatic shifts will also bring about modifications to vegetation patterns (Cannone *et al.*, 2008), and so changes in the expected seasonality of grass vs. snow, as well as changes to the "expected" snow depth during winter (Marty *et al.*, 2017). Both aspects will need further testing in areas with different climates.

It is worth mentioning that, although the choice of these validation datasets allowed us to test the spatial extrapolation abilities of the Random forest, a full evaluation of the

spatio-temporal extrapolation skills was not achieved. The algorithm was trained on all the available years, with a standard out-of-bag validation. This was performed in an effort to maximize the number of training points and climate variability in the training sample. Thus no year was withdraw, to reduce the impact of impoverishment of the sample on the least represented class of random errors.

One critical aspect of these results is the frequently reported underestimation of random errors, like spikes, particularly across the rest-of-Italy data. This may be seen as the natural consequence of the samples being inherently imbalanced towards snow or grass/ground measurements (see Figure 3.3). Moreover, random errors are by definition hard to predict, with the only documented pattern of snowflake interference within the field of view of ultrasonic snow-depth sensors (Avanzi *et al.*, 2020b). A potential solution in this regard is for future applications to specifically target the classification of random errors, by either including more samples of this class or simply extending the analysis to more data. The use of more data is likely the most straightforward option to detect rare random errors. However, other options may prove to be effective. In light of this, the proposed algorithm may be coupled with classical QA/QC procedures imposing a-priori thresholds, like those already proposed by Bavay & Egger (2014b). Such procedures could, e.g., help with the detection of spikes in data using climatological snow-depth thresholds for maximum values.

In recent years, DL has proven successful in dealing with many complex tasks (Camps-Valls *et al.*, 2021). Future research questions may investigate the ability of other algorithms in this classification problem, such as neural networks, which are able to deal with time series and incorporate memory features. One concrete example in this regard is a recurrent neural networks or LSTM. In particular, it would be important to explore the performances of such algorithms in dealing with the recognition of the error class. In any case, the small proportion of random errors over the much more influential systematic issue of grass interference makes the Random forest a promising component of future QA/QC procedures.

# Chapter 4

# Deep learning-based data assimilation in 1D

One of the principal advantages of AI lies in its exceptional capacity to efficiently process, analyse, and interpret vast and complex datasets, far exceeding the capabilities of conventional computational algorithms, often reducing the time efforts.

DA techniques such as the Ensemble Kalman Filter (EnKF) are extremely time and resource consuming and thus often not the first choice when implementing DA in operational modelling chains. This is particularly true in low budget contexts or in areas where the data to be assimilated are sparse; for example, Pagano *et al.* (2014) highlighted that operational hydrological forecasting must deal with limited and uncertain data, imperfect models, and the need to balance forecast accuracy with the constraints of real-time service delivery, all of which complicate the adoption of computationally demanding DA techniques.

AI can work as an emulator to learn from the sounded techniques but to allow to deploy a faster and easy to transfer algorithm.

Here, I present part of the study published, together with co-authors, in the Copernicus journal The Cryosphere, entitled "Learning to filter: snow data assimilation using a Long Short-Term Memory network" (Blandini *et al.*, 2025).

## 4.1 General context

Over the past decade, snow DA has transitioned from a limited number of exploratory case studies to more mature and widely adopted methodologies (Largeron *et al.*, 2020; Girotto *et al.*, 2020; Alonso-González *et al.*, 2022). As mentioned in the previous chapter (see chapter 2), advances in satellite-derived snow products and increasing computational resources (Houser *et al.*, 2012; Aalstad *et al.*, 2018; Deschamps-Berger *et al.*, 2023a; Lievens *et al.*, 2022; Mazzolini *et al.*, 2024) have improved snow modelling yet the efficient assimilation of snow observations at large spatial scales remains a key challenge.

Indeed, practical applicability of ensemble-based methods is often constrained by high computational demands (Girotto *et al.*, 2020), limiting their use in real-time and operational forecasting systems (Pagano *et al.*, 2014). The computational burden primarily arises from the need to propagate large ensembles of model realizations (Evensen *et al.*, 2022), motivating the development of alternative strategies that preserve the benefits of probabilistic DA while reducing computational costs.

In this context, DL techniques could be powerful tools for learning complex, nonlinear system dynamics directly from data, without requiring explicit physical parameterizations (Sit *et al.*, 2020a). Despite the growing body of literature on stand-alone DL approaches for snow modelling (Cui *et al.*, 2023; Daudt *et al.*, 2023), the potential of DL to emulate or accelerate computationally intensive DA algorithms remains largely unexplored. Building on the concept of "Deep Data Assimilation" introduced by Casas *et al.* (2020) and Arcucci *et al.* (2021) as well as a growing literature of related methods (Cheng *et al.*, 2023), this research aims to enhance DA methods in snow hydrology by proposing an alternative approach for assimilating snow-related quantities, specifically SWE and snow depth, through the use of LSTM neural networks (Hochreiter & Schmidhuber, 1997). Developing a supervised learning strategy, the LSTM networks acts as EnKF emulator, with with the goal of improving snowpack estimations while minimizing computational efforts. The research questions guiding this study are: (i) What is the performance of an LSTM network in filtering, particularly in comparison with an EnKF? (ii) How does the network's performance respond to data sparsity? (iii) Is it feasible to transfer an LSTM model trained at one site to other sites without a significant loss in performance? (iv) How does the model's performance vary across different types of water years?

## 4.2 Data

When working with DL algorithms, the quality of the dataset is crucial, as the performance of the trained network will highly depend on it (He *et al.*, 2019). Hence, this study employees high-quality, pre-processed datasets from long-term, internationally acknowledged snow research stations across the northern hemisphere (Figure 4.1). The datasets used where those of precipitation (mm), solar radiation ($W/m^2$), relative humidity (%), air temperature (°C), and daily average temperature (°C) along with SWE (mm) and snow depth (cm) ground measurements.

Here is a list of the station locations, along with their associated reference papers and abbreviations:

- Torgnon, Aosta Valley, Italy, TRG (Filippa *et al.*, 2015).
- Col De Porte, Isère, France, CDP (Lejeune *et al.*, 2019).
- Weissfluhjoch, Davos, Switzerland, WFJ (Wever, 2017).

- Kühtai, Tirol, Austria, KHT (Krajči *et al.*, 2017).
- FMI-ARC Sodankylä Geophysical Observatory, Finnish Lapland, FMI-ARC (Essery *et al.*, 2016).
- Nagaoka, Japan, NGK (Avanzi *et al.*, 2019).
- Reynolds Mountain East, Idaho, USA, RME (Reba *et al.*, 2011).

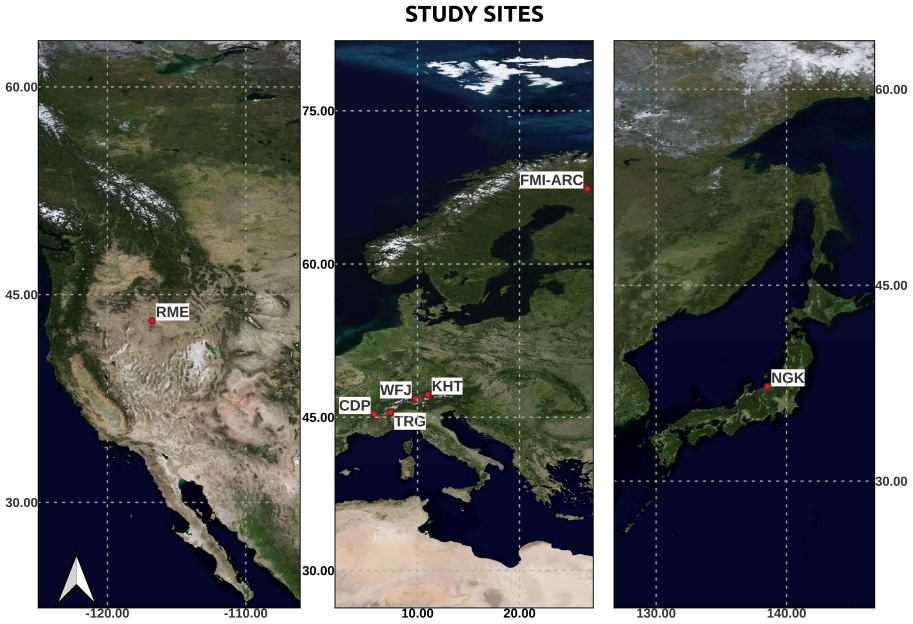


**Figure 4.1:** *Geographical distribution of study sites used for snow modeling and data assimilation: (left) Reynolds Mountain East (RME) in the United States, (center) European sites including Col De Porte, Isère, France , Weissfluhjoch, Davos, Switzerland (WFJ), Torgnon, Aosta Valley, Italy (TRG), Kühtai, Tirol, Austria (KHT) FMI-ARC Sodankylä Geophysical Observatory, Finnish Lapland (FMI-ARC) in Finland, and (right) Nagaoka (NGK) Japan. Map created using the Free and Open Source QGIS.*

The sites were selected to ensure geographic and climatic diversity, spanning various regions that are exposed to a variety of snow climates (Sturm & Liston, 2021a), (see Table 4.1 and Table 4.2). The characteristics of the site vary widely, with elevations ranging from lowland areas such as Sodankylä (179 m) to high alpine environments such as Weissfluhjoch (2540 m). Annual and winter precipitation varies significantly across different locations, ranging from relatively dry areas like Torgnon, with an annual average of 794 mm, to much wetter regions such as Nagaoka, which receives 2773 mm per year. For

this comparative analysis, winter is defined as the meteorological winter in the northern Hemisphere, spanning the months of December through February. Air temperature ranges reflect this environmental diversity, encompassing cold alpine regions, temperate meadows, and wetlands.

| Site | Description | Altitude (m a.s.l) | MAP (mm) | MWP (mm) | MAAT [min,max] (°C) |
|---|---|---|---|---|---|
| TRG | Subalpine grassland | 2160 | 794 | 161 | 3 $[-15, 20]$ |
| CDP | Grassy meadow | 1325 | 1896 | 550 | 6 $[-13, 17]$ |
| WFJ | Almost flat area | 2540 | 1631 | 391 | $-1$ $[-21, 17]$ |
| KHT | Steep alpine valley | 1920 | 1131 | 186 | 3 $[-18, 22]$ |
| FMI-ARC | Large wetland area | 179 | 551 | 125 | 0 $[-35, 27]$ |
| NGK | Flat meadow | 97 | 2773 | 1104 | 12 $[-5, 36]$ |
| RME | Unsheltered mountain area | 2137 | 817 | 350 | 5 $[-20, 30]$ |

**Table 4.1:** *Geographic and climatic characteristics (annual precipitation and air temperature statistics) of the selected study sites. MAP = mean annual precipitation (mm), MWP = mean winter precipitation (mm), and MAAT = mean annual air temperature (°C). Mean values were computed over the entire available time period for each site. The corresponding time frames are reported in Appendix F.2.*

| Site | Peak SWE (mm) | Peak snow depth (cm) | Snow cover duration | Snow Type |
|---|---|---|---|---|
| TRG | 312 | 11 | From October to May | Tundra |
| CDP | 414 | 14 | From November to May | Maritime |
| WFJ | 802 | 23 | From October/November to August | Tundra |
| KHT | 347 | 15 | From October/November to mid June | Tundra |
| FMI-ARC | 197 | 8 | From October to May | Boreal forest |
| NGK | 381 | 14 | From Novemeber to April | Maritime |
| RME | 529 | 17 | From October to May | Montane forest |

**Table 4.2:** *Summary of snow characteristics at the selected study sites. Snow classification by Sturm & Liston (2021a).*

The record period for each dataset varied depending on the time frames available at each site. More info on the first and last year of records available for each site can be found in the appendix at section F.2. All datasets were originally recorded at temporal resolutions ranging from sub-hourly to hourly. Therefore, to ensure uniform application of the algorithm, a 1-hour temporal resolution was selected as the reference, and sub-hourly datasets were resampled to this resolution using linear interpolation. This hourly resolution resolves day-night cycles of melting and refreezing, revealing air temperature fluctuations and their relationship with snowpack outflow. In addition, it enables the evaluation of the precipitation dynamics, the primary mass input to the seasonal snowpack (Avanzi *et al.*, 2014).

Measurement errors used in the DA process (see Section 4.3.2) were assigned according to the specific instrumentation utilized at each site, drawing from a combination of expert knowledge and relevant literature (see Tab. F.1 in the Appendix). These errors

were not directly provided to the DL algorithm; rather, the algorithm learned them indirectly through supervised training on the EnKF analysis results, rather than from the observations themselves.

Based on data sparsity, defined as the presence of 80% or more of the record period containing missing data, or a low temporal data granularity (i.e., temporal frequency coarser than 1 hour), the datasets were categorized into two groups:

- Low data sparsity: NGK, KTH, FMI-ARC, and RME datasets.
- High data sparsity: CDP, TRG, and WFJ datasets.

## 4.3 Methodology

### 4.3.1 Snow model

For this pilot application of a new deep DA scheme, a point-scale version of S3M has been employed (see section 2.1.2 and Avanzi *et al.* (2022a)). This version retains all the features of the original S3M model, such as precipitation-phase partitioning, snow mass balance, snow metamorphism, and hydraulics, but it models snow dynamics at one point rather than in grid cells distributed across the landscape. The model state variables are:

- $\mathrm{SWE}_W$ = wet component of SWE (mm).
- $\mathrm{SWE}_D$ = dry component of SWE (mm).
- $\rho_D$ = dry-snow density (kg $\mathrm{m}^{-3}$).
- $\alpha_S$ = the snow albedo (–).
- $h_G$ = glacier thickness (m).

In the point-scale version, the glacier thickness was not introduced as prognostic variable. The model estimation of SWE and snow depth are to be considered as derived output from a combination of the model state variables.

### 4.3.2 Ensemble Kalman filter assimilation scheme

Aiming at mimicking an established ensemble-based DA algorithm with a DL-based approach, a supervised learning approach was chosen (Murphy, 2022). Hence, the training data had to be derived from the state analysis output by such DA scheme. The assimilation algorithm used as training was designed to focus on retrieving an accurate analysis of the state vector ($\boldsymbol{x} \in \mathbb{R}^n$ with $n$ the number of states), including both the wet and dry components of SWE , the density of dry snow (kg/$\mathrm{m}^3$), and the snow albedo (-). Given the nonlinear nature of S3M, it was decided to use an ensemble method that approximates

the posterior probability density function of the analysis using the mean and covariance matrix (Carrassi *et al.*, 2018; Evensen *et al.*, 2022). Given the high robustness even with a relatively small ensemble (Aalstad *et al.*, 2018), an EnKF scheme was developed in S3M (For a mathematical background see section 2.2.1 and appendix A).

In the present study, a joint DA scheme was developed to update the system state by jointly assimilating ground-based measurements of snow depth and SWE. Despite albedo being another potential data stream to assimilate (Navari *et al.*, 2018), due to the lack of measurements across the 7 sites, assimilation of albedo measurements was not considered herein. Nonetheless, albedo was updated indirectly, based on the assimilation of SWE and snow depth. Moreover, as the model is a point-based simulation, the fractional snow-covered area assimilation was not pursued. Additionally, since the EnKF can not handle binary observations binary snow cover was not an option either. Finally, since S3M does not solve the full energy balance or simulate snow-temperature profiles, no surface temperature proxy was assimilated.

Ensemble generation was performed by perturbing meteorological model forcing, which included total precipitation (mm), solar radiation ($W/m^2$), relative humidity (%), air temperature (°C), and daily average temperature (°C).

To each meteorological forcing data point, an ensemble of multivariate errors was added. These errors were generated as realizations of a multivariate stochastic process designed to have a specified covariance matrix derived from the Gaussianized historical meteorological series.

The objective was to generate a multivariate time series of meteorological values in Gaussian space such that the imposed perturbations respect both temporal coherence and cross-variable coherence. By ensuring that the resulting covariance matrix matches $\mathbf{C}$, the temporal covariance of the historical observations, the generated ensemble perturbations remain physically consistent and statistically representative of the historical variability. Each forcing variable was perturbed such that its distribution remains unbiased and constrained within physically consistent limits.

The procedure for constructing the stochastic process is base on similar approaches implemented by Reichle *et al.* (2007), De Lannoy *et al.* (2010) and Durand & Margulis (2006b) and is described below:

1. **Generation of a random covariance matrix:**
   A random covariance matrix, $\mathbf{C}_o$, was generated.

2. **Construction of a Gaussian stochastic process:**

   (a) A Cholesky decomposition was performed on $\mathbf{C}_o$, yielding:

   $$\mathbf{L}_o = \text{Cholesky}(\mathbf{C}_o).$$

   (b) The multivariate stochastic process was defined as:

   $$\boldsymbol{u}_{k+1} = \boldsymbol{u}_k + \mathbf{L}_o \boldsymbol{\epsilon},$$

where:
$\boldsymbol{\epsilon} \sim \mathcal{N}(0, 0.1)$ represents independent standard normal variables.
$\boldsymbol{u}_k$ represent the vector of the process that has to be transformed into the Gaussian space at time k. In this case it represent the forcing meteo vector.

3. **Calculation of the covariance matrix:** A realization of the stochastic process was generated, and its covariance matrix, $\tilde{\mathbf{C}}$, was computed.

4. **Imposition of the target covariance matrix:**

   (a) The Cholesky decomposition of the target covariance matrix, $\mathbf{C}$, was computed:

   $$\mathbf{L} = \text{Cholesky}(\mathbf{C}).$$

   (b) The stochastic process was constructed to impose the covariance matrix $\mathbf{C}$ as follows:

   $$\tilde{\boldsymbol{u}}_{k+1} = \tilde{\boldsymbol{u}}_k + \mathbf{L}_o \cdot \boldsymbol{\epsilon}$$

   Initially, this process was characterized by the covariance matrix $\tilde{\mathbf{C}}$. To transform it into a process with the covariance matrix $\mathbf{C}$, the following steps were taken:

   (a) The transformation:

   $$\tilde{\mathbf{L}} \cdot \tilde{\boldsymbol{u}}_{k+1}$$

   where $\tilde{\mathbf{L}} = \text{Cholesky}(\tilde{\mathbf{C}})$, was applied, normalizing the covariance matrix $\tilde{\mathbf{C}}$ to the identity matrix $\mathbf{I}$.

   (b) A second transformation was applied:

   $$\boldsymbol{u}_{k+1} = \mathbf{L} \cdot \left(\tilde{\mathbf{L}} \cdot \tilde{\boldsymbol{u}}_{k+1}\right)$$

   which transformed the identity matrix $\mathbf{I}$ into the target covariance matrix $\mathbf{C}$.

5. **Perturbation Calculation:** Finally, to compute the perturbations to be added to the meteorological values, the following expression was used:

   $$\boldsymbol{\Delta} = \tilde{\boldsymbol{u}}_{k+1} - \text{mean}(\tilde{\boldsymbol{u}}_{k+1})$$

   where $\text{mean}(\tilde{\boldsymbol{u}}_{k+1})$ represents the ensemble mean.

The use of a stochastic process to generate the ensemble of errors was pivotal to ensure temporal coherence in meteorological perturbations. This procedure was location-specific, tailored to the 7 study sites. The ensemble size was defined as 100 members. It was determined to be suitable for an EnKF, based on literature (Aalstad *et al.*, 2018) and testing.

To improve filter performance and stability, the forecast model state vector $\boldsymbol{x}_k^f$ at each time step $t_k$ was also perturbed. To obtain the perturbation, a series of multivariate Gaussian random error with imposed process noise covariance matrix $\mathbf{Q}$ was added to each forecast model state vector point. The matrix was retrieved rescaling the state covariance matrix obtained from S3M open loop forecast over the entire historical period for each site. The observation vector $\mathbf{y}_k$ stores, at each time step $k$, the available measurements of SWE and snow depth. As described in Section 4.3.1, these variables are not prognostic model states but are derived from them. Different versions of the observation operator $\mathbf{H}_k$ were constructed to allow assimilation with only one observed variable when necessary. The structure of $\mathbf{H}_k$ is reported in Appendix G. Post-processing was applied to the filter outputs to ensure physical consistency, adjusting corrections to the filter output while maintaining the physical relationships between the elements of the state vector. This included constraining the values within a physical range and modulating them accordingly.

### 4.3.3 Long-short term memory neural network

The development of DA using neural networks was framed as a time series forecasting task, leading to the use of Recurrent Neural Networks Recurrent Neural Network (RNN)s RNNs leverage internal memory to process sequences of data, making them useful for time-dependent analysis. However, they often struggle with long-term dependencies due to vanishing or exploding gradients (Tsantekidis *et al.*, 2022). To address this, LSTM networks introduce gate mechanisms (input, forget, output) to control information flow, effectively managing long-term dependencies (Hochreiter & Schmidhuber, 1997).

#### Data pre-processing

Effective data pre-processing is critical for the successful application of LSTM networks, as it improves prediction accuracy, reduces computational costs, and enhances model robustness and repeatability (Isik *et al.*, 2012). Proper pre-processing not only accelerates network convergence, but also helps the model capture essential patterns in the data. For LSTM networks, which are sensitive to the distribution and scale of the inputs, pre-processing plays a key role in mitigating issues like exploding or vanishing gradients and managing differences in feature magnitudes.

Data pre-processing in this study involved two key steps:

- **Distribution adjustment:** Snow related variables frequently hit the lower physical boundary of 0 mm of SWE or 0 cm of snow depth, posing challenges for the LSTM, which struggled to handle this behavior. To overcome this, the data range was extended by redefining the lower limit to a value below zero.
- **Scaling with historical data:** After adjusting the distribution, the input values were also standardized using the mean and standard deviation calculated from his-

torical records at each site. This standardization ensured that all input features were on a consistent scale.

### Custom loss function

To ensure compliancy of the LSTM predictions to specific problem domain constraints, a custom loss function was developed.

This loss function comprises two main components:

- **Root Mean Square Error RMSE**: This measures the difference between the LSTM predictions $\boldsymbol{x}_k^{a\star}$ and the analysis state vectors generated by the EnKF, $\boldsymbol{x}_k^a$. By minimizing RMSE, the model was trained to closely follow the reference trajectory provided by the EnKF.
- **Physics-based Regularization Term**: An additional U-shaped penalty function was introduced to enforce adherence to physical constraints and guide the model towards specific physical behaviors. This term penalizes the network for making predictions that violate predefined physical boundaries. The function is expressed as:

$$\text{Loss}(\boldsymbol{x}_k^{a\star}) = \frac{1}{\prod_{i=1}^{n} \left| x_{k,i}^{a\star} - a_i \right| \cdot \left| x_{k,i}^{a\star} - b_i \right|} \tag{4.1}$$

where $\boldsymbol{x}_k^{a\star}$ is the analysis state predicted by the LSTM, $n$ is the number of state vector components $x_{k,i}^{a\star}$ and $a_i$ and $b_i$ are the minimum and maximum physical bounds, respectively, of the $i$-th element of the state vector, defined as the minimum historical and maximum historical records.

Furthermore, any LSTM prediction that fell below zero was forced back to zero, effectively managing the intermittent nature of snow data.

This combined loss function is inspired by Physics-Informed DL (Cheng & Zhang, 2021), where domain-specific physical constraints guide the learning process.

### Algorithm development and test configurations

The LSTM algorithm was trained using the analysis state vectors generated by the EnKF -$\boldsymbol{x}_k^a$- to predict the corrected analysis state vector, $\boldsymbol{x}_k^{a\star}$. As a supervised learning task, the training process utilized both input features and target outputs. The input features included meteorological forcing variables, the model's forecast state vector $\boldsymbol{x}_k^f$, and the observation vector, while the target outputs consisted of the analysis state vectors $\boldsymbol{x}_k^a$ from the EnKF. To evaluate its effectiveness, the LSTM predictions were compared to the analysis state vectors generated by the EnKF. To develop the LSTM algorithm, it was used Python 3.9.21 programming language and the open source libraries `Keras v.2.10.0` (Chollet *et al.*, 2015) and `Scikit-learn v.1.1.1` (Pedregosa *et al.*, 2011).

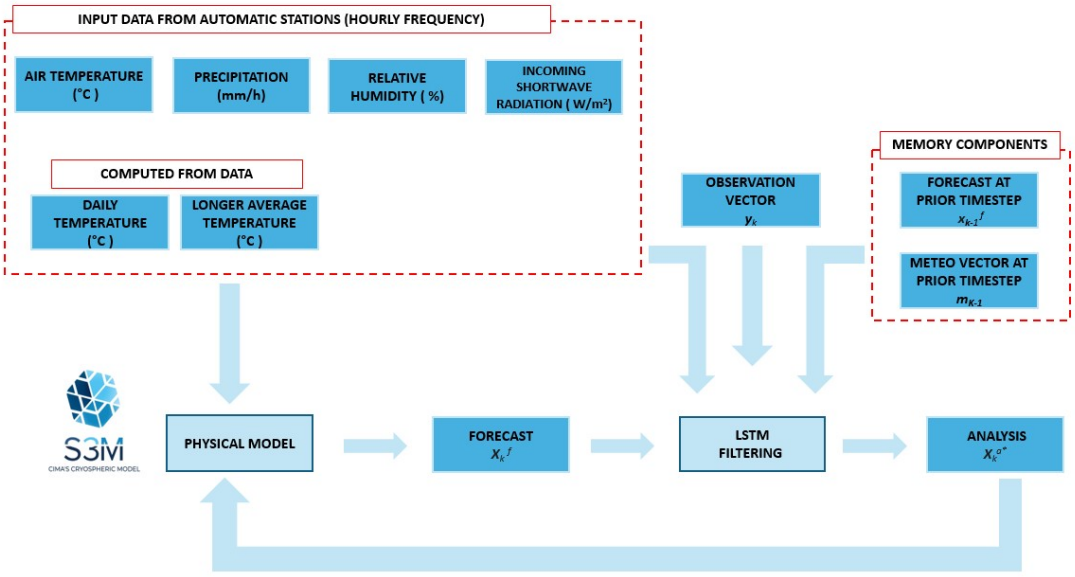


**Figure 4.2:** *Operational Setup for deep data assimilation. This diagram illustrates the operational workflow for integrating observational data with the S3M (CIMA's Cryospheric Model) framework, through DA via a Long-Short-Term-Memory neural network.*

To assess the LSTM robustness and transferability, four experimental setups were tested:

1 **Site-Specific LSTMs for State Correction**
Seven LSTMs were independently trained and tested on each site to optimize hyperparameters. For the site with $> 95\%$ missing SWE data (WFJ), the LSTM was trained using an observation vector containing only snow depth, which instead were not missing. Site-specific limits, derived from historical data, were applied to constrain the training process. Since the training process relies on a cost function that combines the RMSE with a penalty term enforcing physical bounds, the site-specific limits for each state component, namely, the dry and wet components of SWE , snow density, and albedo, were derived from historical data records. These records were pre-processed following the distribution adjustment and scaling procedures described in Section 2.4.1. Since direct historical observations of wet SWE were not available, this variable was assumed to be proportional to the ratio between LWC and total SWE , with dry SWE estimated as the complementary term.

The available data were split by continuous time spans, using the hydrological year

(from the 1st of October to the 30th of September) as the reference unit. Specifically, first 80% of the data, in terms of hydrological years, was allocated for training and testing using a 4:1 ratio, while the remaining 20% was reserved for operational testing. In the operational setup, the framework combined S3M model prediction and state updating with the LSTM (see figure 4.2). At each time step $t_k$, the prior state vector $\boldsymbol{x}_k^f = \mathrm{S3M}\left(\boldsymbol{x}_{k-1}^{a\star}\right)$ from the S3M model's forward simulation was provided as input to the LSTM, along with meteorological forcing and the observation vector $\boldsymbol{y}_k$. The LSTM outputs the updated analysis state vector $\boldsymbol{x}_k^{a\star}$, which served as the initial condition for the subsequent S3M prediction step $\boldsymbol{x}_{k+1}^f = \mathrm{S3M}\left(\boldsymbol{x}_k^{a\star}\right)$ and so on, cycling between S3M prediction and LSTM updating. The framework was validated using root mean square error RMSE metrics for snow depth and SWE between ground observations and model predictions. It is important to stress that, while the training phase was performed in the conventional way of training neural networks -meaning multiple timestep as input to obtain a sequence of outputs - the operational testing phase was performed giving to the LSTM trained models only one timestep at a time, to be coupled with the forward step of the cryospheric model. The metrics were computed for both the test and the operational set; while the first was used to set hyperparameters, the second was used to analyse the performance of the model.

**2 Incorporating Memory to the Site-Specific LSTMs**
The second test configuration introduced an additional feature component to call back on the use of the "long" memory component of the LSTM during the operational test phase. The memory component includes the forecast from the previous timestep $\boldsymbol{x}_{k-1}^f$ as well as the meteorological forcing from the previous time step $k-1$ (relative to the current step $k$). The input vector I at time step $k$, is constructed as follows:

$$\mathbf{I}_k = \left[\mathbf{m}_k\ , \mathbf{m}_{k-1}\ , \mathbf{x}_{k-1}^f\right] \tag{4.2}$$

where:

- $\mathbf{m}_k \in \mathbb{R}^d$: the vector of meteorological forcing variables at time step $k$ where $d = 6$ is the number of forcing variables.
- $\mathbf{m}_{k-1} \in \mathbb{R}^d$: the meteorological forcing at the previous time step $k-1$ (see fig (2) memory component element)
- $\mathbf{x}_{k-1}^f \in \mathbb{R}^n$: the model forecast at the previous time step $k-1$ (see fig (2) memory component element)

**3 Testing Transferability of Site-Specific LSTMs**
While in the Configuration 1, separate LSTM models were trained and tested individually on each site using only site-specific data, in Configuration 3, the spatial transferability of these site-specific models was assed by applying each LSTM trained

on the low data sparsity sites (NGK, KHT, RME, FMI-ARC) to new data from (i) the remaining 20% holdout portion of the low-sparsity sites not used during training, and (ii) high data sparsity sites (CDP and TRG). The WFJ site was excluded from this evaluation due to extensive gaps in its SWE time series. In this test it was chosen to use the LSTM setup with the best performances among prior tests, hence the one with memory components (see point 2)

4 **Multisite LSTM with Global Limits**
A multisite LSTM was trained using data from the four low data sparsity sites (NGK, KHT, RME, FMI-ARC), with global scaling derived from the combined datasets. The training dataset comprised 80% of the data from these four sites, while the remaining 20% alongside all data from the high data sparsity datasets (CDP, WFJ, TRG) were used to test the model generalization capacity over water year type, using the operational setup. Data split was made by randomly sampling whole hydrological years. The water year types were classified based on the total snow depth and include wet years, dry years, and average conditions. Additionally, the results where also analysed comparing the performances per each site with the performances of the best site-specific LSTM-DA algorithm.

Site-specific EnKF results were always used as input for training the LSTM, even in the case of multisite LSTM testing the EnKFs used to generate the training data were always site-specific.

### Long-short term memory structure and hyperparameters

In this study, the hyperparameters of the model were manually tuned, selecting the optimal configuration for each LSTM network. Below are the hyperparameters fine-tuned:

- **Batch size:**
The batch size determines the number of training samples processed in a single forward and backward pass. A critical consideration when choosing the batch size is balancing computational efficiency with the quality of model outputs. To match the size of the observation datasets for each site, we used a standard batch size of 128 for the sites of KHT and NG, and we reduced it each time selecting the most suitable value for optimal training performance on all the other datasets (Bishop & Bishop, 2023).

- **Epochs:**
The number of epochs refers to the total number of complete passes through the training dataset. While a higher number of epochs allows the model to better capture complex patterns in the data, it also increases the risk of overfitting and computational cost. After experimenting with various configurations, we set the number of epochs to 500, allowing for sufficient learning while balancing efficiency.

- **Early Stopping:**
  Early stopping is a technique used to prevent overfitting by halting training when the validation performance fails to improve for a specified number of epochs. In this case,the patience was set to 100, meaning that training would terminate if no improvement was observed in the validation performance for 100 consecutive epochs (Prechelt, 2002).

- **Initial learning Rate:**
  The learning rate controls the step size during the optimization process. A higher learning rate accelerates convergence but may lead to instability, while a lower learning rate can slow down the learning process. Given the relatively small size of the datasets, an initial learning rate of 0.01 was chosen to ensure rapid convergence during the early stages of training (Smith, 2015).

- **learning Rate Decay:**
  To enhance convergence stability and prevent overshooting, it was applied a learning rate decay factor of 1.5 periodically throughout training. This decay reduces the learning rate over time, allowing the model to fine-tune its parameters more effectively in the later stages of training.

- **Dense Layers:**
  Each LSTM network used a single dense layer as the output layer. This dense layer was used to map the LSTM outputs to a fixed-size state vector. The number of neurons in this layer was set to 4, corresponding to the required output dimensions for each network (Murphy, 2023).

- **Hidden LSTM Layers:**
  Two distinct LSTM architectures were employed based on the data sparsity at different sites. For data-dense sites, it was used a single LSTM layer followed by a dense output layer, resulting in a simple 2-layer architecture. This configuration was chosen under the assumption that the data contained enough patterns for the model to learn effectively without requiring excessive model depth. In contrast, for sparse sites, a deeper 3-layer LSTM architecture was implemented, which included two LSTM layers and a dense output layer. This approach aimed to capture more complex dependencies within the data, thereby improving the model's ability to learn from sparser temporal patterns (LeCun *et al.*, 2015a).

- **Hidden units per LSTM Layer:**
  The number of hidden units in each LSTM layer determines the memory capacity of the model. For dense sites, the number was set to 500, allowing the model to learn from more intricate temporal dependencies. For sparse sites, the number was reduced to 100 to prevent overfitting, given the smaller and sparser datasets (Murphy, 2023).

**A note to the reader:** In the following section, the term LSTM refers to the computation of the analysis mean model state vector, denoted as $\boldsymbol{x}_k^{a\star} \in \mathbb{R}^n$, using the LSTM approach. On the other hand, the term EnKF refers to the computation of the analysis mean model state vector, denoted denoted as $\boldsymbol{x}_k^{a} \in \mathbb{R}^n$, using ensemble-based DA via the EnKF scheme.

## 4.4 Results

This section presents the results from the four configuration tests, based on the operational testing setup (see Fig. 4.2). The objective was to replicate the actual algorithm coupling mechanism required in a real-time setup, where the LSTM is used at each time step $k$ to perform filtering.

### 4.4.1 Performance with varying data sparsity

Figures 4.3 and 4.4 show the performance of the LSTM-based analysis compared with the open loop, the EnKF analysis, and the observations. Panels (a–d) display simulations for one hydrological year of both SWE and snow depth at two stations. Panels (e–j) present the values of different performance metrics (RMSE, bias, and KGE) computed across all sites. The metrics are calculated over three years of data; each box represents the distribution of the testing samples over this period, while the individual points correspond to results computed over two years of data, representing the size of the testing sample.

At sites where data is plentiful (that is, available data cover more than 80% of the period of record: NGK, KTH, FMI-ARC, RME), the LSTM demonstrated robust performances, meaning that they were generally comparable to the original EnKF (Figure 4.3). This, however, came with a considerable nearly 70% decrease in computing time. For instance, one year of simulation using the parallelized EnKF took on average 20 minutes, while using the trained LSTM took only 6 minutes.

Only in the case of NGK site, the LSTM-DA was able to outperform both the open loop simulation and the EnKF-DA; At all the other dense sites ( KTH, RME, FMI-ARC), the mean RMSE increase relative to the EnKF for SWE estimation made by site specific LSTMs was within 10 mm (Figure 4.3, panel e). Similarly, the mean RMSE increase- averaged across sites- compared to the EnKF for snow depth estimation made by site specific LSTMs was equal to 6 cm (Figure 4.3, panel f). The only exception is the site of FMI-ARC were the LSTM-DA still underperformed compared to the EnKF, although the absolute values of RMSE are 1 order of magnitude lower than the ones on the other sites. The bias analysis (Figure 4.3, panel g and h) showed that snow depth exhibited a near zero bias, while the LSTM tended to overestimate SWE compared to the EnKF. However, both patterns were consistent in the EnKF and in the S3M open loop.

In the case of datasets with high data sparsity (CDP, WFJ, TRG), the performance of the LSTM was markedly worse than the EnKF estimation of both SWE and snow depth

(+50 mm RMSE for SWE and +19 cm RMSE for snow depth, Figure 4.4 panel e and f).

On the other hand, the timing of SWE and snow depth peaks, as well as the magnitude of snow depth peaks, are generally captured correctly, even in these challenging data sparse scenarios (see Figure4.4 panels a,b,c,d). However, minor discrepancies were noted, even in the case of low data sparsity, including an underestimation of peak snow depth (Figures 4.3, panels c and d) and a slight temporal shift in the SWE peak (Figure 4.3, panel a).

Both the EnKF and LSTM networks improved SWE and snow depth predictions over the Open Loop model, at least in the case of low data sparsity; indeed the LSTM resulted in a reduction of 25 mm in RMSE for SWE , while the EnKF achieved a better reduction of 31 mm. On the other hand, in case of high data sparsity, the LSTM increased the RMSE by 15 mm, while the EnKF reduced the RMSE by 38 mm. For snow depth, the LSTM reduced RMSE by 4 cm in low sparsity, while the EnKF showed a greater reduction of 9 cm. Under high sparsity, the LSTM reduced RMSE by 8 cm, with the EnKF providing a larger reduction of 27 cm.

When it comes to evaluating the Kling-Gupta Efficiency (KGE) (Gupta *et al.*, 2009), for sites with denser measurements (on average 0.72 for both SWE and snow depth), the values are comparable to those obtained with the EnKF-DA (on average 0.75 for SWE and 0.85 for snow depth), supporting the observed improvement trend over the open loop simulation (on average 0.75 for SWE and 0.68 for snow depth). Conversely, in the case of sparse datasets, the lower KGE values (on average -0.4 for SWE and 0.25 for snow depth) highlight the limitations of the LSTM in achieving performances comparable to the EnKF-DA (on average -0.5 for SWE and 0.35 for snow depth). Nevertheless, the LSTM still outperformed the open loop, which recorded even lower KGE scores of -0.50 for SWE and -0.06 for snow depth.

Overall, the LSTM demonstrates a reduction in bias compared to the Open Loop under low data sparsity conditions, with a bias reduction of 7 mm in SWE and 3 cm in snow depth (Figure 4.3, panel h). This improvement becomes even more pronounced in high data sparsity scenarios, where the bias decreases by 15.96 mm in SWE and 5 cm in snow depth (Figure 4.4, panel h). However, despite these improvements, the LSTM still exhibits a higher bias compared to the EnKF.

### 4.4.2 The role of the memory component

Figures 4.5 and 4.6 show the performance of the LSTM-based analysis compared with the open loop, the EnKF analysis, and the observations. Panels (a–d) display simulations for one hydrological year of both SWE and snow depth at two stations. Panels (e–j) present the values of different performance metrics (RMSE, bias, and KGE) computed across all sites. The metrics are calculated over three years of data; each box represents the distribution of the testing samples over this period, while the individual points correspond to results computed over two years of data, representing the size of the testing sample.

For datasets characterized by low data sparsity (NGK, KTH, FMI-ARC, RME), incor-

porating a memory component into the LSTM improved its ability to capture the seasonal dynamics of SWE and snow depth, particularly in accurately representing the timing and magnitude of peak SWE (see Figure 4.5, panels a and b). However, in some instances (see Figure 4.5, panels c and d), the memory component did not lead to a significant performance gain. Instead, it primarily acted as a smoother, dampening short-term fluctuations without substantially enhancing predictive accuracy. Additionally, no significant changes were observed in the snow depth estimation, with a mean RMSE increase of 6 cm compared to the EnKF (Figure 4.5, panel f).

When considering sites with high data sparsity (CDP, WFJ, TRG), a LSTM with the addition of a memory component improved both quantitative and timing estimations of peak SWE and peak snow depth, compared to the LSTM estimates without memory. In fact, it was observed a mean reduction in RMSE equal to 10 mm for SWE estimates and equal to 0.5 cm for snow depth estimates. However, for datasets with extremely high levels of missing data (e.g., 95%, WFJ and TRG – where the assimilated observations consist of manually measured SWE data, as detailed in the corresponding site references), improvements were still insufficient to obtain scores comparable to the EnKF (see Figure 4.6, panels e and f). Nevertheless, the introduction of the memory component reduced model instability and improved snowmelt timing, particularly at sites with sparse observations.

Overall, considering both scenarios, biases (Figure 4.6, panel g) were not affected by the introduction of a memory component.

The inclusion of a memory component narrowed the performance gap between the EnKF and LSTM compared to the Open Loop. For low sparsity, the LSTM reduced RMSE for SWE by 29 mm, while in high sparsity, it limited the increase in SWE RMSE to just 3 mm. In terms of snow depth, the LSTM reduced RMSE by 13 cm in low sparsity and by 7 cm in high sparsity. However, the EnKF still outperformed this LSTM configuration in both cases, highlighting its superior performance despite the added memory and runtime cost.

The KGE values, for both dense and sparse datasets, confirm that the memory component primarily acts as a smoother and enhances performance in most scenarios.

### 4.4.3 Spatial transferability

The LSTM trained on KHT emerged as the only one transferable across sites (Figure 4.7). For SWE estimation this LSTM showed small drops in performances across other sites below 20% and, in some cases, even a performances boost (see LSTM on FMI-ARC, RMSE AND TRG on Tab. 4.3). On the other hand, performance drops for snow depth estimation varied considerably, from 60% to -1 % (Tab. 4.3). Other LSTMs, such as those trained in NGK and FMI-ARC, performed less consistently, showing notable increases in RMSE when transferred to several sites. While recent studies (Kratzert *et al.*, 2024) have strongly advocated for multi-basin training to achieve robust and generalizable LSTM streamflow models, here are intentionally present the single-point case here for snow hydrology to

establish a performance lower bound for snow spatial transferability, highlighting whether even such a constrained model can outperform the open loop and compare with traditional DA approaches.

| | SWE RMSE [mm] | | | | | snow depth RMSE [cm] | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Site** | **$RMSE_{LOCAL}$** | **$\Delta_{\%}$ KTH** | **$\Delta_{\%}$ NGK** | **$\Delta_{\%}$ FMI-ARC** | **$\Delta_{\%}$ RME** | **$RMSE_{LOCAL}$ [cm]** | **$\Delta_{\%}$ KTH** | **$\Delta_{\%}$ NGK** | **$\Delta_{\%}$ FMI-ARC** | **$\Delta_{\%}$ RME** |
| NGK | 14.09 | +8 | - | +125 | +199 | 8 | +34 | - | +90 | +126 |
| KHT | 14.10 | - | +271 | +329 | +155 | 22 | - | -13 | +34 | +2 |
| FMI-ARC | 9.06 | -45 | +119 | - | +32 | 11 | +47 | +25 | - | +78 |
| RME | 39.92 | -51 | +35 | +59 | - | 17 | -54 | +14 | +7 | - |
| CDP | 67.61 | +18 | +68 | +74 | +66 | 12 | +60 | +62 | +253 | +128 |
| TRG | 73.70 | -76 | -58 | -37 | -68 | 22 | -1 | +2 | +19 | +11 |
| Average | - | -29 | +87 | +110 | +77 | - | +17 | +18 | +81 | +69 |

**Table 4.3:** *Percentage change in SWE and snow depth RMSE when using a transferred LSTM assimilation scheme compared to a locally trained LSTM. Positive and negative values indicate improvements or degradation in performance, respectively. $\Delta$ values are obtained as the difference between the RMSE of a locally trained LSTM and that of a transferred LSTM, respectively for SWE and snow depth.*

Tests on correlations between LSTMs performances and biases with various climatological variables showed no statistically significant correlation (see Figure F.1 and F.2 in the appendix).

### 4.4.4 Multi-site long-short term memory

To guarantee a meaningful and practical evaluation of the multi-site LSTM performances, the analysis was performed by comparing RMSE distributions for SWE and snow depth across water year types. Figure 4.8 presents the RMSE distribution for SWE and snow depth under varying water year types, comparing the performance of the S3M open-loop run, the estimates retrieved from the analysis of EnKF, and the LSTM estimates.

A multi-site LSTM generally demonstrated improvements in performance compared to the S3M open-loop run, particularly for SWE . For dry and average years (Figure 4.8, panels b and c), the SWE estimates from the LSTM showed competitive performance over EnKF, with a performance drop of less than 6 mm on average. On the other hand, the LSTM SWE estimation RMSE values were higher during wet years ( +15 mm ). Reduced performances of the Multi-site LSTM simulation on SWE over wet years may be because in wet years, an increased number of snowfall events may introduce additional complexity and uncertainty, both due to the cascading effects of uncertainties in initial conditions and precipitation phase partitioning (Harder & Pomeroy, 2014). Moreover, the formation of several snow layers may not be fully captured by S3M.

For snow depth, the improvements were less clear across all water year types. The RMSE reduction remained modest, with an average loss of 1.8 cm.

Comparing the multi-site LSTM DA with the site-specific LSTM DA trained over KHT, results show comparable performance for SWE , with neither approach consistently outperforming the other (see Figure 4.9). In some cases, the site-specific model achieves

lower errors, while in others the multi-site model performs equally well or slightly better. For snow depth, however, the multi-site LSTM DA tends to outperform the site-specific LSTM DA across most sites, although the improvements are generally modest (e.g. see Figure 4.9 pannel d).

## 4.5 Discussion of results

In snow-dominated regions, accurate snow estimations are crucial for water resources managing, floods forecasting (Andreadis & Lettenmaier, 2006), and for assessing the impact of climate change on the hydrological cycle (Siirila-Woodburn *et al.*, 2021). Nonetheless, significant uncertainties in model predictions and observational data make accurate snow estimates challenging (Blöschl, 1999). Data assimilation, which integrates both sources, is arguably one of the most effective methods for improving snowpack-model reliability. However, state-of-the-art ensemble-based techniques like the EnKF are computationally intensive, potentially limiting their use in operational contexts. Furthermore, one can argue that it is not just the computational expense but also the time and effort required for parameter tuning, setup, and execution that pose significant challenges to their widespread adoption in such applications.

Here is presented an alternative assimilation framework for snow, which relies on having a LSTM neural network (Adnan *et al.*, 2024; Song *et al.*, 2024) to learn how to perform the filtering updates performed by an EnKF.

The key hypothesis underlying this research was that, leveraging DL, it is possible to preserve the skill of an EnKF, while significantly reducing computational efforts.Testing this hypothesis required the development of an EnKF emulator capable of reproducing the behaviour of the DA algorithm. While the use of DL to approximate or replace DA procedures is becoming increasingly common in the literature (Duan *et al.*, 2024; Demil *et al.*, 2025; Patel *et al.*, 2025), the primary motivation for developing such an emulator in this work was not merely to reproduce improvements in observable outputs, such as SWE and snow depth, but rather to maintain physical consistency across the full set of state variables. In the physical model used for the forecast step, internal state variables are dynamically linked to the assimilated observations through the governing model equations. Consequently, the EnKF analysis step propagates observational information through the multivariate structure of the state vector. By training an LSTM network to reproduce the EnKF analysis, the model is expected to learn these implicit relationships directly from the data. In principle, this enables the emulator to approximate the multivariate update performed by the EnKF, preserving the internal coherence among state variables while avoiding the need to explicitly formulate the underlying relationships within the DL architecture.

First, site-specific LSTMs achieved comparable performances to an EnKF, both in predicting SWE and snow depth, as well as their seasonal patterns, with also a significant

reduction in computational time. Besides this temporal efficiency, the LSTM enabled leveraging a complex tool like the EnKF only for initial training, then replicating its capabilities in operational settings using a faster, simpler DA framework.

To evaluate the computational efficiency of the proposed framework, it was benchmarked against a parellized EnKF. Even though the EnKF already benefits from parallelization during the ensemble prediction step using 15 CPU cores, once trained the LSTM-based approach provided a further 70% reduction in computational time. This result underscores the potential of the framework to significantly lower computational overhead, particularly in scenarios with limited resources or parallelization capabilities.

In line with the work of Guidicelli *et al.* (2024), this finding reinforces the potential of DL for DA in snow hydrology. Yet, the LSTM performance was found to be highly sensitive to the temporal resolution of the input data, which is consistent with findings from other ML studies (Xu & Liang, 2021; Gong *et al.*, 2023b). These results emphasize the importance of acquiring high-frequency snow data to ensure optimal performance and accuracy of modern data-assimilation approaches (Dedieu *et al.*, 2016), highlighting the need for investments in this direction (Cui *et al.*, 2023).

Considering potential developments, the LSTM framework could be extended to provide not only deterministic predictions but also probabilistic outputs, enabling explicit estimation of predictive uncertainty. In this context, Klotz *et al.* (2022) show that LSTM models can be designed to directly predict full probability distributions, providing reliable uncertainty estimates in hydrological applications. Their results indicate that approaches such as mixture density networks are particularly effective, as they capture the complex, non-Gaussian, and heteroscedastic nature of hydrological uncertainty more accurately than alternatives such as Monte Carlo dropout,which, however, is easier to implement.

These findings suggest that integrating probabilistic DL within the proposed framework could further enhance its robustness, especially under data limitations, by providing statistically consistent uncertainty estimates alongside state predictions, and thus improving its applicability for operational data assimilation.

Second, the introduction of memory into the algorithm improved both stability and performance, particularly when working with the inherently noisy outputs of the EnKF and in locations where data sparsity was a major issue. Future efforts could explore additional pre-processing of input data to reduce noise (e.g., smoothing or moving averages), though care must be taken to preserve snow intermittency, which is critical in certain hydrological contexts.

Third, the LSTM trained on a long dataset (KHT) demonstrated some potential for spatial transferability with minimal performance loss, opening avenues for distributed applications of deep DA provided that such long datasets are used in training. Although using limited datasets in both temporal and spatial coverage compared to recent studies (Song *et al.*, 2024), the approach discussed so far proved to be effective in speeding up traditional DA techniques while maintaining comparable performance. Additionally, this framework, designed to test the operational viability of a quasi–real-time pilot point

method, still proved the feasibility of an alternative use of LSTM algorithm without loss in performances. The encouraging results provide a foundation for extending this framework to broader, more diverse networks in future research. The lack of statistically significant correlations between performance and specific climatological variables further supports transferability. According to Karniadakis *et al.* (2021), DL, which usually requires a large amount of data to optimally generalize over samples, has a stronger generalization capability, even in small data regimes, if such algorithms are developed with a physics-informed learning approach. In light of this, a soft physical constraints was introduced into the cost function as a way to incorporate an inductive bias. Although this particular approach did not prove effective in significantly enhancing generalization, considerable potential remains in enforcing snow physical constraints in LSTMs (Charbonneau *et al.*, 2024). Further research is needed in this direction to better understand how such constraints can support model generalization and physical consistency. This finding could contribute to the ongoing debate around the unresolved question of DL models transferability (Pakdehi *et al.*, 2024).

With few exceptions, the comparison of RMSE reductions from the Open Loop to the analysis of the LSTM demonstrated substantial improvements. All but 2 out of the 14 site-specific LSTM frameworks significantly outperformed the Open Loop, although none outperformed the EnKF. Nevertheless, the LSTM ability to deliver marked improvements over the Open Loop underscores its promise as a computationally efficient and effective alternative, even under challenging conditions.

Regularization in particular, and uncertainty-quantification more generally, could be improved by using Bayesian DL (Murphy, 2023). For example, a recent cryospheric study used an ensemble Kalman method (rather than stochastic gradient descent) to train Bayesian neural network with an architecture that was tailored to the problem at hand (Pirk *et al.*, 2024). The contrast between this study, where a neural network learns to mimic the EnKF update, and Pirk *et al.* (2024) where an EnKF method trains an uncertainty-aware neural network, are just some recent examples of the synergies that exist between Bayesian DA and DL. The aforementioned study of Guidicelli *et al.* (2024) also explored how DA and DL could be combined for better uncertainty quantification, not only by having a neural network learn the posterior spread from an EnKF method but also by adopting a simple dropout technique for approximate uncertainty quantification in the neural network outputs. The links between DL and Bayesian DA are well established in the literature (Arcucci *et al.*, 2021; Cheng *et al.*, 2023; Murphy, 2023), but this study emphasizes them once more in this discussion because they are perhaps less known to the snow science community.

The fourth and last key aspect that this study highlighted was no dependency of the performance of this algorithm on dry and average water years, despite a diminished robustness in wet years. Nonetheless, this limitation is shared with both the EnKF and S3M in open-loop, as shown by the distributions in these scenarios. Given the predicted decline in snow cover over the coming decades and the emergence of more frequent snow droughts (Larsson Ivanov *et al.*, 2022), the reduced performance of the algorithm in wet years may have a relatively minor overall impact. Under such a fast-paced changing cli-

mate, a climatically robust LSTM could account for physical processes changing faster than scientists change their models (Cowherd *et al.*, 2024a). Additionally, considering a comparison between two approach, neither the site-specific nor the multi-site LSTM-DA consistently outperforms the other. While multi-site training is theoretically expected to improve generalization by exposing the model to a broader range of conditions, this benefit is not clearly observed for SWE . A likely reason could be an uneven representation of sites, combined with variability in snowpack properties, meteorological drivers, and measurement methods, which may bias the model and introduce noise, leading to underfitting. Snow density also plays a crucial role; SWE is defined as $W = d\rho$, where $W$ is SWE [kg/m$^2$], $d$ is snow depth [m], and $\rho$ is bulk snow density [kg/m$^3$]. A site-specific model such as KHT may implicitly capture a representative density evolution that transfers well across sites, whereas a multi-site model must attempt to generalize density dynamics across all environments, often with less accuracy. Overall, the multi-site LSTM-DA and the EnKF-DA perform similarly, with the latter only marginally better. This is encouraging, as it highlights the potential of the multi-site LSTM-DA to achieve comparable performance while substantially reducing the computational cost associated with ensemble-based methods.

The lower performance of the LSTM algorithm compared to the EnKF can be attributed to several factors. When trained as an emulator of the DA procedure, the network inherently depends on the quality of the reference filter used to generate the training data. Consequently, its predictive capability is constrained by the performance of the EnKF, and it cannot realistically exceed the accuracy of the filter it emulates. Improvements in the LSTM architecture could potentially enhance the adequacy of the predictions. For example, modifications to the network structure or the inclusion of additional regularization terms in the loss function, such as an entropy-based component, could help better capture the variability and uncertainty of the assimilated states.

Another relevant aspect concerns the intrinsic limitations of LSTM models in representing extreme values. As highlighted by Kratzert *et al.* (2024), models trained on data from a single basin may struggle to accurately reproduce maximum levels, as the training dataset may not sufficiently represent rare or extreme conditions. These aspects may explain part of the observed performance differences; however, they likely represent only a subset of the possible causes, and additional factors should be investigated in future analyses to better understand the limitations and potential improvements of the proposed framework.sufficiently represent rare or extreme conditions.

It is important to note that the algorithm showed a significant drop in performance when handling missing or sparse data, contrary to an Ensemble Kalman Filter. Future work in this regard should focus on improving performance under circumstances of high data sparsity, exploring advanced smoothing techniques, and extending transferability even to ungauged catchments. Finally, while the results here showed are based on high-quality forcing and observational datasets, it is worth acknowledging that operational applications may involve lower-quality inputs. In such cases, pre-processing strategies (e.g., bias correction, gap-filling) and hybrid DA–AI frameworks could help mitigate performance loss,

with the potential to selectively down-weight unreliable inputs rather than propagating their errors through the model. Recent work by Gauch *et al.* (2025) demonstrates the effectiveness of imputation and correction methods for handling missing or degraded data in operational environments, while generative models such as those explored by Dhoni (2023) offer promising avenues for enriching and augmenting incomplete datasets.

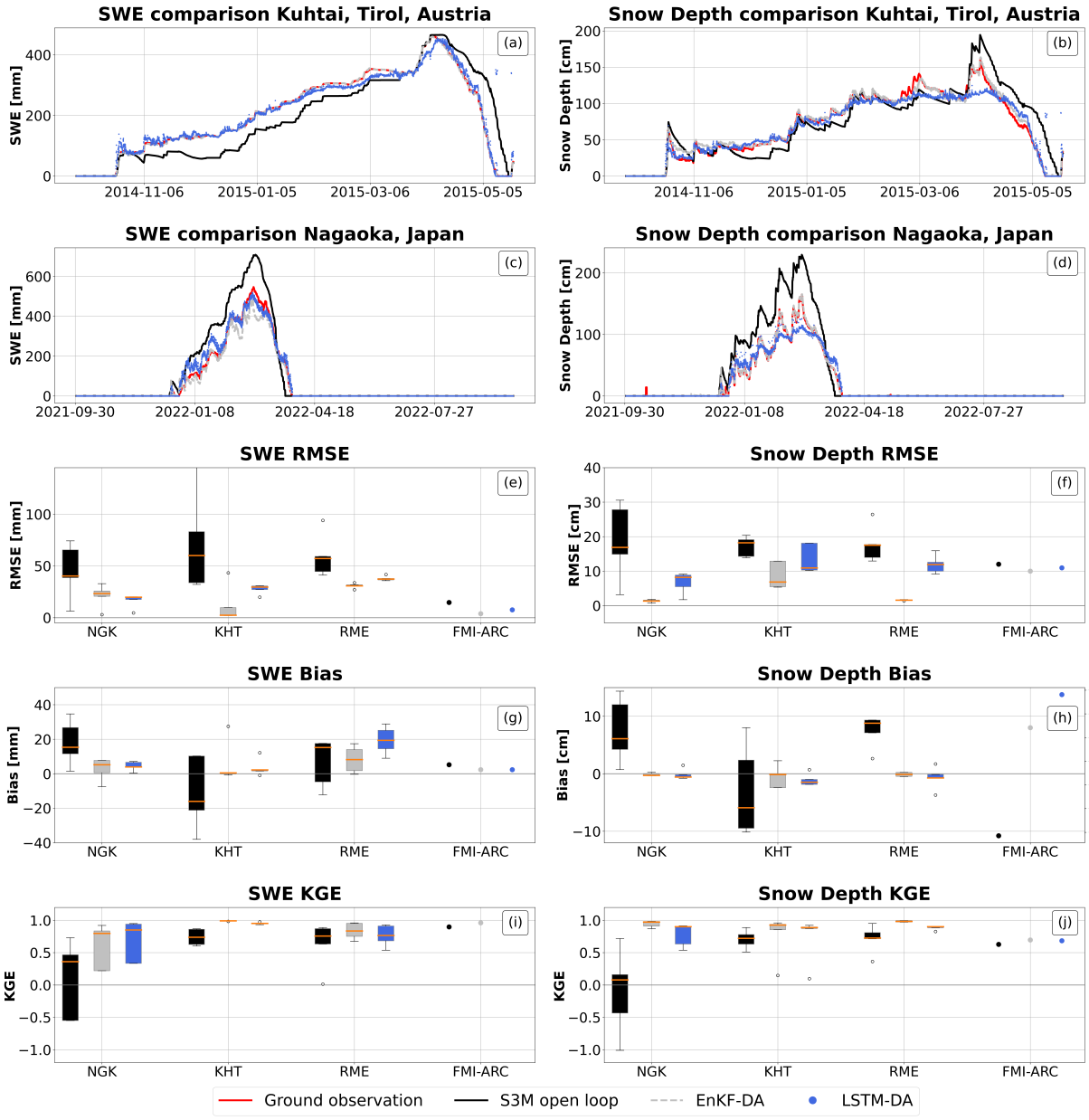


**Figure 4.3:** *Results for sites with low data sparsity, site spefic LSTM. Panels a,b,c,d: comparison between ground observation (red) of SWE (top) and snow depth (bottom) and model estimates by S3M in the open loop (black), using an Ensemble Kalman filter (grey), and using a LSTM neural network (blue) in Kuhtai (row 1) and Nagaoka (row 2). Panels e,f,g,h,i,j: box plots of RMSE, bias and KGE for SWE (pan.e for RMSE, panel f for bias and panel i for KGE) and snow depth (pan.g for RMSE, panel h for Bias and panel j for KGE); fully coloured points represent sites with less than 3 years of validation data, whereas empty circle represents outliers.*

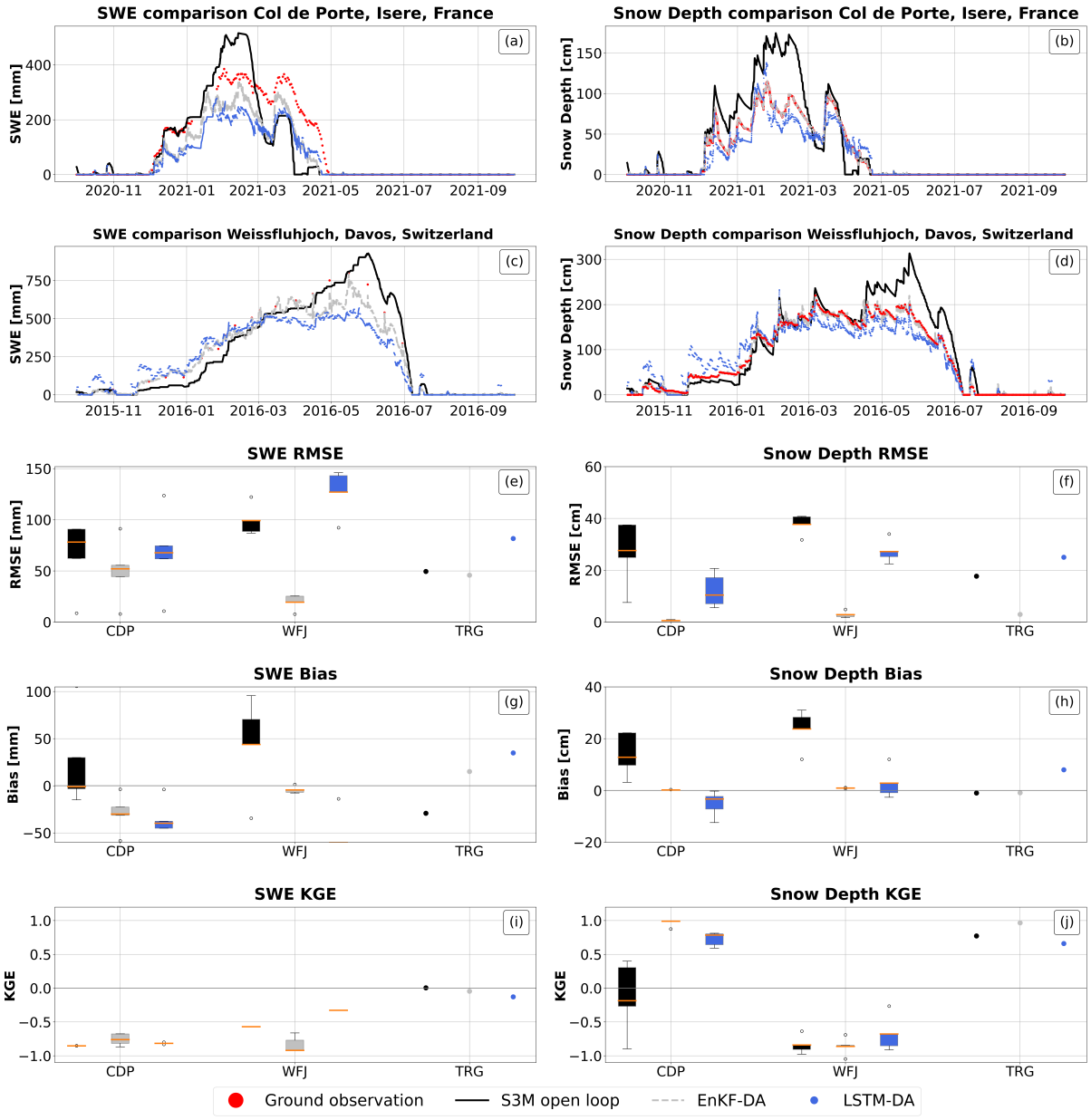


**Figure 4.4:** *Results for sites with high data sparsity, site spefic LSTM. Panel a,b,c,d: comparison between ground observation (red) of SWE (top) and snow depth (bottom) and model estimates by S3M in open loop (black), using an Ensemble Kalman filter (grey), and using a LSTM neural network (blue) in Col de Porte (row 1) and Weissfluhjoch (row 2).Panels e,f,g,h,i,j: box plots of RMSE, bias and KGE for SWE (panel e for RMSE, panel f for bias and panel i for KGE) and snow depth (panel g for RMSE, panel h for Bias and panel j for KGE); fully coloured points represent sites with less than 3 years of validation data, whereas empty circle represents outliers.*

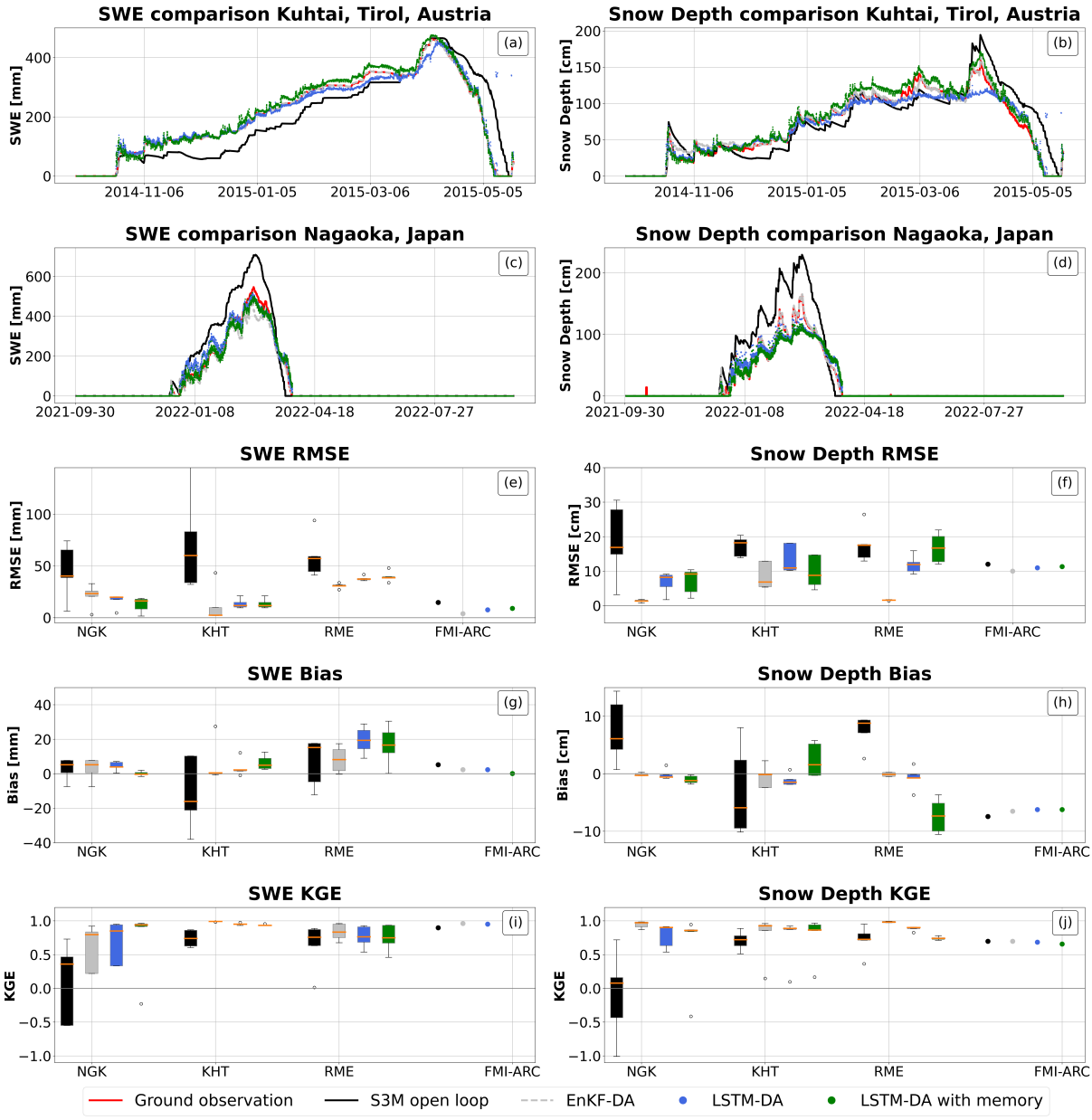


**Figure 4.5:** *Results for sites with low data sparsity. Panel a,b,c,d: comparison between ground observation (red) of SWE (top) and snow depth (bottom) and model estimates by S3M in open loop (black), using an Ensemble Kalman filter (grey), using a LSTM neural network (blue) and using a LSTM neural network with memory (light blue) in Kuhtai(row 1) and Nagaoka (row 2). Panels e,f,g,h,i,j: box plots of RMSE, bias and KGE for SWE (panel e for RMSE, panel f for bias and panel i for KGE) and snow depth (panel g for RMSE, panel h for Bias and panel j for KGE); fully coloured points represent sites with less than 3 years of validation data, whereas empty circle represents outliers.*

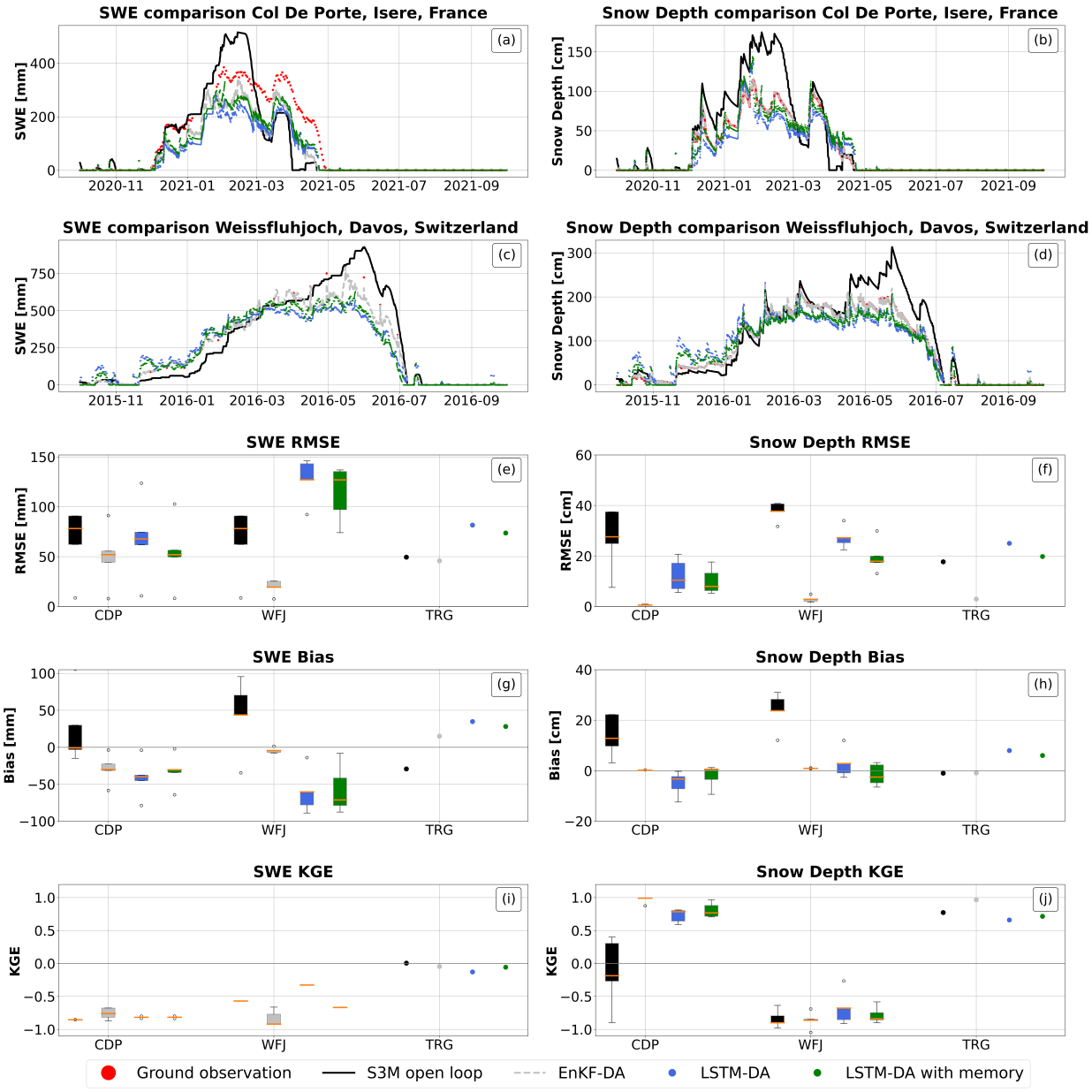


**Figure 4.6:** *Results for sites with high data sparsity. Panel a,b,c,d: comparison between ground observation (red) of SWE (top) and snow depth (bottom) and model estimates by S3M in open loop (black), using an Ensemble Kalman filter (grey), using a LSTM neural network (blue) and using a LSTM neural network with memory (light blue) in Col de Porte (row 1) and Weissfluhjoch (row 2). Panels e,f,g,h,i,j: box plots of RMSE, bias and KGE for SWE (pan.e for RMSE, panel f for bias and panel i for KGE) and snow depth (pan.g for RMSE, panel h for Bias and panel j for KGE); fully coloured points represent sites with less than 3 years of validation data, whereas empty circle represents outliers.*

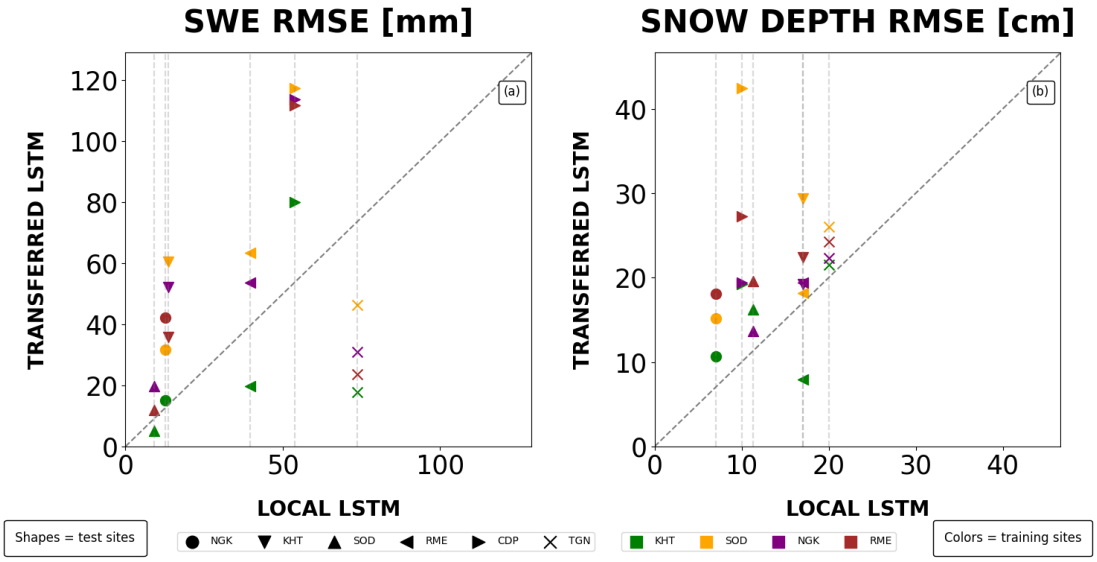


**Figure 4.7:** *Spatial transferability of site-specific LSTMs for SWE and snow depth estimation. Panel (a) shows a comparison between the RMSE for SWE obtained by using each LSTM at the training site (x-axis) and the RMSE obtained when transferring the same LSTM to other sites (y-axis). Panel (b) shows the same information, but for snow depth. The bisectors in the two panels represent the one-to-one lines comparing the RMSE values for SWE and between the site-specific LSTM and the LSTM trained on a different site. The dotted lines in both panels serve as benchmarks, indicating the RMSE values achieved by the site-specific LSTM models. Colors represent training sites, while shapes correspond to the to sites where each LSTM was applied. The lowest granularity site, WFJ, is excluded.*

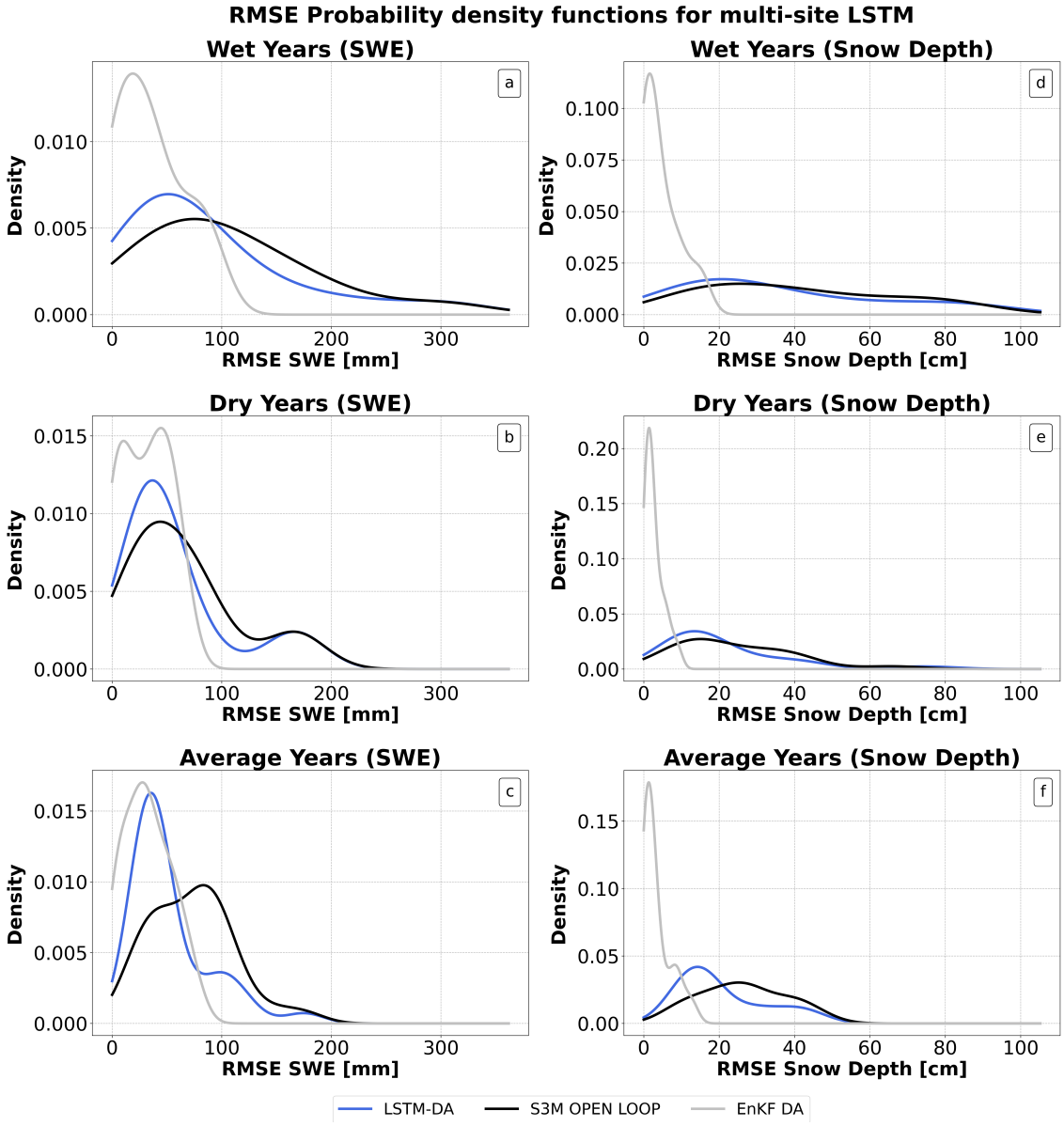


**Figure 4.8:** *RMSE distribution for SWE and snow depth across water year types RMSE distribution for SWE on wet, dry and average years type (panels a,b,c) and snow depth (panels d,e,f) under varying water year types: wet, dry, and average conditions.*

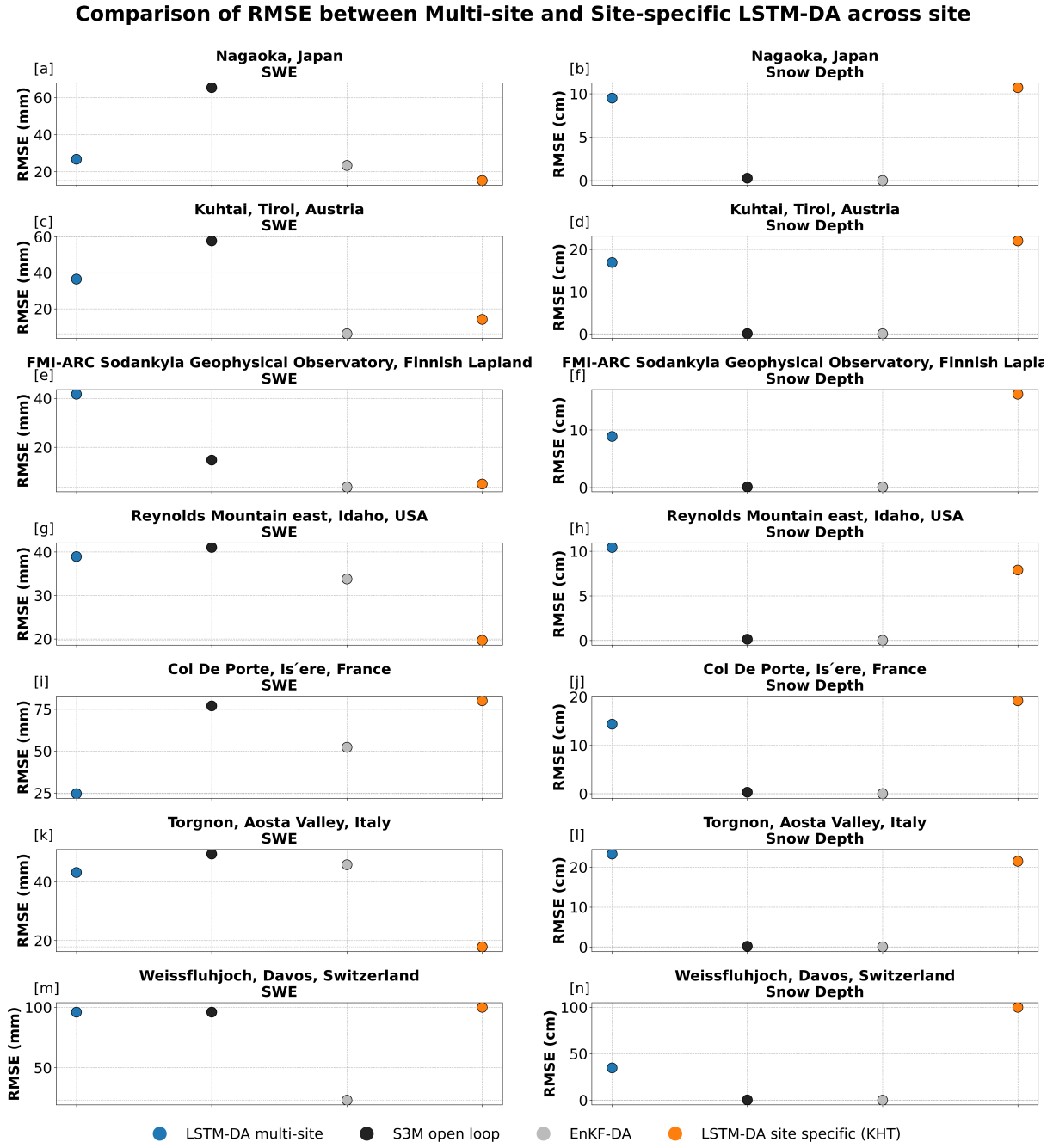


**Figure 4.9:** *Comparison of RMSE for SWE (SWE , left column) and snow depth (right column) across multiple sites and methods. Results are shown for LSTM-DA multisite (blue), S3M open loop (black), EnKF-DA (grey), and LSTM-DA site-specific (orange).*

# Chapter 5

# Deep learning-based data assimilation in 2D

Results from the 1D LSTM-based EnKF emulator, trained across multiple locations, demonstrate the feasibility of developing a spatially distributed emulator from point-scale information within the same catchment. Recent literature further supports this strategy. For instance, Kratzert *et al.* (2024) emphasize the importance of exploiting information from multiple locations when modelling complex hydrological processes, while Lievens *et al.* (2019) highlight the need to explicitly account for both temporal and spatial variability in snow dynamics. Together, these findings motivate the extension of the hybrid AI–DA framework toward a two-dimensional, operational setting, and call for the definition of a clear and robust methodology for doing so with limited resource and computational costs.

Indeed,being conceived as an emulator, the LSTM training process relies on large training datasets, generated through extensive ensemble-based DA experiments. At the 2D scale, the computational cost associated with constructing and evolving such ensembles would rapidly become prohibitive in context of limited resources and low budget, thereby limiting the applicability of fully ensemble-based approaches within an operational modelling chain.

To overcome this limitation while still exploiting both temporal and spatial correlations in snow properties, a hybrid strategy is required.

This chapter is intended as a preliminary exploration, setting out the reasoning and defining the key steps needed to implement the DL-based DA framework in a two-dimensional context.

## 5.1 Point data assimilation for 2D snow variable estimation

The proposed approach is based on performing point-wise DA at a limited number of strategically selected reference locations, followed by the spatial propagation of the re-

sulting corrections over the entire basin. The framework will couple an LSTM-based DA scheme at point scale with a Gaussian Process (GP, Williams & Rasmussen 2006) interpolation algorithm, enabling the transfer of local information to a continuous two-dimensional domain. Gaussian Process models belong to the broader class of *kernel machines*.

To introduce the basic principles underlying Gaussian Process (GP) algorithms, let us consider a supervised learning problem. Let the training dataset be defined as:

$$\mathcal{D} = \{(\mathbf{x_k}, \mathbf{y_k}) \mid k = 1, \ldots, n\}, \tag{5.1}$$

where:

- $n$ is the number of observations.
- $\boldsymbol{x}_k$ represents the $k-th$ input variable.
- $\boldsymbol{y}_k$ is the corresponding output variable.

The objective is to make predictions for new input values that are not included in the training set. This problem can be formulated as the task of inferring a function $f$ that maps input values to outputs, based on the finite dataset $\mathcal{D}$, while remaining valid for all possible inputs. To achieve this, assumptions about the properties of the underlying function must be introduced. Without such assumptions, any function that interpolates the training data would be equally plausible, leading to a risk of overfitting.

Two common modelling strategies can be adopted. The first consists of restricting the hypothesis space by limiting the class of admissible functions, for example to linear functions of the input variables. While conceptually simple, this approach requires the specification of the functional class a priori. If the true target function lies outside this class, the resulting predictions will be inaccurate. Increasing the flexibility of the function class may improve the fit to the training data, but it also increases the risk of overfitting, whereby the model captures noise rather than the underlying signal and performs poorly on unseen data. This approach is the strategy underling the AI-based algorithm developed so far.

The second strategy consists of assigning a prior probability distribution over the space of all possible functions, such that functions deemed more plausible are assigned higher prior probability. This prior is then updated, using Bayesian inference, as observations $(\boldsymbol{x}_k, \boldsymbol{y}_k)$ become available. Gaussian Process algorithms follow this second approach. A GP can be interpreted as a generalization of the Gaussian probability distribution: while a Gaussian distribution characterizes random variables that are scalars or vectors, a stochastic process defines a probability distribution over functions (Zaparoli Cunha *et al.*, 2023).

Given a finite set of input locations, a GP defines a joint probability distribution over the corresponding function values $f(\boldsymbol{x})$. Compared to traditional neural networks, GPs make weaker structural assumptions by directly defining a prior over functions rather than

over model parameters. A GP is fully characterized by its mean function $\mu(\boldsymbol{x})$ and its covariance function, or kernel, $k(\boldsymbol{x}, \boldsymbol{x}')$, such that

$$f(\boldsymbol{x}) \sim \mathcal{GP}\left(\mu(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}')\right). \tag{5.2}$$

Conditioned on the observed data, a GP yields a predictive distribution for unseen inputs, providing both a predictive mean and variance. This feature enables an explicit quantification of predictive uncertainty, which is particularly valuable in geophysical and hydrological applications. The learning process is therefore governed by the choice of the covariance function and by the estimation of its hyperparameters, which are typically inferred by maximizing the marginal likelihood of the posterior distribution (Williams & Rasmussen, 2006).

Despite these advantages, GP methods suffer from significant computational limitations when applied to large datasets (e.g., $n > 10^4$). In particular, GP inference requires the inversion of an $n \times n$ covariance matrix, resulting in a memory complexity of $\mathcal{O}(n^2)$ and a computational complexity of $\mathcal{O}(n^3)$. Consequently, when extending GP-based interpolation to large spatial domains or large catchments, it becomes necessary to consider approximate formulations of the covariance function or sparse GP methods, which will be explored in future implementations.

The conceptual workflow, at the basis of the future step of my research, is illustrated in Figure 5.1.

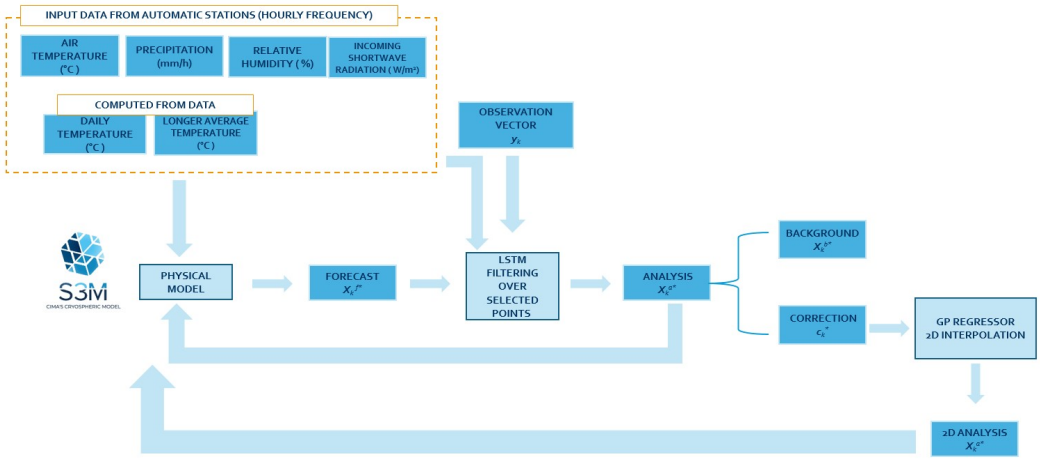


**Figure 5.1:** *Operational Setup Point data assimilation for 2D Snow Variable Estimation. This diagram illustrates the operational workflow for integrating observational data with the S3M (CIMA's Cryospheric Model) framework, through DA via a Long-Short-Term-Memory neural network over a selected subset of points; then the interpolation of the correction is made by using a Gaussian Process regressor.*

The LSTM-based EnKF emulator will initially be trained on a subset of representative points within the catchment. These points will be selected using a KMeans clustering algorithm, which groups locations based on topographic and microtopographic features (Maurer *et al.*, 2021b). Once trained, it will be coupled with forecasts from the physical model to generate pointwise analysis values. The series of correction terms, $\boldsymbol{c}_k^* \ \forall k$, computed as the difference between the analysis $\boldsymbol{x}_k^{a,*} \ \forall k$ and the background $\boldsymbol{x}_k^{b,*} \ \forall k$, will be then used to train a Gaussian Process regressor. This regressor will interpolate the corrections across the entire domain, enabling the derivation of 2D snow depth and SWE estimates from the 1D assimilation outputs.

## 5.2 Validation of the framework

This framework is currently being validated leveraging meteorological forcing data from the Aosta Valley, and assimilating the region's dense network of ground-based snow observations already presented in chapter 3.

## 5.3 Future outlook

By developing and validating this AI-based 2D DA framework, the overarching goal is to enable a more accurate and accessible representation of snow dynamics. Leveraging satellite observations, this approach could provide reliable 2D estimates of snow depth and SWE even in ungauged basins, reducing the dependence on extensive in-situ measurements. This methodology aims to democratize advanced snow modelling, making high-resolution snow simulations more widely available and potentially improving performance beyond the limitations (Tyson *et al.*, 2023) of reanalyses alone (Jiang *et al.*, 2021; Lavers *et al.*, 2022; Kouki *et al.*, 2023). Moreover, applying this framework to underexplored regions would enhance the understanding of snow processes across climatically diverse areas, with important implications for global water resource management.

When selecting basins for future analysis, the main objective has been to consider regions where snow significantly influences water availability, while also prioritizing understudied areas. The Mediterranean and Near East regions, in particular, present compelling candidates for further investigation. These regions share a complex geological history shaped by plate tectonics, orogenic processes, and climatic influences, resulting in comparable snowpack characteristics across their mountain ranges (Cavazza & Wezel, 2003). Situated along the Alpine-Himalayan orogenic belt, they are affected by the convergence of the African, Eurasian, and Arabian plates (Kahle & Mueller, 1998).

Notable mountain ranges shaped by this tectonic activity include:

- The Alps (Europe), Taurus Mountains (Turkey), and Zagros Mountains (Iran/Iraq), characterized by fold-and-thrust structures and extensive carbonate platforms.

- The Pyrenees and Atlas Mountains (North Africa), which share tectonic similarities with the Lebanese Mountains and Anti-Taurus in Turkey due to Neogene compression.
- The Apennines (Italy), Pontic Mountains, and Caucasus, all influenced by Neogene compressional tectonics, with the Caucasus showing structural deformation related to the Arabia-Eurasia collision.

These regions also exhibit similarities in snowpack dynamics, as categorized by the *Seasonal Snow Classification* (Sturm & Liston, 2021b), ranging from Maritime to Ephemeral Snow. Shared climatological influences, such as the Atlantic Ocean, further align these regions with other relevant areas like the Tibetan Plateau (TP). Often referred to as the "Third Pole," the TP is a critical component of regional and global climate systems, influencing atmospheric circulation patterns, particularly the Asian Summer Monsoon, and serving as the source of major rivers sustaining freshwater resources for billions of people (Miao *et al.*, 2024). Its glaciers and snowpacks are highly sensitive to climate change, with rising temperatures driving glacier retreat and altered snow cover, complicating regional hydrology (Zhao *et al.*, 2024a). Interestingly, the TP's seasonal snow cover, shallow, patchy, and short-lived, shares characteristics with Mediterranean and Near East snowpacks, suggesting that AI-based modelling approaches developed in one region may be transferable to others.

Achieving this goal requires a clear strategy, leveraging datasets with high quality and coverage to train AI algorithms capable of generalizing without significant performance loss. The proposed plan is to start with a framework trained on the Mediterranean Mountains and then extend it to other regions. The assimilation of C-SNOW data (Lievens *et al.*, 2022), treated as ground truth, would serve as a representative subset for calibration. Virtual 1D snow gauges would be sampled from C-SNOW data to ensure coverage across different slopes, aspects, and geographical locations. Following assimilation, corrections would be spatially interpolated using a Gaussian Process algorithm. For meteorological forcing, ERA5 or other widely available reanalysis datasets will be employed, enabling open-loop simulations across multiple basins.

# Chapter 6

# General discussion and conclusion

The snow hydrology community broadly agrees on the importance of monitoring snowpack evolution to better capture its implications for water resources and ecosystem functioning (Hatchett *et al.*, 2022), as well as society as a whole (including, for example, winter tourism). This has led to a growing emphasis on advancing both measurement strategies and modelling approaches. AI has demonstrated strong potential in this context (Revuelto *et al.*, 2025), as it enables the integration of heterogeneous data sources, enhances the extraction of meaningful information, and efficiently handles the rapidly increasing volume of available data (Karpatne *et al.*, 2018). Indeed, recent advances in computational capacity and AI have made AI-based methods increasingly attractive within the Earth science paradigm, including cryospheric research (Zhao *et al.*, 2024b).

While working on this dissertation, I viewed AI as a complementary tool capable of enhancing our understanding of complex systems (Shen *et al.*, 2018). From this perspective, AI can reduce technical and computational burdens while opening new opportunities for theoretical advancement rather than replacing it. At the same time, my experience has been that these approaches are better appreciated by the scientific and operational communities when guided by expert knowledge, which is essential for constraining model behaviour, interpreting outputs, and ensuring that AI-based results are physically meaningful and suitable for societal applications (Nearing *et al.*, 2021).

Since the beginning of my PhD in 2022, the sparse literature of AI application in cryosphere science has been increasingly enriched, with studies investigating AI applicability across most cryospheric components. Random forest (RF) methods have been widely used in cryospheric research, including seasonal snow modelling (Ferrarin *et al.*, 2025), snow phenology (Pan *et al.*, 2025a), the estimation of fractional snow cover (Kuter, 2021; Liu *et al.*, 2020b; Rittger *et al.*, 2021b) and snow density (Sun *et al.*, 2024). RF has also been applied to avalanche prediction (Mayer *et al.*, 2022), basin-scale snow depth estimation through the fusion of LiDAR, in situ, and satellite data (Herbert *et al.*, 2025), and comparative assessments of machine-learning models for high-resolution snow variables retrieval

from multisensor observations (Dunmire *et al.*, 2024). On the other hand, DL applications in snow research have so far been more limited and have primarily relied on simple neural network architectures to estimate fractional snow cover (Moosavi *et al.*, 2014; Liu *et al.*, 2020b), snow albedo (Chevrollier *et al.*, 2025), and SWE (Bair *et al.*, 2018b). More advanced architectures, such as recurrent and convolutional neural networks, are emerging, including approaches to improve in situ snow observations (Svoboda *et al.*, 2025), hybrid DL–ensemble DA frameworks (Guidicelli *et al.*, 2024) or glacier modelling (Bolibar *et al.*, 2020; Ren *et al.*, 2024).

In light of this growing body of literature, the use of AI to enhance cryospheric science provided the conceptual foundation of this PhD research, offering a powerful means to advance cryospheric science. The integration of AI into hydrological and cryospheric modelling represents a transformative opportunity to improve the understanding and prediction of complex environmental processes (Taylor *et al.*, 2021). Rather than positioning physically based and data-driven approaches as competing paradigms, this work is grounded in the hypothesis that their synergy is essential for the next generation of environmental models. When informed by domain expertise and physical knowledge, AI augments rather than replaces traditional modelling frameworks (Shen *et al.*, 2018). Such hybrid approaches enhance computational efficiency and scalability, and preserve physical consistency and interpretability. A key outcome of this research is the demonstration that expert-informed data-driven methods can effectively capture the nuanced dynamics of snowpack evolution while substantially reducing computational cost and technical complexity (Tahmasebi *et al.*, 2020).

Building on the considerations discussed above, my PhD research main contribution are: (i) A Random forest automated quality check algorithm for in situ snow depth observations, and (ii) a 1D EnKF emulator based on a LSTM algorithm. Both algorithms aim at reducing the computational and time cost of such expensive procedures, especially in operational chain.

Indeed, the quality and reliability of snow observations play a critical role in both the accuracy of model outputs and the efficiency of operational snow-modelling chains. Noise and inconsistencies in high-resolution snow-depth data substantially limit their effective use in snow models as this poor-quality observations can propagate errors throughout the modelling workflow, increasing computational effort, wasting resources, and raising operational costs. In the most critical cases, these issues may compromise the timely delivery of early-warning information. For instance, snow-depth measurements from ultrasonic sensors are prone to snow vs. grass ambiguity and random noise. Data pre-processing procedures (e.g.quality accuracy and quality control) often considered as one of the most time consuming step of any modelling chain, have been traditionally carried out by visual being not easily reproducible or transferable more so considering the increasing volume of available data. In this context, the Random forest–based quality-control algorithm developed in this research represents an effective example of how expert domain knowledge can be formalized and replicated through a data-driven yet physically informed approach.

The method provides a fully automatic, computationally efficient, and scalable alternative to manual screening, significantly reducing the time and resources required for data quality assurance. The proposed Random forest demonstrates robust performance and strong generalization capabilities across large spatial domains, achieving F1 scores exceeding 90% for the detection of snow versus grass or bare ground, even in regions outside the original training dataset. Furthermore, the algorithm exhibits limited sensitivity to differences in snow-season climatology, indicating its applicability across heterogeneous environmental conditions. Nonetheless, some underestimation of rare and highly irregular random errors persists, highlighting an area for further methodological refinement in future work. Overall, the Random forest approach can be readily integrated into both supervised and unsupervised snow-data processing workflows, providing a reliable and efficient component for operational snow-depth quality control.

Beyond data preprocessing, the operational application of advanced DA techniques remains constrained by their computational and time requirements, particularly for ensemble-based methods. The DL–based DA framework developed in this research addresses these key limitations by leveraging an EnKF emulator.The emulator was designed not only to reproduce observable outputs, such as SWE and snow depth, but also to preserve the physical consistency across all state variables. By training an LSTM to replicate the EnKF analysis, the model learns the implicit multivariate relationships in the data, maintaining the internal coherence of the state vector without explicitly encoding these dependencies in the DL architecture. As a result, the LSTM-based emulator achieves snow depth and SWE estimation performance comparable to the full EnKF, while substantially reducing computational cost. Indeed, when benchmarked against a parallelized EnKF, it achieved an additional 70% reduction in computational time. This highlights its potential for operational deployment, particularly in resource-limited environments or where extensive parallelization is not feasible. Contrary to a traditional EnKF, the proposed configuration deploys ensemble simulations only during the training phase, while inference is performed without ensembles, significantly enhancing operational efficiency. Beyond temporal performance gains, the framework demonstrated notable spatial transferability. When applied to regions outside the training domain, SWE estimation performance decreased by only approximately 20%, indicating robust generalization across spatial contexts and hydroclimatic conditions, including dry and average water years. Preliminary multi-site training experiments further showed that a single LSTM model trained on multiple locations can generalize effectively to unseen sites, supporting the feasibility of scalable, region-wide implementations. Nonetheless, the framework exhibits limitations under sparse or discontinuous observational scenarios. Addressing these challenges represents a key direction for future work, through the development of a distributed framework incorporating advanced input smoothing techniques and the integration of complementary data sources, such as remotely sensed observations. Overall, the proposed approach enables the use of computationally intensive ensemble methods primarily during training, while replicating their assimilation capabilities in operational settings through a faster and more streamlined

data-driven framework.

The results obtained from the pointwise AI-based EnKF emulator open the door to developing a spatially distributed deep DA framework. Building on this idea, the final part of my research extended the application of LSTMs from single-point setups to multiple representative points within a catchment. This approach leverages the LSTM-based DA emulator and incorporates a Gaussian Process interpolation algorithm to propagate corrections across the entire domain. By doing so, it offers a practical compromise between the need for spatially distributed corrections and the computational constraints of full-domain deep DA. The implementation of such an approach is ambitious, as it requires the development of stable algorithms capable of operating on a subset of catchment locations while generalizing effectively across the remaining areas without loss of performance. Thus, in this dissertation I focused on envisioning the theoretical approach and testing all pieces of the algorithm, leaving for future work the full-scale implementation at catchment scale.

It is worth mentioning that, despite its proven achievements, AI has raised scepticism within both the hydrological and cryospheric scientific communities, largely due to its "black-box" nature, which limits transparency regarding the processes leading to a given output. This lack of interpretability is often perceived as undermining the scientific rigour and physical coherence of AI-based approaches (Núñez *et al.*, 2023). In parallel, broader concerns have emerged that the increasing capabilities of AI could supersede human expertise, potentially diminishing the role of domain knowledge and scientific insight (Nearing *et al.*, 2021). Additionally, performance of AI models remains strongly conditioned by the representativeness of the training data, which constrains their extrapolation capability under non-stationary climatic or hydrological regimes (Gröger, 2021). This is particularly critical in the cryospheric domain, where shifts in temperature and precipitation regimes can alter snowpack behaviour beyond the historical range captured in available datasets (Trenberth, 2011).

While the computational efficiency of AI-based surrogates represents a clear advantage for cryosphere and more broadly hydrosphere modelling framework, it should not come at the expense of process transparency and interpretability. Ensuring explainability remains a crucial requirement, particularly in operational and decision-support contexts, where trust, accountability, and physical consistency are essential. Scientists have the role of bridging the gap between time-efficient outputs and interpretable results, improving the readability of AI inputs (Nearing *et al.*, 2021; Shen *et al.*, 2018), while still taking full advantage of the opportunities that AI can offer across the Earth sciences.

Lastly, but equally important, I strongly believe that the cryosphere and hydrology community's efforts toward integrating AI solutions into traditional modelling frameworks have matured beyond the exclusive use of Random forest algorithms. Despite their ease of implementation and relatively transparent, traceable decision-making structure, future research would benefit from adopting more complex, perhaps even more "black-box", models capable of extracting deeper and more nuanced information from the data.

In conclusion, this PhD research advocates for a balanced paradigm in which AI strate-

gically complements, rather than replaces, physically based models. This synergy enables deeper scientific insight and more agile environmental prediction systems. By leveraging the strengths of both approaches, future research can advance robust, interpretable, and scalable models capable of addressing the multifaceted challenges of a changing climate.

# Appendix A

# Kalman filters

Kalman filters (Welch, 1995) are sequential DA techniques that optimally combine model forecasts and observations based on their respective Gaussian error covariances (Särkkä & Svensson, 2023). Using standard error propagation theory (Gupta & Govindaraju, 2019), the analysis state is obtained by correcting the forecast (or prior) state with the Kalman Gain, which weights the difference between forecast and observation according to model and observation error covariances. This produces a statistically optimal estimate of the system state, corresponding to the mean of the posterior probability density function.

Rooted in least-squares analysis, the Kalman Filter relies on assumptions of linearity and Gaussianity. It alternates between a forecast step and an analysis step, sequentially updating both the state estimate and its error covariance. Still based on the Gaussian hypothesis, but relaxing the linear one, the ensemble-based models are based on a Monte Carlo approach. Among these methods, the Ensemble Kalman Filter (Evensen, 2003) EnKF approximates the posterior PDF through an ensemble of model realizations. Instead of computing the error covariance matrices, thus having to solve a sometimes computationally expensive inverse problem, the EnKF estimates the mean and covariance error matrices from the ensemble.

Mathematically, the Kalman filter (Evensen *et al.*, 2022; Särkkä & Svensson, 2023) cycles between two steps:

1. **Prediction step** (also known as the *forecast* or *prior*), where the state is propagated from the previous time $t_{k-1}$ to the current time $t_k$ using a dynamical model.
2. **Update step** (also known as the *analysis* or *posterior*), where the state is corrected by assimilating available observations.

The analysis step can be expressed as:

$$\boldsymbol{x}_{k,a} = \boldsymbol{x}_{k,b} + \mathbf{K}_k \left( \boldsymbol{y}_k - \mathbf{H}_k \boldsymbol{x}_{k,b} \right) \tag{A.1}$$

where:

- $\boldsymbol{x}_{k,b} \in \mathbb{R}^n$ is the background (forecast or prior) model state at time $k$,
- $\boldsymbol{x}_{k,a} \in \mathbb{R}^n$ is the analysis (posterior) model state at time $k$,
- $\boldsymbol{y}_k \in \mathbb{R}^m$ is the observation vector at time $k$,
- $\mathbf{H}_k : \mathbb{R}^n \to \mathbb{R}^m$ is the observation operator mapping model space to observation space at time $k$,
- $\mathbf{K}_k \in \mathbb{R}^{n \times m}$ is the Kalman gain at time $k$, defined as:

$$\mathbf{K}_k = \mathbf{P}_{k,b}\mathbf{H}_k^T \left(\mathbf{H}_k \mathbf{P}_{k,b} \mathbf{H}_k^T + \mathbf{R}_k\right)^{-1}, \tag{A.2}$$

where $\mathbf{P}_b$ is the background error covariance matrix and $\mathbf{R}$ is the observation error covariance matrix.

The Kalman gain $\mathbf{K}_k$ acts as a weighting factor, balancing the contribution of the model forecast against the observations in the innovation term $(\boldsymbol{y}_k - \mathbf{H}\boldsymbol{x}_{k,b})$. This weighting reflects the relative uncertainties in the model (through $\mathbf{P}_b$) and in the observations (through $\mathbf{R}$).

Although the classical Kalman filter is still widely used in signal processing and related fields (Särkkä & Svensson, 2023), it requires linear models with Gaussian error statistics. To overcome this limitation, several non-linear extensions have been developed. Among these, the Ensemble Kalman Filter (EnKF) is particularly well suited for high-dimensional, non-linear geoscientific models (Carrassi *et al.*, 2018; Evensen *et al.*, 2022). Together with particle filtering methods, the EnKF belongs to the class of ensemble-based methods, which currently represent the state of the art for snow DA (Aalstad *et al.*, 2018; Alonso-González *et al.*, 2022).

The EnKF avoids explicit linearization by representing error statistics with an ensemble of model realizations. Through this Monte Carlo approach (Evensen, 2003), the forecast error covariance matrix is approximated empirically from the ensemble:

$$\mathbf{P}_b = \mathrm{Cov}(\mathbf{Ens}_i - \mathbf{Ens}_{mean}), \tag{A.3}$$

where:

- $\mathbf{Ens}_i$ is the $i$-th ensemble member,
- $\mathbf{Ens}_{mean}$ is the ensemble mean.

In principle, an infinitely large ensemble would recover the true error covariance. In practice, finite ensembles introduce sampling errors, with approximation errors decreasing at the rate of $1/\sqrt{N}$, where $N$ is the number of ensemble members.

The ensemble members are propagated forward in time using the dynamical model and are updated whenever new observations become available. The update step, as in

the standard Kalman filter, optimally combines the information from model forecasts and observations, with the Kalman gain determining the relative weighting.

According to Girotto *et al.* (2020), there is common agreement in the snow hydrology community on the advantages of assimilation of SWE or microwave radiance observations through Ensemble Kalman filters; an extensive literature proves overall improvements in estimates of seasonal snow and related variables (e.g., streamflow and snow cover) (Huang *et al.*, 2017b; Alonso-González *et al.*, 2022; Metref *et al.*, 2023a). For examples Huang *et al.* (2017b) improved streamflow prediction trough SWE assimilation via an Ensemble Kalman Filter.

In the snow hydrology community, there is broad agreement on the advantages of assimilating snow-related variables, such as snow water equivalent (SWE), snow depth, and remotely sensed snow cover (Snow Cover (SC)) or microwave radiance observations, through EnKF. The EnKF has been extensively applied in snow DA studies (Dunmire *et al.*, 2026b).

Early work primarily focused on coarse-scale assimilation of in situ SWE and snow depth measurements, demonstrating improvements in snow state estimates over large river basins (Slater & Clark, 2006; Durand & Margulis, 2006a; Andreadis *et al.*, 2009; De Lannoy *et al.*, 2012; Huang *et al.*, 2017a; Pflug *et al.*, 2024).In recent years, research has moved toward high-resolution snow assimilation over complex terrain. Girotto *et al.* (2020) and Alonso-González *et al.* (2022) demonstrated that assimilating SWE and snow depth at finer spatial scales significantly improves representation of snow state variability. The most recent studies (Brangers *et al.*, 2024; De Lannoy *et al.*, 2024; Dunmire *et al.*, 2026a) further advance the field by integrating high-resolution observations and adaptive observation error modelling, better capturing local-scale snow dynamics and uncertainty. Across this literature,EnKF consistently improves seasonal snow estimates and related hydrological variables, including modelled snow depth, SWE, and streamflow, relative to in situ measurements, highlighting its value for both research and operational forecasting (Mirza *et al.*, 2025).

# Appendix B

# Particle filters

Another Monte Carlo approach to the DA problem is provided by The Particle filter (PF) (Moradkhani *et al.*, 2005); the PF allows to draw $N$ samples from the Bayesian posterior (Eq. 2.3) under the assumption that, in the asymptotic limit $N \to \infty$, the conditional density $p(\mathbf{x} \,|\, \mathbf{y})$ can be properly approximated by the ensemble. In its empirical form, the posterior at time $t_{k+1}$ can be written as:

$$p(\boldsymbol{x}_k \,|\, \boldsymbol{y}_k) = \sum_{i=1}^{N} \boldsymbol{w}_k^{(i)} \, \delta\Big(\boldsymbol{x}_k - \boldsymbol{x}_k^{(i)}\Big), \tag{B.1}$$

where $\boldsymbol{x}_k^{(i)}$ denotes the $i$-th particle at time $t_k$, and $\boldsymbol{w}_k^{(i)}$ are the associated normalized, positive weights representing the probability of each particle (Carrassi *et al.*, 2018). The PF therefore approximates the forecast and analysis probability density functions using a discrete set of Dirac delta functions(Schiano Di Cola *et al.*, 2021), one per particle.

A key challenge in particle filtering is weight degeneracy; as the assimilation proceeds, the importance sampling step (the sampling of particles with higher weights) often results in most particles receiving negligible weights, concentrating the probability mass in only a few of them. To mitigate this issue, the sequential Importance Resampling (SIR) (Plaza *et al.*, 2012) algorithm is commonly adopted. SIR periodically replaces low-weight particles with replicas of high-weight ones, thereby stabilizing the ensemble representation of the state distribution (Piazzi *et al.*, 2018a).

However, in snowpack modelling, where state variables such as SWE , density, and albedo may exhibit strong non-linearities and non-Gaussian behaviour, this resampling step poses non-trivial difficulties. In particular:

- Multivariate likelihoods amplify weight collapse, especially when combining uncertainties from several observed snow variables.
- Infrequent observations of mass-related variables (e.g., SWE or bulk density) accelerate sample impoverishment, because particles cannot be sufficiently differentiated

between assimilation events.

- Perturbing atmospheric forcings alone is typically insufficient to maintain adequate ensemble spread.
- Parameter perturbation can partly mitigate this issue by injecting structural variability, but increases sensitivity to the timing and frequency of observations.

At sites with sparse SWE or density data, integrating proxy information through empirical snow density models significantly improves filter performance by providing physically meaningful constraints. Notably, increasing the ensemble size does not substantially enhance the assimilation results: computationally expensive large ensembles fail to compensate for insufficient physical diversity or inadequate observational support (Carrassi *et al.*, 2018).

Overall, the effectiveness of the Particle Filter in snow DA depends less on brute-force sampling, and more on maintaining ensemble diversity through physically justified perturbations, while ensuring a sufficient observational backbone to anchor the evolving state distribution.

Particle Filter (Particle Filter (PF)) approaches have increasingly been applied to snow DA over the last decade, demonstrating improvements in the estimation of snowpack properties. Piazzi *et al.* (2018a) explored the application of PFs for multivariate snow data assimilation, emphasizing challenges such as sample impoverishment and the critical choice of temporal frequency for perturbing meteorological forcing. Smyth *et al.* (2020) first implemented a PF to assimilate snow depth observations, achieving reductions in both snow density and snow water equivalent (SWE) biases across climate zones in both wet and dry years.

Later, Deschamps-Berger *et al.* (2022) extended PF assimilation to spatial snow depth maps in mountainous regions, showing improvements in both the spatial variability of simulated snow depth and precipitation bias correction. Complementary studies further investigated PF approaches for large-scale or spatially distributed snow data. Cantet *et al.* (2019) and Odry *et al.* (2022) implemented spatialized PF to recover the spatial structure of the snowpack, demonstrating the potential for improving distributed SWE estimation. More recent work, such asShrestha & Barros (2025) and Oberrauch *et al.* (2025), has focused on multi-physics and fully distributed frameworks, showing that the performance of PF assimilation depends strongly on model complexity, input data quality, and the representation of snowpack processes. Additionally, Magnusson *et al.* (2025) evaluated methods for estimating SWE from daily snow depth records, illustrating the ongoing interest in operationalizing PF assimilation with limited observational data.

Despite these advances, most studies still rely on offline or retrospective assimilation, often using smoothers or post-processed observations. Sequential DA in near-real-time, where only observations available at the time are used to update model states, remains underexplored for operational snow modelling. This gap motivates the use of PFs and

related sequential frameworks in quasi-operational scenarios, where timely updates are critical for forecasting and hydrological applications.

# Appendix C

# Machine learning

Machine learning (ML) (LeCun *et al.*, 2015a) is a subfield of AI that enables computer programs to automatically improve their performance through experience (Learning, 1997). Considering a supervised learning approach, ML uses data to identify patterns, make predictions, classify information, or uncover trends across large datasets. In this context, human expertise is still necessary to select relevant features and structure inputs, as the overarching goal is to build models that can learn from data and reproduce spatial and temporal patterns without being explicitly programmed (Samuel, 1959).

The development of ML in hydrology has been shaped by successive generations of algorithms, each addressing limitations of its predecessors while opening new avenues of application.

Among the different ML algorithms, Random forest (Breiman, 2001) has become the state of art technology, due to its simplicity of use (Tyralis *et al.*, 2019), as well as its proven effectiveness as a classifier or regressor algorithm. Figure C.1 graphically explains the functioning of a random forest.

Considering the application of such algorithms, several studies have demonstrated the effectiveness of Random Forest methods across different cryospheric and hydrological contexts. Park *et al.* (2020) developed a Random Forest classifier for sea ice detection using Sentinel-1 data. Similarly, Desai & Ouarda (2021) applied Random Forest to flood frequency analysis, showing that the method can be as reliable as more complex models while being computationally more efficient. In the cryospheric domain, Meloche *et al.* (2022) demonstrated the capability of a Random Forest algorithm to predict snow depth distribution from topographic parameters in western Nunavut, Canada, achieving a root mean square error of 8 cm (23%). In the field of hazard prediction, Ponziani *et al.* (2023) showed the effectiveness of Random Forest compared to other ML algorithms by developing a predictive model for debris flows that could be experimentally integrated into the early warning system of the Aosta Valley. More recently, Herbert *et al.* (2025) developed a Random Forest model that combines intermittent airborne lidar snow depth measurements

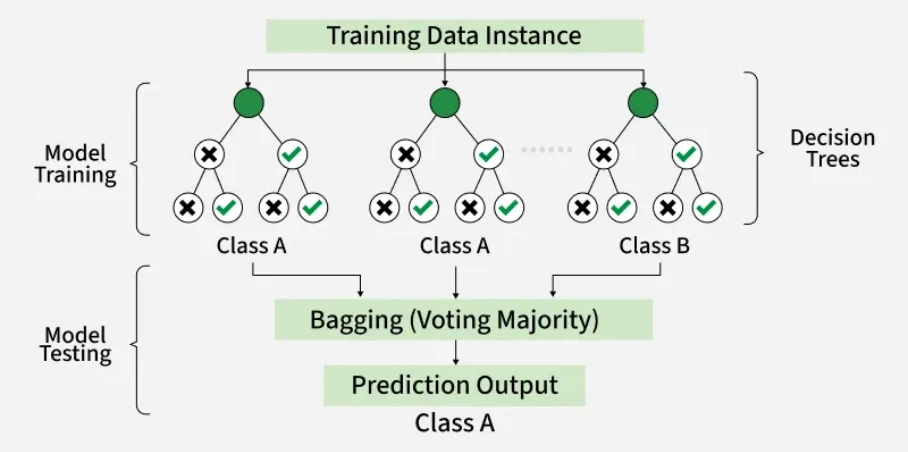


**Figure C.1:** *Schematic of a Random Forest classifier. Multiple decision trees are independently trained on bootstrapped samples of the training data. Each tree outputs a class prediction, and the final prediction is determined through majority voting across all trees, enhancing robustness and reducing overfitting.*

with continuous SNOTEL observations and physiographic variables to produce daily high-resolution (50 m) snow depth estimates across mountain basins in Colorado, effectively bridging the spatial coverage of lidar with the temporal continuity of ground observations. Likewise, Pan *et al.* (2025b) generated a 35-year (1988–2023) snow phenology dataset using a Random Forest model applied to passive microwave satellite data, enabling the detection of basin-wide trends toward earlier snowmelt onset and snow-off associated with increasing temperatures in the Yukon River Basin. Finally, Soltani *et al.* (2026) analysed the spatiotemporal variability of SWE in the Mackenzie River Basin (2000–2020) using in situ observations, remote sensing, reanalysis, and machine-learning datasets. Their results showed that gridded products generally reproduce seasonal and spatial SWE patterns well, with a Random forest–based dataset exhibiting the lowest errors, while also highlighting large regional variability and only a slight, statistically insignificant decreasing trend over time.

# Appendix D

# Deep learning

Deep learning DL (Goodfellow *et al.*, 2016), based on Artificial Neural Networks (Artificial Neural Network (ANN)), is a subset of ML that automates much of the manual human intervention required. Introduced by Hinton & Salakhutdinov (2006), DL algorithms transform raw data representations from one level to a slightly more abstract level, employing non-linear modules to learn complex functions. At the basis of DL, ANN replicate the human brain's function and structure, composed of nodes or artificial neurons. Figure D.1 shows the architecture of an ANN. A brief mathematical introduction is then reported (Nielsen, 2015).

Each artificial neuron is called a **perceptron**. A perceptron can have one or more inputs, which are combined in a **summation function**. To weight the importance of each input, every input is multiplied by a corresponding weight, and a bias term is added:

$$\boldsymbol{z} = \sum_{i=1}^{n} \boldsymbol{w}_i \boldsymbol{x}_i + \boldsymbol{b} \tag{D.1}$$

where:

- $\boldsymbol{z}$ is the output of the neuron's summation function.
- $\boldsymbol{x}_i$ is the $i$-th input to the neuron.
- $\boldsymbol{w}_i$ is the $i$-th weight term.
- $\boldsymbol{b}$ is the bias term.
- $n$ is the number of inputs.
- $i$ is the input index, ranging from 1 to $n$.

An **activation function** $f$ then transforms this weighted sum into the neuron output:

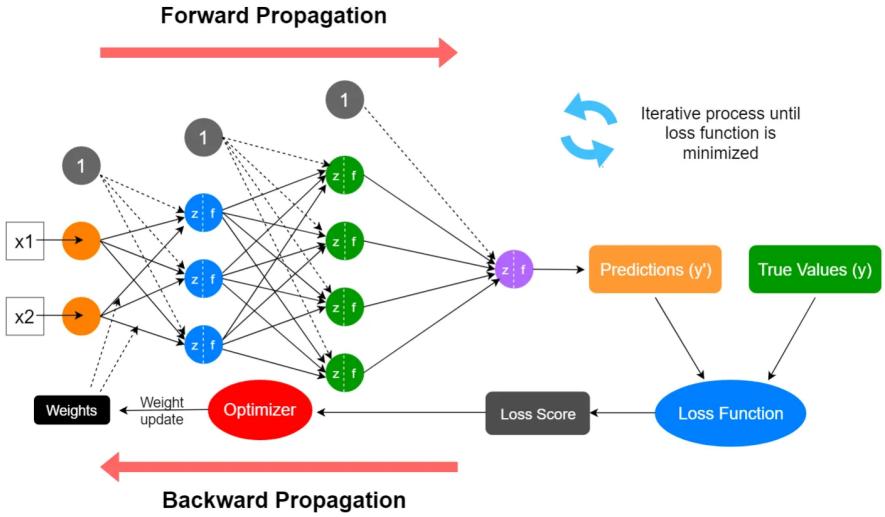


**Figure D.1:** *A schematic of a feedforward neural network training process is illustrated. Input features $(x_1, x_2)$ are propagated through hidden layers during forward propagation to generate predictions $\hat{y}$. The loss function compares these predictions with the true values $y$ to compute a loss score, which is then used by the optimizer to update network weights via backward propagation. This iterative process continues until the loss function is minimized, enabling the network to learn the underlying mapping from inputs to outputs. Source: Pramoditha (n.d.).*

$$\boldsymbol{a} = f(\boldsymbol{z}) \tag{D.2}$$

where:

- $\boldsymbol{a}$ is the neuron output after the activation function.
- $f$ is a differentiable (usually nonlinear) activation function.

The outputs of one layer are then used as inputs to the next layer, repeating this process until the **prediction layer** of the network. The final network output can be written as:

$$\boldsymbol{y}' = \sum_{j=1}^{N} \boldsymbol{w}_j^{\text{out}} \boldsymbol{a}_j + \boldsymbol{b}^{\text{out}} \tag{D.3}$$

where:

- $\boldsymbol{y}'$ is the network's final prediction vector.

- $\boldsymbol{a}_j$ is the output of the $j$-th neuron in the last hidden layer.
- $\boldsymbol{w}_j^{\text{out}}$ is the weight connecting the $j$-th neuron to the output.
- $\boldsymbol{b}^{\text{out}}$ is the output layer bias.
- $N$ is the number of neurons in the last hidden layer.

The loss or cost function quantifies the difference between the network predictions and the target values:

$$\mathbf{L} = \frac{1}{m} \sum_{i=1}^{m} \ell(\boldsymbol{y}_i', \boldsymbol{y}_i, w_i) \tag{D.4}$$

where:

- $\mathbf{L}$ is the general loss function.
- $\ell$ is a differentiable per-sample loss function, which can be Mean Squared Error (MSE), Mean Absolute Error (MAE), Huber, cross-entropy, or a custom loss.
- $w_i$ is an optional weight for the $i$-th target.
- $\boldsymbol{y}_i'$ is the network prediction vector.
- $\boldsymbol{y}_i$ is the target vector.

The training of glsANNs is then performed with the back-propagation algorithm (Rumelhart *et al.*, 1985), adjusting weights and biases iteratively to minimize a loss function's gradient and achieve a global minimum.

$$\nabla \mathbf{L} = \frac{\partial}{\partial \boldsymbol{w}_i} \mathbf{L} + \frac{\partial}{\partial \boldsymbol{b}_i} \mathbf{L} \tag{D.5}$$

where :

- $\mathbf{L}$ is the loss function.
- $\frac{\partial}{\partial \boldsymbol{w}_i}$ is the partial derivative referring to the weight term.
- $\frac{\partial}{\partial \boldsymbol{b}_i}$ is the partial derivative referring to the bias term

Weights and biases are randomly initialized and then updated through back propagation principle. As the loss function is a function of the weights and biases, the aim of the back propagation algorithm is to find weights and biases terms that minimize it. To compute the minimum of the loss function, a gradient descent procedure is used. For each iteration, an update of the weights and biases is computed; then the Neural Network is re-trained using

the new parameters value (forward propagation). Forward propagation and backward propagation will continued for the chosen sample size, until the error reaches minimum value.

Neural networks were pioneered in the 1940s but gained prominence in the 1980s through the work of Rumelhart *et al.* (1985). Advancements in computational power, graphics processing units (GPUs), and AI have given rise to DL. DL involves multi-layer neural networks that overcome the limitations of single-layer networks in handling extensive data and complexity (LeCun *et al.*, 2015b).

Presently, Deep Neural Networks find applications across a wide range of fields, including system control, gaming, medical science, finance, and geoscience and hydrology (LeCun *et al.*, 2015a; Camps-Valls *et al.*, 2021; Sit *et al.*, 2020b). Within hydrology, Artificial Neural Networks (ANNs) have been successfully applied to streamflow prediction (Zealand *et al.*, 1999; Besaw *et al.*, 2010), where their main advantage lies in capturing the strong non-linearities inherent in rainfall–runoff processes. This capability to model complex, non-linear relationships has made ANNs and other architectures increasingly valuable for hydrological forecasting and environmental modelling.

DL has attracted steadily increasing interest among scientists in diverse fields due to its ability to represent complex system dynamics without requiring detailed knowledge of underlying physical processes or relying on rigid structural assumptions (Yu *et al.*, 2024). This trend is particularly evident in Earth and environmental sciences, where DL is increasingly employed to model highly non-linear processes, including hydrological dynamics, climate variability, and other environmental phenomena (Zhao *et al.*, 2024b). By learning patterns directly from data, DL offers a flexible and efficient alternative when physical models are difficult to formulate, computationally demanding, or constrained by incomplete observational coverage.

Based on interconnected neural networks, DL architectures excel at extracting system dynamics from large datasets and can overcome structural limitations that challenge traditional physically-based models (LeCun *et al.*, 2015a; Murphy, 2023). A major advantage of DL is its ability to automate feature extraction, reducing the need for extensive manual feature engineering. Nevertheless, the effectiveness of this automation depends on the specific application and dataset, and careful feature design remains important in many contexts. Through successive layers, DL can identify complex non-linear relationships between inputs and outputs, enabling richer representations of the underlying system dynamics.

Among the most commonly used DL architectures, Long Short-Term Memory (LSTM) networks (Hochreiter & Schmidhuber, 1997), a type of recurrent neural networks, can memorize internal system states and capture long-term dependencies between inputs and outputs. LSTM networks have demonstrated significant success in predicting time-series data, particularly in hydrological applications, where they have shown comparable performance to traditional physically-based models (Fan *et al.*, 2020; Chen *et al.*, 2023; Kratzert *et al.*, 2018, 2019). Due to the strong temporal autocorrelation and memory of the snowpack (Fiddes *et al.*, 2019), these networks appear to be especially well suited for snow.

In the broader field of operational hydrology, Boucher *et al.* (2020) pioneered a novel ensemble-based DA approach leveraging neural networks.

The DL methods are instrumental in addressing challenging problems in the field of hydrology and water resources (Sit *et al.*, 2020b), such as time-series forecasting (Nearing *et al.*, 2021), land use and land cover (LULC) classification (Chen *et al.*, 2017; Vali *et al.*, 2020), hydrodynamic modelling, downscaling (Wang *et al.*, 2020), change and anomaly detection, hurricane tracking (Rodés-Guirao, 2019), water quality assessment (Chen *et al.*, 2024), and extreme weather prediction (Camps-Valls *et al.*, 2025).

However, the use of DL for DA remains largely underexplored in the field of snow hydrology. One exception is the recent study by Guidicelli *et al.* (2024), who combined ensemble-based DA with DL to improve spatio-temporal estimates of SWE using sparse ground track data in the eastern Swiss Alps. This approach utilized an Iterative Ensemble Smoother, an iterative batch-smoother variant of the EnKF, in conjunction with a degree-day model to reconstruct SWE temporal evolution, while a feedforward neural network (FNN) facilitated spatial propagation based on topographic features. As a more recent exception of combining DL and snow data assimilation, Song *et al.* (2024) developed an LSTM-based framework to assimilate lagged observations of SWE or satellite-derived snow cover fraction (SCF) over the western U.S., aiming to improve seasonal snow predictions. While their approach further consolidates the potential of DL for DA in snow hydrology, it relied on a relatively simple assimilation setup, dealing with long lagged time step rather than a consequential and quasi real time approach. Other than these initial attempts, and the body of work on stand-alone DL for snow modelling (Cui *et al.*, 2023; Daudt *et al.*, 2023), the potential of combining advanced DL and DA algorithms for predicting snowpack dynamics remains largely underexplored (Nema & Nagashree, 2024).

# Appendix E

# Random forest test on Italian stations

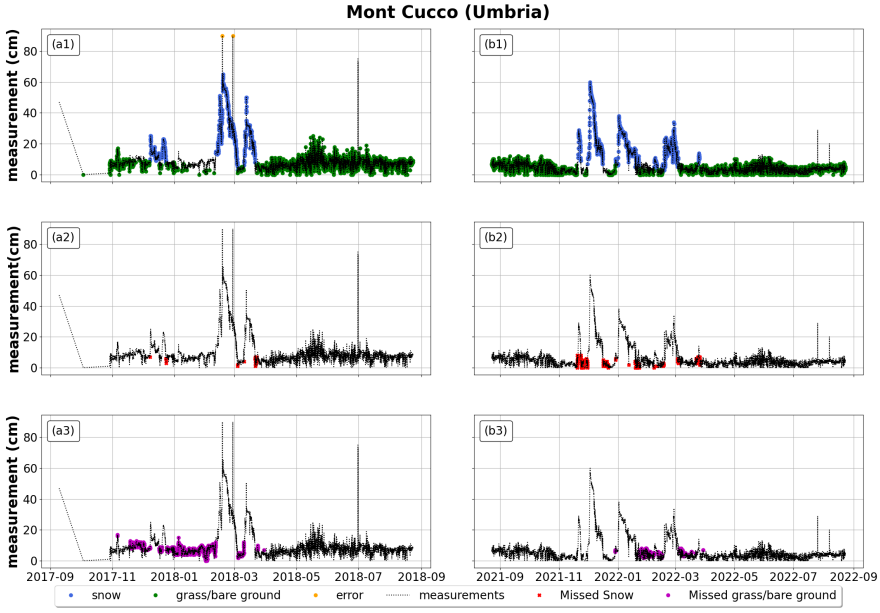


**Figure E.1:** *Monte Cucco (Umbria). Application of Random forest on an Italian station from October 2017 to September 2018 on the left, and from October 2021 to September 2022 on the right. First row reports correct classification of snow, grass/bare ground, and random errors (blue for snow depth, green for grass/ground, orange for random errors); second row reports miss-classified snow depth in red; the third row reports miss-classified grass/bare ground (in purple). All plots also report measured snow depth in black (whether it represents actual snow depth, grass/ground, or random errors)*

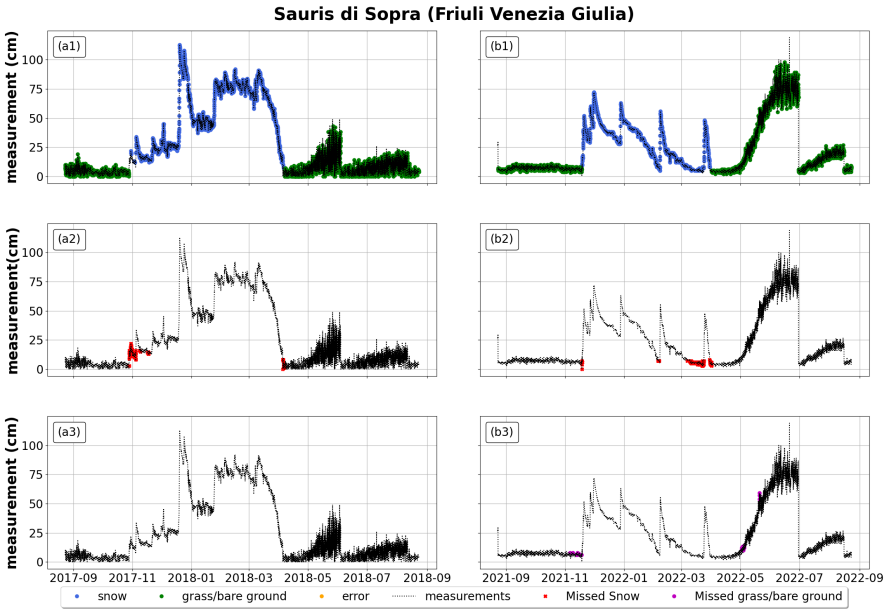


**Figure E.2:** *Sauris di sopra (Friuli Venezia Giulia). Application of Random forest on an Italian station from October 2017 to September 2018 on the left, and from October 2021 to September 2022 on the right. First row reports correct classification of snow, grass/bare ground, and random errors (blue for snow depth, green for grass/ground, orange for random errors); second row reports miss-classified snow depth in red; the third row reports miss-classified grass/bare ground (in purple). All plots also report measured snow depth in black (whether it represents actual snow depth, grass/ground, or random errors)*

# Appendix F

# Training site specifics

## F.1 Coordinates information of the 7 study sites for LSTM

- TRG (Torgnon, Aosta Valley, Italy): 45°50′ N, 7°34′ E
- CDP (Col De Porte, Isère, France): 45°3′ N, 5°77′ E
- WFJ (Weissfluhjoch, Davos, Switzerland): 46°82′ N, 9°8′ E
- KHT (Kühtai, Tirol, Austria): 47°20′71″ N, 11°00′6″ E
- FMI-ARC (FMI-ARC Sodankylä Geophysical Observatory, Finnish Lapland): 67°36′8″ N, 26°63′3″ E
- NGK (Nagaoka, Japan): 37°25′ N, 138°53′ E
- RME (Reynolds Mountain East, Idaho, USA): 43°11′9.36″ N, 116°46′58.9″ W

## F.2 Measurement characteristics across the 7 study sites for the LSTM

| Site | SWE Obs. (mm) | HS Obs. (cm) | Frequency | Error (SWE /HS) [mm/cm] | Time Range |
|---|---|---|---|---|---|
| TRG | 6h, missing (2012–2013, 2014–2015) | ✓ | 30’ | ±15/±10 | Oct 2012–Mar 2023 |
| CDP | From 2002 | ✓ | 1h | ±5/±1 | Oct 1993–Sep 2022 |
| WFJ | Manual, sporadic | ✓ | 60’ | ±10/±20 | Oct 1999–Sep 2018 |
| KHT | ✓ | ✓ | 15’ | ±1/±10 | Oct 1990–Sep 2015 |
| FMI-ARC | Manual, sporadic | ✓ | 60’ | ±15/±10 | Oct 2007–Jul 2014 |
| NGK | ✓ | ✓ | 60’ | ±10/±10 | Oct 2006–Aug 2023 |
| RME | ✓ | From 1999 | 60’ | ±10/±10 | Oct 1984–Sep 2008 |

**Table F.1:** *Measurement Characteristics Across Sites. TRG = Torgnon, Aosta Valley, Italy. CDP = Col de Porte, Isère, France. WFJ = Weissfluhjoch, Davos, Switzerland. KHT = Kühtai, Tirol, Austria. FMI-ARC = FMI-ARC Sodankylä Geophysical Observatory, Finnish Lapland. NGK = Nagaoka, Japanel RME = Reynolds Mountain East, Idaho, USA.*

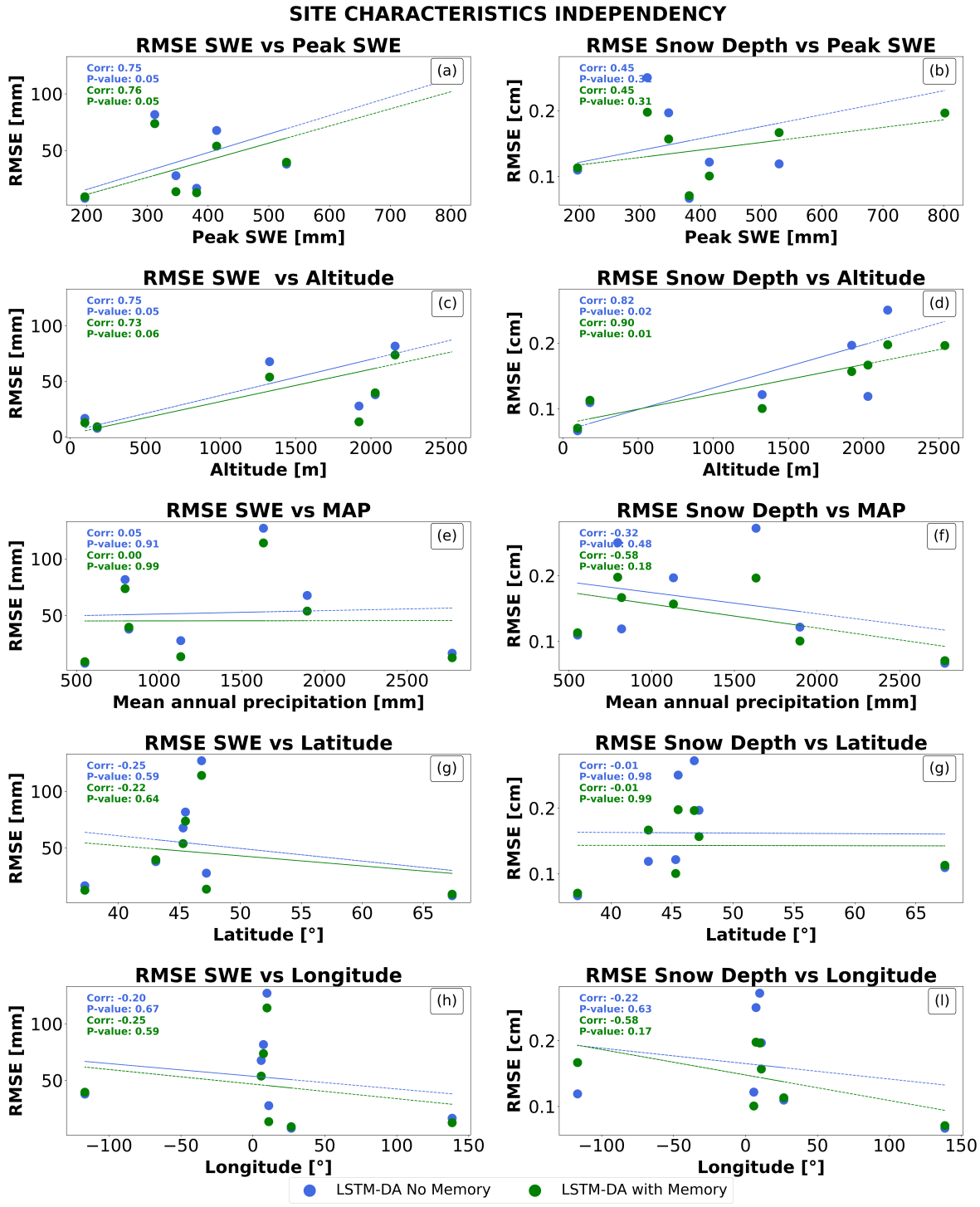


**Figure F.1:** *Analysis of RMSE dependency on site characteristics for SWE and snow depth across different parameters. Subplots (a-b) show RMSE vs. peak SWE , (c-d) vs. altitude, (e-f) vs. annual precipitation, (g-h) vs. latitude, and (i-l) vs. longitude. Blue and cyan markers represent estimations from LSTM with and without memory, respectively. Correlation coefficients, confidence intervals, and p-values indicate weak or negligible dependence of RMSE on these site characteristics, suggesting general independence of model performance from these factors.*

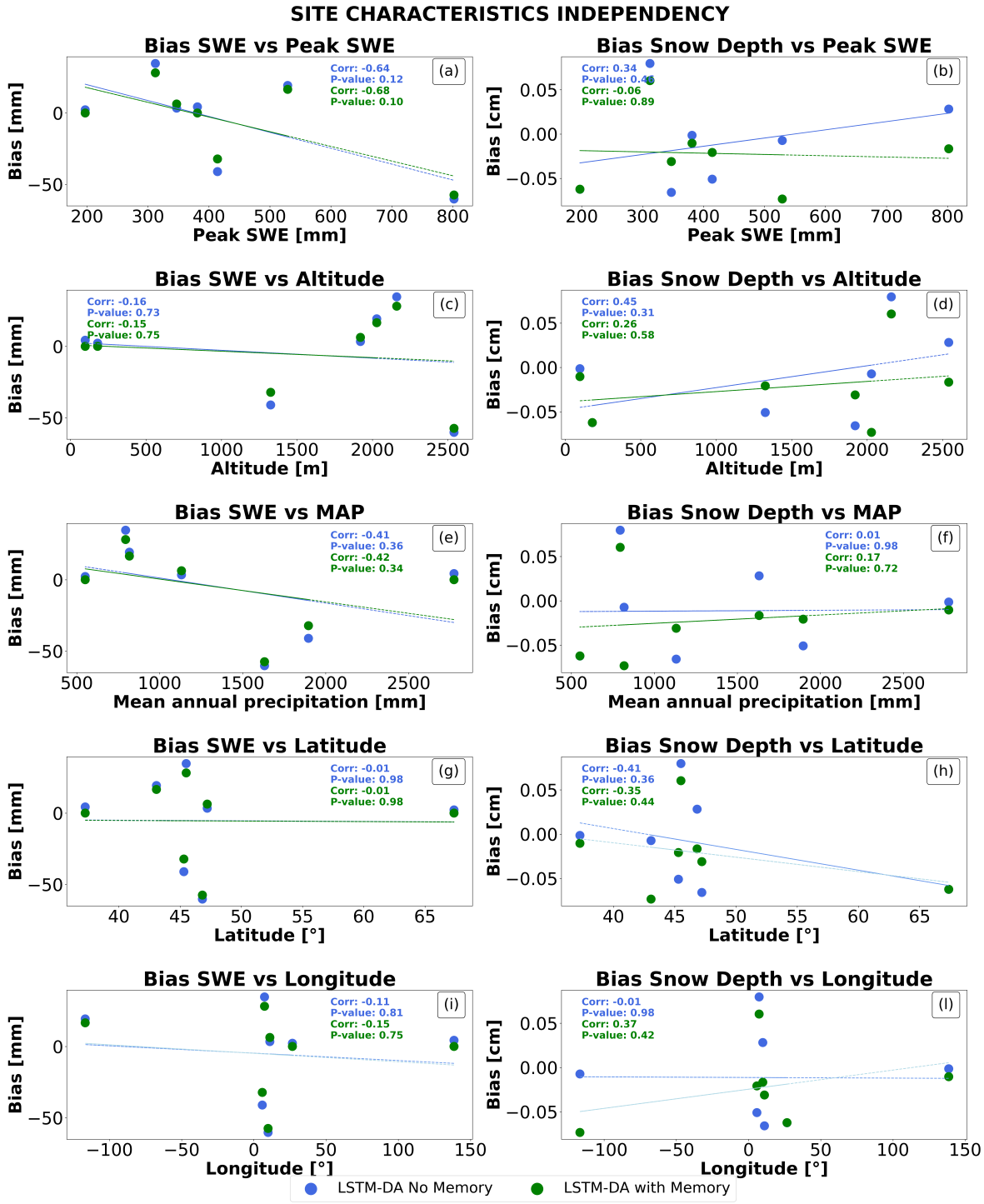


**Figure F.2:** *Bias analysis of SWE and snow depth with respect to site characteristics. Subplots (a-b) illustrate bias vs. peak SWE , (c-d) vs. altitude, (e-f) vs. annual precipitation, (g-h) vs. latitude, and (i-l) vs. longitude. Blue and cyan markers represent estimations from LSTM without and with memory, respectively. Correlation coefficients and p-values suggest minimal or no significant bias dependency on these site characteristics, except for a moderate correlation in specific cases, such as SWE bias with annual precipitation in (e).*

# Appendix G

# Ensemble Kalman filter for S3M

At each time step $k$, the observation operator is defined as:

$$\mathbf{H}_k = \begin{bmatrix} 1 & 1 & 0 & 0 \\ \frac{1}{c} & \frac{1}{\rho_D} & \frac{\partial h}{\partial \rho_D} & 0 \end{bmatrix}.$$

If no measurements of SWE are available:

$$\mathbf{H}_k = \begin{bmatrix} 0 & 0 & 0 & 0 \\ \frac{1}{c} & \frac{1}{\rho_D} & \frac{\partial h}{\partial \rho_D} & 0 \end{bmatrix}.$$

If no measurements of snow depth are available:

$$\mathbf{H}_k = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

# Bibliography


Aalstad, K., Westermann, S., Schuler, T. V., Boike, J., & Bertino, L. 2018. Ensemble-based assimilation of fractional snow-covered area satellite retrievals to estimate the snow distribution at Arctic sites. *The Cryosphere*, **12**(1), 247–270.

Aalstad, Kristoffer, Westermann, Sebastian, & Bertino, Laurent. 2020. Evaluating satellite retrieved fractional snow-covered area at a high-Arctic site using terrestrial photography. *Remote Sensing of Environment*, **239**, 111618.

Abdulameer, Layth, Al-Khafaji, Mahmoud Saleh, Al-Awadi, Aysar Tuma, Al Maimuri, Najah ML, Al-Shammari, Musa, & Al-Dujaili, Ahmed N. 2025. Artificial Intelligence in Climate-Resilient Water Management: A Systematic Review of Applications, Challenges, and Future Directions. *Water Conservation Science and Engineering*, **10**(1), 44.

Acero Triana, Juan S., Chu, Maria L., Guzman, Jorge A., Moriasi, Daniel N., & Steiner, Jean L. 2019. Beyond model metrics: The perils of calibrating hydrologic models. *Journal of Hydrology*, **578**(Nov.), 124032.

Acharki, Hicham, El Alami, Mohammed, Elyoussfi, Haytam, & et al. 2025. Spatial Modeling of Snow Water Equivalent in the High Atlas Using Machine Learning: Implications for the Rheraya Catchment (Morocco). *Scientific Reports*, **15**(1), 14310.

Adhikary, Sunil, Yamaguchi, Yasushi, & Ogawa, Katsuro. 2002. Estimation of snow ablation under a dust layer covering a wide range of albedo. *Hydrological processes*, **16**(14), 2853–2865.

Adnan, Rana Muhammad, Mo, Wang, Kisi, Ozgur, Heddam, Salim, Al-Janabi, Ahmed Mohammed Sami, & Zounemat-Kermani, Mohammad. 2024. Harnessing Deep Learning and Snow Cover Data for Enhanced Runoff Prediction in Snow-Dominated Watersheds. *Atmosphere*, **15**(12), 1407.

Alonso-González, E., Aalstad, K., Baba, M. W., Revuelto, J., López-Moreno, J. I., Fiddes, J., Essery, R., & Gascoin, S. 2022. The Multiple Snow Data Assimilation System (MuSA v1.0). *Geoscientific Model Development*, **15**(24), 9127–9155.

Alonso-González, E., Gascoin, S., Arioli, S., & Picard, G. 2023a. Exploring the potential of thermal infrared remote sensing to improve a snowpack model through an observing system simulation experiment. *The Cryosphere*, **17**(8), 3329–3342.

Alonso-González, E., Aalstad, K., Pirk, N., Mazzolini, M., Treichler, D., Leclercq, P., Westermann, S., López-Moreno, J. I., & Gascoin, S. 2023b. Spatio-temporal information propagation using sparse observations in hyper-resolution ensemble-based snow data assimilation. *Hydrology and Earth System Sciences*, **27**(24), 4637–4659.

Alonso-González, Esteban, Aalstad, Kristoffer, Baba, Mohamed Wassim, Revuelto, Jesús, López-Moreno, Juan Ignacio, Fiddes, Joel, Essery, Richard, & Gascoin, Simon. 2022. The multiple snow data assimilation system (MuSA v1. 0). *Geoscientific Model Development*, **15**(24), 9127–9155.

Anderson, Eric A. 1973. *National Weather Service river forecast system: Snow accumulation and ablation model*. Vol. 17. US Department of Commerce, National Oceanic and Atmospheric Administration.

Andreadis, Konstantinos M., & Lettenmaier, Dennis P. 2006. Assimilating remotely sensed snow observations into a macroscale hydrology model. *Advances in Water Resources*, **29**(6), 872–886.

Andreadis, Konstantinos M., Storck, Pascal, & Lettenmaier, Dennis P. 2009. Modeling snow accumulation and ablation processes in forested environments. *Water Resources Research*, **45**(5).

Arcucci, Rossella, Zhu, Jiangcheng, Hu, Shuang, & Guo, Yi-Ke. 2021. Deep data assimilation: integrating deep learning with data assimilation. *Applied Sciences*, **11**(3), 1114.

Avanzi, F., Gabellani, S., Delogu, F., Silvestro, F., Cremonese, E., Morra di Cella, U., Ratto, S., & Stevenin, H. 2022a. Snow Multidata Mapping and Modeling (S3M) 5.1: a distributed cryospheric model with dry and wet snow, data assimilation, glacier mass balance, and debris-driven melt. *Geoscientific Model Development*, **15**(12), 4853–4879.

Avanzi, Francesco, De Michele, Carlo, Ghezzi, Antonio, Jommi, Cristina, & Pepe, Monica. 2014. A processing–modeling routine to use SNOTEL hourly data in snowpack dynamic models. *Advances in water resources*, **73**, 16–29.

Avanzi, Francesco, Johnson, Ryan Curtis, Oroza, Carlos A, Hirashima, Hiroyuki, Maurer, Tessa, & Yamaguchi, Satoru. 2019. Insights into preferential flow snowpack runoff using random forest. *Water Resources Research*, **55**(12), 10727–10746.

Avanzi, Francesco, Zheng, Zeshi, Coogan, Adam, Rice, Robert, Akella, Ram, & Conklin, Martha H. 2020a. Gap-filling snow-depth time-series with Kalman Filtering-Smoothing

and Expectation Maximization: Proof of concept using spatially dense wireless-sensor-network data. *Cold Regions Science and Technology*, **175**, 103066.

Avanzi, Francesco, Zheng, Zeshi, Coogan, Adam, Rice, Robert, Akella, Ram, & Conklin, Martha H. 2020b. Gap-filling snow-depth time-series with Kalman filtering-smoothing and expectation maximization: Proof of concept using spatially dense wireless-sensor-network data. *Cold Regions Science and Technology*, **175**, 103066.

Avanzi, Francesco, *et al.* 2021a. Learning about precipitation lapse rates from snow course data improves water balance modeling. *Hydrology and Earth System Sciences*, **25**(4), 2109–2131.

Avanzi, Francesco, Ercolani, Giulia, Gabellani, Simone, Cremonese, Edoardo, Pogliotti, Paolo, Filippa, Gianluca, Morra di Cella, Umberto, Ratto, Sara, Stevenin, Hervè, Cauduro, Marco, & Juglair, Stefano. 2021b. Learning about precipitation lapse rates from snow course data improves water balance modeling. *Hydrology and Earth System Sciences*, **25**(4), 2109–2131.

Avanzi, Francesco, Gabellani, Simone, Delogu, Fabio, Silvestro, Francesco, Cremonese, Edoardo, Morra di Cella, Umberto, Ratto, Sara, & Stevenin, Hervé. 2022b. Snow Multidata Mapping and Modeling (S3M) 5.1: a distributed cryospheric model with dry and wet snow, data assimilation, glacier mass balance, and debris-driven melt. *Geoscientific Model Development*, **15**(12), 4853–4879.

Avanzi, Francesco, Gabellani, Simone, Delogu, Fabio, Silvestro, Francesco, Pignone, Flavio, Bruno, Giulia, Pulvirenti, Luca, Squicciarino, Giuseppe, Fiori, Elisabetta, Rossi, Lauro, Puca, Silvia, Toniazzo, Alexander, Giordano, Pietro, Falzacappa, Marco, Ratto, Sara, Stevenin, Hervè, Cardillo, Antonio, Fioletti, Matteo, Cazzuli, Orietta, Cremonese, Edoardo, Morra di Cella, Umberto, & Ferraris, Luca. 2023. IT-SNOW: a snow reanalysis for Italy blending modeling, in situ data, and satellite observations (2010–2021). *Earth System Science Data*, **15**(2), 639–660.

Avanzi, Francesco, Terzi, Stefano, Castelli, Mariapina, Munerol, Francesca, Andreaggi, Margherita, Galvagno, Marta, Galletti, Andrea, Maurer, Tessa, Massari, Christian, Carlson, Grace, Girotto, Manuela, Bertoldi, Giacomo, Cremonese, Edoardo, gabellani, simone, Cella, Umberto Morra di, Altamura, Marco, Rossi, Lauro, & Ferraris, Luca. 2025. Today's snow and tomorrow's water: impacts of Mediterranean snow droughts on mountain socio-ecohydrology. Oct.

Bair, Edward H, Davis, Robert E, & Dozier, Jeff. 2018a. Hourly mass and snow energy balance measurements from Mammoth Mountain, CA USA, 2011–2017. *Earth System Science Data*, **10**(1), 549–563.

Bair, Edward H, Abreu Calfa, Andre, Rittger, Karl, & Dozier, Jeff. 2018b. Using machine learning for real-time estimates of snow water equivalent in the watersheds of Afghanistan. *The Cryosphere*, **12**(5), 1579–1594.

Bakketun, Å., Blyverket, J., & Müller, M. 2026. Ensemble-based snow depth data assimilation for a multi-layer snow scheme over the European Arctic. *The Cryosphere*, **20**(1), 737–756.

Bales, Roger C, Molotch, Noah P, Painter, Thomas H, Dettinger, Michael D, Rice, Robert, & Dozier, Jeff. 2006. Mountain hydrology of the western United States. *Water Resources Research*, **42**(8).

Bavay, M, & Egger, T. 2014a. MeteoIO 2.4. 2: a preprocessing library for meteorological data. *Geoscientific Model Development*, **7**(6), 3135–3151.

Bavay, M., & Egger, T. 2014b. MeteoIO 2.4.2: a preprocessing library for meteorological data. *Geoscientific Model Development*, **7**(6), 3135–3151.

Beck, Hylke E, Zimmermann, Niklaus E, McVicar, Tim R, Vergopolan, Noemi, Berg, Alexis, & Wood, Eric F. 2018. Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific data*, **5**(1), 1–12.

Beniston, Martin, & Stoffel, Markus. 2016. Rain-on-snow events, floods and climate change in the Alps: Events may increase with warming up to 4 C and decrease thereafter. *Science of the total environment*, **571**, 228–236.

Beniston, Martin, Farinotti, Daniel, Stoffel, Markus, Andreassen, Liss M., Coppola, Erika, Eckert, Nicolas, Fantini, Adriano, Giacona, Florie, Hauck, Christian, Huss, Matthias, Huwald, Hendrik, Lehning, Michael, López-Moreno, Juan-Ignacio, Magnusson, Jan, Marty, Christoph, Morán-Tejéda, Enrique, Morin, Samuel, Naaim, Mohamed, Provenzale, Antonello, Rabatel, Antoine, Six, Delphine, Stötter, Johann, Strasser, Ulrich, Terzago, Silvia, & Vincent, Christian. 2018. The European mountain cryosphere: a review of its current state, trends, and future challenges. *The Cryosphere*, **12**(2), 759–794.

Berghuijs, WR, Woods, RA, & Hrachowitz, M. 2014. A precipitation shift from snow towards rain leads to a decrease in streamflow. *Nature climate change*, **4**(7), 583–586.

Besaw, Lance E, Rizzo, Donna M, Bierman, Paul R, & Hackett, William R. 2010. Advances in ungauged streamflow prediction using artificial neural networks. *Journal of Hydrology*, **386**(1-4), 27–37.

Besso, Hannah, Shean, David, & Lundquist, Jessica D. 2024. Mountain snow depth retrievals from customized processing of ICESat-2 satellite laser altimetry. *Remote Sensing of Environment*, **300**, 113843.

Beven, Keith. 2006. A manifesto for the equifinality thesis. *Journal of hydrology*, **320**(1-2), 18–36.

Beven, Keith. 2012. Causal models as multiple working hypotheses about environmental processes. *Comptes rendus geoscience*, **344**(2), 77–88.

Beven, Keith, & Binley, Andrew. 1992. The future of distributed models: model calibration and uncertainty prediction. *Hydrological processes*, **6**(3), 279–298.

Beven, Keith, & Freer, Jim. 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology*, **249**(1), 11–29.

Bishop, Christopher M, & Bishop, Hugh. 2023. *Deep learning: Foundations and concepts*. Springer Nature.

Bitz, Cecilia M, Marshall, Shawn J, & NW, Calgary AB. 2012. Cryosphere models: Ocean and land. *Encyclopedia of sustainability science and technology (section on climate change modeling and methodology)*.

Blandini, Giulia, Avanzi, Francesco, Gabellani, Simone, Ponziani, Denise, Stevenin, Hervé, Ratto, Sara, Ferraris, Luca, & Viglione, Alberto. 2023. A random forest approach to quality-checking automatic snow-depth sensor measurements. *The Cryosphere*, **17**(12), 5317–5333.

Blandini, Giulia, Avanzi, Francesco, Campo, Lorenzo, Gabellani, Simone, Aalstad, Kristoffer, Girotto, Manuela, Yamaguchi, Satoru, Hirashima, Hiroyuki, & Ferraris, Luca. 2025. Learning to filter: Snow data assimilation using a Long Short-Term Memory network. *The Cryosphere*, **19**(10), 4759–4783.

Blöschl, Günter. 1999. Scaling issues in snow hydrology. *Hydrological processes*, **13**(14-15), 2149–2175.

Bolibar, J., Rabatel, A., Gouttevin, I., Galiez, C., Condom, T., & Sauquet, E. 2020. Deep learning applied to glacier evolution modelling. *The Cryosphere*, **14**(2), 565–584.

Boni, Giorgio, Castelli, Fabio, Gabellani, Simone, Machiavello, G, & Rudari, Roberto. 2010. Assimilation of MODIS snow cover and real time snow depth point data in a snow dynamic model. *Pages 1788–1791 of: 2010 IEEE International Geoscience and Remote Sensing Symposium*. IEEE.

Bormann, Kathryn J., Westra, Seth, Evans, Jason P., & McCabe, Matthew F. 2013. Spatial and temporal variability in seasonal snow density. *Journal of Hydrology*, **484**, 63–73.

Born, Andreas, Imhof, Michael A, & Stocker, Thomas F. 2019. An efficient surface energy–mass balance model for snow and ice. *The Cryosphere*, **13**(5), 1529–1546.

Borzì, Iolanda. 2025. Modeling Groundwater Resources in Data-Scarce Regions for Sustainable Management: Methodologies and Limits. *Hydrology*, **12**(1), 11.

Boucher, Marie-Amélie, Quilty, John, & Adamowski, Jan. 2020. Data Assimilation for Streamflow Forecasting Using Extreme Learning Machines and Multilayer Perceptrons. *Water Resources Research*, **56**(06), e2019WR026226.

Bousbaa, Mostafa, Elyoussfi, Haytam, Bargam, Bouchra, Belaqziz, Salwa, Boudhar, Abdelghani, & Chehbouni, Abdelghani. 2024. Establishing Static and Dynamic NDSI Thresholds for Accurate Snow Cover Mapping Using Landsat Imagery. *International Journal of Applied Earth Observation and Geoinformation*, **128**, 103851.

Bozzoli, Michele, Crespi, Alice, Matiu, Michael, Majone, Bruno, Giovannini, Lorenzo, Zardi, Dino, Brugnara, Yuri, Bozzo, Alessio, Berro, Daniele Cat, Mercalli, Luca, & Bertoldi, Giacomo. 2024. Long-term snowfall trends and variability in the Alps. *International Journal of Climatology*, **44**(13), 4571–4591.

Bradley, James A, Anesio, Alexandre M, & Arndt, Sandra. 2017. Microbial and biogeochemical dynamics in glacier forefields are sensitive to century-scale climate and anthropogenic change. *Frontiers in Earth Science*, **5**, 26.

Branco, Paula, Torgo, Luís, & Ribeiro, Rita P. 2016. A survey of predictive modeling on imbalanced domains. *ACM Computing Surveys (CSUR)*, **49**(2), 1–50.

Brangers, Isis, Lievens, Hans, Getirana, Augusto, & De Lannoy, GJM. 2024. Sentinel-1 snow depth assimilation to improve river discharge estimates in the western European alps. *Water Resources Research*, **60**(11), e2023WR035019.

Bravo, David Nogués, Araújo, Miguel B, Lasanta, Teodoro, & Moreno, Juan Ignacio López. 2008. Climate change in Mediterranean mountains during the 21st century. *AMBIO: A Journal of the Human Environment*, **37**(4), 280–285.

Breiman, Leo. 2001. Random Forests. *Machine Learning*, **45**(1), 5–32.

Brunetti, M., Lentini, G., Maugeri, M., Nanni, T., Simolo, C., & Spinoni, J. 2009. 1961–1990 high-resolution Northern and Central Italy monthly precipitation climatologies. *Advances in Science and Research*, **3**(1), 73–78.

Brunt, Kelly M, Neumann, Thomas A, Walsh, Kaitlin M, & Markus, Thorsten. 2013. Determination of local slope on the Greenland Ice Sheet using a multibeam photon-counting Lidar in preparation for the ICESat-2 Mission. *IEEE Geoscience and Remote Sensing Letters*, **11**(5), 935–939.

Buch, S., Gupta, M. K., & Gupta, R. K. 1993. Modeling of the Daily Rainfall-Runoff Relationship with Artificial Neural Networks. *Hydrological Sciences Journal*, **38**(6), 437–448.

Camps-Valls, Gustau, Tuia, Devis, Zhu, Xiao Xiang, & Reichstein, Markus. 2021. *Deep learning for the Earth Sciences: A comprehensive approach to remote sensing, climate science and geosciences.* John Wiley & Sons.

Camps-Valls, Gustau, Fernández-Torres, Miguel-Ángel, Cohrs, Kai-Hendrik, Höhl, Adrian, Castelletti, Andrea, Pacal, Aytac, Robin, Claire, Martinuzzi, Francesco, Papoutsis, Ioannis, Prapas, Ioannis, Pérez-Aracil, Jorge, Weigel, Katja, Gonzalez-Calabuig, Maria, Reichstein, Markus, Rabel, Martin, Giuliani, Matteo, Mahecha, Miguel D., Popescu, Oana-Iuliana, Pellicer-Valero, Oscar J., Ouala, Said, Salcedo-Sanz, Sancho, Sippel, Sebastian, Kondylatos, Spyros, Happé, Tamara, & Williams, Tristan. 2025. Artificial intelligence for modeling and understanding extreme weather and climate events. *Nature Communications*, **16**(1).

Cannone, N., Diolaiuti, G., Guglielmin, M., & Smiraglia, C. 2008. Accelerating climate change impacts on alpine glacier forefield ecosystems in the European Alps. *Ecological Applications*, **18**, 637–648.

Cantet, Philippe, Boucher, MA, Lachance-Coutier, S, Turcotte, R, & Fortin, V. 2019. Using a particle filter to estimate the spatial distribution of the snowpack water equivalent. *Journal of Hydrometeorology*, **20**(4), 577–594.

Canton, Helen. 2021. World meteorological organization—WMO. *Pages 388–393 of: The Europa directory of international organizations 2021.* Routledge.

Carrassi, Alberto, Bocquet, Marc, Bertino, Laurent, & Evensen, Geir. 2018. Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *Wiley Interdisciplinary Reviews: Climate Change*, **9**(5), e535.

Casas, César Quilodrán, Arcucci, Rossella, Wu, Pin, Pain, Christopher, & Guo, Yi-Ke. 2020. A Reduced Order Deep Data Assimilation model. *Physica D: Nonlinear Phenomena*, **412**, 132615.

Cavazza, William, & Wezel, Forese Carlo. 2003. The Mediterranean region—a geological primer. *Episodes*, **26**(3), 160–168.

Change, IPCC Climate, *et al.* 2014. Mitigation of climate change. *Contribution of working group III to the fifth assessment report of the intergovernmental panel on climate change*, **1454**, 147.

Charbonneau, Andrew, Deck, Katherine, & Schneider, Tapio. 2024. A Physics-Constrained Neural Differential Equation Framework for Data-Driven Snowpack Simulation. *arXiv preprint arXiv:2412.06819.*

Chen, Annan, Zhao, Chuanfeng, Zhang, Haotian, Yang, Yikun, Li, Jing, Yu, Yan, Zhang, Qinghong, & Li, Jiefeng. 2025. Weakened snow and ice melting by enhanced cloud short-wave cooling effect in the Arctic. *National Science Review*, **12**(6), nwaf116.

Chen, Hongju, Yang, Jianping, Ding, Yongjian, He, Qingshan, & Ji, Qin. 2021a. Simulation of daily snow depth data in China based on the NEX-GDDP. *Water*, **13**(24), 3599.

Chen, Jingbo, Wang, Chengyi, Yue, Anzhi, Chen, Jiansheng, He, Dongxu, & Zhang, Xiuyan. 2017. Knowledge-guided golf course detection using a convolutional neural network fine-tuned on temporally augmented data. *Journal of Applied Remote Sensing*, **11**(04), 1.

Chen, Jixuan, Wei, Xiaojuan, Liu, Yinxiao, Zhao, Chunxia, Liu, Zhenan, & Bao, Zhikang. 2024. Deep Learning for Water Quality Prediction—A Case Study of the Huangyang Reservoir. *Applied Sciences*, **14**(19), 8755.

Chen, L., Muthu, B., & Subash, C. B. 2021b. Estimating Snow Depth Inversion Model Assisted Vector Analysis Based on Temperature Brightness for North Xinjiang Region of China. *European Journal of Remote Sensing*, **54**(1), 265–274.

Chen, Shengyue, Huang, Jinliang, & Huang, Jr-Chuan. 2023. Improving daily streamflow simulations for data-scarce watersheds using the coupled SWAT-LSTM approach. *Journal of Hydrology*, **622**, 129734.

Cheng, Chen, & Zhang, Guang-Tao. 2021. Deep Learning Method Based on Physics Informed Neural Network with Resnet Block for Solving Fluid Flow Problems. *Water*, **13**(4).

Cheng, Sibo, Quilodrán-Casas, César, Ouala, Said, Farchi, Alban, Liu, Che, Tandeo, Pierre, Fablet, Ronan, Lucor, Didier, Iooss, Bertrand, Brajard, Julien, *et al.* 2023. Machine learning with data assimilation and uncertainty quantification for dynamical systems: a review. *IEEE/CAA Journal of Automatica Sinica*, **10**(6), 1361–1387.

Chevrollier, L.-A., Wehrlé, A., Cook, J. M., Pirk, N., Benning, L. G., Anesio, A. M., & Tranter, M. 2025. Separating the albedo-reducing effect of different light-absorbing particles on snow using deep learning. *The Cryosphere*, **19**(4), 1527–1538.

Chollet, François, *et al.* 2015. *Keras*. `https://keras.io`.

Chu, Haibo, Wei, Jiahua, Wu, Wenyan, Jiang, Yuan, Chu, Qi, & Meng, Xiujing. 2021. A classification-based deep belief networks model framework for daily streamflow forecasting. *Journal of Hydrology*, **595**, 125967.

Clark, Martyn P, Hendrikx, Jordy, Slater, Andrew G, Kavetski, Dmitri, Anderson, Brian, Cullen, Nicolas J, Kerr, Tim, Örn Hreinsson, Einar, & Woods, Ross A. 2011. Representing spatial variability of snow water equivalent in hydrologic and land-surface models: A review. *Water Resources Research*, **47**(7).

Cline, Donald W. 1997. Effect of seasonality of snow accumulation and melt on snow surface energy exchanges at a continental alpine site. *Journal of Applied Meteorology*, **36**(1), 32–51.

Cluzet, B., Magnusson, J., Quéno, L., Mazzotti, G., Mott, R., & Jonas, T. 2024. Exploring how Sentinel-1 wet-snow maps can inform fully distributed physically based snowpack models. *The Cryosphere*, **18**(12), 5753–5767.

Copernicus. 2025a. *Sentinel-2*. `https://sentiwiki.copernicus.eu/web/sentinel-2`. last accessed 2025-11-27.

Copernicus. 2025b. *Sentinel-3*. `https://sentiwiki.copernicus.eu/web/sentinel-3`. last accessed 2025-11-27.

Cowherd, M., Mital, U., Rahimi, S., Girotto, M., Schwartz, A., & Feldman, D. 2024a. Climate change-resilient snowpack estimation in the Western United States. *Commun Earth Environ*, **5**.

Cowherd, Marianne, Mital, Utkarsh, Rahimi, Stefan, Girotto, Manuela, Schwartz, Andrew, & Feldman, Daniel. 2024b. Climate change-resilient snowpack estimation in the Western United States. *Communications Earth & Environment*, **5**(1), 337.

CryoSCOPE. 2026. *CryoSCOPE: Processes · Modeling · Foresight.* `https://cryoscope-project.eu/`. Accessed: 2026-03-10.

Cui, Guotao, Anderson, Michael, & Bales, Roger. 2023. Mapping of snow water equivalent by a deep-learning model assimilating snow observations. *Journal of Hydrology*, **616**, 128835.

Da Ronco, Pierfrancesco, Avanzi, Francesco, De Michele, Carlo, Notarnicola, Claudia, & Schaefli, Bettina. 2020. Comparing MODIS snow products Collection 5 with Collection 6 over Italian Central Apennines. *International Journal of Remote Sensing*, **41**(11), 4174–4205.

Dadrass Javan, Farzaneh, Samadzadegan, Farhad, Toosi, Ahmad, & van der Meijde, Mark. 2024. Unmanned aerial geophysical remote sensing: a systematic review. *Remote Sensing*, **17**(1), 110.

Dai, Liyun, Che, Tao, Xie, Hongjie, & Wu, Xiaodong. 2018. Estimation of Snow Depth over the Qinghai–Tibetan Plateau Based on AMSR-E and MODIS Data. *Remote Sensing*, **10**(12), 1989.

Daudt, Rodrigo Caye, Wulf, Hendrik, Hafner, Elisabeth D., Bühler, Yves, Schindler, Konrad, & Wegner, Jan Dirk. 2023. Snow depth estimation at country-scale with high spatial and temporal resolution. *ISPRS Journal of Photogrammetry and Remote Sensing*, **197**, 105–121.

De Gregorio, Ludovica, Callegari, Mattia, Marin, Carlo, Zebisch, Marc, Bruzzone, Lorenzo, Demir, Begüm, Strasser, Ulrich, Marke, Thomas, Günther, Daniel, Nadalet, Rudi, *et al.* 2019. A novel data fusion technique for snow cover retrieval. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **12**(8), 2862–2877.

De la Fuente, Luis Andres, Bennett, Andrew, Gupta, Hoshin Vijai, & Condon, Laura Elizabeth. 2024. A HydroLSTM-based Machine-Learning Approach to Discovering Regionalized Representations of Catchment Dynamics. *Authorea Preprints*.

De Lannoy, Gabriëlle JM, Reichle, Rolf H, Arsenault, Kristi R, Houser, Paul R, Kumar, Sujay, Verhoest, Niko EC, & Pauwels, Valentijn RN. 2012. Multiscale assimilation of Advanced Microwave Scanning Radiometer–EOS snow water equivalent and Moderate Resolution Imaging Spectroradiometer snow cover fraction observations in northern Colorado. *Water Resources Research*, **48**(1).

De Lannoy, Gabriëlle J.M., Bechtold, Michel, Albergel, Clément, Brocca, Luca, Calvet, Jean Christophe, Carrassi, Alberto, Crow, Wade T., de Rosnay, Patricia, Steele-Dunne, Susan, & More Authors. 2022. Perspective on satellite-based land data assimilation to estimate water cycle components in an era of advanced data availability and model sophistication. *Frontiers in Water*, **4**.

De Lannoy, Gabriëlle JM, Bechtold, Michel, Busschaert, Louise, Heyvaert, Zdenko, Modanesi, Sara, Dunmire, Devon, Lievens, Hans, Getirana, Augusto, & Massari, Christian. 2024. Contributions of irrigation modeling, soil moisture and snow data assimilation to high-resolution water budget estimates over the Po basin: Progress towards digital replicas. *Journal of Advances in Modeling Earth Systems*, **16**(10), e2024MS004433.

De Lannoy, Gabriëlle J. M., Reichle, Rolf H., Houser, Paul R., Arsenault, Kristi R., Verhoest, Niko E. C., & Pauwels, Valentijn R. N. 2010. Satellite-Scale Snow Water Equivalent Assimilation into a High-Resolution Land Surface Model. *Journal of Hydrometeorology*, **11**(2), 352–369.

De Michele, C., Avanzi, F., Ghezzi, A., & Jommi, C. 2013. Investigating the dynamics of bulk snow density in dry and wet conditions using a one-dimensional model. *The Cryosphere*, **7**(2), 433–444.

DeBeer, Chris M, & Pomeroy, John W. 2017. Influence of snowpack and melt energy heterogeneity on snow cover depletion and snowmelt runoff simulation in a cold mountain environment. *Journal of hydrology*, **553**, 199–213.

Dedieu, Jean-Pierre, Carlson, Bradley Z., Bigot, Sylvain, Sirguey, Pascal, Vionnet, Vincent, & Choler, Philippe. 2016. On the Importance of High-Resolution Time Series of Optical Imagery for Quantifying the Effects of Snow Cover Duration on Alpine Plant Habitat. *Remote Sensing*, **8**(6).

Deems, Jeffrey S, Painter, Thomas H, & Finnegan, David C. 2013. Lidar measurement of snow depth: a review. *Journal of Glaciology*, **59**(215), 467–479.

Demil, Getnet, Haghighi, Ali Torabi, Klöve, Björn, & Oussalah, Mourad. 2025. Advances in image-based estimation of snow variable: A systematic literature review on recent studies. *Journal of Hydrology*, 132855.

Desai, Shitanshu, & Ouarda, Taha B.M.J. 2021. Regional hydrological frequency analysis at ungauged sites with random forest regression. *Journal of Hydrology*, **594**, 125861.

Deschamps-Berger, C., Gascoin, S., Shean, D., Besso, H., Guiot, A., & López-Moreno, J. I. 2023a. Evaluation of snow depth retrievals from ICESat-2 using airborne laser-scanning data. *The Cryosphere*, **17**(7), 2779–2792.

Deschamps-Berger, César, Gascoin, Simon, Berthier, Etienne, Deems, Jeffrey, Gutmann, Ethan, Dehecq, Amaury, Shean, David, & Dumont, Marie. 2020. Snow depth mapping from stereo satellite imagery in mountainous terrain: evaluation using airborne lidar data. *The Cryosphere Discussions*, **2020**, 1–28.

Deschamps-Berger, César, Cluzet, B, Dumont, M, Lafaysse, M, Berthier, E, Fanise, Pascal, & Gascoin, S. 2022. Improving the spatial distribution of snow cover simulations by assimilation of satellite stereoscopic imagery. *Water Resources Research*, **58**(3), e2021WR030271.

Deschamps-Berger, César, Gascoin, Simon, Shean, David, Besso, Hannah, Guiot, Ambroise, & López-Moreno, Juan I. 2023b. Supplement of Evaluation of snow depth retrievals from ICESat-2 using airborne laser-scanning data.

Dettinger, Michael. 2014. Impacts in the third dimension. *Nature geoscience*, **7**(3), 166–167.

DeWalle, David R, & Rango, Albert. 2008. *Principles of snow hydrology*. Cambridge University Press.

Dhoni, Pan Singh. 2023. Enhancing Data Quality through Generative AI: An Empirical Study with Data. *Authorea Preprints*.

Di Nunno, Fabio, de Marinis, Giovanni, & Granata, Francesco. 2023. Short-term forecasts of streamflow in the UK based on a novel hybrid artificial intelligence algorithm. *Scientific Reports*, **13**(1), 7036.

Domine, Florent, Fourteau, Kévin, Picard, Ghislain, Lackner, Georg, Sarrazin, Denis, & Poirier, Mathilde. 2022. Permafrost cooled in winter by thermal bridging through snow-covered shrub branches. *Nature geoscience*, **15**(7), 554–560.

Dong, Chunyu. 2018. Remote sensing, hydrological modeling and in situ observations in snow cover research: A review. *Journal of Hydrology*, **561**, 573–583.

Donlon, Craig, Berruti, Bruno, Mecklenberg, S, Nieke, Jens, Rebhan, Helge, Klein, Ulf, Buongiorno, Alessandra, Mavrocordatos, Constantin, Frerick, Johannes, Seitz, Bernd, *et al.* 2012. The sentinel-3 mission: Overview and status. *Pages 1711–1714 of: 2012 IEEE international geoscience and remote sensing symposium*. IEEE.

Douville, H, Peings, Y, & Saint-Martin, D. 2017. Snow-(N) AO relationship revisited over the whole twentieth century. *Geophysical Research Letters*, **44**(1), 569–577.

Dozier, Jeff, Green, Robert O, Nolin, Anne W, & Painter, Thomas H. 2009. Interpretation of snow properties from imaging spectrometry. *Remote Sensing of Environment*, **113**, S25–S37.

Dozier, Jeff, Bair, Edward H, & Davis, Robert E. 2016. Estimating the spatial distribution of snow water equivalent in the world's mountains. *Wiley Interdisciplinary Reviews: Water*, **3**(3), 461–474.

Duan, Shiheng, Ullrich, Paul, Risser, Mark, & Rhoades, Alan. 2024. Using Temporal Deep Learning Models to Estimate Daily Snow Water Equivalent Over the Rocky Mountains. *Water Resources Research*, **60**(4), e2023WR035009. e2023WR035009 2023WR035009.

Duethmann, Doris, Blöschl, Günter, & Parajka, Juraj. 2020. Why does a conceptual hydrological model fail to correctly predict discharge changes in response to climate change? *Hydrology and Earth System Sciences*, **24**(7), 3493–3511.

Dumont, M, Brun, E, Picard, G, Michou, M, Libois, Q, Petit, JR, Geyer, M, Morin, S, & Josse, B. 2014. Contribution of light-absorbing impurities in snow to Greenland's darkening since 2009. *Nature Geoscience*, **7**(7), 509–512.

Dumont, Marie, & Gascoin, Simon. 2016. Optical remote sensing of snow cover. *Pages 115–137 of: Land surface remote sensing in continental hydrology*. Elsevier.

Dunmire, D., Bechtold, M., Boeykens, L., & De Lannoy, G. J. M. 2026a. Advancing snow data assimilation with a dynamic observation uncertainty. *The Cryosphere*, **20**(1), 609–628.

Dunmire, D., Bechtold, M., Boeykens, L., & De Lannoy, G. J. M. 2026b. Advancing snow data assimilation with a dynamic observation uncertainty. *The Cryosphere*, **20**(1), 609–628.

Dunmire, Devon, Lievens, Hans, Boeykens, Lucas, & De Lannoy, Gabrielle JM. 2024. A machine learning approach for estimating snow depth across the European Alps from Sentinel-1 imagery. *Remote Sensing of Environment*, **314**, 114369.

Durand, Michael, & Margulis, Steven. 2006a. Feasibility Test of Multifrequency Radiometric Data Assimilation to Estimate Snow Water Equivalent. *Journal of Hydrometeorology*, **7**(06), 443–457.

Durand, Michael, & Margulis, Steven A. 2006b. Feasibility test of multifrequency radiometric data assimilation to estimate snow water equivalent. *Journal of Hydrometeorology*, **7**(3), 443–457.

Dutra, Emanuel, Balsamo, Gianpaolo, Viterbo, Pedro, Miranda, Pedro MA, Beljaars, Anton, Schär, Christoph, & Elder, Kelly. 2010. An improved snow scheme for the ECMWF land surface model: Description and offline validation. *Journal of Hydrometeorology*, **11**(4), 899–916.

Eberhard, Lucie Anne, Sirguey, Pascal, Miller, Aubrey, Marty, Mauro, Schindler, Konrad, Stoffel, Andreas, & Bühler, Yves. 2020. Intercomparison of photogrammetric platforms for spatially continuous snow depth mapping. *The Cryosphere Discussions*, **2020**, 1–40.

Elyoussfi, Haytam, Boudhar, Abdelghani, Belaqziz, Salwa, Bousbaa, Mostafa, Nifa, Karima, Bargam, Bouchra, & Chehbouni, Abdelghani. 2025a. Leveraging Advanced Deep Learning and Machine Learning Approaches for Snow Depth Prediction Using Remote Sensing and Ground Data. *Journal of Hydrology: Regional Studies*, **57**, 102085.

Elyoussfi, Haytam, Boudhar, Abdelghani, Belaqziz, Salwa, Bousbaa, Mostafa, Nifa, Karima, Bargam, Bouchra, & Chehbouni, Abdelghani. 2025b. Leveraging advanced deep learning and machine learning approaches for snow depth prediction using remote sensing and ground data. *Journal of Hydrology: Regional Studies*, **57**(Feb.), 102085.

Enderlin, Ellyn M, Elkin, Colten M, Gendreau, Madeline, Marshall, HP, O'Neel, Shad, McNeil, Christopher, Florentine, Caitlyn, & Sass, Louis. 2022. Uncertainty of ICESat-2 ATL06-and ATL08-derived snow depths for glacierized and vegetated mountain regions. *Remote Sensing of Environment*, **283**, 113307.

Endrizzi, S, Gruber, Stephan, Dall'Amico, M, & Rigon, Riccardo. 2014. GEOtop 2.0: simulating the combined energy and water balance at and below the land surface accounting for soil freezing, snow cover and terrain effects. *Geoscientific Model Development*, **7**(6), 2831–2857.

Ernakovich, Jessica G, Hopping, Kelly A, Berdanier, Aaron B, Simpson, Rodney T, Kachergis, Emily J, Steltzer, Heidi, & Wallenstein, Matthew D. 2014. Predicted responses of arctic and alpine ecosystems to altered seasonality under climate change. *Global Change Biology*, **20**(10), 3256–3269.

ESA, European Space Agency. 2024. *The Sentinel-1 mission*. `https://www.esa.int/ESA_Multimedia/Images/2024/10/The_Sentinel-1_mission`. last accessed 2025-11-27.

Essery, R. 2015a. A factorial snowpack model (FSM 1.0). *Geoscientific Model Development*, **8**(12), 3867–3876.

Essery, Richard. 2015b. A factorial snowpack model (FSM 1.0). *Geoscientific Model Development*, **8**(12), 3867–3876.

Essery, Richard, Li, Long, & Pomeroy, John. 1999. A distributed model of blowing snow over complex terrain. *Hydrological processes*, **13**(14-15), 2423–2438.

Essery, Richard, Rutter, Nick, Pomeroy, John, Baxter, Robert, Stähli, Manfred, Gustafsson, David, Barr, Alan, Bartlett, Paul, & Elder, Kelly. 2009. SNOWMIP2: An evaluation of forest snow process simulations. *Bulletin of the American Meteorological Society*, **90**(8), 1120–1136.

Essery, Richard, Morin, Samuel, Lejeune, Yves, & Ménard, Cécile B. 2013. A comparison of 1701 snow models using observations from an alpine site. *Advances in water resources*, **55**, 131–148.

Essery, Richard, Kontu, Anna, Lemmetyinen, Juha, Dumont, Marie, & Ménard, Cécile B. 2016. A 7-year dataset for driving and evaluating snow models at an Arctic site (Sodankylä, Finland). *Geoscientific Instrumentation, Methods and Data Systems*, **5**(1), 219–227.

Evensen, Geir. 2003. The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean dynamics*, **53**, 343–367.

Evensen, Geir, Vossepoel, Femke C, & van Leeuwen, Peter Jan. 2022. *Data assimilation fundamentals: A unified formulation of the state and parameter estimation problem*. Springer Nature.

Fan, Hongxiang, Jiang, Mingliang, Xu, Ligang, Zhu, Hua, Cheng, Junxiang, & Jiang, Jiahu. 2020. Comparison of long short term memory networks and the hydrological model in runoff simulation. *Water*, **12**(1), 175.

Fayad, Abbas, Gascoin, Simon, Faour, Ghaleb, López-Moreno, Juan Ignacio, Drapeau, Laurent, Le Page, Michel, & Escadafal, Richard. 2017. Snow hydrology in Mediterranean mountain regions: A review. *Journal of Hydrology*, **551**, 374–396.

Ferrarin, Lucia, Schulz, Karsten, Bocchiola, Daniele, & Koch, Franziska. 2025. Enhancing snow depth estimation with snow cover geometrical descriptors. *Frontiers in Earth Science*, **13**(Oct.).

Ferreira, Luis Eduardo Boiko, Gomes, Heitor Murilo, Bifet, Albert, & Oliveira, Luiz S. 2019. Adaptive random forests with resampling for imbalanced data streams. *2019 International Joint Conference on Neural Networks (IJCNN)*, 1–6.

Ficklin, Darren L, Null, Sarah E, Abatzoglou, John T, Novick, Kimberly A, & Myers, Daniel T. 2022. Hydrological intensification will increase the complexity of water resource management. *Earth's Future*, **10**(3), e2021EF002487.

Fiddes, Joel, Aalstad, Kristoffer, & Westermann, Sebastian. 2019. Hyper-resolution ensemble-based snow reanalysis in mountain regions using clustering. *Hydrology and Earth System Sciences*, **23**(11), 4717–4736.

Fiebrich, Christopher A, Morgan, Cynthia R, McCombs, Alexandria G, Hall, Peter K, & McPherson, Renee A. 2010. Quality assurance procedures for mesoscale meteorological data. *Journal of Atmospheric and Oceanic Technology*, **27**(10), 1565–1582.

Filippa, Gianluca, Cremonese, Edoardo, Galvagno, Marta, Migliavacca, Mirco, Morra di Cella, Umberto, Petey, Martina, & Siniscalco, Consolata. 2015. Five years of phenological monitoring in a mountain grassland: inter-annual patterns and evaluation of the sampling protocol. *International journal of biometeorology*, **59**, 1927–1937.

Flanner, M. G., Shell, K. M., Barlage, M., Perovich, D. K., & Tschudi, M. A. 2011. Radiative forcing and albedo feedback from the Northern Hemisphere cryosphere between 1979 and 2008. *Nature Geoscience*, **4**, 151–155.

Fletcher, Christopher G, Kushner, Paul J, Hall, Alex, & Qu, Xin. 2009. Circulation responses to snow albedo feedback in climate change. *Geophysical Research Letters*, **36**(9).

Follum, Michael L., Downer, Charles W., Niemann, Jeffrey D., Roylance, Spencer M., & Vuyovich, Carrie M. 2015. A radiation-derived temperature-index snow routine for the GSSHA hydrologic model. *Journal of Hydrology*, **529**(Oct.), 723–736.

Franz, Kristie J, Hogue, Terri S, & Sorooshian, Soroosh. 2008. Snow model verification using ensemble prediction and operational benchmarks. *Journal of Hydrometeorology*, **9**(6), 1402–1415.

Frei, Christoph, Davies, Huw C, Gurtz, Joachim, & Schär, Christoph. 2000. Climate dynamics and extreme precipitation and flood events in Central Europe. *Integrated Assessment*, **1**(4), 281–300.

Gacu, Jerome, Monjardin, Cris, Mangulabnan, Ronald, Pugat, Gerald, & Solmerin, Jerose. 2025. Artificial Intelligence (AI) in Surface Water Management: A Comprehensive Review of Methods, Applications, and Challenges. *Water*, **17**(11), 1707.

Ganganwar, Vaishali. 2012. An overview of classification algorithms for imbalanced datasets. *International Journal of Emerging Technology and Advanced Engineering*, **2**(4), 42–47.

Gascoin, Simon, Lhermitte, Stefaan, Kinnard, Christophe, Bortels, Kirsten, & Liston, Glen E. 2013. Wind effects on snow cover in Pascua-Lama, Dry Andes of Chile. *Advances in Water Resources*, **55**, 25–39.

Gascoin, Simon, Hagolle, Olivier, Huc, Mireille, Jarlan, Lionel, Dejoux, J-F, Szczypta, Camille, Marti, Renaud, & Sánchez, R. 2015. A snow cover climatology for the Pyrenees from MODIS snow products. *Hydrology and Earth System Sciences*, **19**(5), 2337–2351.

Gascoin, Simon, Grizonnet, Manuel, Bouchet, Marine, Salgues, Germain, & Hagolle, Olivier. 2019. Theia Snow collection: High-resolution operational snow cover maps from Sentinel-2 and Landsat-8 data. *Earth System Science Data*, **11**(2), 493–514.

Gascoin, Simon, Luojus, Kari, Nagler, Thomas, Lievens, Hans, Masiokas, Mariano, Jonas, Tobias, Zheng, Zhaojun, & De Rosnay, Patricia. 2024. Remote sensing of mountain snow from space: status and recommendations. *Frontiers in Earth Science*, **12**, 1381323.

Gauch, M., Kratzert, F., Klotz, D., Nearing, G., Cohen, D., & Gilon, O. 2025. How to deal with missing input data. *EGUsphere*, **2025**, 1–21.

Girotto, Manuela, Musselman, Keith N, & Essery, Richard LH. 2020. Data assimilation improves estimates of climate-sensitive seasonal snow. *Current Climate Change Reports*, **6**, 81–94.

Girotto, Manuela, Musselman, Keith N., & Essery, Richard L. H. 2023. *Data Assimilation of Seasonal Snow*. Special Publications of the International Union of Geodesy and Geophysics. Cambridge University Press. Page 79–92.

Girotto, Manuela, Formetta, Giuseppe, Azimi, Shima, Bachand, Claire, Cowherd, Marianne, De Lannoy, Gabrielle, Lievens, Hans, Modanesi, Sara, Raleigh, Mark S., Rigon, Riccardo, & Massari, Christian. 2024. Identifying snowfall elevation patterns by assimilating satellite-based snow depth retrievals. *Science of The Total Environment*, **906**, 167312.

Gong, Junfu, Weerts, Albrecht H., Yao, Cheng, Li, Zhijia, Huang, Yingchun, Chen, Yuanfang, Chang, Yifei, & Huang, Pengnian. 2023a. State updating in a distributed hydrological model by ensemble Kalman filtering with error estimation. *Journal of Hydrology*, **620**(May), 129450.

Gong, Youdi, Liu, Guangzhen, Xue, Yunzhi, Li, Rui, & Meng, Lingzhong. 2023b. A survey on dataset quality in machine learning. *Information and Software Technology*, **162**, 107268.

Goodfellow, Ian, Bengio, Yoshua, Courville, Aaron, & Bengio, Yoshua. 2016. *Deep learning*. Vol. 1. MIT press Cambridge.

Gröger, Christoph. 2021. There is no AI without data. *Communications of the ACM*, **64**(11), 98–108.

Group, The World Bank. 2021. *Italy-Climatology*. `https://climateknowledgeportal.worldbank.org/country/italy/climate-data-historical/`. Accessed: 2023-09-15.

Grünewald, Thomas, & Lehning, Michael. 2014. Are flat-field snow depth measurements representative? A comparison of selected index sites with areal snow depth measurements at the small catchment scale. *Hydrological Processes*, **29**(7), 1717–1728.

Guidicelli, Matteo, Aalstad, Kristoffer, Treichler, Désirée, & Salzmann, Nadine. 2024. A combined data assimilation and deep learning approach for continuous spatio-temporal SWE reconstruction from sparse ground tracks. *Journal of Hydrology X*, **25**, 100190.

Gupta, A, & Govindaraju, RS. 2019. Propagation of structural uncertainty in watershed hydrologic models. *Journal of Hydrology*, **575**, 66–81.

Gupta, Hoshin V, Beven, Keith J, & Wagener, Thorsten. 2006. Model calibration and uncertainty estimation. *Encyclopedia of hydrological sciences*.

Gupta, Hoshin V, Kling, Harald, Yilmaz, Koray K, & Martinez, Guillermo F. 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of hydrology*, **377**(1-2), 80–91.

Harder, Phillip, & Pomeroy, John W. 2014. Hydrological model uncertainty due to precipitation-phase partitioning methods. *Hydrological Processes*, **28**(14), 4311–4327.

Harpold, AA, Dettinger, M, & Rajagopal, S. 2017. Defining snow drought and why it matters, Eos, 98. *Eos*, **98**.

Hartman, Robert K, Rost, Andrew A, & Anderson, Donald M. 1995. Operational processing of multi-source snow data. *Proceedings of the Western Snow Conference*, **147**, 151.

Harvey, H Benjamin, & Sotardi, Susan T. 2018. The pareto principle. *Journal of the American College of Radiology*, **15**(6), 931.

Hastie, Trevor, Tibshirani, Robert, Friedman, Jerome H, & Friedman, Jerome H. 2009. *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2. Springer.

Hatchett, B. J., Rhoades, A. M., & McEvoy, D. J. 2022. Monitoring the daily evolution and extent of snow drought. *Natural Hazards and Earth System Sciences*, **22**(3), 869–890.

Hatchett, Benjamin J., & McEvoy, Daniel J. 2018. Exploring the Origins of Snow Drought in the Northern Sierra Nevada, California. *Earth Interactions*, **22**(2), 1–13.

He, Minxue, Hogue, Terri S., Franz, Kristie J., Margulis, Steven A., & Vrugt, Jasper A. 2011. Corruption of parameter behavior and regionalization by model and forcing data errors: A Bayesian example using the SNOW17 model. *Water Resources Research*, **47**(7).

He, Tianxing, Yu, Shengcheng, Wang, Ziyuan, Li, Jieqiong, & Chen, Zhenyu. 2019. From Data Quality to Model Quality: An Exploratory Study on Deep Learning. *In: Proceedings of the 11th Asia-Pacific Symposium on Internetware*. Internetware '19. New York, NY, USA: Association for Computing Machinery.

Helmert, Jürgen, Şensoy Şorman, Aynur, Alvarado Montero, Rodolfo, De Michele, Carlo, De Rosnay, Patricia, Dumont, Marie, Finger, David Christian, Lange, Martin, Picard, Ghislain, Potopová, Vera, *et al.* 2018. Review of snow data assimilation methods for hydrological, land surface, meteorological and climate models: Results from a cost harmosnow survey. *Geosciences*, **8**(12), 489.

Herbert, J. N., Raleigh, M. S., & Small, E. E. 2025. Using a Random Forest Model to Combine Airborne Lidar and Snotel Data for Daily Estimates of Snow Depth Across Mountain Drainage Basins of Colorado. *Water Resources Research*, **61**(8).

Hersbach, Hans, Bell, Bill, Berrisford, Paul, Hirahara, Shoji, Horányi, András, Muñoz-Sabater, Joaquín, Nicolas, Julien, Peubey, Carole, Radu, Raluca, Schepers, Dinand, Simmons, Adrian, Soci, Cornel, Abdalla, Saleh, Abellan, Xavier, Balsamo, Gianpaolo, Bechtold, Peter, Biavati, Gionata, Bidlot, Jean, Bonavita, Massimo, De Chiara, Giovanna, Dahlgren, Per, Dee, Dick, Diamantakis, Michail, Dragani, Rossana, Flemming, Johannes, Forbes, Richard, Fuentes, Manuel, Geer, Alan, Haimberger, Leo, Healy, Sean, Hogan, Robin J., Hólm, Elías, Janisková, Marta, Keeley, Sarah, Laloyaux, Patrick, Lopez, Philippe, Lupu, Cristina, Radnoti, Gabor, de Rosnay, Patricia, Rozum, Iryna, Vamborg, Freja, Villaume, Sebastien, & Thépaut, Jean-Noël. 2020. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **146**(730), 1999–2049.

Hinton, Geoffrey E, & Salakhutdinov, Ruslan R. 2006. Reducing the dimensionality of data with neural networks. *science*, **313**(5786), 504–507.

Hochreiter, Sepp, & Schmidhuber, Jürgen. 1997. Long short-term memory. *Neural computation*, **9**(8), 1735–1780.

Hock, Regine. 1999. A distributed temperature-index ice-and snowmelt model including potential direct solar radiation. *Journal of glaciology*, **45**(149), 101–111.

Hock, Regine. 2003. Temperature index melt modelling in mountain areas. *Journal of hydrology*, **282**(1-4), 104–115.

Horton, S., & Haegeli, P. 2022. Using snow depth observations to provide insight into the quality of snowpack simulations for regional-scale avalanche forecasting. *The Cryosphere*, **16**(8), 3393–3411.

Houser, Paul R, De Lannoy, Gabriëlle JM, & Walker, Jeffrey P. 2012. Hydrologic data assimilation. *Approaches to Managing Disaster-Assessing Hazards, Emergencies and Disaster Impacts*, 41–64.

Huang, C., Newman, A. J., Clark, M. P., Wood, A. W., & Zheng, X. 2017a. Evaluation of snow data assimilation using the ensemble Kalman filter for seasonal streamflow prediction in the western United States. *Hydrology and Earth System Sciences*, **21**(1), 635–650.

Huang, Chengcheng, Newman, Andrew J, Clark, Martyn P, Wood, Andrew W, & Zheng, Xiaogu. 2017b. Evaluation of snow data assimilation using the ensemble Kalman filter for seasonal streamflow prediction in the western United States. *Hydrology and Earth System Sciences*, **21**(1), 635–650.

Hunt, Earl B. 2014. *Artificial intelligence*. Academic Press.

Isik, Fatih, Ozden, Gurkan, & Kuntalp, Mehmet. 2012. Importance of data preprocessing for neural networks modeling: The case of estimating the compaction parameters of soils. *Energy Educ Sci Technol Part A: Energy Sci Res*, **29**, 463–74.

Ismail, Muhammad Fraz, Bogacki, Wolfgang, Disse, Markus, Schäfer, Michael, & Kirschbauer, Lothar. 2023. Estimating degree-day factors of snow based on energy flux components. *The Cryosphere*, **17**(1), 211–231.

Ismail-Zadeh, Alik, Castelli, Fabio, Jones, Dylan, & Sanchez, Sabrina. 2023. *Applications of Data Assimilation and Inverse Problems in the Earth Sciences*. Vol. 5. Cambridge University Press.

Jiang, Qin, Li, Weiyue, Fan, Zedong, He, Xiaogang, Sun, Weiwei, Chen, Sheng, Wen, Jiahong, Gao, Jun, & Wang, Jun. 2021. Evaluation of the ERA5 reanalysis precipitation dataset over Chinese Mainland. *Journal of Hydrology*, **595**(Apr.), 125660.

Jones, Amber Spackman, Horsburgh, Jeffery S, & Eiriksson, David P. 2018. Assessing subjectivity in environmental sensor data post processing via a controlled experiment. *Ecological Informatics*, **46**, 86–96.

Juras, Roman, Blöcher, Johanna R., Jenicek, Michal, Hotovy, Ondrej, & Markonis, Yannis. 2021. What affects the hydrological response of rain-on-snow events in low-altitude mountain ranges in Central Europe? *Journal of Hydrology*, **603**(Dec.), 127002.

Justice, Christopher O, Vermote, Eric, Townshend, John RG, Defries, Ruth, Roy, David P, Hall, Dorothy K, Salomonson, Vincent V, Privette, Jeffrey L, Riggs, George, Strahler, Alan, *et al.* 1998. The Moderate Resolution Imaging Spectroradiometer (MODIS): Land remote sensing for global change research. *IEEE transactions on geoscience and remote sensing*, **36**(4), 1228–1249.

Justice, CO, Townshend, JRG, Vermote, EF, Masuoka, E, Wolfe, RE, Saleous, Nazmi, Roy, DP, & Morisette, JT. 2002. An overview of MODIS Land data processing and product status. *Remote sensing of Environment*, **83**(1-2), 3–15.

Jutz, S, & Milagro-Perez, M Pilar. 2020. Copernicus: the european earth observation programme. *Revista de Teledetección*, V–XI.

Kahle, Hans-Gert, & Mueller, Stephan. 1998. Structure and dynamics of the Eurasian-African/Arabian plate boundary system: Objectives, tasks and resources of the WEGENER group. *Journal of Geodynamics*, **25**(3–4), 303–325.

Karniadakis, George Em, Kevrekidis, Ioannis G, Lu, Lu, Perdikaris, Paris, Wang, Sifan, & Yang, Liu. 2021. Physics-informed machine learning. *Nature Reviews Physics*, **3**(6), 422–440.

Karpatne, Anuj, Atluri, Gowtham, Faghmous, James H, Steinbach, Michael, Banerjee, Arindam, Ganguly, Auroop, Shekhar, Shashi, Samatova, Nagiza, & Kumar, Vipin. 2017. Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Transactions on knowledge and data engineering*, **29**(10), 2318–2331.

Karpatne, Anuj, Ebert-Uphoff, Imme, Ravela, Sai, Babaie, Hassan Ali, & Kumar, Vipin. 2018. Machine learning for the geosciences: Challenges and opportunities. *IEEE Transactions on Knowledge and Data Engineering*, **31**(8), 1544–1554.

Kazama, So, Izumi, Hirokazu, Sarukkalige, Priyantha Ranjan, Nasu, Takayuki, & Sawamoto, Masaki. 2007. Estimating snow distribution over a large area and its application for water resources. *Hydrological Processes*, **22**(13), 2315–2324.

Klotz, D., Kratzert, F., Gauch, M., Keefe Sampson, A., Brandstetter, J., Klambauer, G., Hochreiter, S., & Nearing, G. 2022. Uncertainty estimation with deep learning for rainfall–runoff modeling. *Hydrology and Earth System Sciences*, **26**(6), 1673–1693.

Koehler, Jonas, Dietz, Andreas J., Zellner, Peter, Baumhoer, Celia A., Dirscherl, Mariel, Cattani, Luca, Vlahović, Živa, Alasawedah, Mohammad Hussein, Mayer, Konrad, Haslinger, Klaus, Bertoldi, Giacomo, Jacob, Alexander, & Kuenzer, Claudia. 2022. Drought in Northern Italy: Long Earth Observation Time Series Reveal Snow Line Elevation to Be Several Hundred Meters Above Long-Term Average in 2022. *Remote Sensing*, **14**(23).

Kouki, Kerttu, Luojus, Kari, & Riihelä, Aku. 2023. Evaluation of snow cover properties in ERA5 and ERA5-Land with several satellite-based datasets in the Northern Hemisphere in spring 1982–2018. *The Cryosphere*, **17**(12), 5007–5026.

Krajči, Pavel, Kirnbauer, Robert, Parajka, J, Schöber, Johannes, & Blöschl, Günter. 2017. The Kühtai data set: 25 years of lysimetric, snow pillow, and meteorological measurements. *Water resources research*, **53**(6), 5158–5165.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. 2018. Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks. *Hydrology and Earth System Sciences*, **22**(11), 6005–6022.

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., & Nearing, G. 2019. Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. *Hydrology and Earth System Sciences*, **23**(12), 5089–5110.

Kratzert, Frederik, Gauch, Martin, Klotz, Daniel, & Nearing, Grey. 2024. HESS Opinions: Never train a Long Short-Term Memory (LSTM) network on a single basin. *Hydrology and Earth System Sciences*, **28**(17), 4187–4201.

Krissek, Lawrence, & St. John, Kristen. 2025. *The Earth's Climate System*. Cham: Springer Nature Switzerland. Pages 49–104.

Kuhn, Max, Johnson, Kjell, *et al.* 2013. *Applied predictive modeling*. Vol. 26. Springer.

Kuter, Semih. 2021. Completing the machine learning saga in fractional snow cover estimation from MODIS Terra reflectance data: Random forests versus support vector regression. *Remote Sensing of Environment*, **255**, 112294.

Largeron, Chloé, Dumont, Marie, Morin, Samuel, Boone, Aaron, Lafaysse, Matthieu, Metref, Sammy, Cosme, Emmanuel, Jonas, Tobias, Winstral, Adam, & Margulis, Steven A. 2020. Toward snow cover estimation in mountainous areas using modern data assimilation methods: A review. *Frontiers in Earth Science*, **8**, 325.

Larsson Ivanov, Oskar, Bärring, Lars, & Wilcke, Renate A.I. 2022. Climate change impact on snow loads in northern Europe. *Structural Safety*, **97**, 102231.

Latron, J, Llorens, P, & Gallart, F. 2009. The hydrology of Mediterranean mountain areas. *Geography Compass*, **3**(6), 2045–2064.

Lavers, David A., Simmons, Adrian, Vamborg, Freja, & Rodwell, Mark J. 2022. An evaluation of ERA5 precipitation for climate monitoring. *Quarterly Journal of the Royal Meteorological Society*, **148**(748), 3152–3165.

Leal Filho, Walter, Dinis, Maria Alzira Pimenta, Nagy, Gustavo J, Fracassi, Umberto, & Aina, Yusuf A. 2024. A ticket to where? Dwindling snow cover impacts the winter tourism sector as a consequence of climate change. *Journal of Environmental Management*, **356**, 120554.

Learning, Machine. 1997. Tom mitchell. *Publisher: McGraw Hill.*

LeCun, Yann, Bengio, Yoshua, & Hinton, Geoffrey. 2015a. Deep learning. *nature*, **521**(7553), 436–444.

LeCun, Yann, Bengio, Y., & Hinton, Geoffrey. 2015b. Deep Learning. *Nature*, **521**(05), 436–44.

Lehning, M., Bartelt, P., Brown, B., & Fierz, C. 2002a. A physical SNOWPACK model for the Swiss avalanche warning Part III: meteorological forcing, thin layer formation and evaluation. *Cold Regions Science and Technology*, **35**, 169–184.

Lehning, Michael, Bartelt, Perry, Brown, Bob, Fierz, Charles, & Satyawali, Pramod. 2002b. A physical SNOWPACK model for the Swiss avalanche warning. *Cold Regions Science and Technology*, **35**(3), 147–167.

Lehning, Michael, Bartelt, Perry, Brown, Bob, & Fierz, Charles. 2002c. A physical SNOWPACK model for the Swiss avalanche warning: Part III: meteorological forcing, thin layer formation and evaluation. *Cold Regions Science and Technology*, **35**(3), 169–184.

Lejeune, Yves, Dumont, Marie, Panel, Jean-Michel, Lafaysse, Matthieu, Lapalus, Philippe, Le Gac, Erwan, Lesaffre, Bernard, & Morin, Samuel. 2019. 57 years (1960–2017) of snow and meteorological observations from a mid-altitude mountain site (Col de Porte, France, 1325 m of altitude). *Earth System Science Data*, **11**(1), 71–88.

Li, C. Z., Zhang, L., Wang, H., Zhang, Y. Q., Yu, F. L., & Yan, D. H. 2012. The transferability of hydrological models under nonstationary climatic conditions. *Hydrology and Earth System Sciences*, **16**(4), 1239–1254.

Li, Dongyue, Lettenmaier, Dennis P, Margulis, Steven A, & Andreadis, Konstantinos. 2019. The value of accurate high-resolution and spatially continuous snow information to streamflow forecasts. *Journal of Hydrometeorology*, **20**(4), 731–749.

Lievens, Hans, Demuzere, Matthias, Marshall, Hans-Peter, Reichle, Rolf H, Brucker, Ludovic, Brangers, Isis, de Rosnay, Patricia, Dumont, Marie, Girotto, Manuela, Immerzeel, Walter W, *et al.* 2019. Snow depth variability in the Northern Hemisphere mountains observed from space. *Nature communications*, **10**(1), 4629.

Lievens, Hans, Brangers, Isis, Marshall, Hans-Peter, Jonas, Tobias, Olefs, Marc, & De Lannoy, Gabriëlle. 2022. Sentinel-1 snow depth retrieval at sub-kilometer resolution over the European Alps. *The Cryosphere*, **16**(1), 159–177.

Liu, Changyu, Huang, Xiaodong, Li, Xubing, & Liang, Tiangang. 2020a. MODIS fractional snow cover mapping using machine learning technology in a mountainous area. *Remote Sensing*, **12**(6), 962.

Liu, Changyu, Huang, Xiaodong, Li, Xubing, & Liang, Tiangang. 2020b. MODIS Fractional Snow Cover Mapping Using Machine Learning Technology in a Mountainous Area. *Remote Sensing*, **12**(6).

Liu, ShiYin, Wu, TongHua, Wang, Xin, Wu, XiaoDong, Yao, XiaoJun, Liu, Qiao, Zhang, Yong, Wei, JunFeng, & Zhu, XiaoFan. 2021. Changes in the global cryosphere and their impacts: A review and new perspective. *Sciences in Cold and Arid Regions*, **12**(6), 343–354.

Liu, Yingchun. 2014. Random forest algorithm in big data environment. *Computer modelling & new technologies*, **18**(12A), 147–151.

López-Moreno, Juan I, Fassnacht, SR, Heath, JT, Musselman, KN, Revuelto, Jesús, Latron, Jérôme, Morán-Tejeda, Enrique, & Jonas, Tobias. 2013. Small scale spatial variability of snow density and depth over complex alpine terrain: Implications for estimating snow water equivalent. *Advances in water resources*, **55**, 40–52.

López-Moreno, Juan Ignacio, Pomeroy, JW, Morán-Tejeda, Enrique, Revuelto, Jesús, Navarro-Serrano, FM, Vidaller, Ixeia, & Alonso-González, Esteban. 2021. Changes in the frequency of global high mountain rain-on-snow events due to climate warming. *Environmental Research Letters*, **16**(9), 094021.

Magnusson, Jan, Gustafsson, David, Hüsler, Fabia, & Jonas, Tobias. 2014. Assimilation of point SWE data into a distributed snow cover model comparing two contrasting methods. *Water resources research*, **50**(10), 7816–7835.

Magnusson, Jan, Wever, Nander, Essery, Richard, Helbig, Nora, Winstral, Adam, & Jonas, Tobias. 2015. Evaluating snow models with varying process representations for hydrological applications. *Water Resources Research*, **51**(4), 2707–2723.

Magnusson, Jan, Winstral, Adam, Stordal, {Andreas S.}, Essery, Richard, & Jonas, Tobias. 2017. Improving physically based snow simulations by assimilating snow depths using the particle filter. *Water Resources Research*, Feb.

Magnusson, Jan, Cluzet, Bertrand, Quéno, Louis, Mott, Rebecca, Oberrauch, Moritz, Mazzotti, Giulia, Marty, Christoph, & Jonas, Tobias. 2025. Evaluating methods to estimate the water equivalent of new snow from daily snow depth recordings. *Cold Regions Science and Technology*, **233**, 104435.

Maity, Rajib, Srivastava, Aman, Sarkar, Subharthi, & Khan, Mohd Imran. 2024. Revolutionizing the future of hydrological science: Impact of machine learning and deep learning amidst emerging explainable AI and transfer learning. *Applied Computing and Geosciences*, **24**(Dec.), 100206.

Malek, Sami A., Avanzi, Francesco, Brun-Laguna, Keoma, Maurer, Tessa, Oroza, Carlos A., Hartsough, Peter C., Watteyne, Thomas, & Glaser, Steven D. 2017a. Real-Time Alpine Measurement System Using Wireless Sensor Networks. *Sensors*, **17**(11).

Malek, Sami A., Avanzi, Francesco, Brun-Laguna, Keoma, Maurer, Tessa, Oroza, Carlos A., Hartsough, Peter C., Watteyne, Thomas, & Glaser, Steven D. 2017b. Real-Time Alpine Measurement System Using Wireless Sensor Networks. *Sensors*, **17**(2583).

Margulis, Steven A, Cortés, Gonzalo, Girotto, Manuela, & Durand, Michael. 2016. A Landsat-era Sierra Nevada snow reanalysis (1985–2015). *Journal of Hydrometeorology*, **17**(4), 1203–1221.

Marks, Danny, Kimball, John, Tingey, Dave, & Link, Tim. 1998. The sensitivity of snowmelt processes to climate conditions and forest cover during rain-on-snow: A case study of the 1996 Pacific Northwest flood. *Hydrological Processes*, **12**(10-11), 1569–1587.

Marsh, Christopher B, Harder, Phillip, & Pomeroy, John W. 2023. Validation of FABDEM, a global bare-earth elevation model, against UAV-lidar derived elevation in a complex forested mountain catchment. *Environmental Research Communications*, **5**(3), 031009.

Martinec, J. 1975. Snowmelt-runoff model for stream flow forecasts. *Hydrology Research*, **6**(3), 145–154.

Marty, Christoph, Tilg, Anna-Maria, & Jonas, Tobias. 2017. Recent Evidence of Large-Scale Receding Snow Water Equivalents in the European Alps. *Journal of Hydrometeorology*, **18**(4), 1021–1031.

Matiu, M., Crespi, A., Bertoldi, G., Carmagnola, C. M., Marty, C., Morin, S., Schöner, W., Cat Berro, D., Chiogna, G., De Gregorio, L., Kotlarski, S., Majone, B., Resch, G., Terzago, S., Valt, M., Beozzo, W., Cianfarra, P., Gouttevin, I., Marcolini, G., Notarnicola, C., Petitta, M., Scherrer, S. C., Strasser, U., Winkler, M., Zebisch, M., Cicogna, A., Cremonini, R., Debernardi, A., Faletto, M., Gaddo, M., Giovannini, L., Mercalli, L., Soubeyroux, J.-M., Sušnik, A., Trenti, A., Urbani, S., & Weilguni, V. 2021. Observed snow depth trends in the European Alps: 1971 to 2019. *The Cryosphere*, **15**(3), 1343–1382.

Maurer, Tessa, Avanzi, Francesco, Oroza, Carlos A, Glaser, Steven D, Conklin, Martha, & Bales, Roger C. 2021a. Optimizing spatial distribution of watershed-scale hydrologic models using Gaussian Mixture Models. *Environmental Modelling & Software*, **142**, 105076.

Maurer, Tessa, Avanzi, Francesco, Oroza, Carlos A., Glaser, Steven D., Conklin, Martha, & Bales, Roger C. 2021b. Optimizing spatial distribution of watershed-scale hydrologic models using Gaussian Mixture Models. *Environmental Modelling & Software*, **142**, 105076.

Mayer, S., van Herwijnen, A., Techel, F., & Schweizer, J. 2022. A random forest model to assess snow instability from simulated snow stratigraphy. *The Cryosphere*, **16**(11), 4593–4615.

Mazzolini, M., Aalstad, K., Alonso-González, E., Westermann, S., & Treichler, D. 2024. Spatio-temporal snow data assimilation with the ICESat-2 laser altimeter. *EGUsphere*, **2024**, 1–29.

Meloche, Julien, Langlois, Alexandre, Rutter, Nick, McLennan, Donald, Royer, Alain, Billecocq, Paul, & Ponomarenko, Serguei. 2022. High-resolution snow depth prediction using Random Forest algorithm with topographic parameters: A case study in the Greiner watershed, Nunavut. *Hydrological Processes*, **36**(3), e14546.

Metref, S., Cosme, E., Le Lay, M., & Gailhard, J. 2023a. Snow data assimilation for seasonal streamflow supply prediction in mountainous basins. *Hydrology and Earth System Sciences*, **27**(12), 2283–2299.

Metref, Sammy, Cosme, Emmanuel, Le Lay, Matthieu, & Gailhard, Joël. 2023b. Snow data assimilation for seasonal streamflow supply prediction in mountainous basins. *Hydrology and Earth System Sciences*, **27**(12), 2283–2299.

Mhangara, Paidamwoyo, & Mapurisa, Willard. 2019. Multi-Mission Earth Observation Data Processing System. *Sensors*, **19**(18), 3831.

Miao, Chiyuan, Immerzeel, Walter W., Xu, Baiqing, Yang, Kun, Duan, Qingyun, & Li, Xin. 2024. Understanding the Asian water tower requires a redesigned precipitation observation strategy. *Proceedings of the National Academy of Sciences*, **121**(23).

Mirza, B. N., Small, E. E., & Raleigh, M. S. 2025. Evaluating the Utility of Sentinel-1 in a Data Assimilation System for Estimating Snow Depth in a Mountainous Basin. *EGUsphere*, **2025**, 1–31.

Moghaddasi, Pouya, Gavahi, Keyhan, & Moradkhani, Hamid. 2025. Snow drought to hydrologic drought progression using machine learning and probabilistic analysis. *Scientific Reports*, **15**(1), 20918.

Mohammed, Sedir, Budach, Lukas, Feuerpfeil, Moritz, Ihde, Nina, Nathansen, Andrea, Noack, Nele, Patzlaff, Hendrik, Naumann, Felix, & Harmouch, Hazar. 2025. The effects of data quality on machine learning performance on tabular data. *Information Systems*, **132**, 102549.

Montiforte, Vivian A., Ngodock, Hans E., & Souopgui, Innocent. 2024. A comparison of two nonlinear data assimilation methods. *Nonlinear Processes in Geophysics*, **31**(4), 463–476.

Moosavi, Vahid, Malekinezhad, Hossein, & Shirmohammadi, Bagher. 2014. Fractional snow cover mapping from MODIS data using wavelet-artificial intelligence hybrid models. *Journal of Hydrology*, **511**(Apr.), 160–170.

Moradkhani, Hamid, Hsu, Kuo-Lin, Gupta, Hoshin, & Sorooshian, Soroosh. 2005. Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. *Water resources research*, **41**(5).

Murphy, Kevin P. 2022. *Probabilistic Machine Learning: An introduction*. MIT Press.

Murphy, Kevin P. 2023. *Probabilistic Machine Learning: Advanced Topics*. MIT Press.

Murphy, RE, Ardanuy, Phillip, Deluccia, Frank J, Clement, JE, & Schueler, Carl F. 2006. The visible infrared imaging radiometer suite. *Pages 199–223 of: Earth Science Satellite Remote Sensing: Vol. 1: Science and Instruments*. Springer.

Muñoz-Sabater, Joaquín, Dutra, Emanuel, Agustí-Panareda, Anna, Albergel, Clément, Arduini, Gabriele, Balsamo, Gianpaolo, Boussetta, Souhail, Choulga, Margarita, Harrigan, Shaun, Hersbach, Hans, Martens, Brecht, Miralles, Diego G., Piles, María, Rodríguez-Fernández, Nemesio J., Zsoter, Ervin, Buontempo, Carlo, & Thépaut, Jean-Noël. 2021. ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, **13**(9), 4349–4383.

Naghibi, Seyed Amir, Ahmadi, Kourosh, & Daneshi, Alireza. 2017. Application of Support Vector Machine, Random Forest, and Genetic Algorithm Optimized Random Forest Models in Groundwater Potential Mapping. *Water Resources Management*, **31**(9), 2761–2775.

Navari, Mahdi, Margulis, Steven A, Tedesco, Marco, Fettweis, Xavier, & Alexander, Patrick M. 2018. Improving Greenland Surface Mass Balance Estimates Through the Assimilation of MODIS Albedo: A Case Study Along the K-Transect. *Geophysical Research Letters*, **45**(13), 6549–6556.

Nearing, Grey S, Kratzert, Frederik, Sampson, Alden Keefe, Pelissier, Craig S, Klotz, Daniel, Frame, Jonathan M, Prieto, Cristina, & Gupta, Hoshin V. 2021. What role does hydrological science play in the age of machine learning? *Water Resources Research*, **57**(3), e2020WR028091.

Nema, Manish K, & Nagashree, GE. 2024. A Review of Approaches and Applications for Streamflow Forecasting Using AI-Based Models. *Applications of Machine Learning in Hydroclimatology*, 17–33.

Neumann, Thomas A, Martino, Anthony J, Markus, Thorsten, Bae, Sungkoo, Bock, Megan R, Brenner, Anita C, Brunt, Kelly M, Cavanaugh, John, Fernandes, Stanley T, Hancock, David W, *et al.* 2019. The Ice, Cloud, and Land Elevation Satellite–2 Mission: A global geolocated photon product derived from the advanced topographic laser altimeter system. *Remote sensing of environment*, **233**, 111325.

Nielsen, Michael A. 2015. *Neural networks and deep learning*. Vol. 25. Determination press San Francisco, CA, USA.

Niwano, Masashi, Aoki, Teruo, Kuchiki, Katsuyuki, Hosaka, Masahiro, & Kodama, Yuji. 2012. Snow Metamorphism and Albedo Process (SMAP) model for climate studies: Model validation using meteorological and snow impurity data measured at Sapporo, Japan. *Journal of Geophysical Research: Earth Surface*, **117**(F3).

Núñez, Jorge, Cortés, Catalina B., & Yáñez, Marjorie A. 2023. Explainable Artificial Intelligence in Hydrology: Interpreting Black-Box Snowmelt-Driven Streamflow Predictions in an Arid Andean Basin of North-Central Chile. *Water*, **15**(19), 3369.

Oberrauch, Moritz, Cluzet, Bertrand, Magnusson, Jan, & Jonas, Tobias. 2024. Improving Fully Distributed Snowpack Simulations by Mapping Perturbations of Meteorological Forcings Inferred From Particle Filter Assimilation of Snow Monitoring Data. *Water Resources Research*, **60**(12), e2023WR036994. e2023WR036994 2023WR036994.

Oberrauch, Moritz, Cluzet, Bertrand, Magnusson, Jan, & Jonas, Tobias. 2025. The performance gains of assimilating limited information into a fully distributed snowpack model depend on model complexity and input data quality. *Water Resources Research*, **61**(11), e2025WR040681.

Odry, J., Boucher, M.-A., Lachance-Cloutier, S., Turcotte, R., & St-Louis, P.-Y. 2022. Large-scale snow data assimilation using a spatialized particle filter: recovering the spatial structure of the particles. *The Cryosphere*, **16**(9), 3489–3506.

Ohmura, Atsumu. 2001. Physical basis for the temperature-based melt-index method. *Journal of applied Meteorology*, **40**(4), 753–761.

Pagano, Thomas C., Wood, Andrew W., Ramos, Maria-Helena, Cloke, Hannah L., Pappenberger, Florian, Clark, Martyn P., Cranston, Michael, Kavetski, Dmitri, Mathevet, Thibault, Sorooshian, Soroosh, & Verkade, Jan S. 2014. Challenges of Operational River Forecasting. *Journal of Hydrometeorology*, **15**(4), 1692–1707.

Painter, Thomas H, Berisford, Daniel F, Boardman, Joseph W, Bormann, Kathryn J, Deems, Jeffrey S, Gehrke, Frank, Hedrick, Andrew, Joyce, Michael, Laidlaw, Ross, Marks, Danny, *et al.* 2016. The Airborne Snow Observatory: Fusion of scanning lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo. *Remote Sensing of Environment*, **184**, 139–152.

Pakdehi, M., Ahmadisharaf, E., Nazari, B., & Cho, E. 2024. Transferability of machine-learning-based modeling frameworks across flood events for hindcasting maximum river water depths in coastal watersheds. *Natural Hazards and Earth System Sciences*, **24**(10), 3537–3559.

Pan, C. G., Lasko, K., Griffin, S. P., Kimball, J. S., Du, J., Meehan, T. G., & Kirchner, P. B. 2025a. A random-forest-derived 35-year snow phenology record reveals climate trends in the Yukon River Basin. *The Cryosphere*, **19**(8), 2797–2819.

Pan, Caleb G., Lasko, Kristofer, Griffin, Sean P., Kimball, John S., Du, Jinyang, Meehan, Tate G., & Kirchner, Peter B. 2025b. A random-forest-derived 35-year snow phenology record reveals climate trends in the Yukon River Basin. *The Cryosphere*, **19**(8), 2797–2819.

Parajka, J., & Blöschl, G. 2006. Validation of MODIS snow cover images over Austria. *Hydrology and Earth System Sciences*, **10**(5), 679–689.

Park, J.-W., Korosov, A. A., Babiker, M., Won, J.-S., Hansen, M. W., & Kim, H.-C. 2020. Classification of sea ice types in Sentinel-1 synthetic aperture radar images. *The Cryosphere*, **14**(8), 2629–2645.

Patel, Akansha, Mark, Bryan G., Haritashya, Umesh K., & Bawa, Arun. 2025. Twenty first century snow cover prediction using deep learning and climate model data in the Teesta basin, eastern Himalaya. *Climate Dynamics*, **63**(3).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, **12**, 2825–2830.

Pellicciotti, Francesca, Brock, Ben, Strasser, Ulrich, Burlando, Paolo, Funk, Martin, & Corripio, Javier. 2005. An enhanced temperature-index glacier melt model including the shortwave radiation balance: development and testing for Haut Glacier d'Arolla, Switzerland. *Journal of glaciology*, **51**(175), 573–587.

Pflug, J. M., Wrzesien, M. L., Kumar, S. V., Cho, E., Arsenault, K. R., Houser, P. R., & Vuyovich, C. M. 2024. Extending the utility of space-borne snow water equivalent observations over vegetated areas with data assimilation. *Hydrology and Earth System Sciences*, **28**(3), 631–648.

Piazzi, G., Thirel, G., Campo, L., & Gabellani, S. 2018a. A particle filter scheme for multivariate data assimilation into a point-scale snowpack model in an Alpine environment. *The Cryosphere*, **12**(7), 2287–2306.

Piazzi, G., Thirel, G., Perrin, C., & Delaigue, O. 2021. Sequential Data Assimilation for Streamflow Forecasting: Assessing the Sensitivity to Uncertainties and Updated Variables of a Conceptual Hydrological Model at Basin Scale. *Water Resources Research*, **57**(4).

Piazzi, Gaia, Campo, Lorenzo, Gabellani, Simone, Castelli, Fabio, Cremonese, Edoardo, Cella, Umberto Morra di, Stevenin, Hervé, & Ratto, Sara Maria. 2018b. An Enkf-Based Scheme for Snow Multivariable Data Assimilation at an Alpine Site. *Journal of Hydrology and Hydromechanics*, **67**(1), 4–19.

Piazzi, Gaia, Campo, Lorenzo, Gabellani, Simone, Castelli, Fabio, Cremonese, Edoardo, Cella, Umberto Morra di, Stevenin, Hervé, Ratto, Sara Maria, *et al.* 2019. An EnKF-based scheme for snow multivariable data assimilation at an Alpine site. *Journal of Hydrology and Hydromechanics*, **67**, 4–19.

Pirazzini, Roberta, Leppänen, Leena, Picard, Ghislain, Lopez-Moreno, Juan Ignacio, Marty, Christoph, Macelloni, Giovanni, Kontu, Anna, Von Lerber, Annakaisa, Tanis, Cemal Melih, Schneebeli, Martin, *et al.* 2018. European in-situ snow measurements: practices and purposes. *Sensors*, **18**(7), 2016.

Pirk, Norbert, Aalstad, Kristoffer, Mannerfelt, Erik Schytt, Clayer, François, de Wit, Heleen, Christiansen, Casper T., Althuizen, Inge, Lee, Hanna, & Westermann, Sebastian. 2024. Disaggregating the Carbon Exchange of Degrading Permafrost Peatlands Using Bayesian Deep Learning. *Geophysical Research Letters*, **51**(10), e2024GL109283.

Plaza, D. A., De Keyser, R., De Lannoy, G. J. M., Giustarini, L., Matgen, P., & Pauwels, V. R. N. 2012. The importance of parameter resampling for soil moisture data assimilation into hydrologic models using the particle filter. *Hydrology and Earth System Sciences*, **16**(2), 375–390.

Ponziani, M, Ponziani, D, Giorgi, A, Stevenin, H, & Ratto, SM. 2023. The use of machine learning techniques for a predictive model of debris flows triggered by short intense rainfall. *Natural Hazards*, 1–20.

Potin, Pierre, Bargellini, Pier, Laur, Henri, Rosich, Betlem, & Schmuck, Siegfried. 2012. Sentinel-1 mission operations concept. *Pages 1745–1748 of: 2012 IEEE international geoscience and remote sensing symposium*. IEEE.

Pramoditha, Rukshan. *Overview of a Neural Network's Learning Proce*. Accessed: 2025-11-24.

Prechelt, Lutz. 2002. Early stopping-but when? *Pages 55–69 of: Neural Networks: Tricks of the trade*. Springer.

Qin, Dahe, Yao, Tandong, Ding, Yongjian, & Ren, Jiawen. 2021. *Introduction to cryospheric science*. Springer Nature.

Räisänen, Jouni. 2008. Warmer climate: less or more snow? *Climate Dynamics*, **30**(2), 307–319.

Ramyachitra, D, & Manikandan, Parasuraman. 2014. Imbalanced dataset classification and solutions: a review. *International Journal of Computing and Business Research (IJCBR)*, **5**(4), 1–29.

Rasul, Golam, & Molden, David. 2019. The global social and economic consequences of mountain cryospheric change. *Frontiers in Environmental Science*, **7**, 91.

Reba, Michele L, Marks, Danny, Winstral, Adam, Link, Timothy E, & Kumar, Mukesh. 2011. Sensitivity of the snowcover energetics in a mountain basin to variations in climate. *Hydrological Processes*, **25**(21), 3312–3321.

Reichle, Rolf H, Koster, Randal D, Liu, Ping, Mahanama, Sarith PP, Njoku, Eni G, & Owe, Manfred. 2007. Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR). *Journal of Geophysical Research: Atmospheres*, **112**(D9).

Reichstein, Markus, Camps-Valls, Gustau, Stevens, Bjorn, Jung, Martin, Denzler, Joachim, Carvalhais, Nuno, & Prabhat, F. 2019a. Deep learning and process understanding for data-driven Earth system science. *Nature*, **566**(7743), 195–204.

Reichstein, Markus, Camps-Valls, Gustau, Stevens, Bjorn, Jung, Martin, Denzler, Joachim, Carvalhais, Nuno, & Prabhat. 2019b. Deep learning and process understanding for data-driven Earth system science. *Nature*, **566**(7743), 195–204.

Reinking, Adele K., Højlund Pedersen, Stine, Elder, Kelly, Boelman, Natalie T., Glass, Thomas W., Oates, Brendan A., Bergen, Scott, Roberts, Shane, Prugh, Laura R., Brinkman, Todd J., Coughenour, Michael B., Feltner, Jennifer A., Barker, Kristin J., Bentzen, Torsten W., Pedersen, Åshild Ø., Schmidt, Niels M., & Liston, Glen E. 2022. Collaborative wildlife–snow science: Integrating wildlife and snow expertise to improve research and management. *Ecosphere*, **13**(6).

Ren, Weiwei, Zhu, Zhongzheng, Wang, Yingzheng, Su, Jianbin, Zeng, Ruijie, Zheng, Donghai, & Li, Xin. 2024. Comparison of Machine Learning Models in Simulating Glacier Mass Balance: Insights from Maritime and Continental Glaciers in High Mountain Asia. *Remote Sensing*, **16**(6), 956.

Revuelto, J, Alonso-González, E, Deschamps-Berger, C, Gutmann, ED, & López-Moreno, JI. 2025. Recent Advances in Snow Monitoring from Local to Global Scales. *Current Climate Change Reports*, **11**(1), 10.

Revuelto, Jesús, López-Moreno, Juan I, Azorín-Molina, César, Zabalza, J, Arguedas, G, & Vicente-Serrano, Sergio M. 2014. Mapping the annual evolution of snow depth in a small catchment in the Pyrenees using the long-range terrestrial laser scanning. *Journal of Maps*, **10**(3), 379–393.

Rew, Lisa J, McDougall, Keith L, Alexander, Jake M, Daehler, Curtis C, Essl, Franz, Haider, Sylvia, Kueffer, Christoph, Lenoir, Jonathan, Milbau, Ann, Nuñez, Martin A, *et al.* 2020. Moving up and over: redistribution of plants in alpine, Arctic, and Antarctic ecosystems under global change. *Arctic, Antarctic, and Alpine Research*, **52**(1), 651–665.

Rittger, Karl, Painter, Thomas H, & Dozier, Jeff. 2013. Assessment of methods for mapping snow cover from MODIS. *Advances in Water Resources*, **51**, 367–380.

Rittger, Karl, Bormann, Kat J, Bair, Edward H, Dozier, Jeff, & Painter, Thomas H. 2021a. Evaluation of VIIRS and MODIS snow cover fraction in high-mountain Asia using landsat 8 OLI. *Frontiers in Remote Sensing*, **2**, 647154.

Rittger, Karl, Krock, Mitchell, Kleiber, William, Bair, Edward H, Brodzik, Mary J, Stephenson, Thomas R, Rajagopalan, Balaji, Bormann, Kat J, & Painter, Thomas H. 2021b. Multi-sensor fusion using random forests for daily fractional snow cover at 30 m. *Remote Sensing of Environment*, **264**, 112608.

Robinson, David A. 1989. Evaluation of the collection, archiving and publication of daily snow data in the United States. *Physical Geography*, **10**(2), 120–130.

Rodés-Guirao, Lucas. 2019. Deep Learning for Digital Typhoon : Exploring a typhoon satellite image dataset using deep learning.

Rohli, Robert V, & Li, Chunyan. 2021. Energy Transfer and Electromagnetic Radiation. *Pages 27–47 of: Meteorology for Coastal Scientists*. Springer.

Rouzegari, Nazak, Bolboli Zadeh, Mohammad, Jimenez Arellano, Claudia, Afzali Gorooh, Vesta, Nguyen, Phu, Meng, Huan, Ferraro, Ralph R, Kalluri, Satya, Sorooshian, Soroosh, & Hsu, Kuolin. 2025. Passive Microwave Imagers, Their Applications, and Benefits: A Review. *Remote Sensing*, **17**(9), 1654.

Rudari, Roberto, Entekhabi, Dara, & Roth, Giorgio. 2005. Large-scale atmospheric patterns associated with mesoscale features leading to extreme precipitation events in Northwestern Italy. *Advances in Water Resources*, **28**(6), 601–614.

Rumelhart, David E, Hinton, Geoffrey E, Williams, Ronald J, *et al.* 1985. *Learning internal representations by error propagation.*

Ryan, Wendy A, Doesken, Nolan J, & Fassnacht, Steven R. 2008. Preliminary results of ultrasonic snow depth sensor testing for National Weather Service (NWS) snow measurements in the US. *Hydrological Processes: An International Journal*, **22**(15), 2748–2757.

Réveillet, MARION, VINCENT, CHRISTIAN, SIX, DELPHINE, & RABATEL, ANTOINE. 2016. Which empirical model is best suited to simulate glacier mass balances? *Journal of Glaciology*, **63**(237), 39–54.

Sabzipour, Behmard, Arsenault, Richard, Troin, Magali, Martel, Jean-Luc, & Brissette, François. 2023. Sensitivity analysis of the hyperparameters of an ensemble Kalman filter application on a semi-distributed hydrological model for streamflow forecasting. *Journal of Hydrology*, **626**(Nov.), 130251.

Sahu, Mukul Kumar, Shwetha, H. R., & Dwarakish, G. S. 2023. State-of-the-art hydrological models and application of the HEC-HMS model: a review. *Modeling Earth Systems and Environment*, **9**(3), 3029–3051.

Samuel, Arthur L. 1959. Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, **3**(3), 210–229.

Sanders-DeMott, Rebecca, McNellis, Risa, Jabouri, Maroua, & Templer, Pamela H. 2018. Snow depth, soil temperature and plant–herbivore interactions mediate plant response to climate change. *Journal of Ecology*, **106**(4), 1508–1519.

Särkkä, S., & Svensson, L. 2023. *Bayesian Filtering and Smoothing.* 2 edn. Cambridge University Press.

Savenije, H. H. G. 2009. HESS Opinions "The art of hydrology"*. *Hydrology and Earth System Sciences*, **13**(2), 157–161.

Schiano Di Cola, Vincenzo, Cuomo, Salvatore, & Severino, Gerardo. 2021. Remarks on the numerical approximation of Dirac delta functions. *Results in Applied Mathematics*, **12**, 100200.

Schmidt, Lennart, Schaefer, David, Geller, Juliane, Lünenschloss, Peter, Palm, Bert, Rinke, Karsten, & Bumberger, Jan. 2018. System for automated Quality Control (SaQC) to enable traceable and reproducible data streams in environmental science. *Available at SSRN 4173698*.

Shen, Chaopeng. 2018. A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resources Research*, **54**(11), 8558–8593.

Shen, Chaopeng, Laloy, Eric, Elshorbagy, Amin, Albert, Adrian, Bales, Jerad, Chang, Fi-John, Ganguly, Sangram, Hsu, Kuo-Lin, Kifer, Daniel, Fang, Zheng, Fang, Kuai, Li, Dongfeng, Li, Xiaodong, & Tsai, Wen-Ping. 2018. HESS Opinions: Incubating deep-learning-powered hydrologic science advances as a community. *Hydrology and Earth System Sciences*, **22**(11), 5639–5656.

Shen, Chaopeng, Chen, Xingyuan, & Laloy, Eric. 2021. *Broadening the use of machine learning in hydrology.*

Sheridan, Iain. 2019. *Drone Navigation in Polar and Cryospheric Regions.* Ph.D. thesis.

Shrestha, Prabhakar, & Barros, Ana P. 2025. Multi-Physics Data Assimilation Framework for Remotely Sensed Snowpacks to Improve Water Prediction. *Water Resources Research*, **61**(2), e2024WR037885. e2024WR037885 2024WR037885.

Siirila-Woodburn, Erica R, Rhoades, Alan M, Hatchett, Benjamin J, Huning, Laurie S, Szinai, Julia, Tague, Christina, Nico, Peter S, Feldman, Daniel R, Jones, Andrew D, Collins, William D, *et al.* 2021. A low-to-no snow future and its impacts on water resources in the western United States. *Nature Reviews Earth & Environment*, **2**(11), 800–819.

Singh, Bhupinderjeet, Ferdousi, Tanvir, Abatzoglou, John T., Swarup, Samarth, Adam, Jennifer C., & Rajagopalan, Kirti. 2024. Sensitivity of snow magnitude and duration to hydrology model parameters. *Journal of Hydrology*, **645**, 132193.

Sit, Muhammed, Demiray, Bekir Z, Xiang, Zhongrun, Ewing, Gregory J, Sermet, Yusuf, & Demir, Ibrahim. 2020a. A comprehensive review of deep learning applications in hydrology and water resources. *Water Science and Technology*, **82**(12), 2635–2670.

Sit, Muhammed, Demiray, Bekir Z., Xiang, Zhongrun, Ewing, Gregory J., Sermet, Yusuf, & Demir, Ibrahim. 2020b. A comprehensive review of deep learning applications in hydrology and water resources. *Water Science and Technology*, **82**(12), 2635–2670.

Skiles, S McKenzie, Donahue, Christopher P, Hunsaker, Adam G, & Jacobs, Jennifer M. 2023. UAV hyperspectral imaging for multiscale assessment of Landsat 9 snow grain size and albedo. *Frontiers in Remote Sensing*, **3**, 1038287.

Skovsholt, Louis J, Pastor, Ada, Docherty, Catherine L, Milner, Alexander M, & Riis, Tenna. 2020. Changes in hydrology affects stream nutrient uptake and primary production in a high-Arctic stream. *Biogeochemistry*, **151**(2), 187–201.

Slater, Andrew G, & Clark, Martyn P. 2006. Snow data assimilation via an ensemble Kalman filter. *Journal of Hydrometeorology*, **7**(3), 478–493.

Slater, Louise, Blougouras, Georgios, Deng, Liangkun, Deng, Qimin, Ford, Emma, Hoek van Dijke, Anne, Huang, Feini, Jiang, Shijie, Liu, Yinxue, Moulds, Simon, Schepen, Andrew, Yin, Jiabo, & Zhang, Boen. 2025. Challenges and opportunities of ML and explainable AI in large-sample hydrology. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **383**(2302).

Smith, J., & Eli, R. N. 1995. Neural Network Models of Rainfall-Runoff Processes. *Journal of Hydrology*, **171**(1-4), 1–11.

Smith, Leslie N. 2015. Cyclical Learning Rates for Training Neural Networks.

Smyth, Eric J, Raleigh, Mark S, & Small, Eric E. 2020. Improving SWE estimation with data assimilation: The influence of snow depth observation timing and uncertainty. *Water Resources Research*, **56**(5), e2019WR026853.

Soltani, Mohsen, Fletcher, Christopher G., Erler, Andre, Bunn, Melissa I., & Russell, Hazen A.J. 2026. Assessment of snow water equivalent characteristics in time and space over the Mackenzie River basin. *Canadian Water Resources Journal / Revue canadienne des ressources hydriques*, Feb., 1–24.

Song, Yalan, Tsai, Wen-Ping, Gluck, Jonah, Rhoades, Alan, Zarzycki, Colin, McCrary, Rachel, Lawson, Kathryn, & Shen, Chaopeng. 2024. LSTM-based data integration to improve snow water equivalent prediction and diagnose error sources. *Journal of Hydrometeorology*, **25**(1), 223–237.

Soomro, Shan-e-hyder, Soomro, Abdul Razzaque, Batool, Sahar, Guo, Jiali, Li, Yinghai, Bai, Yanqin, Hu, Caihong, Tayyab, Muhammad, Zeng, Zhiqiang, Li, Ao, *et al.* 2024. How does the climate change effect on hydropower potential, freshwater fisheries, and hydrological response of snow on water availability? *Applied Water Science*, **14**(4), 65.

Spoto, Francois, Sy, Omar, Laberinti, Paolo, Martimort, Philippe, Fernandez, Valerie, Colin, Olivier, Hoersch, Bianca, & Meygret, Aime. 2012. Overview of sentinel-2. *Pages 1707–1710 of: 2012 IEEE international geoscience and remote sensing symposium.* IEEE.

Steele, Hannah, Small, Eric E., & Raleigh, Mark S. 2024. Demonstrating a Hybrid Machine Learning Approach for Snow Characteristic Estimation Throughout the Western United States. *Water Resources Research*, **60**(6).

Strasser, Ulrich, Warscher, Michael, Rottler, Erwin, & Hanzer, Florian. 2024. openAMUNDSEN v1. 0: an open-source snow-hydrological model for mountain regions. *Geoscientific Model Development*, **17**(17), 6775–6797.

Strobl, Carolin, Boulesteix, Anne-Laure, Zeileis, Achim, & Hothorn, Torsten. 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC bioinformatics*, **8**(1), 1–21.

Sturm, Matthew. 2015. White water: Fifty years of snow research in WRR and the outlook for the future. *Water Resources Research*, **51**(7), 4948–4965.

Sturm, Matthew, & Liston, Glen E. 2021a. Revisiting the global seasonal snow classification: An updated dataset for earth system applications. *Journal of Hydrometeorology*, **22**(11), 2917–2938.

Sturm, Matthew, & Liston, Glen E. 2021b. Revisiting the Global Seasonal Snow Classification: An Updated Dataset for Earth System Applications. *Journal of Hydrometeorology*, **22**(11), 2917 – 2938.

Sturm, Matthew, Taras, Brian, Liston, Glen E, Derksen, Chris, Jonas, Tobias, & Lea, Jon. 2010. Estimating snow water equivalent using snow depth data and climate classes. *Journal of Hydrometeorology*, **11**(6), 1380–1394.

Sturm, Matthew, Goldstein, Michael A, & Parr, Charles. 2017. Water and life from snow: A trillion dollar science question. *Water Resources Research*, **53**(5), 3534–3544.

Sun, Liyang, Zhang, Xueliang, Wang, Huadong, Xiao, Pengfeng, & Wang, Yunhan. 2024. Estimating Daily Snow Density Through a Spatiotemporal Random Forest Model. *Water Resources Research*, **60**(7).

Sun, Ziheng, Sandoval, Laura, Crystal-Ornelas, Robert, Mousavi, S Mostafa, Wang, Jinbo, Lin, Cindy, Cristea, Nicoleta, Tong, Daniel, Carande, Wendy Hawley, Ma, Xiaogang, *et al.* 2022. A review of earth artificial intelligence. *Computers & Geosciences*, 105034.

Svoboda, J., Ruesch, M., Liechti, D., Jones, C., Volpi, M., Zehnder, M., & Schweizer, J. 2025. Towards deep-learning solutions for classification of automated snow height measurements (CleanSnow v1.0.2). *Geoscientific Model Development*, **18**(5), 1829–1849.

Taheri, Mercedeh, & Mohammadian, Abdolmajid. 2022. An overview of snow water equivalent: Methods, challenges, and future outlook. *Sustainability*, **14**(18), 11395.

Tahmasebi, Pejman, Kamrava, Serveh, Bai, Tao, & Sahimi, Muhammad. 2020. Machine learning in geo- and environmental sciences: From small to large scale. *Advances in Water Resources*, **142**, 103619.

talk science, Let's. 2021. *Ice on Earth - The Cryosphere.* Accessed: 2025-11-12.

Tarboton, David G, Luce, Charles H, *et al.* 1996. *Utah energy balance snow accumulation and melt model (UEB).* Citeseer.

Taylor, Liam S, Quincey, Duncan J, Smith, Mark W, Baumhoer, Celia A, McMillan, Malcolm, & Mansell, Damien T. 2021. Remote sensing of the mountain cryosphere: Current capabilities and future opportunities for research. *Progress in Physical Geography: Earth and Environment*, **45**(6), 931–964.

Terzago, S., Andreoli, V., Arduini, G., Balsamo, G., Campo, L., Cassardo, C., Cremonese, E., Dolia, D., Gabellani, S., von Hardenberg, J., Morra di Cella, U., Palazzi, E., Piazzi, G., Pogliotti, P., & Provenzale, A. 2020. Sensitivity of snow models to the accuracy of meteorological forcings in mountain environments. *Hydrology and Earth System Sciences*, **24**(8), 4061–4090.

Thackeray, Chad, Derksen, Chris, Fletcher, Christopher, & Hall, Alex. 2019. Snow and Climate: Feedbacks, Drivers, and Indices of Change. *Current Climate Change Reports*, **5**(12).

Thackeray, Chad W, & Fletcher, Christopher G. 2016. Snow albedo feedback: Current knowledge, importance, outstanding issues and future directions. *Progress in Physical Geography*, **40**(3), 392–408.

Toreti, A., *et al.* 2022a. Drought in Europe - April 2022.

Toreti, A., Bavera, D., Avanzi, F., Cammalleri, C., De Felice, M., De Jager, A., Di Ciollo, C., Gabellani, S., Maetens, W., Magni, D., Manfron, G., Masante, D., Mazzeschi, M., Mccormick, N., Naumann, G., Niemeyer, S., Rossi, L., Seguini, L., Spinoni, J., & Van Den Berg, M. 2022b. Drought in northern Italy - March 2022: GDO analytical report.

Torres, Ramon, Snoeij, Paul, Geudtner, Dirk, Bibby, David, Davidson, Malcolm, Attema, Evert, Potin, Pierre, Rommen, BjÖrn, Floury, Nicolas, Brown, Mike, *et al.* 2012. GMES Sentinel-1 mission. *Remote sensing of environment*, **120**, 9–24.

Toth, E, Brath, A, & Montanari, A. 2000. Comparison of short-term rainfall prediction models for real-time flood forecasting. *Journal of hydrology*, **239**(1-4), 132–147.

Treichler, Désirée, & Kääb, Andreas. 2017. Snow depth from ICESat laser altimetry—A test study in southern Norway. *Remote sensing of environment*, **191**, 389–401.

Trenberth, KE. 2011. Changes in precipitation with climate change. *Climate Research*, **47**(1), 123–138.

Tsantekidis, Avraam, Passalis, Nikolaos, & Tefas, Anastasios. 2022. Chapter 5 - Recurrent neural networks. *Pages 101–115 of:* Iosifidis, Alexandros, & Tefas, Anastasios (eds), *Deep Learning for Robot Perception and Cognition*. Academic Press.

Tyralis, Hristos, Papacharalampous, Georgia, & Langousis, Andreas. 2019. A brief review of random forests for water scientists and practitioners and their recent history in water resources. *Water*, **11**(5), 910.

Tyson, Conor, Longyang, Qianqiu, Neilson, Bethany T., Zeng, Ruijie, & Xu, Tianfang. 2023. Effects of meteorological forcing uncertainty on high-resolution snow modeling and streamflow prediction in a mountainous karst watershed. *Journal of Hydrology*, **619**, 129304.

Valence, Eole, Baraer, Michel, Rosa, Eric, Barbecot, Florent, & Monty, Chloe. 2022. Drone-based ground-penetrating radar (GPR) application to snow hydrology. *The Cryosphere*, **16**(9), 3843–3860.

Vali, Ava, Comai, Sara, & Matteucci, Matteo. 2020. Deep Learning for Land Use and Land Cover Classification Based on Hyperspectral and Multispectral Earth Observation Data: A Review. *Remote Sensing*, **12**(15), 2495.

van Hateren, Theresa C, Jongen, Harro J, Al-Zawaidah, Hadeel, Beemster, Joris GW, Boekee, Judith, Bogerd, Linda, Gao, Sijia, Kannen, Christin, van Meerveld, Ilja, de Lange, Sjoukje I, *et al.* 2023. Where should hydrology go? An early-career perspective on the next IAHS scientific decade: 2023–2032. *Hydrological Sciences Journal*, **68**(4), 529–541.

van Huissteden, J. 2020. *Thawing Permafrost: Permafrost Carbon in a Warming Arctic*. Springer International Publishing.

Van Leeuwen, Peter Jan. 2015. Representation errors and retrievals in linear and nonlinear data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **141**(690), 1612–1623.

Van Rijsbergen, C. 1979. Information retrieval: theory and practice. *In: Proceedings of the Joint IBM/University of Newcastle upon Tyne Seminar on Data Base Systems*, vol. 79.

Viallon-Galinier, Léo, Hagenmuller, Pascal, & Lafaysse, Matthieu. 2020. Forcing and evaluating detailed snow cover models with stratigraphy observations. *Cold Regions Science and Technology*, **180**, 103163.

Vionnet, V, Brun, E, Morin, S, Boone, A, Faroux, S, Le Moigne, P, Martin, E, & Willemet, J-M. 2012. The detailed snowpack scheme Crocus and its implementation in SURFEX v7. 2. *Geoscientific model development*, **5**(3), 773–791.

Vitasse, Yann, Rebetez, Martine, Filippa, Gianluca, Cremonese, Edoardo, Klein, Geoffrey, & Rixen, Christian. 2017. 'Hearing'alpine plants growing after snowmelt: ultrasonic snow sensors provide long-term series of alpine plant phenology. *International Journal of Biometeorology*, **61**(2), 349–361.

Viviroli, Daniel, Dürr, Hans H, Messerli, Bruno, Meybeck, Michel, & Weingartner, Rolf. 2007. Mountains of the world, water towers for humanity: Typology, mapping, and global significance. *Water resources research*, **43**(7).

Voordendag, Annelies, Réveillet, Marion, MacDonell, Shelley, & Lhermitte, Stef. 2021. Snow model comparison to simulate snow depth evolution and sublimation at point scale in the semi-arid Andes of Chile. *The Cryosphere Discussions*, **2021**, 1–25.

Walvoord, Michelle A, & Kurylyk, Barret L. 2016. Hydrologic impacts of thawing permafrost—A review. *Vadose Zone Journal*, **15**(6), vzj2016–01.

Wang, Gongxue, Jiang, Lingmei, Xiong, Chuan, & Zhang, Yongsheng. 2022a. Characterization of NDSI variation: Implications for snow cover mapping. *IEEE Transactions on Geoscience and Remote Sensing*, **60**, 1–18.

Wang, Jiwen, Yuan, Qiangqiang, Shen, Huanfeng, Liu, Tingting, Li, Tongwen, Yue, Linwei, Shi, Xiaogang, & Zhang, Liangpei. 2020. Estimating snow depth by combining satellite data and ground-based observations over Alaska: A deep learning approach. *Journal of Hydrology*, **585**, 124828.

Wang, Xiaoyuan, Lu, Haibo, & Yuan, Wenping. 2022b. Inter-Annual Variations of Precipitation Modulate the Dry Spell Length. *GeoHealth*, **6**(4), e2022GH000611.

Webster, Clare, & Jonas, Tobias. 2018. Influence of canopy shading and snow coverage on effective albedo in a snow-dominated evergreen needleleaf forest. *Remote Sensing of Environment*, **214**, 48–58.

Welch, G. 1995. An Introduction to the Kalman Filter.

Wever, Nico. 2017. *Weissfluhjoch dataset for ESM-SnowMIP*.

Wieder, William R, Kennedy, Daniel, Lehner, Flavio, Musselman, Keith N, Rodgers, Keith B, Rosenbloom, Nan, Simpson, Isla R, & Yamaguchi, Ryohei. 2022. Pervasive alterations to snow-dominated ecosystem functions under climate change. *Proceedings of the National Academy of Sciences*, **119**(30), e2202393119.

Wikipedia contributors. 2004. *Wikipedia, The Free Encyclopedia*. [Online; accessed 27-November-2025].

Williams, Christopher KI, & Rasmussen, Carl Edward. 2006. *Gaussian processes for machine learning*. Vol. 2. MIT press Cambridge, MA.

Xu, Tianfang, & Liang, Feng. 2021. Machine learning for hydrologic sciences: An introductory overview. *Wiley Interdisciplinary Reviews: Water*, **8**(5), e1533.

Yaseen, Zaher Mundher. 2023. A New Benchmark on Machine Learning Methodologies for Hydrological Processes Modelling: A Comprehensive Review for Limitations and Future Research Directions. *Knowledge-Based Engineering and Sciences*, **4**(3), 65–103.

Yasin, Hajar Maseeh, & Khorsheed, Aso Kareem. 2025. Automated Data Cleaning in Large Databases Using Machine Learning Methods. *Asian Journal of Research in Computer Science*, **18**(5), 364–386.

Yu, Manzhu, Huang, Qunying, & Li, Zhenlong. 2024. Deep learning for spatiotemporal forecasting in Earth system science: a review. *International Journal of Digital Earth*, **17**(1), 2391952.

Zanotti, Fabrizio, Endrizzi, Stefano, Bertoldi, Giacomo, & Rigon, Riccardo. 2004. The GEOTOP snow module. *Hydrological Processes*, **18**(18), 3667–3679.

Zaparoli Cunha, Barbara, Droz, Christophe, Zine, Abdel-Malek, Foulard, Stéphane, & Ichchou, Mohamed. 2023. A review of machine learning methods applied to structural dynamics and vibroacoustic. *Mechanical Systems and Signal Processing*, **200**(Oct.), 110535.

Zealand, Cameron M, Burn, Donald H, & Simonovic, Slobodan P. 1999. Short term streamflow forecasting using artificial neural networks. *Journal of hydrology*, **214**(1-4), 32–48.

Zhang, Tingjun. 2005. Influence of the seasonal snow cover on the ground thermal regime: An overview. *Reviews of Geophysics*, **43**(4).

Zhang, Z., Glaser, S., Bales, R., Conklin, M., Rice, R., & Marks, D. 2017. Insights into mountain precipitation and snowpack from a basin-scale wireless-sensor network. *Water Resources Research*, **53**(8), 6626–6641.

Zhao, F., Gong, W., Bianchini, S., & Yang, Z. 2024a. Linking glacier retreat with climate change on the Tibetan Plateau through satellite remote sensing. *The Cryosphere*, **18**(12), 5595–5612.

Zhao, Tianjie, Wang, Sheng, Ouyang, Chaojun, Chen, Min, Liu, Chenying, Zhang, Jin, Yu, Long, Wang, Fei, Xie, Yong, Li, Jun, Wang, Fang, Grunwald, Sabine, Wong, Bryan M., Zhang, Fan, Qian, Zhen, Xu, Yongjun, Yu, Chengqing, Han, Wei, Sun, Tao, Shao, Zezhi, Qian, Tangwen, Chen, Zhao, Zeng, Jiangyuan, Zhang, Huai, Letu, Husi, Zhang, Bing, Wang, Li, Luo, Lei, Shi, Chong, Su, Hongjun, Zhang, Hongsheng, Yin, Shuai, Huang, Ni, Zhao, Wei, Li, Nan, Zheng, Chaolei, Zhou, Yang, Huang, Changping, Feng, Defeng, Xu, Qingsong, Wu, Yan, Hong, Danfeng, Wang, Zhenyu, Lin, Yinyi, Zhang, Tangtang, Kumar, Prashant, Plaza, Antonio, Chanussot, Jocelyn, Zhang, Jiabao, Shi, Jiancheng, & Wang, Lizhe. 2024b. Artificial intelligence for geoscience: Progress, challenges, and perspectives. *The Innovation*, **5**(5), 100691.

Zhong, Shifa, Zhang, Kai, Bagheri, Majid, Burken, Joel G, Gu, April, Li, Baikun, Ma, Xingmao, Marrone, Babetta L, Ren, Zhiyong Jason, Schrier, Joshua, *et al.* 2021. Machine learning: new ideas and tools in environmental science and engineering. *Environmental Science & Technology*, **55**(19), 12741–12754.

Zwally, H Jay. 2002. *ICESat: Ice, Cloud, and land Elevation Satellite*. NASA Goddard Space Flight Center.