

# **IDENTIFYING AND PREDICTING PATTERNS OF SNOWPACK RIPENING WITH MACHINE LEARNING METHODS**

by

# CLÉMENT CHERBLANC

Bachelor's Degree in Geosciences, Université Savoie Mont Blanc,  
Le Bourget-du-Lac, France, 2022

Thesis  
presented in partial fulfillment of the requirements for the  
Degree of

# Master's Thesis in Geosciences

University of Montana

July 2024

Approved by

Ashby Kinch  
(Graduate School)

Dr. Joel T. Harper  
(UMT, Geosciences)

Dr. Javier Pérez Álvaro  
(UMT, Mathematics)

Dr. Jesse V. Johnson  
(UMT, Computer Sciences)

Dr. W. Payton Gardner  
(UMT, Geosciences)

I confirm that this Master's thesis is my own work and I have documented all sources and material used.

© COPYRIGHT

by

Clément Cherblanc

2024

All rights reserved

Munich, 29.06.2024

Clément Cherblanc

## Identifying and Predicting Patterns of Snowpack Ripening with Machine Learning Methods

Chairperson: Dr. Joel T. Harper

The timing of water release from the snowpack plays key roles in ecosystem services, groundwater recharge, and water resource management. However, two internal barriers in a standing snowpack must be overcome before runoff can outflow from the base: 1) the cold content must be exhausted, and 2) the interconnected network of snow grains must be filled with liquid water to residual saturation. Expressing the liquid water as latent heat allows the two barriers to be grouped as an energy ( $J/m^2$ ) to define a snowpack's Runoff Energy Hurdle (REH). The growth and loss of REH is driven by evolution of pore space and heat exchange with the atmosphere, both of which are modulated by the seasonal cycle. The snowpack is commonly referred to as 'ripe' after energy input to the snowpack from heat/water has exhausted REH. Here, we define activation as the switch to REH=0 of a given mass of snow, at a location and time, indicating that the snowpack is capable of runoff outflow from the base. The spring activation is defined as the highest magnitude of snow-water-equivalent activated in a water year. First, we seek to quantify the spatial and temporal variability of activations and spring activations within the snowpack of the South Fork of the Flathead River, Montana, a 4341 km<sup>2</sup> mountainous watershed. We use the 2005-2023 daily outputs of the SNODAS model to compute distributed REH and activations. For north aspects, spring activation occurs 3.2 days later per 100 m elevation gain, and for southwest aspects it is 4.8 days later per 100 m elevation gain. We classify yearly REH time series in 4 clusters using K-Means. Cluster 1 contains the lowest REH locations, always including the valley floor but extending to higher elevations during warm and low snow-accumulation years. During low snow years, up to 92% of the watershed is in cluster 1 and the snowpack is activated numerous times throughout the winter season. In contrast, cluster 2 is represented all years and cluster 3 15/19 years, they characterize a snowpack with greater REH having 69% of the basin's snow water equivalent volume, sitting at higher elevations and ripening within a two-month period. Finally, cluster 4 only appears during exceptionally high-accumulation years, drawing the locations with the greatest snow accumulation from cluster 3 and activating within a one-month period. Second, we assess predictive models for spring activation timing for clusters 2 through 4, based on partial REH time series until 1 March. We find that K-Means clustering yields the best predictions of spring activation timing for the high-REH clusters (3 and 4) that can only ripen once energy inputs become great in the spring; the two respective mean absolute errors (MAE) are 12.3 and 8.8 days, which is an improvement from the historical mean of 16.9 days. A Long Short-Term Memory (LSTM) neural network is deployed, taking advantage of the temporal correlation and embracing the non-linearity of REH, but it only outperforms K-Means (MAE=21.3 days) for cluster 2 with a mean absolute error of 19.3 days, barely outperforming the cluster 2's historical mean. Years mostly classified in cluster 1 are expected to become more frequent in a warmer world, bringing numerous activations throughout the winter that cannot be predicted, unlike higher clusters' activations, therefore reducing the predictability of spring activation timing for flood forecasters and dam managers. Conversely, increased water vapor in the atmosphere with the rising temperatures will likely promote continued development of cluster 4 in heavy accumulation years, which has a narrow window for spring activation timing.

# Contents

1 Introduction .....	1
2 Background .....	4
2.1 Components and Processes of the Runoff Energy Hurdle .....	4
2.1.1 Cold Content .....	4
2.1.2 Residual Saturation Deficit .....	6
2.1.3 Physical Processes to Exhaust the REH .....	6
2.2 Spatio-Temporal Variability of the Snowmelt Metrics .....	7
2.3 Snowmelt and Climate Change .....	9
3 Methods .....	10
3.1 Study Site .....	10
3.2 Snowpack Analysis .....	12
3.2.1 SNODAS Model .....	12
3.2.2 The Runoff Energy Hurdle Computation .....	13
3.2.3 REH Dataset .....	14
3.2.4 Defining the Spring Activation .....	15
3.3 Statistical Methods, Machine and Deep Learning .....	16
3.3.1 K-Means Clustering of REH Time Series .....	16
3.3.2 Statistical Methods .....	17
3.3.3 Yearly Mean Cluster Number and Correlations .....	18
3.3.4 Predicting the Timing of Spring Activations from Partial REH Time Series .....	18
3.3.5 Interpretable Machine Learning to Quantify Feature Importance . 20	
3.3.6 Shapley Values and Model Selection .....	21
3.3.7 Trends in Widespread Activation Events .....	22
4 Results .....	24
4.1 REH Behavior During a Water Year .....	24

4.1.1 Choice of the Number of Clusters .....	26
4.1.2 K-Means Clustering of the REH Time Series .....	27
4.1.3 Temporal Variability of Each Cluster .....	29
4.1.4 Spatial Variability of Each Cluster .....	35
4.2 Cold Content and Residual Saturation Deficit .....	39
4.2.1 Fraction of REH Due to Cold Content .....	39
4.2.2 Cluster Preferences for Cold Content and Residual Saturation Deficit Values .....	42
4.3 Activation of SWE .....	43
4.4 Predictive Models of the Timing of Spring Activation .....	46
4.4.1 Spring Activation Timing Forecasting from Partial REH Time Series .....	47
4.4.2 Terrain Feature Importance .....	50
4.5 Anatomy of Widespread Spring Activation Events .....	54
5 Discussion .....	55
5.1 Limitations .....	55
5.1.1 SNODAS .....	55
5.1.2 Uncertainties on the Residual Saturation Deficit .....	56
5.1.3 REH to Capture the Outflow of Water .....	57
5.1.4 Correlation of REH to SWE .....	57
5.1.5 Discrete Clustering in a Continuous System .....	57
5.2 Current State of REH and Activations .....	58
5.2.1 Influences on REH and Cluster Number .....	58
5.2.2 Consequences of REH Values on Activations .....	59
5.2.3 Implications of REH for Rain-on-snow Events .....	61
5.3 Implications for a Warmer World .....	62
5.4 Practical Predictions of Spring Activation Timing for Dam and Flood Managing .....	63
6 Conclusion .....	64
Bibliography .....	67

# 1 Introduction

The mountain snowpack is a key part of ecosystems. Snow is a source of water resources and contributes to hydrological processes (e.g., Broxton et al., 2019), and it stores water equivalent that will gradually melt throughout the spring and summer providing sustained runoff to the rivers, aquifers and ecosystems (e.g., Loe et al., 2016). Human population relies on the snowpack for drinking water, agriculture, cattle, hydroelectric energy production in dams, etc. Changes in the behavior of the snowpack have critical consequences downstream and during the following summer. Even though an early-season snowmelt is associated with lower melt rates due to lower sensible heat and solar radiation (Lundquist & Flint, 2006), an early snow runoff causes increased drought (Stewart, 2009) and increased wild-fires in the Arctic (Scholten et al., 2022) relative to later snowmelt. Two aspects of snowmelt runoff are important: 1) the amount of water stored in the snow that will run off, and 2) the timing of release of the water. The two factors are not independent as the amount of snow often dictates the timing. Furthermore, a snowpack must evolve to become capable of releasing meltwater for a given energy input, a process commonly called ripening.

Here we consider runoff capability in terms of both timing and amount with an ‘activation’ metric (defined in detail in section 4.2.4) which incorporates both the timing and amount of snow capable of yielding runoff. To understand the activation of a snowpack, we are first required to make the distinction between a non-runoff-capable snowpack and a runoff-capable snowpack. A non-runoff-capable snowpack is a snowpack that is cold and dry and so there is an energy hurdle to initiate runoff. Activation will happen when that snowpack warms up and reaches a certain amount of liquid water content in the snow, and thus becomes runoff capable. This warming up and wetting is equivalent to exhausting the Cold Content (CC) of the snowpack and reaching its Residual Saturation (RS). A ripened snowpack has no hurdle to runoff, while a cold and dry, winter snowpack has some hurdle to runoff. This is introduced as a snowpack’s Runoff Energy Hurdle

(REH), which is the sum of the cold content and the deficit to residual saturation or residual saturation deficit (Colbeck, 1976), expressed in J/m<sup>2</sup>.

The timeline of events is summarized as follows:

1. a typical winter snowpack accumulates cold content, and dry pore space. The REH and SWE (Snow Water Equivalent, in m<sup>3</sup>/m<sup>2</sup>=m) generally increase step wise due to snowfall and weather events.
2. surface energy inputs become sufficient to exhaust the cold content by conduction and turbulent heat transfer. Liquid water input at the snow surface, either by rainfall or snowmelt, infiltrates into the snow and refreezing, adding latent heat. Eventually all cold content is lost and the liquid water content in the pore spaces increases until residual saturation (nominally 7% of pore space) is reached. The snowpack in this state is commonly referred to as a “ripe” snowpack. Here, we define the transition from unripe to ripe (timing and SWE) as ‘activation’ of the snowpack. We use a new term to avoid ambiguity since the literature is consistent regarding the meaning of a snowpack in the “ripe” state, but not specific regarding the internal conditions required for transition. Further, our use of activation captures the time component of the transition and a magnitude component, which are our focus here.
3. Once the snowpack is completely activated, continued energy and water input will yield mass loss and runoff from the snowpack. From that point on, the liquid water at the interface of the snow and the ground is fractioned into soil and groundwater recharge, vegetation uptake, and runoff. Flow along the base or a layer of the snowpack can be on the order of meters to tens of meters per hour according to Eiriksson et al. (2013). Liquid water flow through saturated snow was found to reach 6 m/h (Singh et al., 1997).

The establishment of REH creates a new metric of the overall state of the snowpack. The REH quantifies the amount of energy needed to be input to a given snowpack at a given time to start runoff. When the REH drops to 0, the snowpack is activated, or ripened. A snowpack may be extremely thick yet have close to no REH because it is warm and contains some liquid water, by opposition to a thin

but cold snowpack. Two locations may exhibit very similar SWE curves but very different REH signals.

In summary, REH is the Energy Hurdle that needs to be overcome in a snowpack before it can produce snowmelt runoff. A ripened snowpack is an ambiguous definition used in the literature that we simplify to be a snowpack that has no REH. Snowpack activation is the timing and the amount of SWE transitioning from  $\text{REH}>0$  to  $\text{REH}=0$ . The spring activations corresponds to the greatest SWE activation and its timing.

A watershed spanning thousands of meters of elevation, with variable slope, aspect and vegetation interactions does not ripen at once. A distributed analysis of the REH and the loss of that REH is necessary to understand how topography influences snowpack ripening. That understanding is critical for dam management, soil infiltration and flood forecasting because different areas of a watershed are expected to activate at different moments and in response to different mechanisms. Precisely, south aspects receive more solar radiation, especially in the spring or during temperature inversions, while usual temperature gradients affect lower areas earlier in the season, regardless of aspect. Rain on a cold snowpack will not create runoff if the REH is sufficient, while rain on a ripened snow will (Schirmer et al., 2022). The timing of activation in a complex watershed is therefore crucial and requires understanding. Different watersheds are expected to behave differently, especially in different climates. For example, a wet and warm Washington coastal watershed will not have similar REH curves and activations to a cold and dry Northern Rockies continental watershed.

Prior work related to snowpack ripening and runoff generation is explored in more detail in section 4 with general conclusions and findings summarized as follows. Numerous studies consider the runoff behavior of entire watersheds using stream gauges, thus ignoring within-watershed variability. Other works only consider the timing of peak SWE at individual locations, without having distributed understanding of the behavior of the internal snowpack properties. Work has been done on cold content or snowpack liquid water content at the point scale. Just one

study addressed cold content and liquid water content combined but was limited to spring refreezing events at the point scale. Finally, melt emerging from sea ice and ice sheets has been quantified, but those environments are not directly upstream from millions of humans and mostly span flat topography or ocean. Although the consequences are important, scientific literature lacks a focused understanding of where and when a mountain snowpack activates and becomes runoff capable.

Predicting the activations before they occur allows for better managing of dams and water resources. This work aims to quantify the spatio-temporal variability of the activation of a snowpack in a continental watershed using machine learning methods. We utilize the SNODAS model output to compute REH past history, then deduce the activation of the snowpack history and use machine learning methods to quantify influences and differences in the behavior of the snowpack. We finally use more machine learning to make practical predictions of the activation of the snowpack using partial REH time series.

## 2 Background

### 2.1 Components and Processes of the Runoff Energy Hurdle

The Runoff Energy Hurdle (REH) has been used by Seligman et al. (2014) for the delaying of snowmelt during cold spring events under the name of spring energy hurdle. Seligman et al. (2014) considered that the spring energy hurdle was the sum of cold content and residual saturation deficit, but applied this expression to spring storms instead of the general snowpack behavior. The use of REH is more appropriate in our work since it considers the runoff ability of any snowpack.

#### 2.1.1 Cold Content

Multiple studies have focused on cold content. They have determined cold content differently: 1) as a function of air temperature during precipitation (Jennings et al., 2018), 2) as the physical expression used in this work and described further (Dewalle & Rango, 2008; Jennings et al., 2018; Jennings & Molotch, 2020; Selig-

man et al., 2014) , and 3) as the residual of the combination of mass and surface energy budgets Gordon et al. (2022). The physical expression is preferred for its precision and reliability, and the availability of data.

Using the model SNOWPACK, for an alpine and a sub-alpine location, Jennings & Molotch (2020) quantified the advancing of the day of the year when snow surface net fluxes are sufficient to exhaust the cold content under a warmer climate. However, they omitted treatment of the residual saturation deficit. They found that the melt onset advanced by 6.2 days for the alpine and 8.8 days in the subalpine per 1°C warming of the climate. The greater shift in subalpine could be due to the greater vegetation and/or higher temperature as both tend to drive heat into the snowpack. Jennings et al. (2018) acknowledged the relative importance of residual saturation deficit but ignore it in their focus on correlations between cold content and the timing of peak SWE. They found that snowfall cold content was responsible for the great majority of a snowpack cold content, and that the timing of peak cold content was highly correlated to the timing of peak SWE. Additionally, the highest melt days were days with 0 cold content at 6am, confirming the finding of Seligman et al. (2014) that REH is a key factor of snowmelt at an hourly time scale.

A dataset of snowpack cold content field observations was collected at 17 sites from the USGS Rocky Mountain Snowpack Chemistry Program through the US Rockies and was then analyzed by Schmidt (2017). Schmidt attempted to find trends between the topographic properties of the sites and the cold content, with no success. The cold content was extremely heterogeneous, through time and space. Thirty field measurements of snow temperature gradient were collected in a small basin in the Bridger Mountain Ranger by Deems et al. (2002). They performed simple linear regression between the measured snow temperature gradients and various properties (elevation, slope, aspect, etc.). They found correlations ranging from 0 to 70% on a given date, but pooled data from the whole season yielded low correlations, suggesting a very high temporal variability.

Parajuli et al. (2021) compared observed and modeled cold content in a humid boreal forest. Their three study sites were a few hundred meters apart and at similar elevations, aspects, and slopes. Their analysis therefore isolated vegetation influence in detail. They found greater SWE and cold content around shorter vegetation. In the context of rain on snow, Marks et al. (1998) and Marks et al. (2001) considered cold content in the energy budget of a snowpack during a rainstorm. Even though they omitted the residual saturation deficit, they found good agreement between observed and modeled SWE over the course of the storm.

Multiple papers use trends in the cold content changes of ice sheet firn to track sensible heat fluxes in and out of the ice sheets (e.g., Cox et al., 2015; Humphrey et al., 2012; Vandecrux et al., 2020). This is because most of the time, the firn cold content is sufficient to exhaust any liquid water intrusion. In rare occasions, liquid water sits in the firn for extended periods of time because of physical limitations to water percolation and the slow heat diffusion necessary to refreeze large volumes of water (Humphrey et al., 2021).

### **2.1.2 Residual Saturation Deficit**

Surface tension causes liquid water to stagnate in the meniscus of the snowpack pores. Before snow runoff can occur from the base of the snowpack, the pores must be filled to residual saturation, also called irreducible water content in the literature. Residual saturation has been estimated at 7% of pore space volume (Colbeck, 1974; Harris & Morrow, 1964), and as ranging 6.5 to 8.5% of pore space volume or 5 to 15% of SWE (Coléou & Lesaffre, 1998). Typically, if liquid water content throughout the snowpack exceeds the residual saturation, base runoff is initiated and transient values of liquid water content up to 30% of mass can be observed, especially in the spring when high liquid water input faces a limited runoff due to a saturated soil (Bonnell et al., 2021; Colbeck, 1976).

### **2.1.3 Physical Processes to Exhaust the REH**

In practice, exhausting REH is a complicated non-linear process. Rain and surface melting leads to liquid water percolating from the surface of the snowpack downward through "matrix" and/or "preferential" flow (Waldner et al., 2004). Matrix

(wetting front) or preferential (piping) flow depends on the flux of water input and the grain arrangement of the snowpack grains (Conway & Benedict, 1994). The water quickly exhausts the cold content due to the much greater latent heat of water than cold content (sensible heat) before building up to residual saturation.

The downward propagation of liquid water is even further complicated by piping, snow stratigraphy and occasional ice layers. Transition in snow layers are associated with surface tension changes, water head accumulates before bursting as a flow finger or a pipe (Marsh & Woo, 1984a). Especially, the fine over coarse boundaries impede water downward flow due to capillary effects (Waldner et al., 2004) but also foster ice layers when cold content is sufficient (McDowell et al., 2023). Ice layers restrict downward water flow in similar ways (Marsh & Woo, 1984a). While lateral flow mostly occurs at the base of snowpack, it remains occasional that lateral flow happens along snow stratigraphy when those are highly marked in a snowpack (Eriksson et al., 2013; Evans et al., 2016), sometimes resulting in water outflow from the stratigraphy into streams without reaching the base of the snowpack.

## 2.2 Spatio-Temporal Variability of the Snowmelt Metrics

Significant work has been done on various proxies of the timing of the snowpack activation. Those proxies include the timing of the disappearance of snow (e.g., Evan & Eisenman, 2021), the timing of disappearance of half of the SWE (e.g., McCabe et al., 2007), the amount of SWE (e.g., Kalra et al., 2008; Regonda et al., 2005), and blooming of flowers (e.g., Cayan et al., 2001). These studies all find trends towards an earlier loss of the snowpack in recent decades (Gordon et al., 2022). Moore et al. (2007) confirms and emphasizes that the peak river discharge is a complex non-linear response to higher temperatures. But little work is available on the activation of the snowpack, especially regarding spatial trends.

Remote sensing methods are the most common for distributed snowpack wetting studies. Zheng et al. (2022) used passive microwave data to detect the presence of liquid water and onset of snowmelt on sea ice, ice sheets, and seasonal snow for latitudes 50 to 90 degrees north and south. They observe significant wetting of

the snowpack, but with no absolute certainty or quantification that liquid water percolates to the base. They find a trend towards earlier melt onset for the period 1979-2018. Using ERA5 data, they found that main melt events are attributed to abnormal circulation conditions advecting warm air. The complexity of the topography associated with mountain snowpack resulted in conclusions being mainly drawn from the much flatter sea ice and ice sheets. Additional work has been done on sea ice melt onset (e.g., Anderson & Drobot, 2001; Arndt & Haas, 2019), but mountain snowpack studies are limited.

The timing of snow wetting can be found using Synthetic Aperture Radar (SAR). Diurnal wetting of the snow surface can be seen at high temporal resolution (Lund et al., 2022). Over greater areas and larger time scale, the timing of activation has its own signature. SAR signal over the Pacific Northwest stratovolcanoes revealed a strong relationship to elevation and a mostly secondary relationship to diurnal anisotropic heating index, which is a metric of cumulative potential solar radiation (Gagliano et al., 2023).

Gordon et al. (2022) acknowledges the high spatio-temporal variability of the snowmelt and river discharge (indirectly correlated to snowmelt) and suggests a framework to accommodate it. The framework is in three steps and considers 1) the overall water vapor budget during snow season, 2) the liquid water inputs during snow season and 3) the synchrony of precipitation and surface energy deficits that allow for solid precipitation. This purely physical framework is not complete yet but suggested as the first step towards large scale understanding of trends in the highly spatio-temporally variable snow processes. The framework is tested on a dataset combining observed streamflow and modeled snowpack properties. They estimate the potential changes in snowpack properties and the duration of the snow cover with regards to global warming in different key areas.

Using data from remote telemetry stations and PRISM, Musselman et al. (2021) found that the timing of the snowmelt onset was significantly more affected by the mean air temperature than by the amount of SWE. Warmer temperatures

implied a greater amount of snowmelt before 1 April, and a greater fraction of the total snowmelt amount had runoff by that time.

## 2.3 Snowmelt and Climate Change

Climate change is relevant to the snowpack in two ways: 1) general trends (temperatures, precipitation) and 2) the increase of the intensity, duration and frequency of extreme events.

With an earlier snowmelt onset, snowmelt rates are expected to be lower (Musselman et al., 2017) due to the lower shortwave radiation and sensible heat available earlier in the spring. With increasing temperatures, and even with increasing precipitation, snowmelt appears to happen earlier in most mountain regions of North America and Eurasia (Stewart, 2009). In a warmer climate, earlier and more unpredictable snowpack and snowmelt is expected to lower the ability to predict droughts (Livneh & Badger, 2020).

Climate change affects the spatiotemporal variability of the onset of snowmelt and its peak. The climate trends will on average cause snowmelt to happen significantly earlier (Jennings & Molotch, 2020; Rauscher et al., 2008; Stewart, 2009; Stewart et al., 2004) and at a lower rate (Lundquist & Flint, 2006; Musselman et al., 2017). This supports the dead reckoning concept that there must be an aspect and slope related variability of the snowpack ripening. However, the latter may not have a simple characterization due to influences from elevation, vegetation, topographic shading, etc. In other words, snowmelt onset will not occur at the same time for all same elevations, not even for the same elevations, aspects and slopes. Gottlieb & Mankin (2024) found that the response of snow depth and melt to temperature is highly non-linear, and much greater snow losses and earlier snowmelt are to be expected than what was seen so far.

The increasing intensity, duration and frequency of extreme events will have consequences on the already slower-melting snow. Indeed, "heavy precipitation over land" events are expected to increase (IPCC, 2021, fig. SPM. 6). The global trend of atmospheric rivers is extremely variable (Payne et al., 2020). Espinoza

et al. (2018) found that for the northern mid-latitudes (including the USA) will endure stronger and longer-lasting atmospheric rivers but little change in their frequency.

With regards to water resources and dam management, the expected intensification of winter and early spring snowmelt runoff and the limited availability of water in summer will result in the demand for water resources not to be met in areas with significant population and economic activity (Barnett et al., 2005).

## 3 Methods

### 3.1 Study Site

This work focuses on the Hungry Horse watershed, or South Fork of the Flathead River, shortened as SFF, presented on Figure 1. SFF is categorized as a HUC8 watershed in the Watershed Boundary Dataset (<https://www.usgs.gov/national-hydrography/watershed-boundary-dataset> ; ID: 17010209) by the US Geological Survey (<https://usgs.gov/> ; USGS). SFF is in northwest Montana, south of Glacier National Park, east of Flathead Lake and north of the Mission Mountains and its area is 4340 km<sup>2</sup>. The elevations in the basin range from 920 to 2650 masl (meters above sea level).

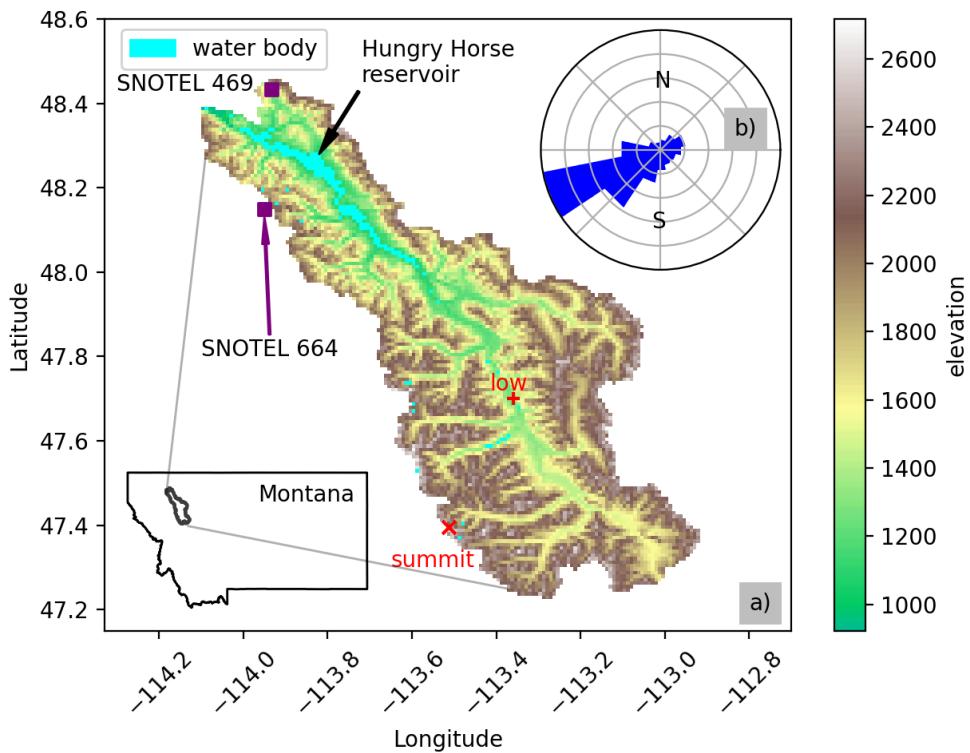


Figure 1: a) location, digital elevation model of the study site and key locations; b) wind rose for 19 years of ERA5 reanalysis (Hersbach et al. 2020).

SFF contains 2 SNOTEL (SNOW TELemetry network, operated by Natural Resources Conservation Service: NRCS) that record multiple snowpack properties hourly and act as a meteorological station, recording precipitation and air temperature (Emery Creek, ID 469, <https://wcc.sc.egov.usda.gov/nwcc/site?sitenum=469>; and Noisy basin, ID 664, <https://wcc.sc.egov.usda.gov/nwcc/site?sitenum=664>). SNOTEL 469 sits at 1326 masl, recorded a max SWE of 65.3 cm for a median of 20 cm. SNOTEL 664 sits at 1841 masl, recorded a max SWE of 204.2 cm for a median of 59.9 cm. They respectively record 5.7 and 7.9 months of snow per year on average, and  $-3.2$  and  $-4.1^{\circ}\text{C}$  for the mean air temperature of January, February and March.

## 3.2 Snowpack Analysis

### 3.2.1 SNODAS Model

This project uses SNODAS (NSIDC, 2004) model output, produced by National Operational Hydrologic Remote Sensing Center of National Oceanic and Atmospheric Administration and distributed by NSIDC (National Snow and Ice Data Center). The model output is freely available online. The output variables that we used are shown in Table 1.

Parameter	Abbreviation	Unit
Snow Depth	depth	Meters
Snow Water Equivalent	SWE	Meters
Snow Melt Runoff at the Base of the Snowpack (“Base Runoff”)	BR	Meters
Snowpack average Temperature	Temp	°C

Table 1: SNODAS model parameters and their properties that are used for this research project.

SNODAS grows a snowpack in a grid of locations throughout the Contiguous US. The grid has a resolution of 30 arc seconds (1 kilometer nominally) and the state of the snowpack is computed every hour for the center of the cell. The model output is only available for 06:00 UTC though (NSIDC, 2004). The model is designed as follows.

Rapid Update Cycle 2 (RUC) numerical weather prediction model was used until September 2018 to force the snow model. After that, SNODAS is forced by a combination of Rapid Refresh (RAP) and High-Resolution Rapid Refresh (HRRR), however the study site for this project is forced by HRRR only. The snow model physics are based on SNTHERM.89 (Barrett, 2003; NSIDC, 2004), and the model includes 3 snow layers and 2 soil layers. Finite differences compute the evolving snow properties. The surface forcing is done using fluxes from the weather model including temperature, humidity, radiative exchange, turbulent and latent heat fluxes. The time-step procedure for SNODAS is coarser than SNTHERM’s: 1 hour instead of an adaptive 5 to 900 seconds. The propagation of water through the snowpack occurs after exhaustion of the cold content and achievement of the residual saturation, originally set at 4% in SNTHERM (Jordan, 1990).

The SWE is then corrected using the ground observations (SNOWTEL mainly) and satellite data is used to determine the no-snow boundary. Large differences between observed and modeled snow properties are manually inspected and corrections are applied to the model grid by nudging over multiple hours (Barrett, 2003). Different procedures are in place to address various issues like the lack of satellite images due to cloud cover. The output is clipped with the watershed shapefile and transformed into a netCDF dataset containing all the relevant variables through time and space. On 1 October 2013, NSIDC employed a coordinate shift to round coordinate values, all data prior to that date is interpolated onto the grid post-2013.

Water Year (WY) is considered at multiple occasions through this work and follows SNODAS' delimitation; water year Y is defined from 1 October from the calendar year Y-1 to September of the calendar year Y. For example, water year 2024 goes from 1 October 2023 to 30 September 2024.

### 3.2.2 The Runoff Energy Hurdle Computation

The Runoff Energy Hurdle (REH; Joules/m<sup>2</sup>) is the sum of the Cold Content (CC) and the liquid water content to achieve Residual Saturation or Residual Saturation deficit (RS<sub>deficit</sub>) expressed as its latent heat of refreezing. REH (Joules/m<sup>2</sup>) is computed using Equation 1.

$$REH = \underbrace{-SWE \rho_i C_i T_s}_{CC} + \underbrace{(RS_{frac} - LWC_{frac}) \phi \rho_w L_i}_{RS_{deficit}} \quad (1)$$

With:

- SWE: Snow Water Equivalent (m<sup>3</sup>/m<sup>2</sup>=m).
- $\rho_i = 917 \text{ kg.m}^{-3}$ : density of ice.
- $C_i = 2090 \text{ J.kg}^{-1}K^{-1}$ : heat capacity of ice.
- $T_s$ : temperature of the snow (°C).
- CC: Cold Content (Joules/m<sup>2</sup>).
- $RS_{frac} = \frac{m_{\text{liquid water}}^3}{m_{\text{pore space}}^3}$ : dimensionless, maximum fraction of the pore volume that can be filled with liquid water before it yields runoff.

- $LWC_{frac} = \frac{m^3_{liquid\ water}}{m^3_{pore\ space}}$ : dimensionless, actual percent of pore volume filled with liquid water, this is assumed to always be smaller than 1.
- $\phi$ : available pore space in the snow column (m).
- $\rho_w = 1000 \text{ kg.m}^{-3}$ : density of water.
- $L_i = 334 \text{ kJ.kg}^{-3}$ : latent heat of fusion of ice.
- $RS_{deficit}$ : Residual Saturation deficit (Joules/m<sup>2</sup>).

In practice, REH has a unit of energy per surface area. This is consistent with the definition of cold content, but a clarification is needed for the residual saturation deficit. To express the REH as energy, the residual saturation deficit must be expressed as energy too. To do so, we compute the latent heat that would be needed to melt the amount of water considered to reach residual saturation. No distinction is made regarding whether the water originated as surface melt or rainfall. As a heat energy, REH can therefore be compared to other heat amounts in the surface energy balance of the snowpack.

### 3.2.3 REH Dataset

The REH is computed daily from the output of SNODAS for the 19 years of model output available (WY 2005 to 2023) and for the 7889 cells composing the study basin. There are some caveats, however, described in this section.

Equation 1 uses the liquid water content of a snowpack. While SNODAS simulates liquid water content the values are not output. The point at which the snowpack reaches residual saturation therefore cannot be determined directly. A workaround is to track the presence of the ‘Base Runoff’ term in the outputs of SNODAS. Base runoff can occur only if the snowpack has reached residual saturation. Combining the snow temperature at or below 0°C and the presence or not of base runoff allows for 4 possible scenarios, presented in Table 2. Each scenario’s rate of occurrence is in parenthesis.

- If the snowpack is cold and there is no base runoff, it is safe to assume that the snowpack has negligible liquid water content. This occurs 44% of the time (lower right cell of Table 2).

- As the snowpack cold content is exhausted, the snowpack may sit at 0°C but still have no base runoff because there is a lag in the wetting of the base of the snowpack. This occurs 16% of the time (lower left cell of Table 2). In this situation, the liquid water content may be anywhere between 0 and  $RS_{frac}$  and no additional information on the  $LWC_{frac}$  can be gained from SNODAS. The best estimation that we can do is to consider the snowpack halfway between 0 and  $RS_{frac}$ , we therefore consider to be at  $0.5RS_{deficit}$ .
- Melt reaches the bed, and the snowpack is both at 0°C and has base runoff, there is no REH at all. This occurs 38% of the time (top left cell of Table 2).
- 2% of the time, (top right cell of Table 2), the snowpack may have base runoff but be colder than 0°C. This may happen when the top of a ripen snowpack freezes and the lag time to stop base runoff is long enough to be witnessed. This rare scenario can also occur in case of great geothermal flux (hot ground) on cold snow, the base will melt first. In this situation, in addition to the cold content, the best estimation of liquid water content is to consider that the snowpack is sitting at half of the  $RS_{frac}$ , which is  $0.5RS_{frac}$ .

	$T_s = 0^\circ C$	$T_s < 0^\circ C$
Base Runoff > 0	REH = 0 (38%)	$REH = CC + \frac{1}{2}RS_{deficit}$ (2%)
Base Runoff = 0	$REH = \frac{1}{2}RS_{deficit}$ (16%)	$REH = CC + RS_{deficit}$ (44%)

Table 2: practical computation of REH for different scenarios of snow temperature and base runoff.

### 3.2.4 Defining the Spring Activation

A snowpack becomes runoff-capable once its REH is at 0. Again, we define the transition from  $REH > 0$  to  $REH = 0$  as ‘activation’, which will have an associated time. We further define the ‘activated SWE’ as the amount of SWE experiencing an activation. Finally, we define the spring activation as follows. At a given location-year, the spring activation is the greatest amount of SWE that activates.

While the snowpack may experience marginal activations in the fall, the spring activation metric is intended to capture the seasonal transition of the snowpack commensurate with snowpack accumulation switching to snowpack melt. This

signal may not be discernible for the valley floor that experiences repeated cold snowfall and melting events during the whole snow season. In general, in case of low snow and warm winters, the spring activation is not defined in those cases.

### 3.3 Statistical Methods, Machine and Deep Learning

#### 3.3.1 K-Means Clustering of REH Time Series

Clustering methods group data points into labeled clusters, then minimize the within-cluster variance and maximize the between-cluster variance by shifting the label of data points. Once a solution is attained by a clustering method, the labeled data points within a given cluster will be relatively close compared to the data points outside of that given cluster. We use the clustering method K-Means with Lloyd's algorithm, which is described further by (James et al., 2022, Chapter 12.4.1).

The number of clusters for K-Means must be defined a priori. Multiple metrics exist to select the number of clusters, among which we used: inertia, silhouette score, silhouette plots, performances of the other methods tested on the K-Means output and physical intuition of the ideal number of divisions needed. Specifically, the following decomposition is observed as ideal: low REH values (all years); medium REH values (all years); high REH values (average and high accumulation years) and extreme REH values (extreme accumulation years).

Inertia is the sum of distances between all data points and their cluster center, this number cannot increase with more clusters, and it is much greater for few clusters, it then plateaus and slowly converges to 0. Visual inspection of inertia plots allows to identify the ‘elbow’, when an additional cluster minimally affects the inertia.

The silhouette score compares the distance of a data point to its cluster center to the distance to other cluster centers, a higher score signifies that a data point is close to its cluster center relative to the center of the other clusters. The mean of the silhouette score of all data points is used for choosing the number clusters

because the mean silhouette score is a function of the cluster number. The higher the mean silhouette score, the better the classification.

The silhouette plots are a graphical representation of all individual data point's silhouette score, there should be no cluster with negative silhouette scores (closer to other cluster centers than its own cluster center), and all should have the approximate same maximum silhouette score, corresponding to the overall mean silhouette score.

We applied the clustering algorithm to the time series of REH. Other studies have performed clustering on time series (e.g., Farmer et al., 2010; Solgaard et al., 2022) and extracted meaningful information, in contrast to clustering of properties manually extracted from time series (Farmer et al., 2010). Spectral clustering (on a correlation matrix of the time series instead of the time series themselves) can be used for highly dimensional datasets, especially when the number of data points is smaller than the number of dimensions, which is not our case. Although the clustering happened on the REH time series only, each time series' location in the watershed was tracked. From that tracking and for each cluster, we computed histograms of the elevation, slope, aspect, volume of water equivalent, average SWE, area and activations. For a given cluster, its centroid was computed as the mean behavior of all locations in that cluster. This provides qualitative relationships between trends in the features and the overall REH behavior.

### 3.3.2 Statistical Methods

Several statistics are used in this work, including  $R^2$ , which is defined in detail in Section 3.1.3 of James et al. (2022). The t-statistic is the number of standard deviations away that the slope is from 0, given by  $t = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)}$ , where the hat symbolizes that we make a statistical estimation of  $\beta_i$ , and the p-value of the t-statistic indicates the likeliness of the considered relationship to be due to chance. Those are defined in 3.1.2 of James et al. (2022).

Practically speaking, a significant relationship will have a  $R^2$  close to 1, a t-statistic  $> 2$  and a p-value  $< 0.1$ . The greater the  $R^2$  and the t-statistic and the smaller the p-value, the more reliable the relationship is.

### **3.3.3 Yearly Mean Cluster Number and Correlations**

To quantify the consequences of low snow years, a metric of “good” and “poor” is established using the mean cluster number for a given year, hereafter called mean cluster number. If a year is dominated by cluster 1 its mean cluster number will be very close to 1, while a high snow year could theoretically be close to 4. This number allows for quantitative comparisons to 1) the mean drought index (described next paragraph), 2) the mean air winter temperature (January, February, March), and 3) the mean winter value of the ENSO index (Sea Surface Temperature anomalies, dataset Niño4; Reynolds et al., 2002).

The National Integrated Drought Information System (<https://www.drought.gov/data-maps-tools/us-drought-monitor>) releases drought indices for multiple areas. We focus on the index of Flathead County. This index is greater when a greater area spends a greater time in the following categories: “Abnormally Dry or D0, (a precursor to drought, not actually drought), and Moderate (D1), Severe (D2), Extreme (D3) and Exceptional (D4) Drought”. Whenever an index is greater than 0 (some of the area is classified in some Dx for some time), the score is increased by 1 for the week considered. The mean drought index over the months of July, August and September gives a quantitative representation of more extreme droughts for longer periods of time, and is compared to the mean cluster number.

### **3.3.4 Predicting the Timing of Spring Activations from Partial REH Time Series**

Inter-annual variability must be considered to predict the timing of the spring activation. Three methods are investigated here, and both are applied on partial REH time series. The first method is a linear regression on the REH values to predict the timing of activation, the second is to make estimations of the cluster number, and the third one involves a neural network.

Using the REH values of X days before 1 March as features of a linear regression model (Described in section 3.1.5), the timing of spring activation is predicted and the mean absolute error for X days considered is gathered.

The mean of past spring activation timings is a first naïve model of upcoming activations, but in an operational context, this mean from past history can be refined down to the per cluster mean. Estimating the cluster number from a partial REH time series is done using the previously determined cluster centroids. Both those centroids and REH time series are shortened to end on a chosen date such that they become “shrunken” centroids and REH time series. Then shrunken REH time series are ingested into this shrunken K-Means and their predicted cluster number is compared to the actual one. By shifting the threshold date, the accuracy of this method can be estimated for a different date.

Long Short-Term Memory (LSTM) neural networks (Hochreiter & Schmidhuber, 1997) use LSTM cells to perform significantly better on time series. The model ingests REH time series from 1 October through 1 March (151 days), the output variable is the timing of spring activation. We elected to train 19 models, each of them would be trained on all years but one and tested on that one year left aside. This situation is 1) more realistic of the problem we try to tackle, and 2) it resolves the following problem: REH is spatially correlated, the REH in one cell is correlated to the REH in its neighbor cells, and if one of those cells is in the training data and the other in the testing data, the performances of the model are artificially improved. Each model is composed of:

1. an LSTM layer with 151 features.
  2. a 20% dropout layer.
  3. a fully connected layer (151 neurons, ReLU activation).
  4. a 25% dropout layer.
  5. a fully connected layer (50 neurons, ReLU activation).
  6. a 25% dropout layer.
  7. a fully connected layer (10 neurons, ReLU activation).
  8. a 25% dropout layer.
  9. a one neuron output fully connected layer to yield the spring activation timing.
- The neural network must be trained. In other words, its parameters, also called weights, need to be adjusted to the problem considered. Gradient descent is used

to tune the weights at each training step towards smaller error of the model, and the size of that step is partially dictated by a constant that we could pick called the learning rate. We utilize Adam, an adaptive method for gradient descent that changes the learning rate at each step and uses momentum and is common method for faster learning (Kingma & Ba, 2017). Training must be done according to a loss function that Adam will attempt to minimize. We use the mean absolute error between the modeled and observed spring activation timing. The training set was randomly split into 70% training and 30% validation data at each training epoch. Final performance was determined using the Mean Absolute Error (MAE) on the testing data, and the MAE for each cluster in the testing data (done by sub-sampling the testing dataset by cluster number).

### 3.3.5 Interpretable Machine Learning to Quantify Feature Importance

We attempt to improve the within-cluster timing of spring activation predictions using a group of models described in this section. Table 3 summarizes the models used and the consideration of inter-annual variability using cluster number.

Method to consider inter-annual variability in the model	Models tested
Not considering inter-annual variability	Simple Linear Regression, Random Forest, Gradient Boosting
Considering inter-annual variability by including it as a model feature	Simple Linear Regression, Random Forest, Gradient Boosting, Neural Network
Considering inter-annual variability by fitting one model per cluster	Simple Linear Regression, Random Forest, Gradient Boosting, Neural Network

Table 3: summary of the model tested to understand spatio-temporal variability of the spring activation timing.

The simple linear regression makes predictions according to the following equation:  $\hat{y}_i = \hat{\beta}_0 + \sum_{i=1}^{14} \hat{\beta}_i x_i$ , with 15 different estimated coefficients  $\hat{\beta}_i$  (one intercept and one coefficient per feature: elevation, slope, aspect and the 10 different values of land cover type),  $\hat{y}_i$  being the prediction of SAS and  $x_i$  being the value of a given feature. The coefficient calculation is described in Chapter 3.1.1 of Sohil et al. (2022).

The Random Forest and Gradient Boosting are both ensemble methods based on decision trees, described in detail in Chapter 8.2 of James et al. (2022). A decision tree minimizes the entropy of a dataset at each branch of the tree by dividing the labels or target variable such that it maximizes explanation from a specific explanatory variable. Random Forest and Gradient Boosting add a stochastic component by respectively using numerous parallel trees, and by applying recursively trees to the residuals of previous trees.

The neural networks are composed of two fully connected layers of 13 and 10 neurons. More complex models show instability or simply learn the data and simpler models defy the purpose of complex non-linear learning of neural networks. The models use mean absolute error as their loss function, ADAM (Kingma & Ba, 2017) for Gradient Descent, and are trained for 1000 epochs. Simple Linear Regression and Random Forest and Gradient Boosting are all fitted using the default parameters of the Python library scikit-learn (Pedregosa et al., 2011). All models are trained on a training set, itself randomly split between training and validation subsets at each epoch of the neural network, and tested on a testing set that is completely unseen before by the model. Metrics are computed on the testing set as are the Shapley values. The models all use at least aspect, elevation, land cover type (categorical variable with 10 different values) as features, then is added cluster number as a numeric feature for the relevant models.

### **3.3.6 Shapley Values and Model Selection**

To define the performance of the models, their utility in other words, two methodologies are involved: Shapley values and overall model performance using the Mean Absolute Error.

Shapley, values are originally defined in Shapley (1953). More intuition and applications are available in more recent publications (<https://christophm.github.io/interpretable-ml-book/shapley.html>, etc.). Shapley values represent the relative contribution of a given feature to a given output. The mean Shapley value of a given feature on all outputs is now considered the “Shapley value” of that feature. The Shapley values of all features for a given data point will sum up to the differ-

ence between the output and the overall mean. In other words, if a given location activates 15 days later than the average, the sum of its Shapley values will be 15. As a result, the contribution of aspect, elevation, land cover type and slope to the prediction of spring activation timing for a given cluster can be quantified.

The Mean Absolute Error (MAE) represents the mean distance between the modeled and observed spring activation timing. MAE is defined as follows:

$MAE = \frac{1}{n} \sum_{i=1}^n |D_{i, \text{model}} - D_{i, \text{obs}}|$ , with  $n$  samples of the Day  $D$  at which spring activation occurred, modeled (model) and observed (obs). A useful model would have a lower MAE than the other models. The MAE is chosen over a squared metric for its ability to focus on the bulk values relatively close to the mean instead of being more influenced by outliers. MAE also gives a result in days which is interpretable by a human, unlike “squared days”.

For each feature of each model in Section 3.3.5, Shapley values are the respective contributions of each feature for a given model output. Using multiple model types allowed us to detect for unstable predictions and unreliable models. If Shapley values are uniform among models, the physical relationships between terrain property and spring activation timing will be stronger and more reliable.

The most naive and simplest model is to predict the mean, this is called the “baseline case”, and all models designed and tested here must improve predictions from that baseline case. Linear models (Linear Regression), or non-linear (Random Forest and Gradient Boosting) may have very different performances (different MAE) than the neural network on the testing data for example. This is why investigating them all is necessary to isolate the most useful model, and to ensure the robustness of the Shapley values.

### 3.3.7 Trends in Widespread Activation Events

To study the spatial propagation of spring activation timing through the watershed during a period of greater energy input (e.g., rain, warmer temperature than average, low cloud cover resulting in high solar radiation hitting the ground), we identify large areas of the watershed that activate over a multi-day window (ranging from 4 to 17 days). We manually identified 31 widespread activation

events. Each location's elevation, northing, SWE, and the potential incident solar radiation are recorded during the widespread activation events. The mean of the locations activating will change from day to day because new locations are activating. It is expected that ripening will propagate upward and towards locations with less solar radiation locations for example. Trends are identified using a simple linear regression and their goodness of fit is evaluated using the  $R^2$ , the t-statistic, and the p-value of the t-statistic. Particularly, we use the p-value of the t-statistic to find significant trends through time, and to find if this behavior is consistent through widespread activation events.

The variables we consider in this analysis are:

- Elevation, taken from the DEM.
- The mean SWE getting activated from SNODAS.
- Northing, defined as a symmetrical linear relationship between South ( $=0$ ) and North ( $=1$ ). This only partially accounts for other aspect trends (South-West being the warmest for example).
- Solar Radiation. Calculated for each cell. Potential incident solar radiation on the ground surface with an azimuth and slope, is computed from the sun's altitude and azimuth at a given time according to Hock (1999). The equation selected here is:

$$I = I_0 \left( \frac{R_m}{R} \right)^2 \Psi_a^{\frac{P}{P_0 \cos(Z)}} \cos(\Theta)$$

Using:

- $I_0 = 1368 \text{ W.m}^{-2}$ : solar constant
- $\left( \frac{R_m}{R} \right)^2$ : Earth eccentricity correction factor.
- $\Psi_a = 0.75$ : mean atmospheric clear-sky transmissivity (Dimensionless according to Hock, 1999).
- $P = P(z) = -7.4305(z + 610) + 108900 \text{ Pa}$ : atmospheric pressure at that location (With  $z$  in masl, adapted from [https://en.wikipedia.org/wiki/International\\_Standard\\_Atmosphere](https://en.wikipedia.org/wiki/International_Standard_Atmosphere)).
- $P_0 = 101325 \text{ Pa}$ : the standard atmospheric pressure.
- $Z$ : local zenith angle.

- $\Theta$ : the angle of incidence between the normal of the terrain surface and the solar beam,  $\cos(\Theta)$  is approximated as:  $\cos(\Theta) = \cos(\beta) \cos(Z) + \sin(\beta) \sin(Z) \cos(\varphi_{\text{sun}} - \varphi_{\text{slope}})$ , with:
  - $\beta$ : the slope angle.
  - $\varphi$ : the respective azimuths.

## 4 Results

### 4.1 REH Behavior During a Water Year

REH time series can be analyzed for individual locations or as bulk behavior of a watershed.

To demonstrate the high spatial and temporal variability, we compare two end-member locations identified in Figure 1 during two end-member years (Figure 2). During the water year 2005 (extreme low) at the low location (valley floor), there are only a few centimeters of SWE and near zero REH in January; the summit location has up to 20 cm of SWE early April and about 5 MJ/m<sup>2</sup> of REH. The low location has more SWE and more REH overall during the water year 2011 (extreme high) than the summit during the low year: SWE gets close to 25 cm and REH reaches 10 MJ/m<sup>2</sup>. The summit location during water year 2011 is an order of magnitude greater, reaching 1.5 m of SWE and 45 MJ/m<sup>2</sup>, peaking May 1 before a 10-day REH loss.

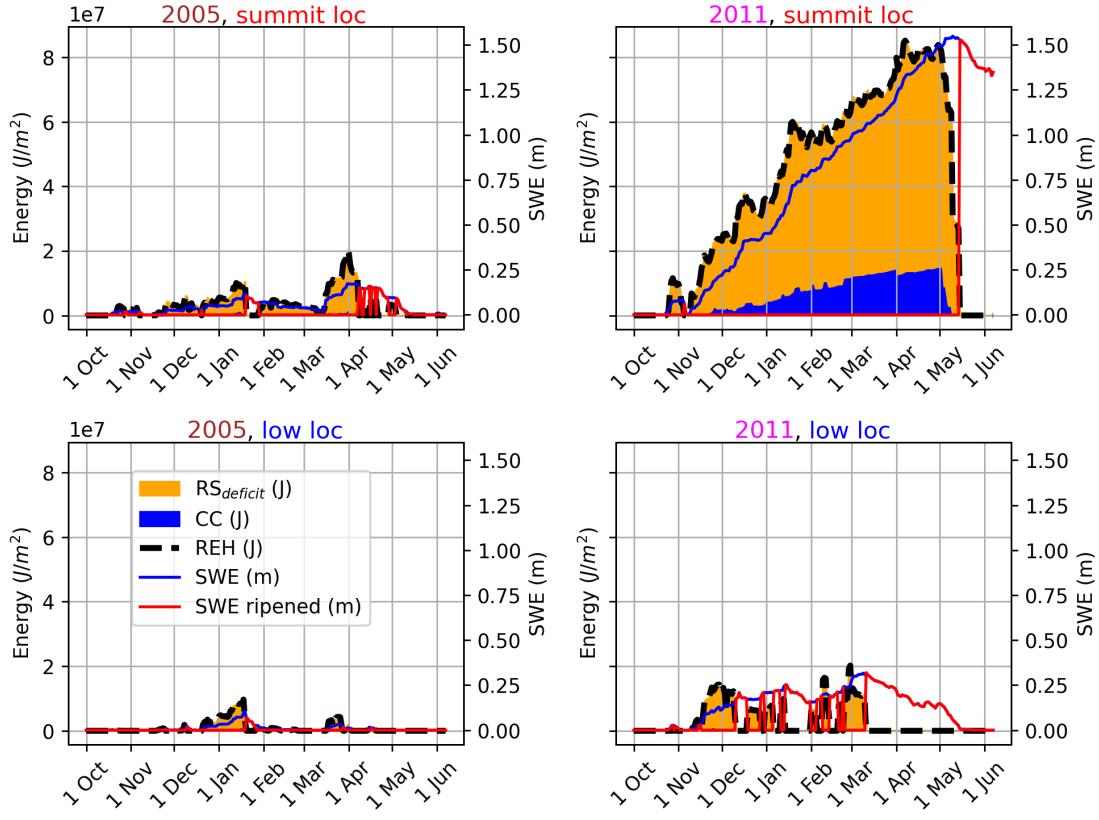


Figure 2: Comparison of  $RS_{\text{deficit}}$  (orange area) and CC (blue area) that sum up to REH (black dashed line, all in  $J/m^2$  and on the left axis), and of SWE in blue and the ripened SWE (SWE with  $REH=0$ , both on right axis) in red, for 2 different locations and years. Those snow properties are compared between a low snow year (left column) and a high snow year (right column) and a high location (top row) and a low location (bottom row). A high location during a low-SWE year can have an overall behavior similar to the valley floor on a high-SWE year, suggesting the strength of the inter-annual variability.

Watershed-wide bulk behavior is also highly variable over 19 years (Figure 3). While 2005 and 2011 have extreme SWE values, the REH in these years are not extreme. Other years in the record achieve similar values. This is confirmation that REH carries additional information, relative to SWE. Interquartile and interdecile of SWE and REH values give information on the spread of the series. The interquartile area is relatively small for SWE relative to the REH one.

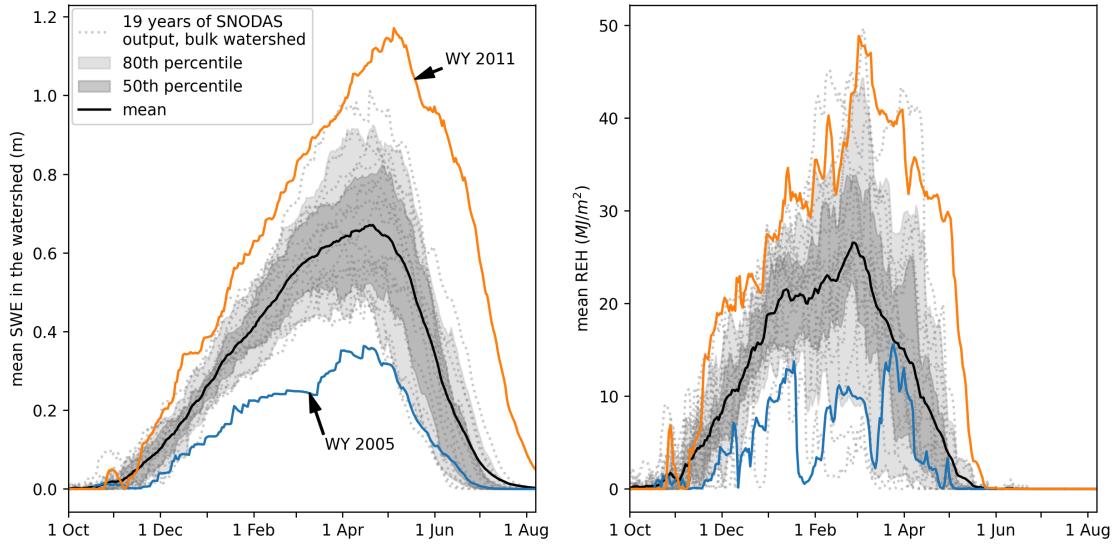
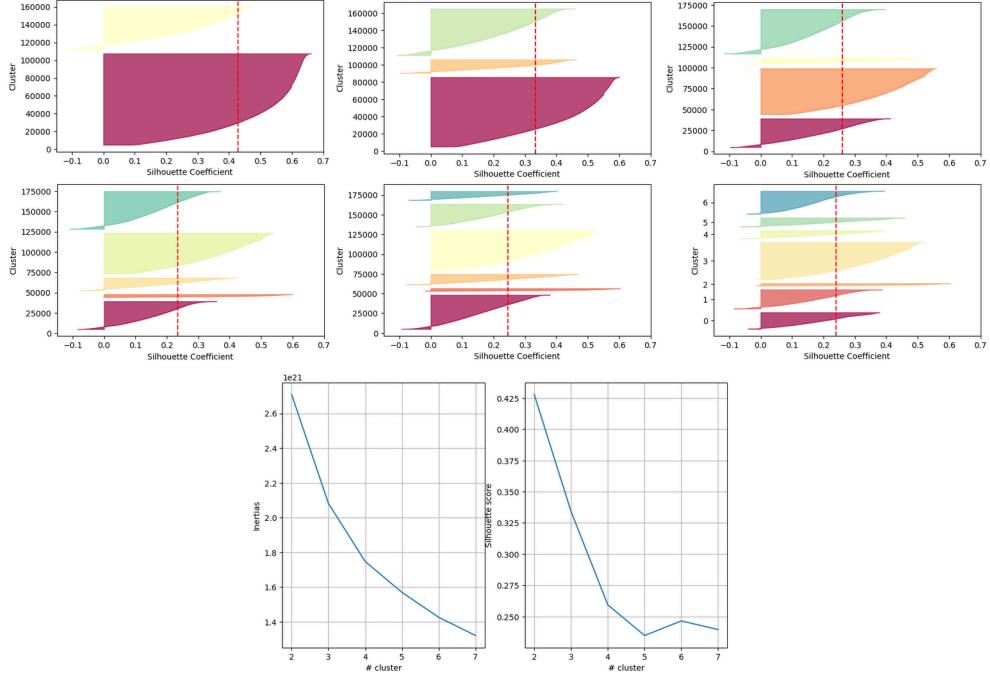


Figure 3: **Outlier years in SWE are not outliers in REH.** Watershed-wide mean SWE (m, left) and mean REH ( $\text{MJ}/\text{m}^2$ , right), 19 years of model output displayed in fine dotted lines, are also displayed the mean in black and interquartile and interdecile in shaded grey. Two exceptional years are annotated in blue (2005, low year) and orange (2011, high year).

On average, REH is about an order of magnitude smaller than the latent heat needed to melt the snowpack though. Therefore, exhausting REH happens in significantly less time (days to weeks) than melting the snowpack requires (weeks to months).

#### 4.1.1 Choice of the Number of Clusters

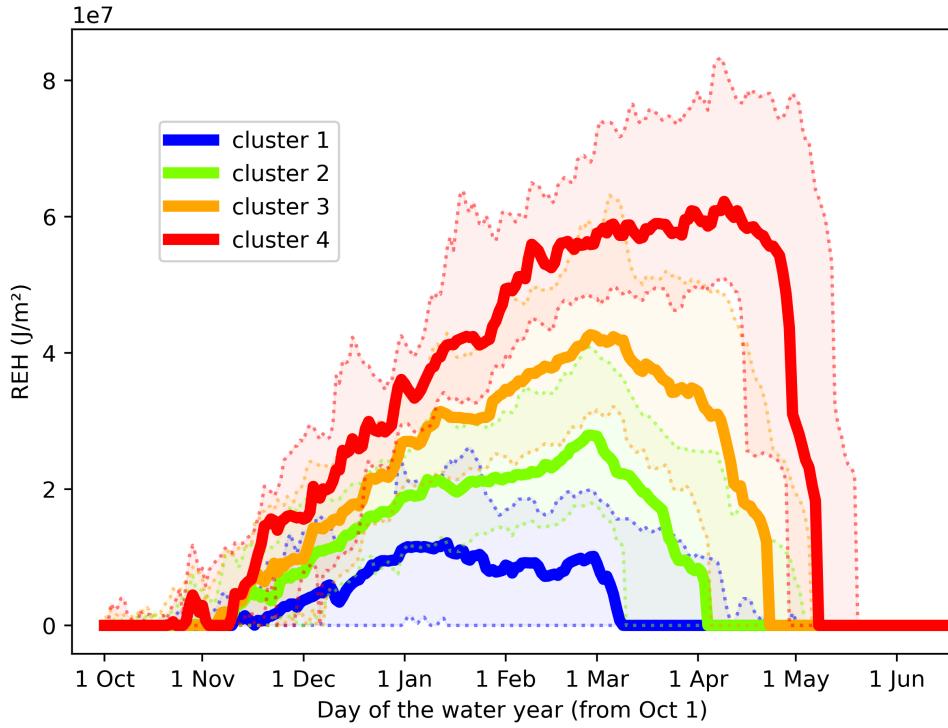
Five methods were described in section 5.3.1 to pick the number of clusters. The inertias, silhouette scores and silhouette plots are shown in Figure 4. There is no strong elbow in the inertia curves, limiting our interpretation abilities. While 2 or 3 clusters may have the highest silhouette score, having too few clusters limits the results found in other sections. Cluster 1 remains relatively unaffected by the change of number of clusters, but other clusters mix up and hide portions of the signals and trends found. The silhouette curves suggest that 5 is too much. Silhouette curves show gradually more negative values and do not converge to the mean silhouette score (vertical line) for a number of clusters greater than 3. The optimum is 4 clusters because it mitigates all problems raised here.



**Figure 4: 4 clusters is the best compromise between the statistics presented here and the rest of the analysis.** Mosaic of plots useful to determine the most appropriate number of clusters for K-Means. Top row : Silhouette plots for 2, 3, 4, 5, 6 and 7 clusters. Lower row: inertias and silhouette scores VS number of clusters used. From the silhouette plots: poor classification (negative values) for any cluster number, too large values for high cluster numbers. Strong elbow in the inertia plot for 3 cluster, higher silhouette score than for 4 clusters, but using 3 clusters eventually hides a part of signal for other results.

#### 4.1.2 K-Means Clustering of the REH Time Series

Clustering finds REH time series that behave similarly. For example, the low location and the high location in 2005 (Figure 2) would be clustered together, while the two locations would be clustered differently in 2011. The REH time series were clustered in 4 clusters (Figure 4 and Figure 5) and their REH overall values are gradually increasing. Cluster 1 has very limited REH growth and REH goes null repeatedly throughout the winter. The higher order clusters have gradually longer growth periods and later and more rapid REH losses. Clusters 2 and 3 peak late February and cluster 4 peaks mid-April.

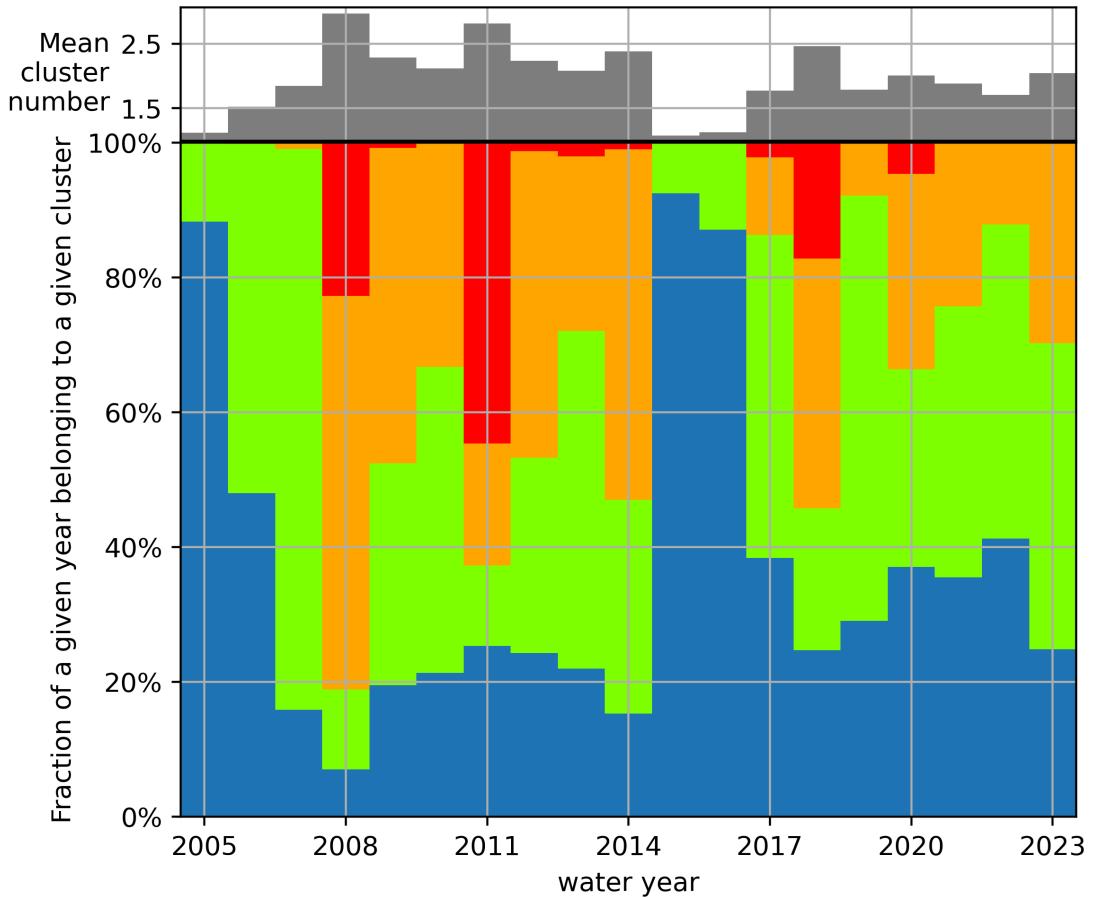


**Figure 5: The entire REH time series contributes to the clustering instead of specific values.** Median REH values (thick colored line) for each cluster with their respective interdecile envelopes (shaded area, bounded by the dashed lines), through the relevant part of the water year. Around 1 March, some time series in cluster 3 may exceed cluster 4's median, yet they are not classified in cluster 4, because the rest of the time series is relevant and the time series in cluster 3, although high on 1 March, decline earlier than the ones in cluster 4.

The magnitude of growth and peak REH increases for higher clusters. The low REH values of cluster 1 are a combination of 1) low snowpack conditions, due to low SWE or a warm/wet snowpack, and 2) mid-winter activations of the snow, sometimes repeatedly. Mid-winter cyclical snowfall and melt typically occurs at the valley floor. This behavior is also observed at higher elevations in certain years. The high clusters do not activate mid-winter because the snowpack gradually gains REH with increasing SWE without warming up significantly. Eventually, sensible heat and longwave radiation become sufficient to activate the snow.

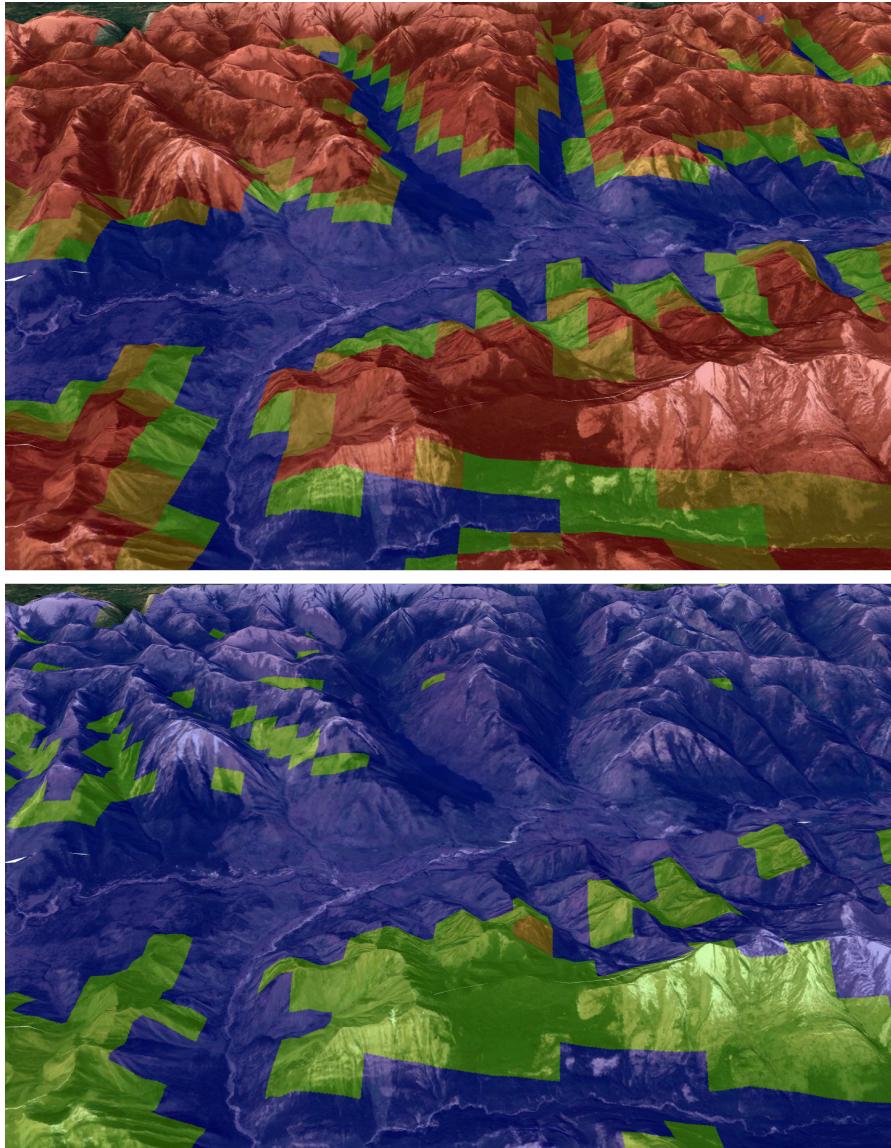
#### **4.1.3 Temporal Variability of Each Cluster**

The annual partitioning of the basin to the four possible clusters is shown in Figure 6. The representation of a given cluster varies from year-to-year. Low REH years like 2005, 2015 and 2016 respectively had 88, 92 and 87% of the watershed area clustered in cluster 1. Low years also have a lower Yearly Snow Cluster Index, which is the mean cluster number in the watershed for a given year. Cluster 2 is also represented every year, some years as little as 5% and up to 60%. Cluster 3 is represented two-thirds of the years and is only missing during extreme low years. Cluster 4 is only represented half of the years, but it is relevantly represented (>5%) during 4 years (2008, 2011, 2018 and 2020).



**Figure 6: The predominance of the cluster number in the study domain is highly variable from year to year.** Top: bar plots of the mean cluster number, ranging between 1.08 to 2.97. Bottom: stacked bar plots of the fraction of the watershed area belonging to a given cluster on a given water year.

Figure 6 reveals a strong inter-annual variability, most of the watershed changes classification from year to year. As seen before, cluster 1 is characterized by very low REH values and/or mid-winter activations, therefore, low REH years (2005, 2015 and 2016 for example) have watershed-wide snowpack very close to activation mid-winter, or even watershed-wide activations mid-winter. Confirmation of the proper behavior of K-Means and intuition about the inter-annual variability can be gained by comparing maps of the watershed during the end member years 2005 and 2011 (Figure 7).



**Figure 7: End member years (2011 and 2005) result in a drastically different classification of the locations in different clusters.** Overlayed cluster clustering classification (consistent colors with Figure 5) over satellite imagery and a vertically exaggerated (3 times) digital elevation model. In 2011, cluster 3 and 4 are common, they span the highest elevations, and a clear gradient is observed from cluster 1 to 4 along the mountain walls. In 2005, some of the highest locations are cluster in cluster 2 and one north-facing location in this snapshot is in cluster 3, but most locations are in cluster 1, signifying a relatively low REH throughout the watershed.

The mean cluster number is compared to other metrics of the local and global climate. Those simple comparisons provide relevance to the mean cluster num-

ber metrics and the clustering method in general. This analysis is not exhaustive and simplifies the complex relationships considered, we simply aim to register the cluster number in the real world. The correlation between the mean cluster number and the mean drought index has a  $R^2$  of 0.40 with a p-value  $< 0.01$ , giving strong evidence of the significance of the relationship (Figure 8). The trend is  $-4.5$  drought index points per increase of 1 mean cluster number. Beyond quantifying the timing of activations for a given year, the mean cluster number therefore gives indication of the water resource state of the following summer; greater mean cluster number ensures late and high availability of water. The mean cluster number was also compared to the winter air temperature in Figure 9. There is evidence (p-value  $\approx 0.01$ ) of a loss of 0.23 mean cluster number per degree C increase. Finally, the mean cluster number is correlated to the El Niño South Oscillation (ENSO) index on Figure 10 with strong evidence of a 0.42 of mean cluster number per ENSO index increase.

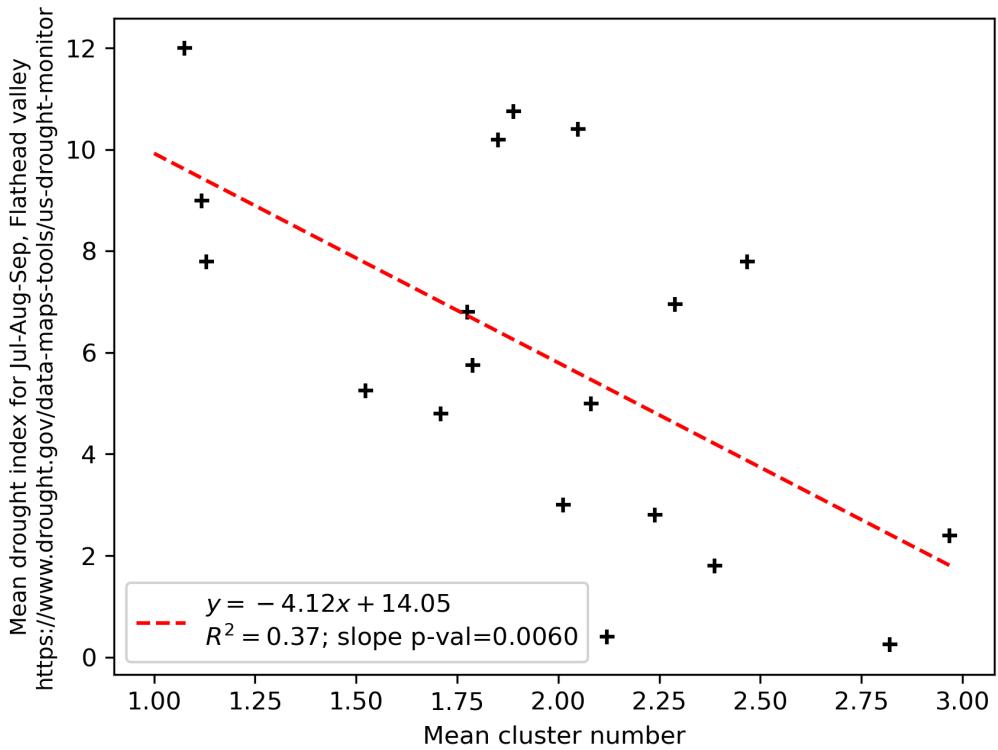


Figure 8: **A low mean cluster number snowpack strengthens the droughts the following summer.** Mean summer (Jul-Aug-Sep) drought index as a function of the mean cluster number, low correlation ( $R^2=0.37$ ) but significant trend ( $p\text{-val} = 0.0060$ ).

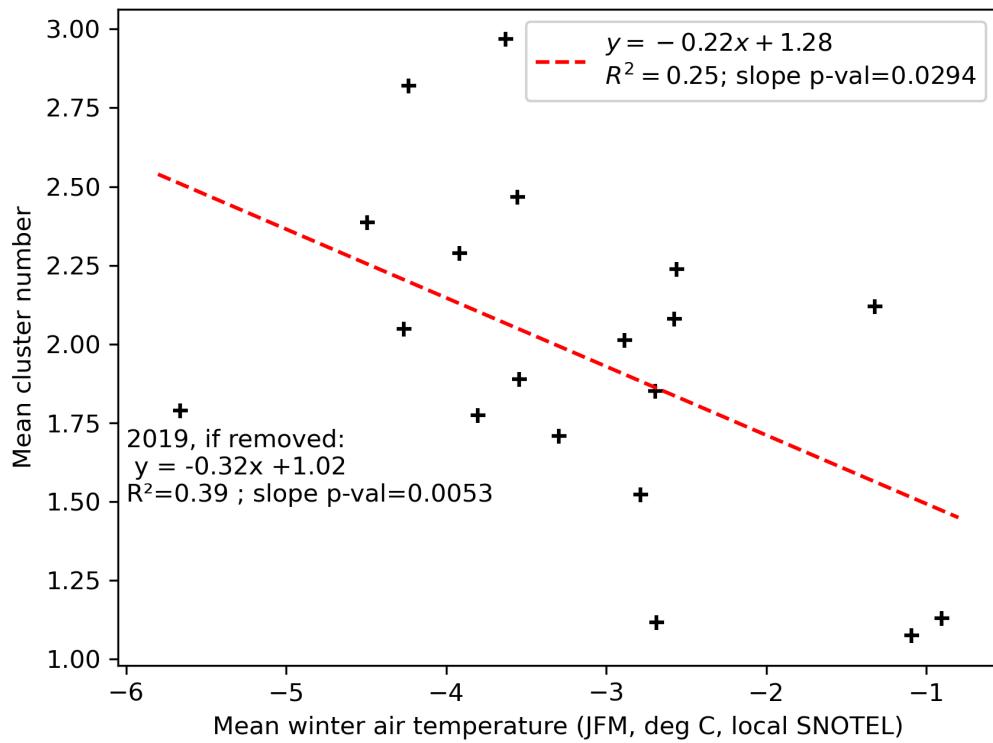


Figure 9: **Warmer air temperatures lower the mean cluster number.** Mean cluster number VS Mean winter air temperature with evidence of a significant ( $p\text{-value}=0.0294$ ) negative relationship (colder years bring higher cluster numbers on average). The removal of the outlier year 2019 increases the  $R^2$  to 0.39 and lowers the p-value to 0.0053, strengthening the relationship.

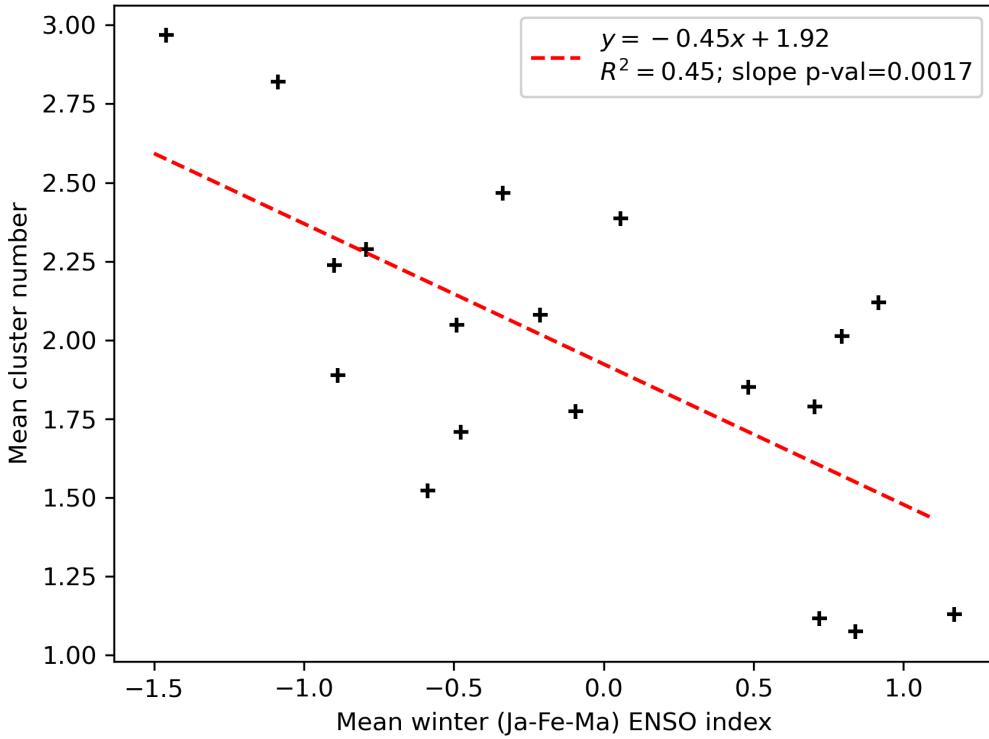


Figure 10: **El Niño years have lower cluster number.** Comparison of the mean cluster number VS winter ENSO index (Jan-Feb-Mar), not a great correlation ( $R^2 = 0.45$ ) but a significant trend ( $p\text{-val}=0.0017$ ).

#### 4.1.4 Spatial Variability of Each Cluster

Per cluster histograms of a variety of properties extract the spatial variability of clusters. The different clusters span different elevations with some overlap (Figure 11), but they are differently centered: they have gradually higher mean elevations (1490, 1831, 2004 and 2121 masl respectively). While clusters 1 through 3 have visibly different mean elevations, the difference in mean elevation between clusters 3 and 4 is much smaller. Given the uneven representation of Cluster 4 through years (Figure 6), cluster 4 is considered a sub-sample of cluster 3 during extreme years, therefore explaining their similar means. A t-test statistic had a p-value  $< 10-16$ , providing overwhelming evidence that the two means are different. Figure 6 also confirms that the inter-annual variability is the cause for a part of

the overlap in other figures. A location may be classified in different clusters from one year to another.

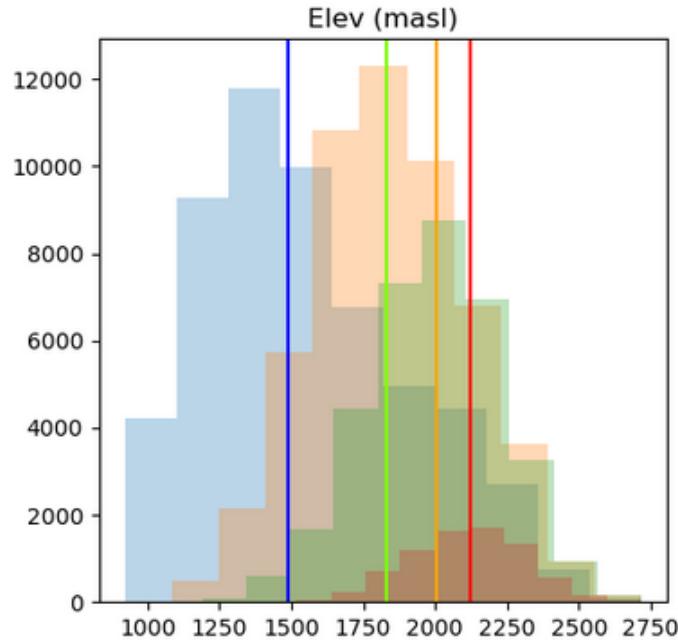
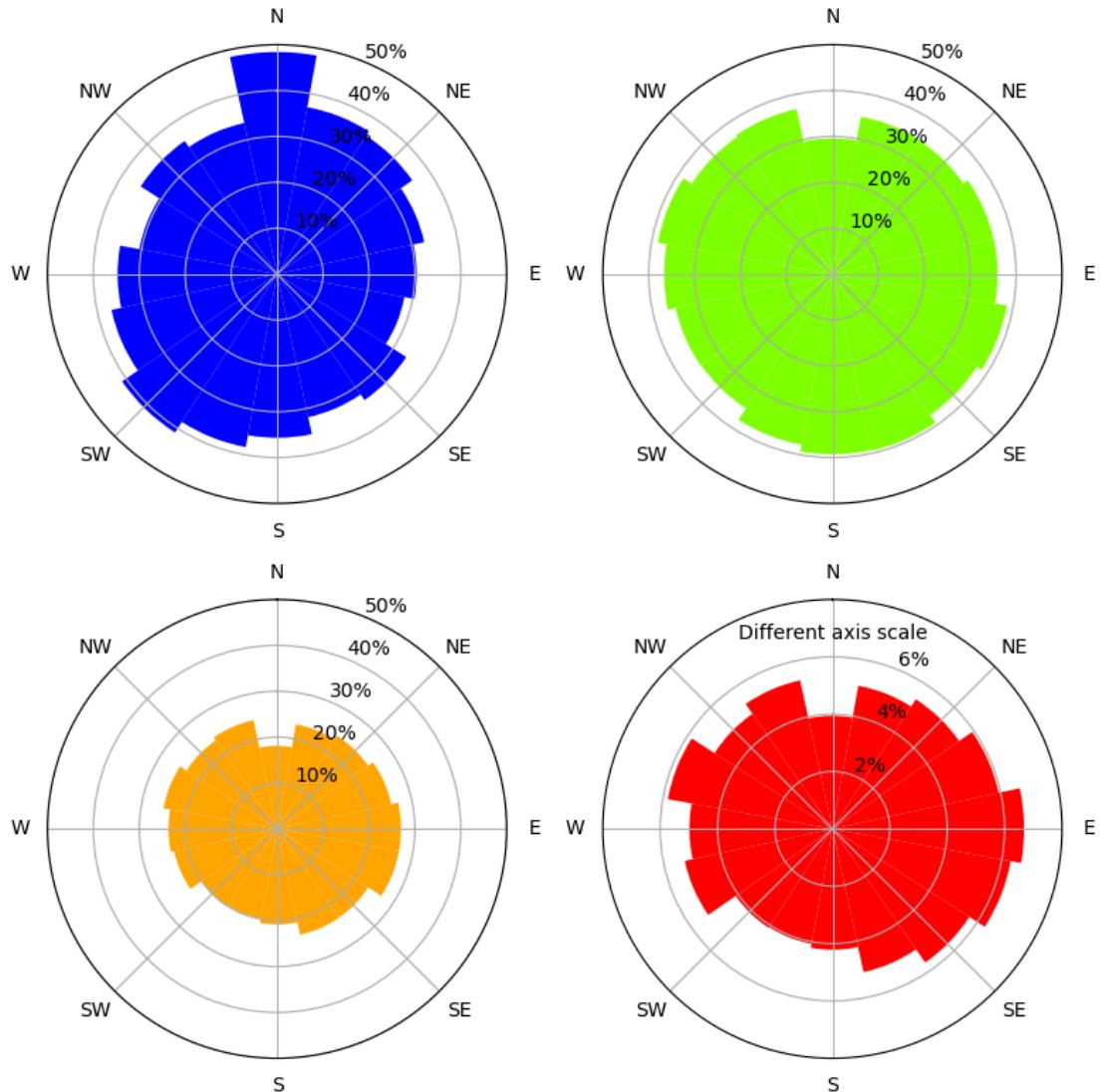


Figure 11: **Higher elevations are in higher clusters on average.** Histograms of elevations for each cluster with the mean as a vertical bar. Nineteen years of data are represented here.

Per cluster aspect histograms are presented in Figure 12. Cluster 1 is more common on SW aspects, cluster 2 is more evenly distributed, and clusters 3 and 4 are concentrated on the E aspect and less common on S/SW aspects.



**Figure 12: Lower clusters are concentrated on SW aspects, while higher clusters prefer E aspects.** Percent of the watershed area belonging in an aspect bin and to a given clusterbin. For example, 41% of the SW facing terrain belongs to cluster 1. If these 4 bar plots are stacked, they sum up to 100% in each aspect bin.

Forests and grassland are the most common land cover types in the study basin. They are relatively evenly represented throughout clusters (Figure 13), with a tendency of shorter vegetation to be in higher cluster (grassland, versus shrubland, versus forest). The other land cover types, that are less common, have extreme behaviors: barren lands and snow and ice are mostly in high clusters, while water, urban, wetland and subtropical forest are mostly in cluster 1. The latter land cover

types are most common at low elevations though, by opposition to barren land and snow and ice that are only found at high elevation. This confirms that there is some degree of correlation between the topography.

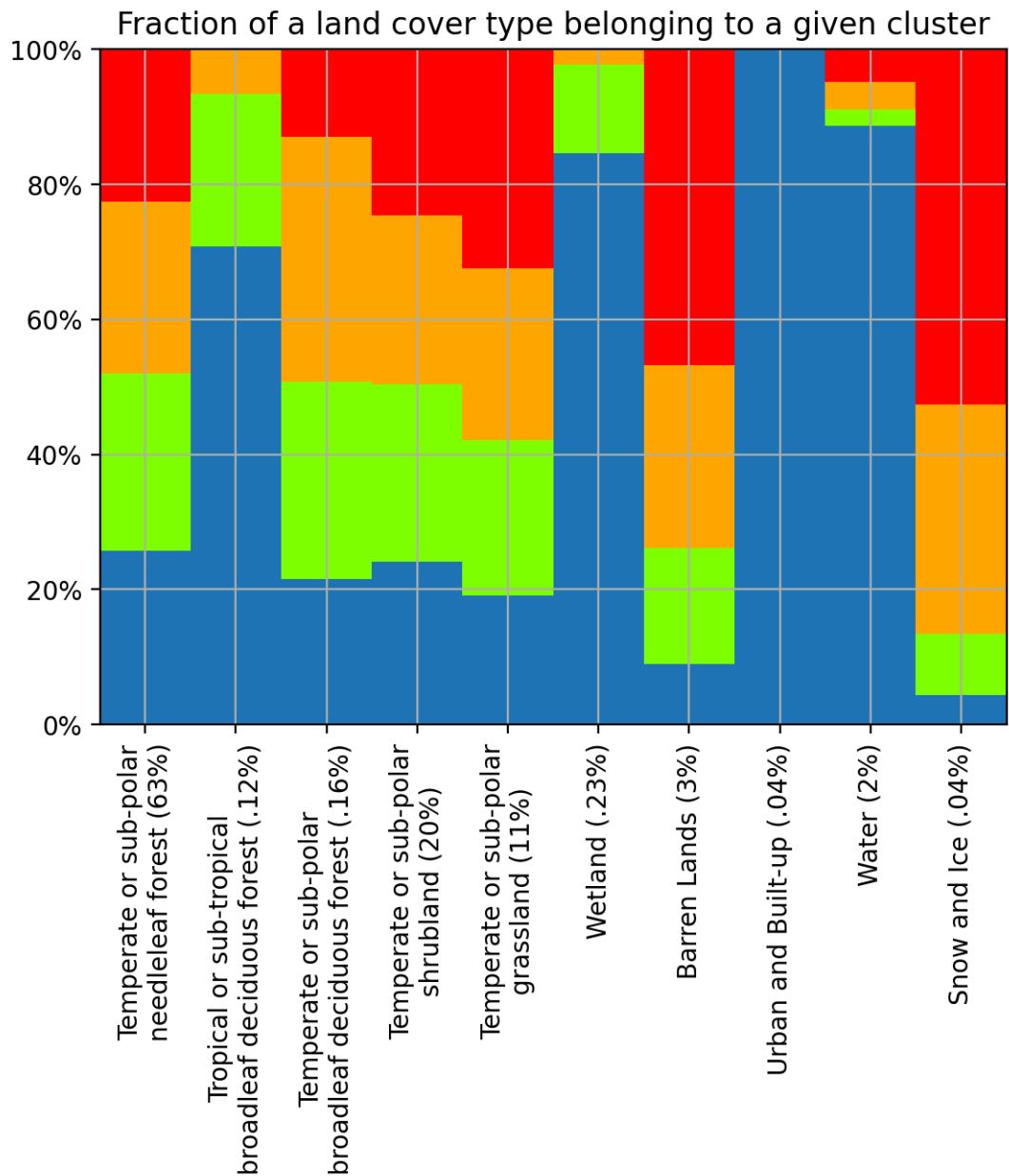


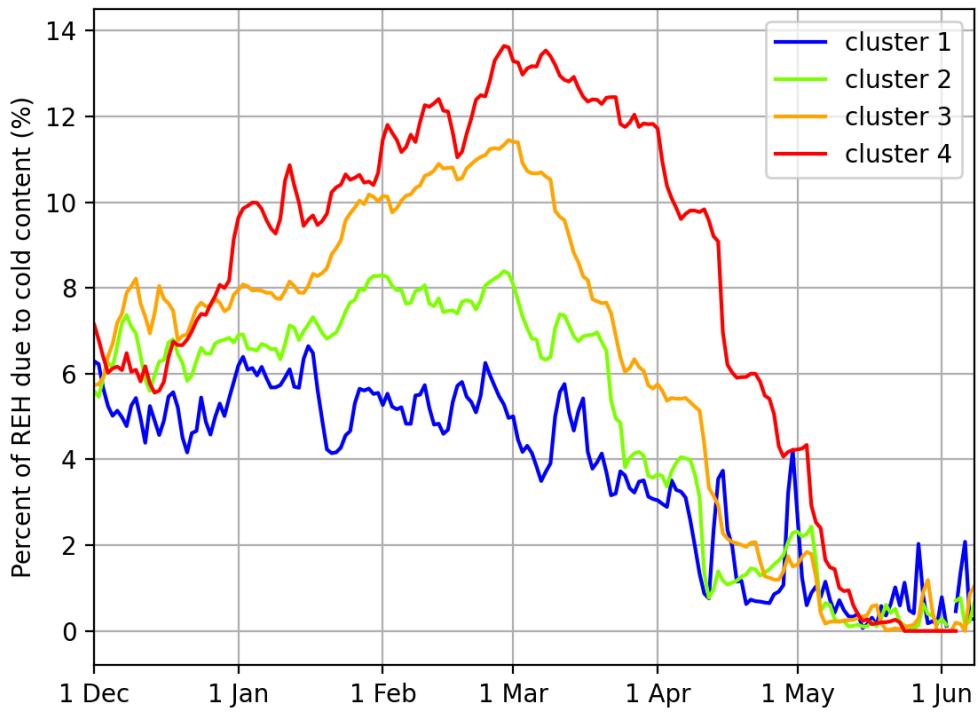
Figure 13: Clustering preference for a given land cover type. In brackets is the area of the watershed represented by a given land cover type.

## 4.2 Cold Content and Residual Saturation Deficit

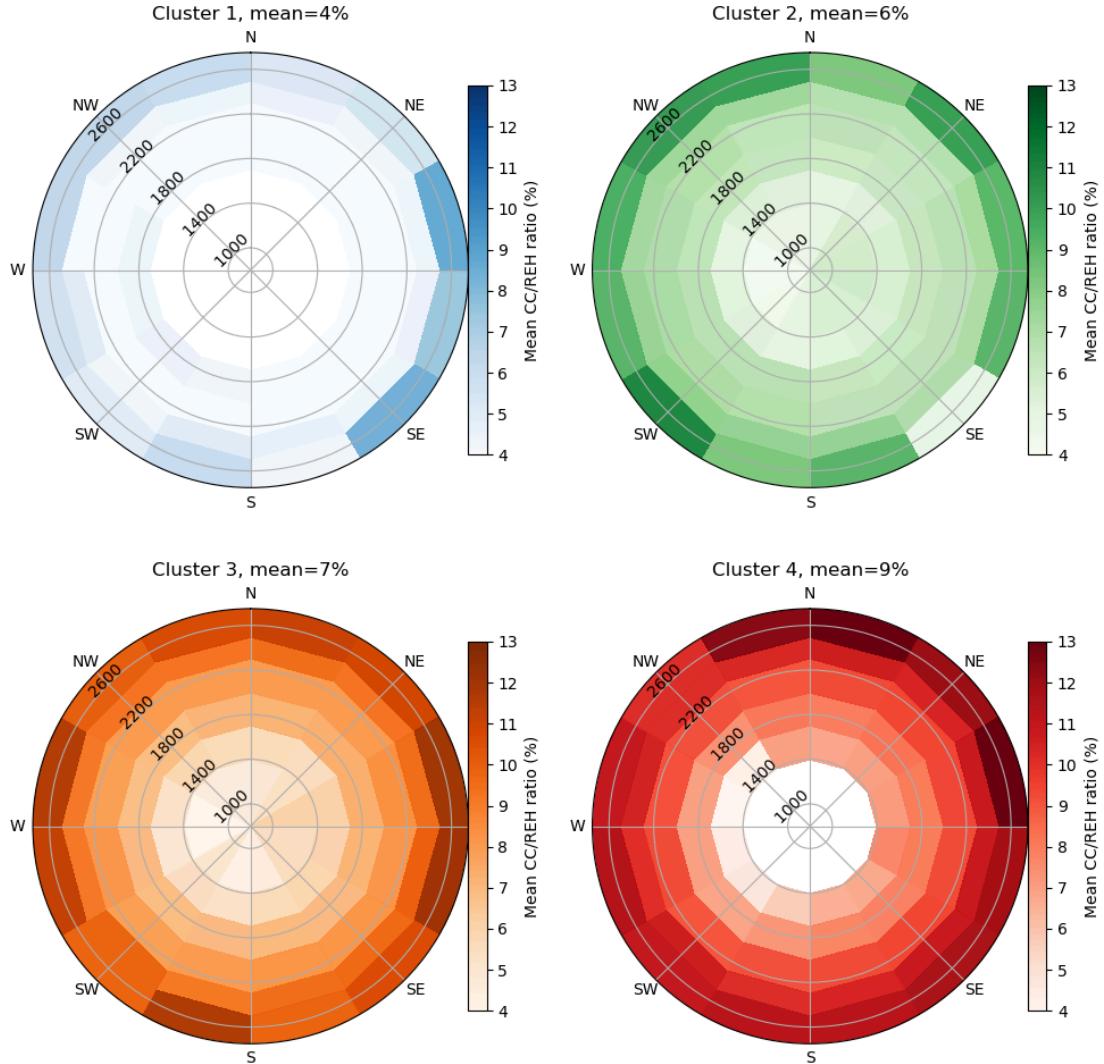
Cold content and residual saturation deficit sum up to REH, but they can take very different values. REH increases by about 6.25% per 1°C decrease. It is more frequent to have residual saturation deficit than cold content (Table 2) and cold content appears later and disappears earlier on the samples shown in Figure 2.

### 4.2.1 Fraction of REH Due to Cold Content

The fraction of REH due to cold content is a metric of the presence or absence of cold content. The fraction of REH due to cold content drops earlier for lower clusters (Figure 14), confirming that cold content is exhausted before residual saturation deficit is. Generally, the mean fraction of REH due to cold content declines through the snow season and is a function of elevation (Figure 15) and cluster number (Figure 14), increasing from 2% to 10% for lower clusters and gradually up to 15% for higher clusters. Clusters 1 through 4 have a greater cold content fraction ranging consistently from 4 to 9% on average (Figure 15). A greater fraction of a greater REH is due to cold content for higher clusters and elevations, resulting in a positive relationship between REH and cold content. The relationship between cold content fraction and aspect is subtle (Figure 15). For clusters 3 and 4, greater cold content fraction is found at high the highest elevations and for north aspects (darker colors for the highest elevation bin on Figure 15), which is consistent with the expected high REH to be found in those bins (Figure 10 and Figure 11). The cold content over residual saturation deficit ratio is poorly correlated with the mean cluster number though ( $R^2=0.10$ ) and the relationship is not significant ( $p\text{-value} > 0.10$ ).



**Figure 14: Cold content is a greater component of REH for high clusters but is exhausted before residual saturation deficit, at earlier timing for higher clusters.** Cold content divided by REH for each cluster, through the winter and spring. Before 1 December, REH is consistently close to 0 and the fraction of REH due to CC is and highly variable, it is not representative of any process, we therefore disregarded those values.



**Figure 15: Higher clusters have a greater fraction of the REH due to cold content (CC) than to residual saturation deficit ( $RS_{deficit}$ ), there is a greater fraction of cold content at high elevation.** For each cluster, elevation and aspect bin, the mean of CC/REH is computed and displayed on those polar 2D histograms. For example, on average, 13% of REH is due to cold content in cluster 4, at high elevation north facing terrain. The mean of each of those 4 diagrams (in their respective subtitles, 4 through 9%) corresponds to the mean of each curve of Figure 14. Higher elevations have a greater fraction of their REH due to CC. For a given elevation, greater CC fraction is observed for northern aspects, but not consistently for all clusters, due to the unequal area represented in each cell.

#### 4.2.2 Cluster Preferences for Cold Content and Residual Saturation Deficit Values

Cold content and residual saturation deficit do not result in the same clustering classification. Looking at the REH, SWE and temperature at peak REH reveals a gradation in the preferred classification during clustering (Figure 16). The transition from clusters 1 to 2 and from clusters 2 to 3 is largely independent of SWE, but the transition from cluster 3 to 4 depends on both SWE and REH (Figure 16).

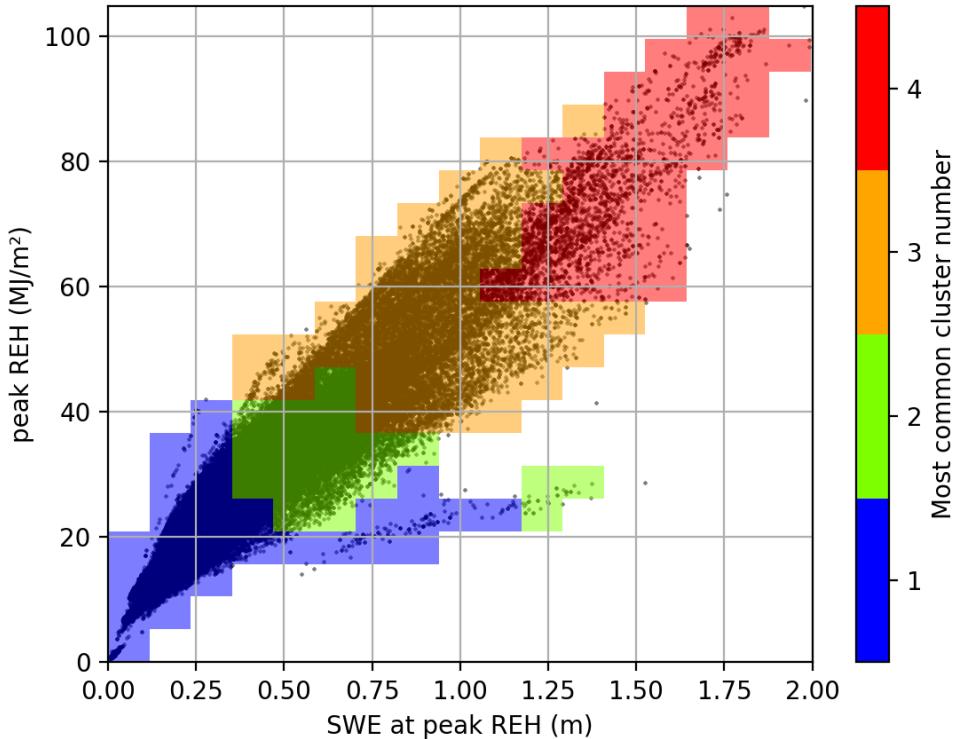
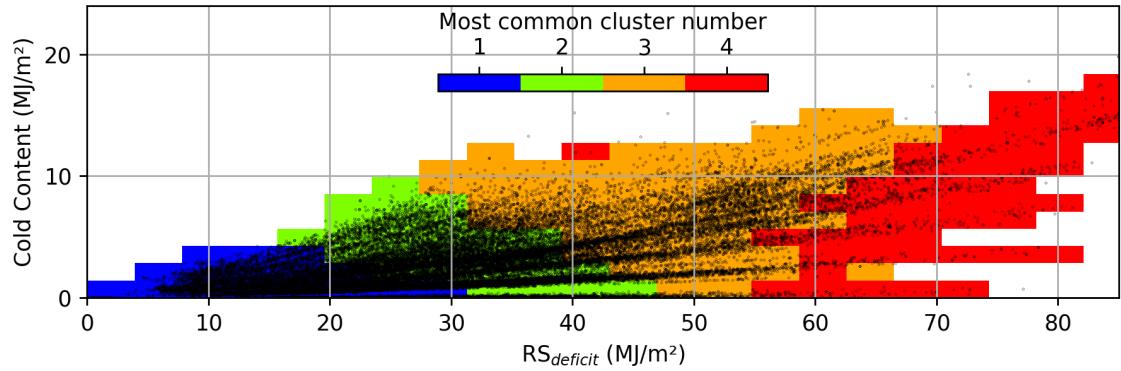


Figure 16: **The relationship between REH and SWE is modulated by the snow temperature.** At Peak REH, REH VS SWE with the most likely cluster. Peak REH and SWE at peak REH are highly correlated (Pearson correlation coefficient  $r=0.95$ ), but during the ablation season, REH and SWE are totally disconnected. The transition from cluster 1 to cluster 2 is a result of both SWE and REH, but the transition from cluster 2 to cluster 3 disregards SWE, visible thanks to the horizontal boundaries. The transition from cluster 3 to cluster 4 is a combination of REH and of SWE (oblique boundary).

The cold content and residual saturation deficit also contribute to different extents. The lower transitions (cluster 1 to 2 and 2 to 3) follow the sum of cold content and residual saturation deficit which is REH (Figure 17). The transition from cluster 3 to 4 is different again though. It is mostly affected by the availability of pore space, which is SWE. Cluster 4 mostly exists during extreme SWE years.



**Figure 17: Cluster number depends on a combination of cold content and residual saturation deficit.** At peak REH, cold content VS residual saturation deficit, with colored most likely cluster for a given CC/RS<sub>deficit</sub> bin. The transition from cluster 1 to cluster 2 and from cluster 2 to cluster 3 is more sensitive to cold content (proxy for air temperature) than the residual saturation deficit (proxy for snow accumulation), as shown by the close-to-flat boundary. The transition from cluster 3 to cluster 4 is more vertical, therefore depends on snow accumulation (RS<sub>deficit</sub>). Axes are on the same scale to allow bias-free visual analysis of this figure.

### 4.3 Activation of SWE

The mean SWE activated on a given day-cluster (m/day) is displayed in Figure 18. Critically, a mass of snow can be activated multiple times in cases where REH again rises above zero due to new snowfall with open pore space and/or growth of cold content. The time series show that cluster 1 has a different behavior than the higher clusters. Cluster 1 is marked by activations throughout the snow season that never exceed 4 cm/day and activations are greater mid-January through early April. Higher clusters have no activations during extended periods of time: late December through early March for cluster 2, early December through late March for cluster 3 and early December through late April for cluster 4. The length of the

spring activation season is shorter for higher clusters, lasting 2, 1.5 and 1 month respectively for clusters 2, 3 and 4. The maximum and overall magnitude of the activations grows with the cluster number: 4 cm, 8 cm and 12 cm for the maximum for the clusters 2, 3 and 4, respectively. Thus, as the cluster number increases, the magnitude of SWE activation increases, and the timing of activation is delayed and tightened.

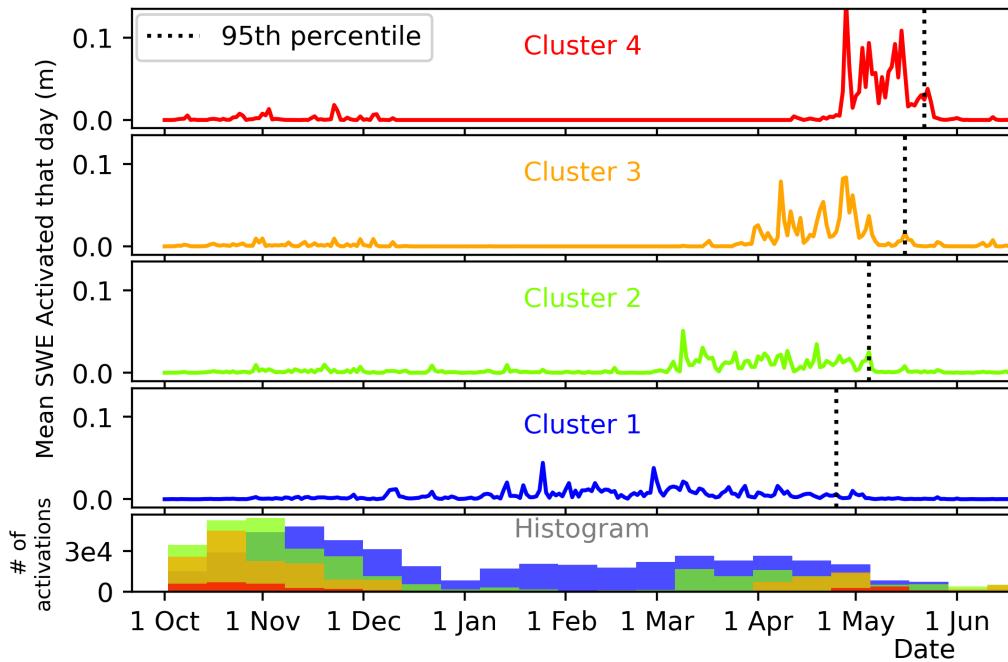


Figure 18: **High clusters have a distinct spring activation season while cluster 1 activates throughout the snow season.** For each cluster, the mean SWE that gets activated on a given day is computed and plotted, along with the 95th percentile (95% of the area under the curve is on the left of the bar).

Gathering the activation of SWE from all places allows to further extract common behavior within a cluster and different patterns between clusters. It also smooths the individually jagged activation curves. The low number of years observed on a watershed where there is a strong spatial correlation of the behavior causes the jaggedness to not entirely disappear as can be seen on Figure 18. Numerous similar locations were clustered together and underwent activation at the same time,

yielding spikes in the curves. Smoothing would falsify the data and is therefore not done. Cluster 4 has the smallest area and fewest activations (Figure 19), yet the area under the curve on Figure 18 (total SWE activated) is comparable to other clusters. This is because cluster 4 activates the greatest amounts of SWE, sometimes exceeding 2 meters.

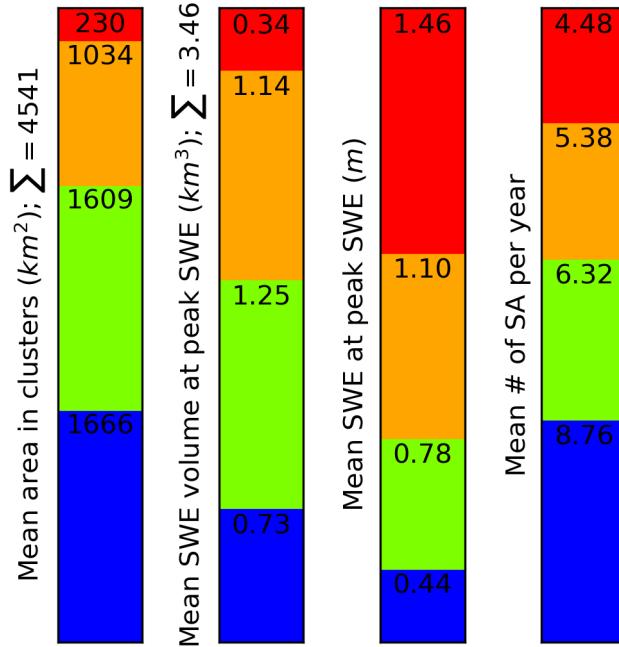


Figure 19: area, water mass, mean SWE and mean number of activations per year for each cluster. Per cluster numbers sum up to the number in the label ( $\sum$ ).

A comparison of time versus activated SWE with histograms reveals the partitioning between activated SWE and number of events (Figure 20). SWE activations occur October through June, but most happen in the fall (before 1 January) and rarely exceed 25 cm, due to the combination of relatively low REH and relatively high energy inputs. The activations that occur in the fall show little pattern with regards to cluster number, while in the spring, higher clusters experience fewer, later and greater activations (Figure 20). The maximum amount of SWE that can be activated on a given day increases 1 October through mid-May, before dropping to 0 by 1 July. The greatest magnitude of SWE activations is achieved by cluster 4 in the spring. Activations greater than  $\approx .3\text{m}$  and past early March mainly belongs

to cluster 2 and higher; activations greater than .7 m and past early April mainly belongs to cluster 3 and up; and activations greater than 1 m of SWE past late April mainly belongs to cluster 4. Most activations occur for cluster 1 throughout the year, confirmed by Figure 19. Higher clusters demonstrate fewer activations per year: 8.76, 6.33, 5.38 and 4.48, respectively (Figure 19).

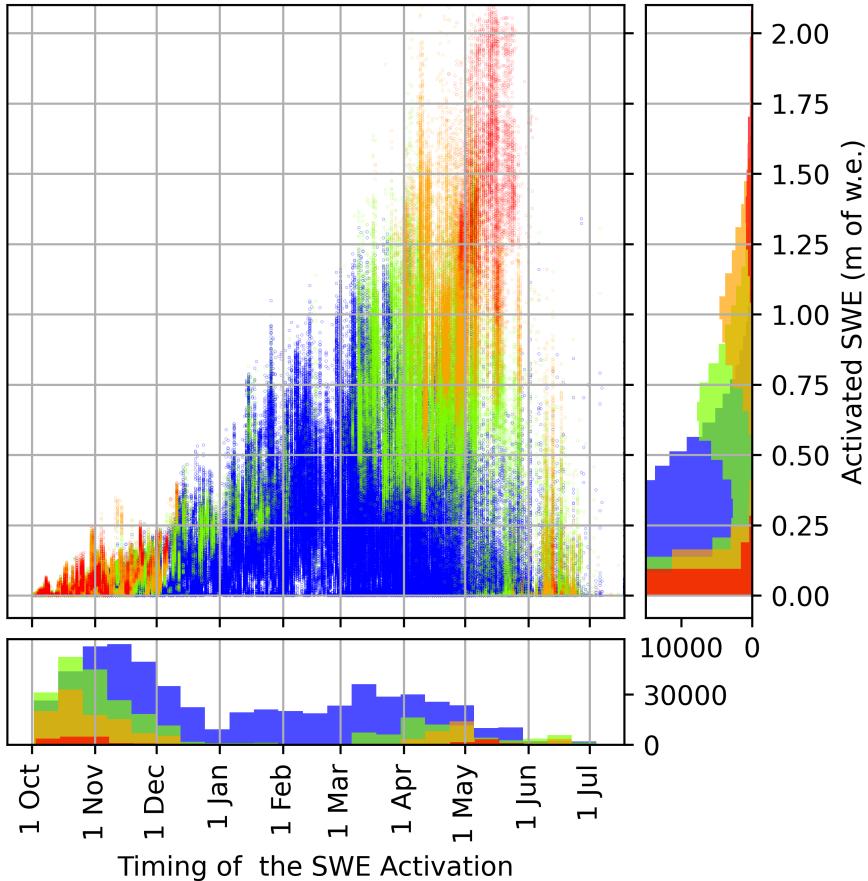


Figure 20: amount VS timing of activated SWE and their respective histograms, for all location-years.

#### 4.4 Predictive Models of the Timing of Spring Activation

spring activation was computed as the greatest SWE amount activated per location-year and its day of the year. A visual inspection of 200 individual time series found that on rare occasions, a location would endure its greatest SWE activation, but recover and reactivate a couple of weeks later with a slightly lower amount of SWE. In those situations, even if it has a lower amount of SWE activated, the

second activation should be picked by the algorithm employed to identify spring activations as the spring activation since spring snow runoff starts there and not before, but the algorithm would pick the first activation with a greater magnitude. This scenario was only present 3.5% of time series though, yielding a 96.5% accuracy for the simple spring activation identification method proposed here. Due to the numerous activations of very variable magnitudes in cluster 1, spring activation was not computed for it.

There are two objectives with predicting the timing of spring activation: 1) forecasting for operation purposes, and 2) quantifying the importance of terrain properties.

#### **4.4.1 Spring Activation Timing Forecasting from Partial REH Time Series**

Spring activations were gathered for clusters 2 and higher. The mean difference between all spring activation dates and their mean is 16.9 days. In other words, the most naive model, which is to take the observed historical mean, yields a mean absolute error of 16.9 days. Our objective is to improve those predictions using information from the REH history or from the terrain properties. The first attempted method is to refine the historical mean prediction to per-cluster predictions.

Spring activation happens later for higher cluster with a smaller variability (Figure 21). Cluster 1 represents the locations that activate mid-winter with unstructured timing. Computing spring activation makes less physical sense for this cluster. Consequently, no spring activation analysis is done on cluster 1. Higher clusters have a later spring activation timing (5 April, 22 April and 8 May for clusters 2, 3 and 4 respectively) with smaller MAE (17.3, 9.8 and 7 days respectively). Simply knowing the cluster number of a given location early in the year has predictive power. Using partial REH time series, cluster number is estimated for the entire time series. The accuracy of this classification is a function of the length of the record considered and is shown on Figure 22. A random classification would yield an accuracy of 25% for 4 clusters. By clustering until 1 January, 60%

of the clusters are correctly classified. Improvement of the classification with later dates then follows a relatively linear trend as day-of-year increases, until 99% of accuracy is attained by clustering until 1 May. The accuracy of this partial data clustering can then be propagated onto the MAE of individual clusters to estimate uncertainties of the model predictions.

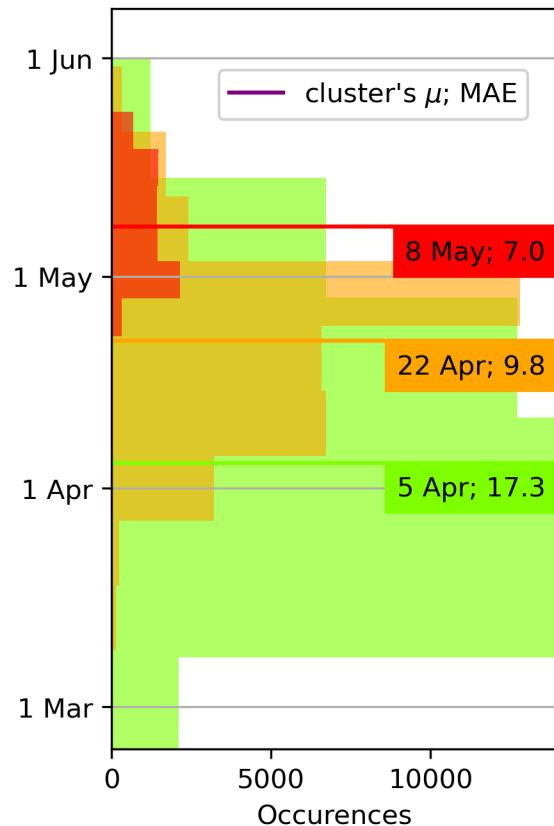


Figure 21: histogram of the spring activation for each cluster, with their mean and Mean Absolute Error (MAE).

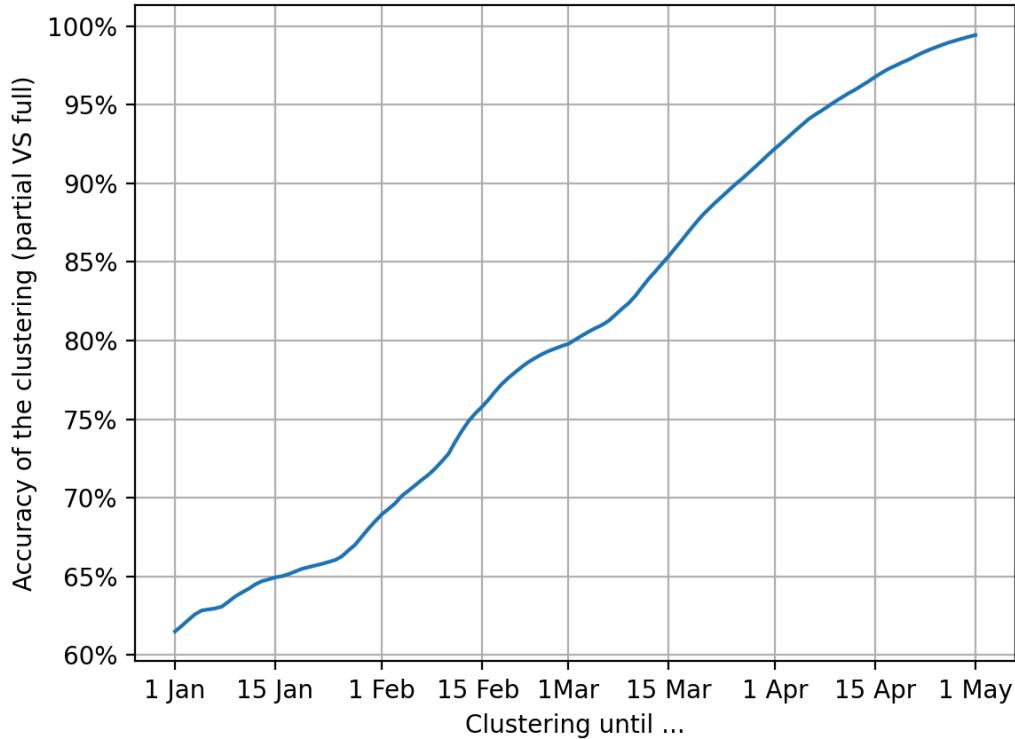


Figure 22: accuracy of clustering until a certain day. Clustering of partial time series interrupted on 1 March will result in correct classification with 80% accuracy.

Considering the clustering accuracy on 1 March being 80%, with the associated MAE from each cluster (Figure 21), the real K-Means prediction uncertainties are 21.6, 12.3 and 8.8 days for clusters 2, 3 and 4 respectively.

The ensemble of Long Short-Term Memory (LSTM) neural networks fully embraces the complexity of the signal. The LSTM uses the fact that the REH is temporally correlated: REH on a given day is correlated to REH on previous days, and REH is correlated to the timing of spring activation. This model ingests REH values from 1 October through 1 March and estimates the timing of spring activation that could be up to two months later. A set of predictions is done for each year individually and the performances for a given cluster, throughout all years, are compared. The overall MAE of the LSTM is 17.0 days, which is not better than taking the historical mean. More precisely, MAE are 19.3, 17.3 and

23.2 days for clusters 2, 3 and 4, respectively. The LSTM suffers from strong biases, especially, it underpredicts higher clusters' spring activation timings. Results are gathered on Table 4.

	Cluster 2	Cluster 3	Cluster 4	Overall
K-Means' MAE	21.6	12.3	8.8	16.9
LSTM's MAE	19.3	17.3	23.2	17.0

Table 4: Mean Absolute Errors (MAE) of considering the mean of each cluster, or overall, and the LSTM's predictions on each cluster. K-Means outperforms all alternatives for clusters 3 and 4, the LSTM outperforms K-Means for the highly-spread cluster 2.

#### 4.4.2 Terrain Feature Importance

A total of 18 models were fitted. Their MAE are considered for overall model performance and the bee swarm and violin Shapley plot provide information on the influence of individual features on the model outputs. The MAEs are gathered on Table 5 and the plotting of Shapley values is done on Figure 23. The results here may look better than the ones in Table 4, but they are the results of a different method. Precisely, they suffer from spatial correlation: the testing dataset is selected randomly instead of the per-year operational sorting of the LSTM ensemble. The reason for examining those results is because they may reveal spatial trends, although they are not relevant for operational predictions.

Model	Mean (base-line)	Simple Lineare Regres- sion	Random Forest	Gradient Boosting	Artificial Neural Network
No inter-annual variability considered (only terrain features)	25.5	13.11	14.22	13.13	-
Inter-annual variability considered through cluster number as a model feature		11.45	12.8	11.42	-
Model for cluster 4 only	6.5	6.2	6.81	6.35	8.68
Model for cluster 3 only	8.2	8.23	9.32	8.21	9.85
Model for cluster 2 only	14.3	13.97	15.12	13.96	15.43

Table 5: Performances (MAE) of numerous models, with spatial correlation between data. The per cluster baseline model (mean) is not significantly outperformed by any other model, no improvement can be obtained.

# Models summary

# Results

Inter-annual variability considered

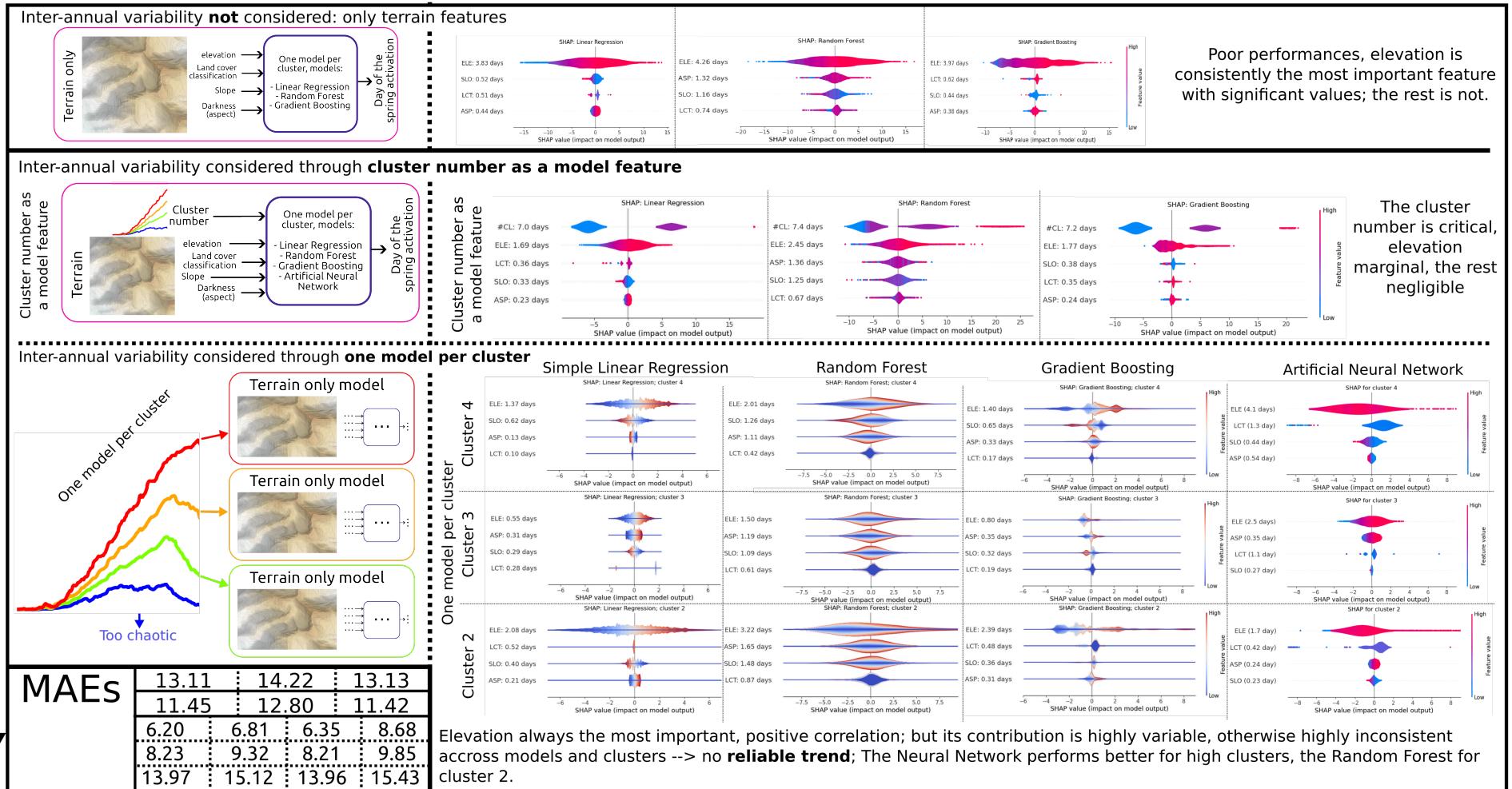


Figure 23: summary of the models and findings for all model architectures. Left column are model summaries and describe the right column, which is the different violin plots and Shapley values of each feature in each model. The violin plots display all datapoints on the x axis, the y axis shows the number of data points and the color is the positive or negative impact on the model output. For example, a positive correlation would have warmer colors on the right side of the x axis. On the bottom left are included the Mean Absolute Errors (MAE) of each model, for comparison purposes, identical to Table 5.

The baseline model is simply the mean, which yields a model with a Mean Absolute Error (MAE) of 25.5 days overall. Thus, on average, this model's prediction of the Spring Activation will be 25.5 days off. Cluster-by-cluster filtering yields significantly lower MAEs of 6.5, 8.2 and 14.3 days for clusters 4, 3 and 2 respectively.

Models incorporating only terrain features and considering cluster number as a model feature yield MAEs ranging from 11.42 to 14.22 days, corresponding to a doubling of the prediction accuracy. With no cluster number, the Shapley plot shows that elevation is the most important feature with contributions ranging 3.8 to 4.3 days depending on the model, until cluster number is added as a categorical feature and elevation becomes marginal, like the other terrain features. Cluster number contributes 7.0 to 7.4 days depending on the model and outweighs all other features.

For a given cluster, no model significantly outperforms the baseline case of taking the historical mean. With one model per cluster (Figure 23, row 3), performances of the various models are very limited. The various models perform 5% better to 33% worse than the naive models.

Neglecting inter-annual variability, another approach to quantify the spatial variability of spring activation timing is done on Figure 24. For elevation and aspect bins, the mean spring activation timing is computed with colors and contour lines. There is a strong elevation component to the variability of mean spring activation timing shown by the gradual later mean spring activation at higher elevations (4.8 days per 100 m). But the relationship with aspect is more subtle. For a given elevation location, the mean spring activation is later for a north aspect than on a south-west aspect. The mean difference is in the order of a couple of days. With regards to elevation, the mean equivalence of behavior between a north facing terrain and a south-west facing terrain is about 400 m at 2000 m. For north aspect, spring activation occurs 21 m higher per day on average and for southwest aspects, 31 m of elevation per day.

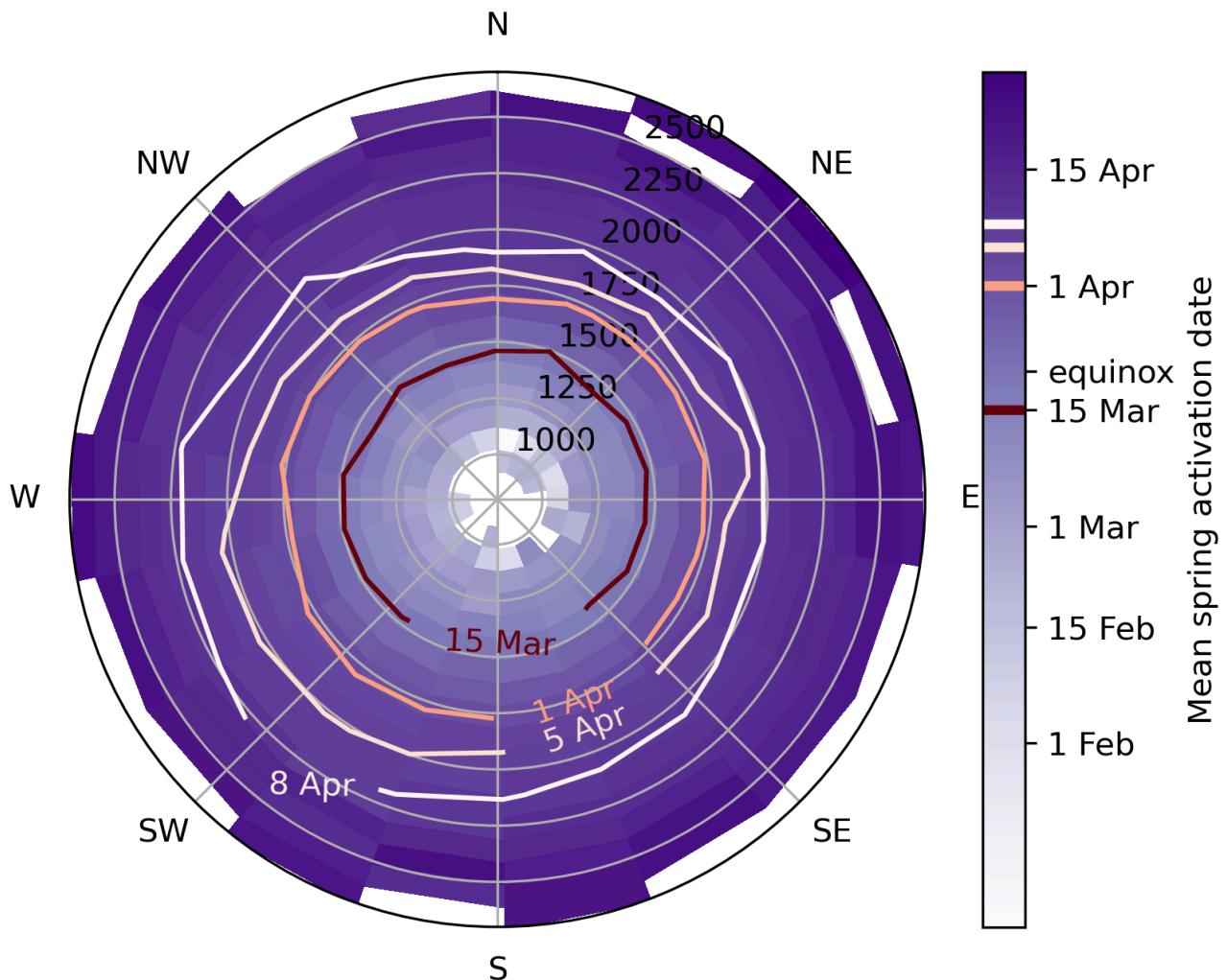


Figure 24: **Elevation is the main indicator of spring activation timing; aspect can have a strong influence later in the season when the solar radiation gets greater.** Colors show the mean spring activation date for aspect and elevation bins, lines are contours for visualization of the skewness towards SW aspects. For north aspect, spring activation propagates 21 m of elevation per day, for southwest aspects, 31 m of elevation per day.

#### 4.5 Anatomy of Widespread Spring Activation Events

31 widespread spring activation events were identified, ranging 4 to 17 days and between 261 km<sup>2</sup> and 3192 km<sup>2</sup> (5% to 70% of the watershed area) activated in a short time window. On average, events are 8 days long and spanning 1602 km<sup>2</sup> (35% of the watershed area). During those widespread spring activation events, trends in elevation, spring activation, northing and solar radiation trends were collected and analyzed (Figure 25). During the widespread activation events, there is overwhelming evidence that elevation increases by 32 m per day ( $t\text{-statistic}>1$ ,  $p\text{-value}=0.00023$ ), there is weak evidence of an increase of the

northing (t-statistic 1, p-value = 0.09), there is overwhelming evidence of an increase in SWE of 3.7 cm/day (t-stat>1, p-value=  $10^{-5}$ ) and there is no evidence of any trend with incident solar radiation (t-statistic < 1, p-value=0.7). This is consistent with the findings using Shapley values.

During 31 ripening events, the temporal evolution of the following coeffs in unit/day is:  
(This includes the mean of the distribution and its standard deviation)

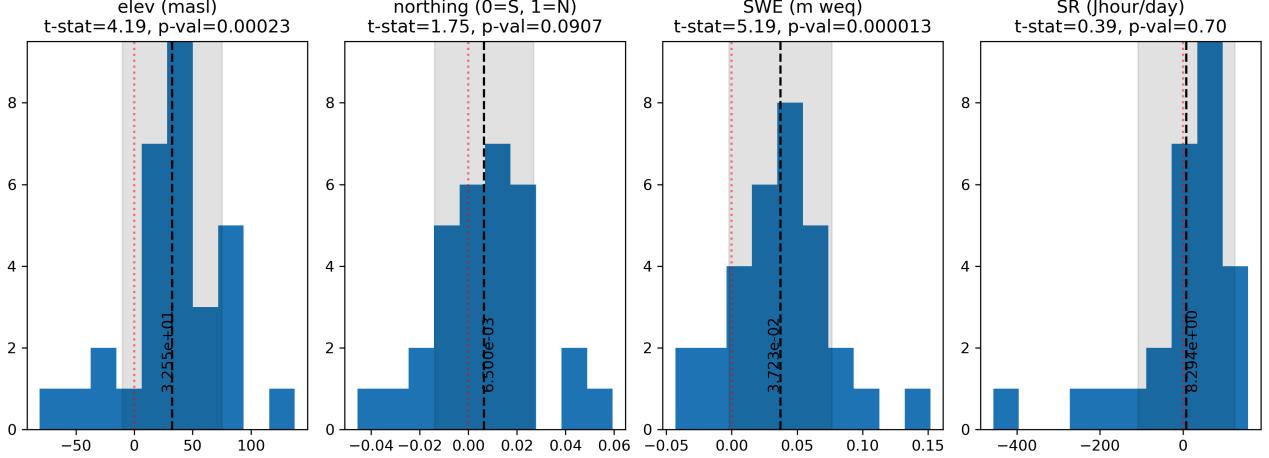


Figure 25: histograms, trend and significance of the trend over multiple properties during multi-day watershed-wide activation events. Spring activations progress 33 m/day,  $6.5 \times 10^{-5}$  northing index/day, an additional 3.7 cm SWE/day and with no significant trend regarding solar radiation.

## 5 Discussion

### 5.1 Limitations

#### 5.1.1 SNODAS

The data used in this work are the output of the operational model SNODAS and not observed values. The analysis here is subject to all the model's strengths and weaknesses, and numerous simplifying assumptions of the model. SNODAS uses coarse meteorological data as inputs, leading to coarse model resolution. While the model can be improved with airborne survey and SNOTEL data, the model output's accuracy remains poorly validated.

A key model output used here is snowpack temperature and there are no available observations of snow temperature for validation (i.e., this is not a measured parameter at SNOTEL sites). Additionally, SNODAS outputs the snow temperature as an integer. Given that SNODAS uses 3 layers of snow, it is unclear whether the final snow temperature is a rounding or a truncation of the mean of the 3 layers.

In a study of the Bitterroot Range, Montana, Kellner (2013) found that SNODAS has a greater SWE variability than any other model or dataset considered. They found that

SNODAS highly underpredicted SWE at high elevations, which is consistent with a study in Colorado by Clow et al. (2012). A more recent study in Alberta by Lv et al. (2019) found that SNODAS treated wind and its interactions with the tree canopy poorly. They also found that the precipitation used to force SNATHERM was gradually overestimated SWE during the years 2011-2015, resulting in excessive snow accumulation values. They found that assimilating SNODAS data into a physics-based hydrological model corrected a significant part of those biases. Due to difficulties with assessment studies stemming from a lack of observations, and the different locations and time intervals of assessment studies, generalized conclusions about the fidelity of SNODAS model output are elusive.

Nonetheless, SNODAS includes physical complex processes, sophisticated treatment of inputs, and is the only available operational snowpack model. SNODAS is therefore used for decision making for water resources forecasting, dam management and other purposes. Thus, even with its limitations and uncertain fidelity, SNODAS outputs are relevant for society and must be understood, thus providing relevancy to our work. Also, this work is transferable to a dataset originating from better snowpack modeling, once available.

### 5.1.2 Uncertainties on the Residual Saturation Deficit

While SNODAS simulates liquid water content, the values are used internally and not output. This number is needed to compute REH (Equation 1). We therefore use base runoff as an indicator that the model is simulating a snowpack at residual saturation (Section 3.2.3). This resulted in no certainty in the actual value of  $RS_{\text{deficit}}$ ; we only know if  $RS_{\text{deficit}}$  is null or not. The impact of this workaround is limited, since this limitation was consistent across all time series.

The liquid water content of a snowpack may significantly exceed 7% ephemerally when liquid water inputs overcome the runoff capacity of the snowpack, reaching over 20% on rare occasions (Bonnell et al., 2021; Colbeck, 1974; 1976; Coléou & Lesaffre, 1998) before declining to  $RS_{\text{frac}}$ . The of a snowpack is not necessarily a constant 7% either, it has been found to range 6.5 to 8.5% (Coléou & Lesaffre, 1998) and Colbeck (1974) picked 7% as the ideal value, which we used here. A sensitivity analysis was conducted by clustering REH time series with varying from 6.5 to 9.5% with 0.5% increments. Observed changes in the clustering with alternative values were less than 0.01% change, suggesting that our analysis remains robust to our assumption.

The two limitations described above are also mitigated using K-Means. Indeed, K-Means is scale-invariant and so it would produce the same output on a scaled dataset. The relative values of REH between time series and through time are the critical numbers here, not the

absolute amount of REH. Because the lack of and the uncertainty on are consistent through the time series, we conclude that those flaws only have limited impact on our results.

### **5.1.3 REH to Capture the Outflow of Water**

REH is a metric for full ripening of the snowpack and is an imperfect proxy for outflow from the base of the snowpack due to inhomogeneous infiltration processes. Studies have found that occasionally melt may exit the snowpack without reaching its base due lateral flow along snow stratigraphy (Eriksson et al., 2013). Furthermore, water can penetrate to the ground surface locally in preferential flow paths before the entire snowpack has ripened (e.g., Conway & Benedict, 1994; Marsh & Woo, 1984; Marsh & Woo, 1985; Waldner et al., 2004). This is not accounted for here and thus our results represent a later stage in the runoff regime where the bulk snowpack is contributing to water input to the ground.

### **5.1.4 Correlation of REH to SWE**

While REH and SWE are correlated, each offers unique information. For example, two SWE curves may be identical, but have different REH curves if one has less cold content or if one activates mid-winter when the other one does not. When REH is nonzero, the Pearson correlation coefficient between REH and SWE is  $r=0.85$  (Figure 16). However, REH and SWE are completely uncorrelated during the ablation season because REH is almost always 0.

In addition, the loss of REH has been found to be the critical measure for this study instead of the magnitude of REH. An example of that is the water year 2023 that has the third highest mean cluster number value while being among the lowest SWE years. Furthermore, while SWE decreases progressively with lower mean cluster number, the REH does not decrease similarly. Instead, REH time series within cluster 1 abruptly oscillate between 0 and relatively low values. While we must acknowledge the correlation between REH and SWE, REH analysis is more than a study of SWE.

### **5.1.5 Discrete Clustering in a Continuous System**

K-Means proves to be a powerful tool to identify behaviors and trends in the data, but this method discretizes a continuous system. The behavior of the snow is assumed to be spatially correlated and so the REH time series at a location resembles the time series at neighbor locations. This leads to a continuous spectrum of REH time series. This results in a high resemblance between the highest time series of a given cluster and the lowest time series of the next higher cluster, which is confirmed by negative silhouette values (Figure 4). One solution to this would be to use fuzzy clustering (Warren Liao, 2005), which yields a real-number cluster number instead of an integer but restricts us from doing any classification of other properties.

## 5.2 Current State of REH and Activations

### 5.2.1 Influences on REH and Cluster Number

In the available history, the 4 clusters are not equally represented and do not contribute equally to water resources or water release timing. Cluster 1 is the most wide-spread cluster and covers 37% of the watershed area but it has the least SWE on average (0.44 m), resulting in a total volume of water of 0.77 km<sup>3</sup> (21% of total volume; Figure 19). Cluster 2 covers 35% of the watershed and has 0.77 m of SWE on average, therefore containing the greatest volume of water with 1.32 km<sup>3</sup> (36% of total volume). The higher clusters are gradually less represented and represent a declining volume of water because of their inconsistent representation, especially cluster 4 which is only represented 9/19 years. Cluster 4 is a subsample of lower clusters during high accumulation years.

The spatial properties of the clusters have different trends and patterns. Higher clusters (with higher REH values) have on average higher elevation, are more frequent on east aspects and shorter/no vegetation, and most importantly, are associated with later, more temporally constrained and greater magnitude spring activation (Figures 5 through 12, 15 and 19 through 21). We quantified those trends using Shapley values on a wide range of models. Using the Shapley values, we found that overall, elevation affects the spring activation timing output by up to 4 days (Figure 23), which is a smaller contribution than any other methodology found.

REH is the sum of the cold content and the residual saturation deficit (Equation 1), which have heterogeneous contributions through time and space. The fraction of the REH represented by cold content is greater for higher clusters. For example, cluster 4 has on average 13% of its REH due to cold content while cluster 1 has 9% (Figure 15). Generally, the contribution of cold content to REH is high until peak REH and becomes marginal past it (Figures 15 and 16). Greater fractions due to cold content are observed at high elevations and in high clusters.

The representation of the clusters through years emphasizes the strong inter-annual variability (Figure 6). From year to year, 11 to 96% of the watershed can be clustered in cluster 1, cluster 4 only occurs appears 7 out of 19 years and the mean cluster number ranges from 1.03 to 2.68, showing that weather is a key part of the snowpack state. This inter-annual variability is a key component of this analysis and shows the extremes that could occur.

The extreme years are a result of weather forcing which can be decomposed in the precipitation amounts (highly linked to SWE, pore space or RS<sub>deficit</sub>) and in the air temperature (highly linked to CC). The transitions from clusters 1 to 2 and from 2 to 3 are a result of the

combination of cold content and residual saturation deficit, while cluster 4 is mostly insensitive to cold content, it only presents during years with high amounts of SWE (Figure 16). The general dependency on air temperature is confirmed by comparisons to SNOTEL temperature measurements. Warmer years result in lower mean cluster number with a loss of 0.23 mean cluster number per increase in the January, February and March mean air temperature (Figure 9). Also, January, February and March El Niño indices cause a loss of 0.42 mean cluster number per increase in El Niño index (Figure 10).

The type of snowpack and its properties determines the cluster number. Colder air temperatures cause a higher cold content in the snowpack, therefore classifying that snowpack in a higher cluster. A greater snow depth will bring more pore space, therefore increasing the RS deficit, but resulting in a smaller increase in the cluster number than cold content does. Although the contribution of pore space remains secondary for lower clusters, it is the main condition to transition from cluster 3 to cluster 4, which only occurs during extreme accumulation years.

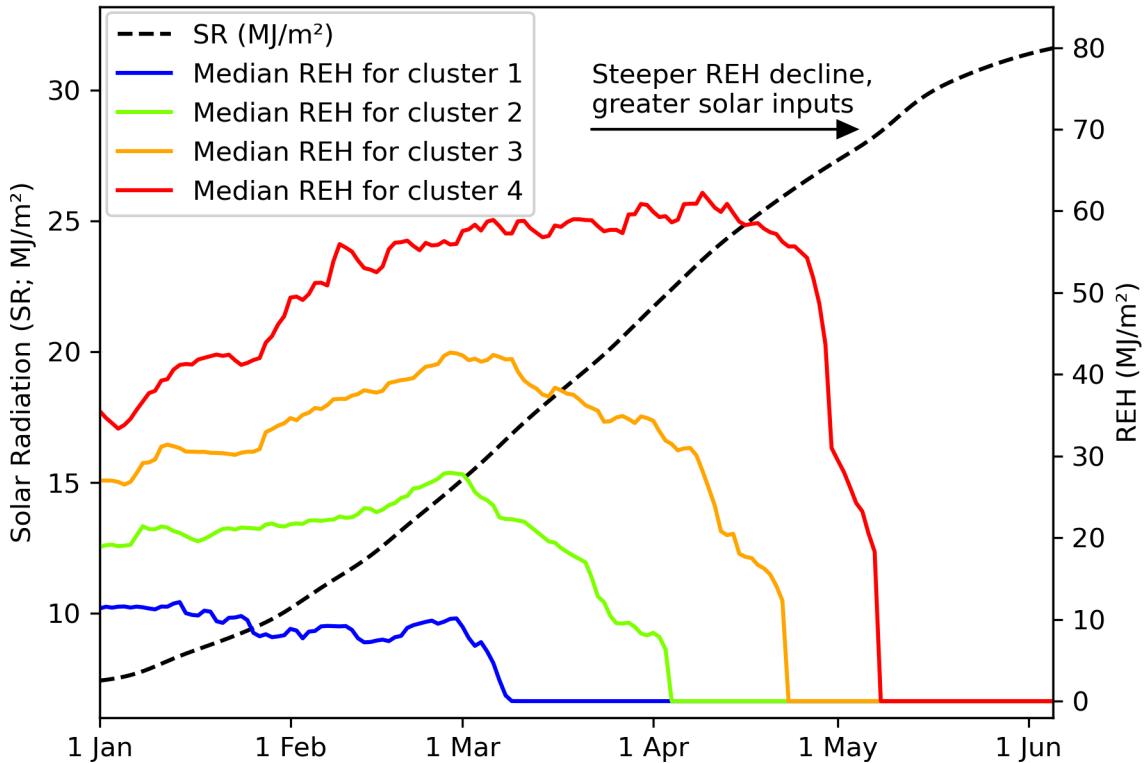
### **5.2.2 Consequences of REH Values on Activations**

The patterns of activation are REH-dependent, and the spatio-temporal variability of spring activation timing is high. Computing simple correlation coefficients only is efficient in a very limited way; we risk oversimplifying reality.

The timing of spring activation has high interannual variability and we embraced the inter-annual variability by considering the cluster number as the major indicator for accurate predictions. No addition of other terrain properties in the multiple models tested improved our predictions. While the high clusters are associated with a later, tighter, and greater magnitude spring activation, the behavior of the first cluster suggests that consistent mid-winter activations are possible under certain conditions. Cluster 1's activations are spread over the entire snow season, while the other clusters have extended duration that are completely activation-free. Due to the frequent activations throughout the winter of cluster 1, we did not compute a spring activation. Cluster 1 exhibits significant mid-winter activations, and a preponderance of cluster 1 results in droughts the following summer (Figure 8).

The presence of REH is bounded by the exponentially increasing cumulative surface energy inputs. Even though the interaction between incoming shortwave radiation from the sun and the snow cover depends on elevation, slope, aspect, and canopy, the watershed-wide mean cumulative solar radiation brings exponentially more energy to ripen and melt the snow (Figure 26). A greater decrease in REH is witnessed later in the season relative to earlier in the spring, the different cluster's median REH declines faster later in the season (Figure 26),

which is consistent with higher melt rates reported later in the season due to greater solar radiation inputs (Lundquist & Flint, 2006; Musselman et al., 2017). Cluster 1 shows numerous activations throughout the winter because those are due to the highly variable weather. On the other hand, higher clusters' activations happen later, when the solar inputs become greater, therefore causing the narrowing of the spring activation window, converging towards the spring activation boundary.



**Figure 26: The loss of REH accelerates later in the season when energy inputs, especially the shortwave radiation, increase significantly.** Semi-conceptual plot comparing the REH to solar radiation, a proxy of the increasing energy inputs in the spring. In the winter, long nights and low sensible heat overcome daily energy inputs, keeping the snow cold. But in the spring, the snowpack transitions to energy intake and the later the REH is exhausted, the faster.

To complement and expand our understanding of the spatial variability of spring activation, we developed an alternative approach to the problem. We disregard the cluster number and consider the mean spring activation for elevation and aspect bins. We found that the timing of spring activation increases by 4.8 days per 100 m of elevation or propagates upward by 21 m per day (Figure 25). The relationship between spring activation timing and aspect is more subtle and seems to only be revealed later in the spring season, when the solar radiation becomes greater relative to other energy sources. The south-west aspects activate later at

high elevations. More precisely, north and south-west facing locations will activate on 8 April if they have a 400 m elevation difference. Those results are consistent with the findings of Gagliano et al. (2023) regarding elevation and aspect trends. Earlier in the year and at lower elevations, we suspected that weather conditions are more likely to activate the snowpack than solar radiation, which is still comparatively low, proven by the near-circular contours for 1 April and before (Figure 24). Later in the season, when solar radiation becomes a major source of energy, activations become more sensitive to it, resulting in the observed trends. The seasonal sensitivity of spring activation timing to solar radiation is coherent with the lower melt rates observed early in the melt season by Lundquist & Flint (2006).

The detection of watershed-wide spring activation events complements the previous findings to understand the timeline of events (Figure 25). Overall, Spring Activations start at lower clusters and propagate to higher clusters. Within a cluster, spring activations start at one extreme: low elevation, high slope, south facing; and propagate upward, to flatter slopes and to more north facing terrain until they reach the other extreme: high north-facing relatively flat slopes. The reason why steep north faces do not emerge in this trend is because steep slopes are a minority. Slopes greater than  $30^\circ$  only represent 6% of the watershed area and slopes greater than  $35^\circ$  represent 2.8% of the watershed area. This trend is general and the rareness of  $>30$  degrees slopes does not emphasize the trends on the steeper north-facing slopes. Those slopes cannot be too steep however, slopes over 30% tend to experience gravity transport of snow (Dewalle & Rango, 2008, Chapter 10.4.9), reducing its SWE and REH and forwarding the spring activation. This does not quantify the timing between the two extremes though, which presumably depends on the cause of the widespread spring activation event and the timing of the year.

### 5.2.3 Implications of REH for Rain-on-snow Events

Rain-on-snow may or may not cause a flood, depending on the energy inputs and the snowpack REH. A rainstorm can input significant amounts of heat into the snowpack via turbulent heat fluxes, longwave radiation from the clouds and latent heat of the rain. While some studies acknowledge that the snowpack state matters (e.g., Marks et al., 1998, 2001; Musselman et al., 2018), most studies focus on the rain and neglect the snowpack REH (e.g., Li et al., 2019; McCabe et al., 2007). The rainstorm studied by Marks et al. (1998) was “warm, very humid” and lead to major flood events in February. That storm energy inputs into the snowpack peaked at approximately  $300 \text{ W/m}^2$ , corresponding to approximately  $8.6 \text{ MJ/m}^2$  of energy input over 12 h. Yet, the REH exceeds that amount in many situations (Figure 5). For example, the centroid of cluster 2 peaks at approximately  $12 \text{ MJ/m}^2$ . The higher REH

curves in cluster 4 even exceed 40 MJ/m<sup>2</sup>. In years with a low mean cluster number, the snowpack frequently has less REH than what a rainstorm may bring, resulting in activations (Figure 18). The lower the REH, the lower the cluster number and the more likely that a storm's energy input overcomes the snowpack REH. This demonstrates the importance of REH to studies of rain on snow events.

### 5.3 Implications for a Warmer World

Very low SWE and a warm winter are necessary to classify nearly the entire watershed in cluster 1. We found that lower clusters are more frequent during warm years (Figure 9) and associated with a lower cold content (Figure 15). The relationship between cold content and cluster number also depends on the RS deficit (Figure 17); but with fixed available pore space, a loss of cold content directly implies a reduction in the number of clusters. The contribution of pore space and cold content to the loss of cluster number is generally equal; a loss of one may be compensated by an identical gain of the other, except for the transition to cluster 4. That transition from extreme REH, represented by cluster 4, to cluster 3 is mostly dependent on the snow amount and not on cold content (Figure 17). In the Pacific Northwest, middle-to-upper elevation snowpack has high SWE but low REH because of the frequent wetting of the coastal snowpack. This low REH results in already visible advancement of the snowpack ablation (Musselman et al., 2021). Any loss of cold content or SWE results in a lower mean cluster number, the combination of both accelerates the decline of mean cluster number, but only a loss of SWE can cause the transition from cluster 4 to cluster 3. A sufficient loss of cold content and pore space will result in severe loss of mean cluster number, which may classify the entire study area in cluster 1.

Although there is no absolute certainty about the future of snowpack in the mountain west, it is commonly assumed that the future's average snowpack will shift towards today's low-end extremes. Although the shift in snowmelt onset from global warming has been documented by multiple studies (Clow, 2010; Evan & Eisenman, 2021; Gottlieb & Mankin, 2024; Musselman et al., 2017; 2021), we find no significant trend in our data and most years do not appear to have large mid-winter activations. Specifically, 3/19 very low REH years have been witnessed in this study (Figure 6) and are the tails of the current distribution of mean cluster number. The increase in droughts following low cluster winters seen in this study is consistent with previous findings (Livneh & Badger, 2020), and the increase in January–February stream runoff volume observed (Hamlet et al., 2007, p. 200) corresponds to an increase in activations in that time frame. The remaining years will shift towards the lower tail end of snow behavior. Both high and low extremes were observed in 19 years, revealing

the magnitude of the shift that could occur between those and extreme lows. In other words, extremes from today may be tomorrow's average simply by changing the high and average years into low and extreme low REH years. While the predicted slight increase in overall precipitation predicted does not provide clear trend on whether those precipitation will be solid or not and whether it will contribute to an increase or loss of mean cluster number, the predicted warming of the Northern Rockies (IPCC, 2021, fig. SPM. 5) will contribute to an overall drop in mean cluster number.

## 5.4 Practical Predictions of Spring Activation Timing for Dam and Flood Managing

Mean of past history, the K-Means-refined mean of past history and a Long Short-Term Memory (LSTM) neural network were proposed to predict the timing of spring activation, the later two use the partial REH time series, taking advantage of the temporal correlation of REH. In addition to K-Means, we attempted to further refine the predictions using additional models, with poor performances (rows 2 and 3 of Figure 23). Those models were linear regression, random forest, gradient boosting and a simple fully connected neural network, with terrain properties as model inputs (elevation, slope, aspect and land cover type). Each model was fitted 1) with cluster as a feature or, 2) for each cluster. None of the models significantly outperformed simply taking the mean of the historical spring activation within each cluster.

Overall, K-Means clustering is a powerful method to embrace complex signals and extract trends. Inter-annual variability of the snowpack properties is a strong obstacle to our understanding of the ripening process. The metrics addressed in this work include REH, snowpack activation, and spring activation timing. A common method to account for the inter-annual variability is to compute correlations between a given feature and the observed property, yielding various quantitative results (e.g., Gagliano et al., 2023, for spring activation timing). However, the findings are not necessarily transferrable to new locations nor new years. Gordon et al. (2022) proposed a general and transferrable framework to quantify the complex interannually variable snow properties. Our framework is statistical, intuitive to humans, and is performed using K-Means clustering, offering high performances in parsing out snowpack behavior and trends and robustness to a warming climate. Operational predictions can be done with greater accuracy than with the naive model of taking the historical mean. When done on 1 March, the accuracy of the clustering is 80%; it functions on never-seen-before data (i.e.,new years) and yields the best performances. The K-Means classification does excel for classifying new years. We compared the classification of a randomly removed

year between 1) K-Means fitted on all years, and 2) K-Means fitted on all years but the one year removed. The classification differs by 4.3% on average. This means that the cluster centroids are barely changing with the addition of new years and this method is useful and reliable for future predictions.

For all clusters, the LSTM yielded predictions with mean absolute errors (17.0 days) comparable to simply taking the historical mean (16.9 days). This may be due to the limitations of the complex algorithms that are neural networks, they cannot predict anything, or this error corresponds to the still-existing variability in weather in the spring. We investigated numerous model architectures, including alternative approaches to the problem like fitting one model to all years but one and testing on that last one, and we always varied the number of layers and nodes at each later. This comforts us in the validity and strength of our methods, and that the system studied is extremely complex.

## 6 Conclusion

From the output of the operational model SNODAS, we have computed the Runoff Energy Hurdle (REH) over 19 years for a watershed of the Northern Rockies and at daily resolution. The REH represents the constraint on energy input to a snowpack that must be satisfied before a snowpack is runoff capable. REH starts developing in early winter, peaks between March and May and declines in a couple of days or weeks (Figures 2, 3 and 5). First, our work investigates the time and space distribution of REH around the watershed. Second, we identified ‘activations’ as the timing and amount of SWE achieving REH=0, and the spring activation as the greatest SWE activation and its timing for a given location-year. We quantify the character of these activations over the period of record. Finally, we investigate machine learning methods for classifying and predicting the timing of activation events.

*Clusters* – To understand spatio-temporal trends in the behavior of REH and activations, we used the K-Means clustering method on the REH time series due to its ability to utilize the entire time series instead of a single value. We found that 4 clusters was the ideal number to satisfy statistical metrics and physical requirements. During the classification, we tracked each location-year’s properties. Lower clusters have lower REH values and from the lowest to the highest cluster, median peak REH increases from 1.1 to 6.2 MJ/m<sup>2</sup>, the mean elevation increases from 1490 to 2121 masl and higher cluster are more frequent on east and north aspect than on south aspects. While the lower clusters span a greater area (37% of the watershed area for cluster 1 and 5% for cluster 4), they have a lower mean peak SWE (0.44m for cluster 1 and 1.46m for cluster 4), resulting in the greatest volume of water sitting

in clusters 2 and 3 with 1.32 and 1.20km<sup>3</sup> respectively (36 and 33% of the total volume). Higher cluster's REH is also due to a greater fraction of cold content, from 5% to 10% for cluster 1 versus cluster 4.

The clusters are not equally represented through time and the watershed tends towards lower cluster numbers during warmer El Niño years. The mean cluster number for the watershed ranges from 1.08 to 2.97 with a reduction of 0.22 mean cluster number per degree C increase in the January-February-March air temperature at the local SNOTELs. 5 to 92% of the watershed has been clustered in cluster 1 during the 19 years observed, showing the wide extent and the strong inter-annual variability of REH and clustering.

*Activations* – Snowpack activations are most frequent in the fall, before January and due to the low REH and high energy inputs, but because of the small amount of SWE, they are usually insignificant. Mid-winter activations only occur with sufficiently low REH, which is usually seen by the preponderance of cluster 1 in the watershed. Years with significant mid-winter activations are usually followed by increased droughts in the county directly downstream of the study site. The spring activations are identified for clusters 2 through 4 and occur later for higher clusters: early March through early May, early April through mid-May and late April through late May, respectively. Spring activations generally propagate within the watershed from lower elevations to higher elevations at a rate between 21 and 33 m/day and towards north aspects with a more subtle trend. Spring activation timing seems to be unaffected by aspect until approximately 4 April, it then creates a gap between north and southwest facing slopes. A north facing location will experience an identical spring activation timing as a southwest location if it is approximately 400 m lower. Early in the spring, the combination of low solar radiation and weather makes activations aspect independent. Late in the spring, solar radiation becomes strong and south-west aspects at high elevations, where sensible heat remains low, endure significant solar irradiance relative to north aspects. Spring activation and the presence of REH are also bounded in the late spring by the exponential increase in surface energy, mostly due to the exponentially increasing cumulative solar radiation.

*Predictions* – Both for understanding of spatial relationship and for operational prediction purposes, we implemented linear regression, random forest, gradient boosting and fully connected neural networks and the Long Short-Term Memory neural network to predict the spring activation timing from terrain properties and REH values. We concluded that K-Means already extracted the complex interactions between weather and terrain. In other words, no significant additional variability can be accounted for by the listed models. K-Means

is a powerful, intuitive method that can replace complex physical frameworks or location-specific trends. The extremes low snow witnessed over the 19 years studied may represent the future average. While the difference in REH between a high and average snowpack may only lie on a spectrum of similar activation patterns, low REH years exhibit significant mid-winter activations with increased droughts the following summer. During a water year, if the watershed is widely classified in cluster 1, the snowpack has fewer REH and experiences significant mid-winter activations. In a warmer world, the rising temperatures will increase the representation of cluster 1, with the associated consequences, which is consistent with findings of others. The timing of the spring activation being the critical number to address the availability of water resources for dam management and flood forecasting, we compared the prediction abilities of K-Means clustering and of a Long Short-Term Memory (LSTM) neural network on partial REH time series. The LSTM did not outperform simply taking the historical mean of spring activation timings, but K-Means clustering's ability to subsample from past history yielded the best performance, especially for clusters 3 and 4. As a result, spring activation timing predictions can be made with a mean absolute error inferior to 17.0 days, and under 12.3 days for clusters 3 and 4.

## Bibliography

- Anderson, M. R., & Drobot, S. D. (2001). Spatial and Temporal Variability in Snowmelt Onset over Arctic Sea Ice. *Annals of Glaciology*, 33, 74–78. <https://doi.org/10.3189/172756401781818284>
- Arndt, S., & Haas, C. (2019). Spatiotemporal Variability and Decadal Trends of Snowmelt Processes on Antarctic Sea Ice Observed by Satellite Scatterometers. *The Cryosphere*, 13(7), 1943–1958. <https://doi.org/10.5194/tc-13-1943-2019>
- Barnett, T. P., Adam, J. C., & Lettenmaier, D. P. (2005). Potential Impacts of a Warming Climate on Water Availability in Snow-Dominated Regions. *Nature*, 438(7066), 303–309. <https://doi.org/10.1038/nature04141>
- Barrett, A. P. (2003). *NSIDC Special Report #11*.
- Bonnell, R., McGrath, D., Williams, K., Webb, R., Fassnacht, S. R., & Marshall, H.-P. (2021). Spatiotemporal Variations in Liquid Water Content in a Seasonal Snowpack: Implications for Radar Remote Sensing. *Remote Sensing*, 13(21), 4223–4224. <https://doi.org/10.3390/rs13214223>
- Broxton, P. D., Leeuwen, W. J. D., & Biederman, J. A. (2019). Improving Snow Water Equivalent Maps With Machine Learning of Snow Survey and Lidar Measurements. *Water Resources Research*, 55(5), 3739–3757. <https://doi.org/10.1029/2018WR024146>
- Cayan, D. R., Dettinger, M. D., Kammerdiener, S. A., Caprio, J. M., & Peterson, D. H. (2001). Changes in the Onset of Spring in the Western United States. *Bulletin of the American Meteorological Society*, 82(3), 399–415. [https://doi.org/10.1175/1520-0477\(2001\)082<0399:CITOOS>2.3.CO;2](https://doi.org/10.1175/1520-0477(2001)082<0399:CITOOS>2.3.CO;2)
- Clow, D. W. (2010). Changes in the Timing of Snowmelt and Streamflow in Colorado: A Response to Recent Warming. *Journal of Climate*, 23(9), 2293–2306. <https://doi.org/10.1175/2009JCLI2951.1>
- Clow, D. W., Nanus, L., Verdin, K. L., & Schmidt, J. (2012). Evaluation of SNODAS Snow Depth and Snow Water Equivalent Estimates for the Colorado Rocky Mountains, USA: EVALUATION OF SNODAS. *Hydrological Processes*, 26(17), 2583–2591. <https://doi.org/10.1002/hyp.9385>
- Colbeck, S. C. (1974). The Capillary Effects on Water Percolation in Homogeneous Snow. *Journal of Glaciology*, 13(67), 85–97. <https://doi.org/10.3189/S002214300002339X>
- Colbeck, S. C. (1976). An Analysis of Water Flow in Dry Snow. *Water Resources Research*, 12(3), 523–527. <https://doi.org/10.1029/WR012i003p00523>

- Coléou, C., & Lesaffre, B. (1998). Irreducible Water Saturation in Snow: Experimental Results in a Cold Laboratory. *Annals of Glaciology*, 26, 64–68. <https://doi.org/10.3189/1998AoG26-1-64-68>
- Conway, H., & Benedict, R. (1994). Infiltration of Water into Snow. *Water Resources Research*, 30(3), 641–649. <https://doi.org/10.1029/93WR03247>
- Cox, C., Humphrey, N., & Harper, J. (2015). Quantifying Meltwater Refreezing along a Transect of Sites on the Greenland Ice Sheet. *The Cryosphere*, 9(2), 691–701. <https://doi.org/10.5194/tc-9-691-2015>
- Deems, J. S., Birkeland, K. W., & Hansen, K. J. (2002). *Topographic Influence on the Spatial Patterns of Snow Temperature Gradients in a Mountain Snowpack*. 8–9.
- Dewalle, D. R., & Rango, A. (2008). *PRINCIPLES OF SNOW HYDROLOGY*.
- Eriksson, D., Whitson, M., Luce, C. H., Marshall, H. P., Bradford, J., Benner, S. G., Black, T., Hetrick, H., & McNamara, J. P. (2013). An Evaluation of the Hydrologic Relevance of Lateral Flow in Snow at Hillslope and Catchment Scales. *Hydrological Processes*, 27(5), 640–654. <https://doi.org/10.1002/hyp.9666>
- Espinoza, V., Waliser, D. E., Guan, B., Lavers, D. A., & Ralph, F. M. (2018). Global Analysis of Climate Change Projection Effects on Atmospheric Rivers. *Geophysical Research Letters*, 45(9), 4299–4308. <https://doi.org/10.1029/2017GL076968>
- Evan, A., & Eisenman, I. (2021). A Mechanism for Regional Variations in Snowpack Melt under Rising Temperature. *Nature Climate Change*, 11(4), 326–330. <https://doi.org/10.1038/s41558-021-00996-w>
- Evans, S. L., Flores, A. N., Heilig, A., Kohn, M. J., Marshall, H.-P., & McNamara, J. P. (2016). Isotopic Evidence for Lateral Flow and Diffusive Transport, but Not Sublimation, in a Sloped Seasonal Snowpack, Idaho, USA. *Geophysical Research Letters*, 43(7), 3298–3306. <https://doi.org/10.1002/2015GL067605>
- Farmer, C., Nelson, T., Wulder, M., & Derksen, C. (2010). Identification of Snow Cover Regimes through Spatial and Temporal Clustering of Satellite Microwave Brightness Temperatures. *Remote Sensing of Environment*, 114(1), 199–210. <https://doi.org/10.1016/j.rse.2009.09.002>
- Gagliano, E., Shean, D., Henderson, S., & Vanderwilt, S. (2023). Capturing the Onset of Mountain Snowmelt Runoff Using Satellite Synthetic Aperture Radar. *Geophysical Research Letters*, 50(21), e2023GL105303. <https://doi.org/10.1029/2023GL105303>

- Gordon, B. L., Brooks, P. D., Krogh, S. A., Boisrame, G. F. S., Carroll, R. W. H., McNamara, J. P., & Harpold, A. A. (2022). Why Does Snowmelt-Driven Streamflow Response to Warming Vary? A Data-Driven Review and Predictive Framework. *Environmental Research Letters*, 17(5), 53004–53005. <https://doi.org/10.1088/1748-9326/ac64b4>
- Gottlieb, A. R., & Mankin, J. S. (2024). Evidence of Human Influence on Northern Hemisphere Snow Loss. *Nature*, 625(7994), 293–300. <https://doi.org/10.1038/s41586-023-06794-y>
- Hamlet, A. F., Mote, P. W., Clark, M. P., & Lettenmaier, D. P. (2007). Twentieth-Century Trends in Runoff, Evapotranspiration, and Soil Moisture in the Western United States\*. *Journal of Climate*, 20(8), 1468–1486. <https://doi.org/10.1175/JCLI4051.1>
- Harris, C. C., & Morrow, N. R. (1964). Pendular Moisture in Packings of Equal Spheres. *Nature*, 203(4946).
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Hock, R. (1999). A Distributed Temperature-Index Ice- and Snowmelt Model Including Potential Direct Solar Radiation. *Journal of Glaciology*, 45(149), 101–111. <https://doi.org/10.3189/S0022143000003087>
- Humphrey, N. F., Harper, J. T., & Meierbachtol, T. W. (2021). Physical Limits to Meltwater Penetration in Firn. *Journal of Glaciology*, 67(265), 952–960. <https://doi.org/10.1017/jog.2021.44>
- Humphrey, N. F., Harper, J. T., & Pfeffer, W. T. (2012). Thermal Tracking of Meltwater Retention in Greenland's Accumulation Area: THERMAL TRACKING OF MELTWATER RETENTION. *Journal of Geophysical Research: Earth Surface*, 117(F1), n/a–n/a. <https://doi.org/10.1029/2011JF002083>
- IPCC. (2021). Summary for Policymakers. In V. Masson-Delmotte, P. Zhai, A. Pirani, S. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. Matthews, T. Maycock, T. Waterfield, O. Yelekçi, R. Yu, & B. Zhou (Eds.), *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 3–32). Cambridge University Press. <https://doi.org/10.1017/9781009157896.001>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2022). An Introduction to Statistical Learning with Applications in R: By Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani, New York, Springer Science and Business Media, 2013, \$41.98, eISBN:

978-1-4614-7137-7. *Statistical Theory and Related Fields*, 6(1), 87–88. <https://doi.org/10.1080/24754269.2021.1980261>

Jennings, K. S., & Molotch, N. P. (2020). Snowfall Fraction, Cold Content, and Energy Balance Changes Drive Differential Response to Simulated Warming in an Alpine and Subalpine Snowpack. *Frontiers in Earth Science*, 8, 186–187. <https://doi.org/10.3389/feart.2020.00186>

Jennings, K. S., Kittel, T. G. F., & Molotch, N. P. (2018). Observations and Simulations of the Seasonal Evolution of Snowpack Cold Content and Its Relation to Snowmelt and the Snowpack Energy Budget. *The Cryosphere*, 12(5), 1595–1614. <https://doi.org/10.5194/tc-12-1595-2018>

Jordan, R. (1990). *A One-Dimensional Temperature Model for a Snow Cover*.

Kalra, A., Piechota, T. C., Davies, R., & Tootle, G. A. (2008). Changes in U.S. Streamflow and Western U.S. Snowpack. *Journal of Hydrologic Engineering*, 13(3), 156–163. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2008\)13:3\(156\)](https://doi.org/10.1061/(ASCE)1084-0699(2008)13:3(156))

Kellner, F. (2013). *Ensemble Modeling of SWE Distribution in the Bitterroot Mountains, Montana, USA*. 77–78.

Kingma, D. P., & Ba, J. (2017). *Adam: A Method for Stochastic Optimization* (Issue arXiv:1412.6980). arXiv.

Li, D., Lettenmaier, D. P., Margulis, S. A., & Andreadis, K. (2019). The Role of Rain-on-Snow in Flooding Over the Conterminous United States. *Water Resources Research*, 55(11), 8492–8513. <https://doi.org/10.1029/2019WR024950>

Livneh, B., & Badger, A. M. (2020). Drought Less Predictable under Declining Future Snowpack. *Nature Climate Change*, 10(5), 452–458. <https://doi.org/10.1038/s41558-020-0754-8>

Loe, L. E., Hansen, B. B., Stien, A., D. Albon, S., Bischof, R., Carlsson, A., Irvine, R. J., Meland, M., Rivrud, I. M., Ropstad, E., Veiberg, V., & Mysterud, A. (2016). Behavioral Buffering of Extreme Weather Events in a high-Arctic Herbivore. *Ecosphere*, 7(6). <https://doi.org/10.1002/ecs2.1374>

Lund, J., Forster, R. R., Deeb, E. J., Liston, G. E., Skiles, S. M., & Marshall, H.-P. (2022). Interpreting Sentinel-1 SAR Backscatter Signals of Snowpack Surface Melt/Freeze, Warming, and Ripening, through Field Measurements and Physically-Based SnowModel. *Remote Sensing*, 14(16), 4002–4003. <https://doi.org/10.3390/rs14164002>

Lundquist, J. D., & Flint, A. L. (2006). Onset of Snowmelt and Streamflow in 2004 in the Western United States: How Shading May Affect Spring Streamflow Timing in a Warmer

World. *Journal of Hydrometeorology*, 7(6), 1199–1217. <https://doi.org/10.1175/JHM539>.

1

Lv, Z., Pomeroy, J. W., & Fang, X. (2019). Evaluation of SNODAS Snow Water Equivalent in Western Canada and Assimilation Into a Cold Region Hydrological Model. *Water Resources Research*, 55(12), 11166–11187. <https://doi.org/10.1029/2019WR025333>

Marks, D., Kimball, J., Tingey, D., & Link, T. (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. [https://doi.org/10.1002/\(SICI\)1099-1085\(199808/09\)12:10/11<1569::AID-HYP682>3.0.CO;2-L](https://doi.org/10.1002/(SICI)1099-1085(199808/09)12:10/11<1569::AID-HYP682>3.0.CO;2-L)

Marks, D., Link, T., Winstral, A., & Garen, D. (2001). Simulating Snowmelt Processes during Rain-on-Snow over a Semi-Arid Mountain Basin. *Annals of Glaciology*, 32, 195–202. <https://doi.org/10.3189/172756401781819751>

Marsh, P., & Woo, M.-K. (1984a). Wetting Front Advance and Freezing of Meltwater within a Snow Cover: 1. Observations in the Canadian Arctic. *Water Resources Research*, 20(12), 1853–1864. <https://doi.org/10.1029/WR020i012p01853>

Marsh, P., & Woo, M.-K. (1985b). Meltwater Movement in Natural Heterogeneous Snow Covers. *Water Resources Research*, 21(11), 1710–1716. <https://doi.org/10.1029/WR021i011p01710>

McCabe, G. J., Clark, M. P., & Hay, L. E. (2007). Rain-on-Snow Events in the Western United States. *Bulletin of the American Meteorological Society*, 88(3), 319–328. <https://doi.org/10.1175/BAMS-88-3-319>

McDowell, I. E., Keegan, K. M., Wever, N., Osterberg, E. C., Hawley, R. L., & Marshall, H.-P. (2023). Firn Core Evidence of Two-Way Feedback Mechanisms Between Meltwater Infiltration and Firn Microstructure From the Western Percolation Zone of the Greenland Ice Sheet. *Journal of Geophysical Research: Earth Surface*, 128(2), e2022JF006752. <https://doi.org/10.1029/2022JF006752>

Moore, J. N., Harper, J. T., & Greenwood, M. C. (2007). Significance of Trends toward Earlier Snowmelt Runoff, Columbia and Missouri Basin Headwaters, Western United States. *Geophysical Research Letters*, 34(16), 2007–2008. <https://doi.org/10.1029/2007GL031022>

Musselman, K. N., Addor, N., Vano, J. A., & Molotch, N. P. (2021). Winter Melt Trends Portend Widespread Declines in Snow Water Resources. *Nature Climate Change*, 11(5), 418–424. <https://doi.org/10.1038/s41558-021-01014-9>

- Musselman, K. N., Clark, M. P., Liu, C., Ikeda, K., & Rasmussen, R. (2017). Slower Snowmelt in a Warmer World. *Nature Climate Change*, 7(3), 214–219. <https://doi.org/10.1038/nclimate3225>
- Musselman, K. N., Lehner, F., Ikeda, K., Clark, M. P., Prein, A. F., Liu, C., Barlage, M., & Rasmussen, R. (2018). Projected Increases and Shifts in Rain-on-Snow Flood Risk over Western North America. *Nature Climate Change*, 8(9), 808–812. <https://doi.org/10.1038/s41558-018-0236-4>
- NSIDC. (2004). *Snow Data Assimilation System (SNODAS) Data Products at NSIDC, Version 1*. <https://doi.org/10.7265/N5TB14TC>
- Parajuli, A., Nadeau, D. F., Anctil, F., & Alves, M. (2021). Multilayer Observation and Estimation of the Snowpack Cold Content in a Humid Boreal Coniferous Forest of Eastern Canada. *The Cryosphere*, 15(12), 5371–5386. <https://doi.org/10.5194/tc-15-5371-2021>
- Payne, A. E., Demory, M.-E., Leung, L. R., Ramos, A. M., Shields, C. A., Rutz, J. J., Siler, N., Villarini, G., Hall, A., & Ralph, F. M. (2020). Responses and Impacts of Atmospheric Rivers to Climate Change. *Nature Reviews Earth & Environment*, 1(3), 143–157. <https://doi.org/10.1038/s43017-020-0030-5>
- 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. (2011). Scikit-Learn: Machine Learning in Python. *MACHINE LEARNING IN PYTHON*.
- Rauscher, S. A., Pal, J. S., Diffenbaugh, N. S., & Benedetti, M. M. (2008). Future Changes in Snowmelt-Driven Runoff Timing over the Western US. *Geophysical Research Letters*, 35(16), L16703. <https://doi.org/10.1029/2008GL034424>
- Regonda, S. K., Rajagopalan, B., Clark, M., & Pitlick, J. (2005). Seasonal Cycle Shifts in Hydroclimatology over the Western United States. *Journal of Climate*, 18(2), 372–384. <https://doi.org/10.1175/JCLI-3272.1>
- Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C., & Wang, W. (2002). An Improved In Situ and Satellite SST Analysis for Climate. *Journal of Climate*, 15(13), 1609–1625. [https://doi.org/10.1175/1520-0442\(2002\)015<1609:AIISAS>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<1609:AIISAS>2.0.CO;2)
- Schirmer, M., Winstral, A., Jonas, T., Burlando, P., & Peleg, N. (2022). Natural Climate Variability Is an Important Aspect of Future Projections of Snow Water Resources and Rain-on-Snow Events. *The Cryosphere*, 16(9), 3469–3488. <https://doi.org/10.5194/tc-16-3469-2022>
- Schmidt, J. (2017). *CU Boulder Geography Undergraduate Honors Thesis*.

- Scholten, R. C., Coumou, D., Luo, F., & Veraverbeke, S. (2022). Early Snowmelt and Polar Jet Dynamics Co-Influence Recent Extreme Siberian Fire Seasons. *Science*, 378(6623), 1005–1009. <https://doi.org/10.1126/science.abn4419>
- Seligman, Z. M., Harper, J. T., & Maneta, M. P. (2014). Changes to Snowpack Energy State from Spring Storm Events, Columbia River Headwaters, Montana. *Journal of Hydrometeorology*, 15(1), 159–170. <https://doi.org/10.1175/JHM-D-12-078.1>
- Shapley, L. S. (1953). 17. A Value for n-Person Games. In H. W. Kuhn & A. W. Tucker (Eds.), *Contributions to the Theory of Games (AM-28), Volume II* (pp. 307–318). Princeton University Press. <https://doi.org/10.1515/9781400881970-018>
- Singh, P., Spitzbart, G., Hübl, H., & Weinmeister, H. (1997). Hydrological Response of Snowpack under Rain-on-Snow Events: A Field Study. *Journal of Hydrology*, 202(1–4), 1–20. [https://doi.org/10.1016/S0022-1694\(97\)00004-8](https://doi.org/10.1016/S0022-1694(97)00004-8)
- Solgaard, A. M., Rapp, D., Noël, B. P. Y., & Hvidberg, C. S. (2022). Seasonal Patterns of Greenland Ice Velocity From Sentinel-1 SAR Data Linked to Runoff. *Geophysical Research Letters*, 49(24). <https://doi.org/10.1029/2022GL100343>
- Stewart, I. T. (2009). Changes in Snowpack and Snowmelt Runoff for Key Mountain Regions. *Hydrological Processes*, 23(1), 78–94. <https://doi.org/10.1002/hyp.7128>
- Stewart, I. T., Cayan, D. R., & Dettinger, M. D. (2004). Changes in Snowmelt Runoff Timing in Western North America under a ‘Business as Usual’ Climate Change Scenario. *Climatic Change*, 62(1–3), 217–232. <https://doi.org/10.1023/B:CLIM.0000013702.22656.e8>
- Vandecrux, B., Fausto, R. S., van As, D., Colgan, W., Langen, P. L., Haubner, K., Ingeman-Nielsen, T., Heilig, A., Stevens, C. M., MacFerrin, M., Niwano, M., Steffen, K., & Box, J. (2020). Firn Cold Content Evolution at Nine Sites on the Greenland Ice Sheet between 1998 and 2017. *Journal of Glaciology*, 66(258), 591–602. <https://doi.org/10.1017/jog.2020.30>
- Waldner, P. A., Schneebeli, M., Schultze-Zimmermann, U., & Flühler, H. (2004). Effect of Snow Structure on Water Flow and Solute Transport. *Hydrological Processes*, 18(7), 1271–1290. <https://doi.org/10.1002/hyp.1401>
- Warren Liao, T. (2005). Clustering of Time Series Data—a Survey. *Pattern Recognition*, 38(11), 1857–1874. <https://doi.org/10.1016/j.patcog.2005.01.025>
- Zheng, L., Cheng, X., Chen, Z., Wang, S., Liang, Q., & Wang, K. (2022). Global Snowmelt Onset Reflects Climate Variability: Insights from Spaceborne Radiometer Observations. *Journal of Climate*, 35(10), 2945–2959. <https://doi.org/10.1175/JCLI-D-21-0265.1>