SNOTEL Representativeness and Post-Fire Forecast Bias: Evaluating M^4 Streamflow Predictions in the McKenzie River Basin

by

Reilly Seeley

A Project submitted to Dr. Jamon Van Den Hoek

GEOG 581: Satellite Image Analysis

In partial fulfillment of

the requirements for the

degree of

Master of Science in Geography and Geospatial Science

Presented March, 2024

Abstract

A warming climate highlights the necessity for better hazard mitigation strategies and regional water supply forecasts, particularly in the mountainous western U.S., where snow plays a pivotal role as a freshwater reservoir in the local hydrologic cycle. The multi-model machine-learning metasystem (M^4) was developed by the U.S. Department of Agriculture Natural Resources Conservation Service (NRCS), marking a significant advancement over traditional water supply forecasting methods. By employing a multi-model ensemble and algorithmic logic approach, the M^4 model demonstrates enhanced forecast skill, robustness, and interpretability, addressing challenges of equifinality and compensating for variability in snowpack data. This breakthrough reflects the potential of machine learning technologies to revolutionize water supply forecasting across diverse geophysical environments, offering a more realistic, automated, and physically informed prediction framework. Using the 2003 B&B Complex Fire in Oregon as a natural experiment, this study blends satellite-derived burn-severity metrics with the ensemble-learning model M^4 to test how wildfire-disturbed snow water equivalent (SWE) records reshape seasonal water-supply forecasts. Remote-sensing evidence of canopy loss and high burn severity corresponded to persistent, post-fire underprediction of April–July streamflow, demonstrating how burned SNOTEL sites can mislead M^4 model predictions when basin conditions shift. The findings highlight the need for multi-site SWE inputs and post-disturbance recalibration, offering a roadmap for future M^4 applications that better account for wildfire-driven changes in snowpack, melt timing, and downstream hydrology.

Introduction

In the western United States, the precision of water supply forecasting relies heavily on an accurate assessment of snow water equivalent (SWE). Historically, traditional forecasting methods have been driven by statistical analyses of sparse SWE data from snow pillow sites (Deschamps-Berger et al., 2023). A snow pillow is a large membrane filled with a liquid antifreeze solution to prevent freezing and to ensure that the pressure readings are not affected by ice formation. Snow accumulates on top of the pillow, where the weight of snow exerts pressure directly proportional to the weight of the snow and allows the system to calculate how much water would produce the same amount of pressure over the known area of the pillow. Accurately representing the spatial variability of snow over large areas poses significant challenges due to factors like topography, vegetation cover, and microclimatic conditions - all of which interact in complex ways.

Snow depth also varies at microscale levels due to topographic effects and wind redistribution. Snowpack evolves continuously through accumulation, compaction, melt, refreezing, and wind redistribution, so a sparse network of point-scale SWE observations offers only a momentary view of basin-wide water storage. Additional process choices - how to parameterize saturation-excess runoff (overland flow that begins once the whole soil column is saturated) or macropore flow (rapid transport via root channels and fractures), how to let bedrock topography and permeability shape subsurface storage, how much detail is needed to model the penetration and scattering of short-wave solar radiation through a patchy forest canopy, and whether to resolve spatially variable flow paths explicitly - all add complex structural decisions that can quickly overwhelm an otherwise physics-based modeling framework (Clark et al., 2011). These compounded uncertainties propagate into stream-flow forecasts and seasonal water-supply estimates, complicating reservoir operations, allocation planning, and flood-mitigation strategies in water-resource management.

The absence of specific models dedicated to representing the effects of fires and forest recovery on eco-hydrologic processes hinders the precise understanding of how fires impact water supply forecasting. The abundance of wildfire and SNOTEL data on the west coast provides promising research opportunities to enhance this understanding through a machine

learning lens to help researchers better understand the interactive factors shaping snow dynamics post-wildfire (Barnard et al., 2023).

This study investigates the impacts of wildfire on snow dynamics through a machine learning lens. To directly address the research question of SNOTEL representativeness, this analysis will evaluate how well the SWE measurements from the Hogg Pass SNOTEL site capture the broader basin-wide snow water contributions that influence streamflow at the McKenzie River outlet. By comparing model predictions based on site-specific data with observed streamflow patterns, particularly in the context of wildfire-affected conditions, this study aims to clarify whether single-point SNOTEL measurements can reliably stand in for the complex, spatially variable snow processes occurring across a watershed. This activity aligns with Bureau of Reclamation funding guidelines by: (1) Demonstrating emerging snow monitoring technologies; (2) Showcasing improvements to existing snow monitoring approaches; and (3) Enhancing the use of snow monitoring data to improve water supply forecasting skill. As climate uncertainty increases the likelihood of evolving trends in wildfire severity and extent, robust model predictions can inform stakeholder decisions regarding optimal snow monitoring strategies. Developing an effective approach for studying wildfire effects on model predictions requires a thorough understanding of initial results related to the representation of observed streamflow changes in training data, which will be examined in this analysis.

Literature Review

M^4 Machine Learning Framework for Water Supply Forecasting

The volume of snow-water storage and the timing of snowmelt is essential for controlling downstream water resources, where headwater snow storage depends on a variety of factors to keep the accumulated snowpack reservoir cold into spring (Smoot & Gleeson, 2021). River forecast centers have traditionally produced water supply forecasts (WSFs) using statistical techniques and physics-based models reliant on limited snow-pillow SWE data. However, advanced machine learning (ML) algorithms have shown promise in capturing the variability of snowpack from sparse historical records. The NRCS recently introduced M^4, an ensemble

artificial intelligence prototype that improved the coefficient of determination (R^2) and Ranked Probability Skill Score (a metric that measures the accuracy of probabilistic forecasts) by over 50% compared to previous WSF benchmarks (Fleming et al., 2021). Designed to meet practical criteria, the M^4 model delivers enhanced forecast skill, more realistic prediction intervals, robustness, automation, and interpretability grounded in physical hydrologic processes, making it applicable across a range of geophysical environments. By employing a multi-model ensemble approach, the system effectively addresses the problem of equifinality, the possibility of different model structures or parameter sets producing similar outcomes by compensating for poor performance in some models with stronger results from others. Additionally, automated algorithmic logic fine-tunes the system's parameters to produce the most reliable forecasts and a range of possible outcomes. Although M^4 core predictors are snow-water equivalent and accumulated precipitation, it can also ingest ancillary hydrologic and management variables, such as antecedent streamflow, soil-moisture indices, reservoir storage, climate teleconnections, and evaporative demand (Fleming et al., 2021) to sharpen and contextualize its seasonal water-supply forecasts.

Coupled with the complex, nonlinear dynamics of climate, topography, and land cover, the 'black box' nature of an AI model (its internal workings being opaque and difficult to interpret) make it challenging to understand exactly how inputs translate into outputs. Thorough study is needed to determine how this new snow monitoring technology can improve water supply forecasts. Ensemble forecasting techniques continue to be developed to combine multiple models for more robust and reliable predictions (Fleming et al., 2023). Furthermore, studies are evaluating the economic impacts of improved forecasting, emphasizing the value of accurate predictions to sectors such as hydroelectric power generation and agriculture. Through these research endeavors, hydrologists aim to enhance water supply forecasting for better water resource management in the face of changing environmental conditions.

Wildfire Impact on Water Supply Forecasting in the Western United States

Wildfire activity in the western United States expanded by more than 1,100% between 1984 and 2020 (Williams et al., 2022), raising new concerns about the reliability of snow-based water-supply forecasts. The M^4 model, whose primary predictors are snow-water equivalent (SWE) and precipitation, is typically calibrated on records that already embed topographic

influences like fire disturbances to topography. Whether those calibrations remain valid once a basin burns is largely untested, creating a critical knowledge gap in operational forecasting.

Forested catchments exhibit steep snowpack gradients because canopy structure, interception efficiency, wind redistribution, understory shading, and topographic shading all contribute to snow accumulation and melt. Paired forested and open plots in Oregon's Cascade Range illustrate this sensitivity: mean canopy-interception efficiencies reached 79% and 76% of event snowfall at low- and mid-elevation sites but only 31% at a sparser high-elevation site. The energy balance also shifts with canopy density - long-wave radiation provided 93%, 92%, and 47% of total snowpack energy at those same sites, showing how dense canopy cover both shields snow from short-wave input and enhances long-wave loading (MacDonald & Putz, 2011). Disturbances like wildfire add another layer of complexity. Combustion can make soil water-repellent, reducing its capacity to absorb water (soil hydrophobicity). This reduction in infiltration capacity may result in increased surface runoff from snowmelt and elevate the risk of soil erosion (MacDonald & Putz, 2011). Charred needles and ash deposited on the snowpack also lower its albedo (surface reflectance), allowing the snow to absorb more solar energy and melt faster. Subtle variations in canopy density or structure can also alter the amount and timing of snow that reaches and persists on the forest floor, with measurable impacts on both snowpack longevity and the timing of meltwater release. Recognizing and representing these post-fire changes is therefore critical for reliable snowmelt forecasts and the water-resource decisions that depend on them.

Empirical evidence confirms that these disturbances matter. An analysis of 72 forested basins across the West showed that mean streamflow rose above climate-based expectations for roughly six years after fire, with no dependence on basin size, elevation, slope, aspect, precipitation, or forest cover (Williams et al., 2022). Likewise, a study of 78 SNOTEL stations located in burned areas found an average 72 mm decline in peak SWE and a five-day advance in peak-SWE timing during the decade following fire (Smoot & Gleeson, 2021). These findings imply that wildfire signals can be extracted from streamflow and SWE without heavy reliance on topographic corrections, presenting an opportunity to integrate burn metrics directly into machine-learning frameworks. To investigate this, the present study analyzes M^4 predictions made with pre-burn training data and post-burn testing data to determine whether omitting fire

extent and severity systematically degrades forecast skill. By observing the degree to which a single, fire-affected SNOTEL record misrepresents basin-wide SWE, we evaluate the need for additional predictors, such as high-resolution burn-severity maps, to maintain model accuracy as wildfires grow in size and intensity.

Philosophical Foundations of Machine Learning in Scientific Modeling

In an era where machine learning and complex models shape decision-making across scientific domains, the philosophical stance of scientific realism is particularly significant. As technological methods become more intricate, questions about how well these models correspond to actual phenomena grow increasingly relevant. This concern motivates not only clarity in model interpretation, but also broader philosophical debates about the nature of truth and representation in science. The interpretability challenges posed by the 'black box' nature of machine learning models are significant; their inherent complexity often makes it difficult to trace the reasoning behind specific predictions and decisions. Philosophers of science have long debated the criteria for realism in scientific modeling. Some posit that aspects of a model should correspond to genuine features of reality, thus offering direct insight into the world. Others suggest that realism may reside in the structural or relational components of scientific theories, advocating that *even as theories evolve, their referencing capacity toward real-world structures or entities sustains their realist character* (Frigg & Hartmann, 2020). Such arguments gain traction in applied research fields, like snow hydrology, where extensive data analyses within sophisticated modeling systems often generate new lines of inquiry. The M^4 metasystem exemplifies these philosophical concerns by integrating advanced techniques such as feature extraction, nonlinear regression, and ensemble modeling to analyze interactions between snow and topography (Fleming & Goodbody, 2019). While skepticism towards innovations like M^4 is acknowledged due to its black box nature, successful applications and the persistent need for more accurate water supply forecasting highlight a need for further study. Thus, technical considerations in M^4 decision-making procedures are deeply intertwined with discussions of scientific and structural realism. Embracing scientific realism within this project underscores the commitment to ensuring that M^4 modeled outputs are not only predictive, but also meaningfully anchored to the underlying hydrologic realities shaping snow, wildfire, and streamflow interactions in the study basin. This philosophical stance extends beyond technical rigor - it reinforces the importance of communicating results transparently and effectively to both

scientific and public audiences. By grounding findings in observable and interpretable processes, the research aims to enhance public trust in scientific outcomes and support the broader sharing of reliable information essential for informed water resource management decisions.

Research Question - *To what extent does the Hogg Pass SNOTEL record represent watershed-scale snow-water equivalent and maintain its predictive link to April–July streamflow at the McKenzie River outlet after a wildfire disturbance?*

Hypothesis - SWE measurements from the Hogg Pass SNOTEL site will misrepresent basin-wide snow conditions, resulting in inaccurate predictions of McKenzie River streamflow from the pre-fire trained M^4 model.

Study Area

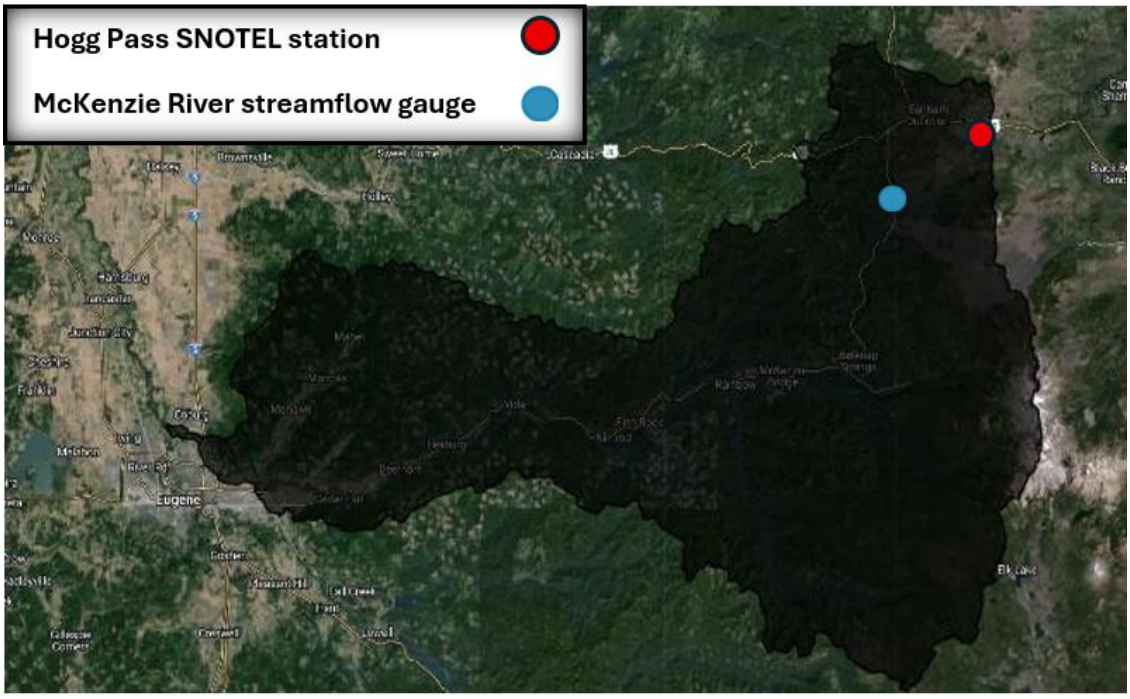

Fig. 1 *McKenzie River subbasin (HUC-8: 17090004)*

The Hogg Pass SNOTEL site was the ideal candidate for this case study due to its proximity from Oregon State University and site familiarity. It was engulfed by the B&B

Complex Fire in August 2003. The Air & Water Database Report Generator from the NRCS contains filters to identify daily start-of-day SWE values and water year (10/1 - 9/31 the following year) precipitation accumulation in inches (NRCS, 2024). The NRCS database contains information from 1980 to present day, allowing for both an ample training period for M^4 (1980 - 2002) and comparative output analysis (2003 - 2023). The Data Basin mapping and analysis platform from the United States Geological Survey (USGS) was used to identify the McKenzie River at the Outlet of Clear Lake streamflow gauge, which is located inside the of the watershed containing the Hogg Pass SNOTEL and is not blocked by any dams that might further influence streamflow values (USGS Data Basin, 2024). This was confirmed using the Global Watersheds App (MGHydro, 2024).

Methodology

M^4 Training and Application

The USGS National Water Information System (NWIS) contains monthly statistics for USGS streamflow gauges. NWIS reports discharge in cubic feet per second, so each daily mean within that window was multiplied by the number of seconds in the day and summed to yield a single April-July volume (ft^3). Volumes from 1980 to the present form the predictand used to train and validate the model. Coinciding with higher water demand and flood-control decisions, March 1st was chosen as a forecast 'issue' date. Expressing the objective as a four-month volume rather than instantaneous discharge aligned predictions directly with contract accounting, hydropower scheduling, and regulated flow objectives (U.S. Bureau of Reclamation, 2009; USACE et al., 2024).The modelling task was defined as predicting the total flow that would pass the gage between April 1st and July 31st of the same water year using March 1st SWE and cumulative water-year precipitation data.

M^4 was implemented in RStudio through three straightforward steps that followed the agency user manual. First, the run-control file "*MMPE_RunControlFile.txt*" was opened and the flag "*RunTypeFlag = BUILD*" was specified, leaving default parameters unchanged (see **Disclaimer**). A tab-delimited file named "*MMPEInputData_ModelBuildingMode.txt*" was then prepared: the first column listed the water year, the second the April-July volume just calculated,

and subsequent columns the chosen predictors (March 1st SWE and cumulative water-year precipitation). Running the source code started an automated process that searches for the best combination of predictor variables. This process creates an ensemble of six calibrated sub-models and saves them in the working directory. On a modern laptop, building a model for a single site only takes a few minutes.

For the operational forecast, the user prepares "*MMPEInputData_ForecastingMode.txt*", a file that contained current-year predictor values in the identical order used during model training (see **Disclaimer**). M^4 must be informed exactly which predictors each sub-model should keep. The "BUILD" step saves this information in small summary files called bit strings that specify whether a predictor was selected (value of 1) or dropped (value of 0). The bit strings from the "*BUILD*" step summary file are used to match the "*VariableSelection_Frwrd_Model1*" line in the run-control file to ensure the "*FORECAST*" run applies exactly the predictor sets that proved optimal during model training. This run writes a result table in CSV format for interpretation. Within this CSV is the ensemble-mean seasonal flow, the average of the six model predictions, as well as the 10%, 30%, 70%, and 90% exceedance values (the chance the true flow is higher).

Satellite Image Analysis

A spatial analysis of the Hogg Pass SNOTEL site, McKenzie River streamflow gauge, watershed area, burn extent, and burn severity, were interpreted through Google Earth Engine (GEE):

(1) NDVI Change

After locating the Hogg Pass SNOTEL site and McKenzie River streamflow gauge, the watershed was loaded in using its known ID from the GEE Feature Collection 'HUC08: USGS Watershed Boundary Dataset of Subbasins'. A circular buffer was manually established around the SNOTEL site in an attempt to more accurately represent the area contributing snowmelt to the measured discharge. This buffer had a 1400 meter radius around the SNOTEL site and was clipped to include only the area in the watershed. The Landsat 5 satellite image collection was then uploaded to calculate the mean Normalized Difference Vegetation Index (NDVI) change 6 years before and after the fire occurred. NDVI is a widely used remote sensing metric that

quantifies vegetation health and density based on the difference between near-infrared wavelength electromagnetic radiation (which vegetation strongly reflects) and red light (which vegetation absorbs). Tracking NDVI is significant as it offers insight into post-fire ecosystem recovery, land cover changes, and the impact of fire on vegetation (Pettorelli, 2013). Functions to add NDVI bands to the image and mask cloud cover were also integrated into this code. The difference between mean NDVI in 2002 and 1997, as well as between 2009 and 2003, was calculated and mapped - where a red-color indicated a decrease in NDVI and a blue color indicated an increase. This process was repeated for 2010 and 2016, utilizing available Landsat 7 satellite imagery instead.

(2) FIRMS Burn Severity

In a separate script, active-fire detections were retrieved from NASA's Fire Information for Resource Management System (FIRMS) database. Of note, burn severity is not consistently standardized and varied by study, so 2 metrics were explored in this study. The FIRMS database contained brightness temperature ($T_{(\beta)}$) information using Moderate Resolution Imaging Spectroradiometer (MODIS) data channels representing the centroid of a 1-kilometer pixel at near real-time of active fire (NASA LANCE FIRMS, 2024). Brightness temperature ($T_{(\beta)}$) is the effective temperature a perfect blackbody (a theoretical object that absorbs all incoming energy and emits radiation at the maximum possible efficiency) would need to emit to match the observed thermal radiance, making it a convenient proxy for surface fire intensity (Schroeder et al., 2014). Although brightness temperature is not a universally standardized metric, higher post-fire temperatures have been linked to earlier peak SWE dates, accelerated melt, and shorter snow-cover duration (Smoot & Gleeson, 2021).

Because the hydrological model is calibrated with data through 2003, the earliest FIRMS records (November 2000 - February 2003) were first queried to verify that no large active-fire pixels intersected the basin during the training period. To visualize post-training burns, a burn-perimeter shapefile from the Monitoring Trends in Burn Severity (MTBS) program was uploaded and overlaid so that the maximum $T_{(\beta)}$ of every MODIS cell lying within the study area perimeter could be recorded. To summarize fire intensity, pixels hotter than 325K were averaged to obtain a basin-wide mean, and a second count was made of all pixels exceeding 300K. The latter threshold is consistent with the Collection-6 MODIS active-fire algorithm, which

dynamically sets its potential-fire test but constrains it to 300K - 330K (Giglio et al., 2016). These temperature-based metrics provided a physically grounded indicator of the parts of the watershed most likely to influence snowmelt timing and downstream streamflow.

(3) CBI Random Forest Burn Severity classification

In a third script, the Landsat-based Composite Burn Index (CBI) Random-Forest model developed by Parks et al. (2019) to map burn severity at 30 m resolution was loaded in, providing a sharper comparison to the coarser FIRMS brightness-temperature proxy. The CBI is a field-based metric quantifying fire severity, ranging from 0 to 3 with higher values representing more substantial ecological change. It incorporates observations of vegetation and canopy changes at multiple strata. Field-based CBI measurements serve as training data for a Random Forest regression model, which is an ensemble machine learning technique that builds multiple decision trees and aggregates their results for more accurate predictions. This approach supports the prediction of CBI values based on a suite of spectral, geographic, and climatic variables, enabling the spatial mapping of fire severity across extensive geographic regions.

In addition to producing a higher resolution dataset than FIRMS, the model corrects for bias to mitigate overprediction of burn severity at low CBI values and underprediction at high CBI values. Raw predictions are shifted the 1:1 line of a bias-correction curve embedded in the model, eliminating the systematic over-estimation of lightly burned plots and under-estimation of severely burned plots (Parks et al., 2019).

The B&B Complex perimeter used in the previous script was uploaded as a GEE asset with its attributes (fire ID, name, year, start and end Julian days) formatted to match model requirements. Running the model produced rasters of raw CBI values, bias-corrected CBI values, and relative burn ratio (RBR, not used in this analysis) values. Threshold functions (e.g., CBI $\geq$ 1.99) were then applied to estimate the proportion of high burn-severity pixels in the study area for further analysis.

Results and Analysis

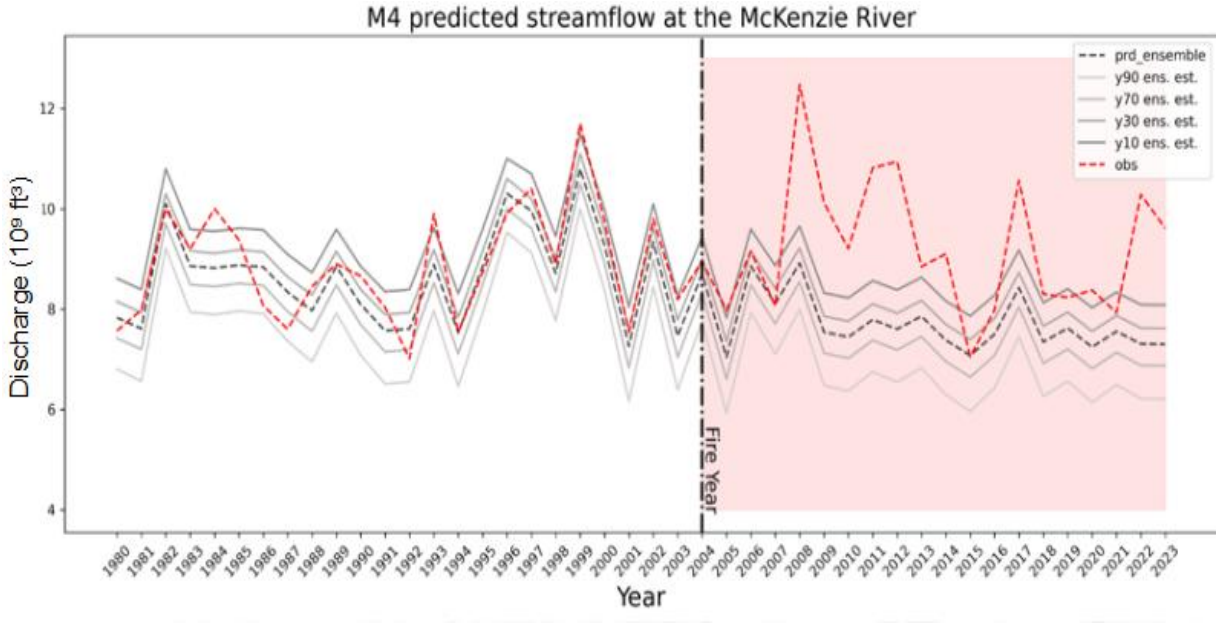

Fig. 2 *Average April-July cumulative streamflow, predicted by six M4 models each year at four confidence intervals, compared to observed flows at McKenzie River Outlet. Red shading marks the transition from training to forecast period.*

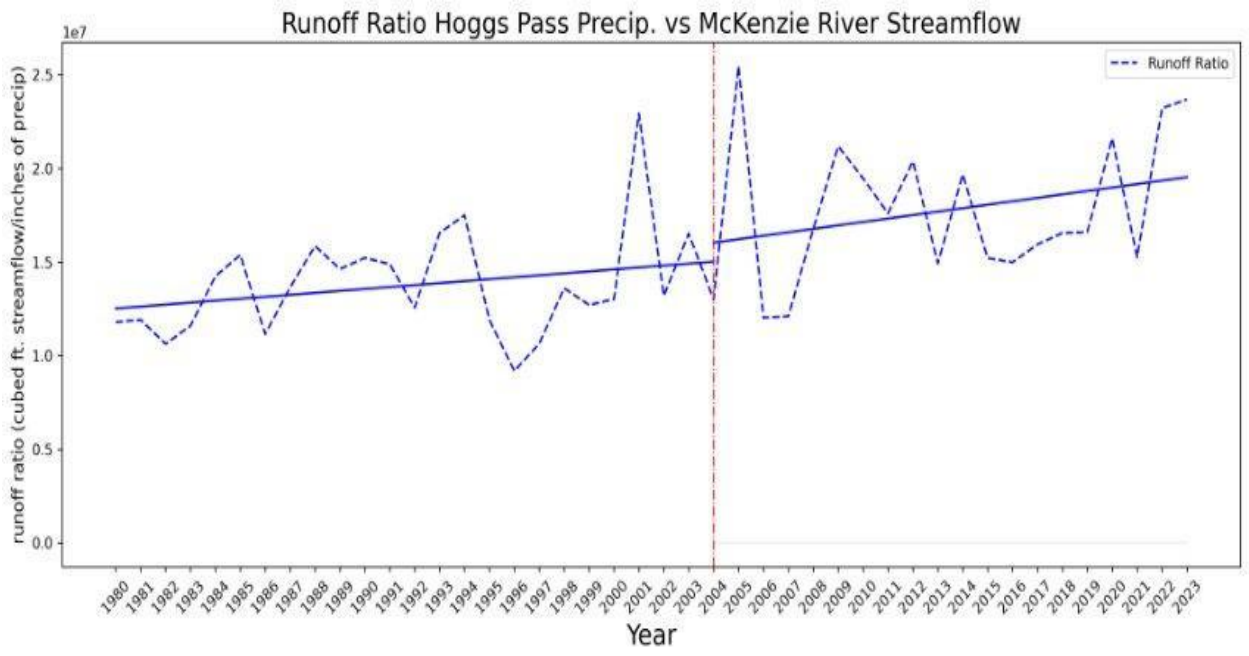

Fig. 3 *McKenzie River seasonal streamflow volume divides by Hogg Pass seasonal precipitation (Runoff Ratio) with a solid line representing the linear trend through the time series before and after the B&B Complex Fire (red vertical line).*

The satellite image analysis provided much needed context for the M^4 predictions. **Fig. 2** revealed a persistent negative bias in the M^4 forecasts beginning in water-year 2006 (roughly three spring runoff seasons after the 2003 B&B Complex Fire occurred). The ensemble forecasts climb after 2003, but observed April–July flows do not rise in step, opening a persistent gap in which predictions run high by roughly 0.5–1.5 × 10^8 ft^3 most years. Every March-issue prediction fell short of the April-July volume measured at the McKenzie River outlet, supporting the hypothesis that the Hogg Pass SNOTEL data would misrepresent McKenzie River Outlet streamflow after the B&B Complex wildfire. When this underprediction is viewed alongside **Fig. 3**, where both the runoff ratio and SNOTEL-measured SWE rise sharply after the burn, it is evident that spring flows were not accurately predicted by M^4 based on the pre-fire relationships on which the model was trained (1980-2002).

Between 1997 and 2002, mean NDVI rose by ~0.024, whereas between 2003 and 2009 it fell by only ~0.0003. On an index scale that spans -1 to +1, a ±0.05 shift is typically the minimum considered ecologically meaningful - values smaller than that often fall within normal phenological variability or sensor noise (Pettorelli, 2013). Both the +0.024 pre-fire NDVI increase and the post-fire NDVI decrease are minor, and the accompanying histograms' tight clustering around zero confirms that broad-scale greenness hardly changed in the study area (see **Fig. 4 and Fig. 5**). Landsat 7 images from 2010-2016 also show a small mean decline of ~0.014 and a wider spread of NDVI change in the accompanying histogram, but interpretation is limited because the data are marred by scan-line gaps from the 2003 Scan Line Corrector (SLC) failure. Since these NDVI shifts fall well below thresholds that flag substantial canopy loss, the analysis focused more on burn-severity metrics better suited to capture post-fire impacts.

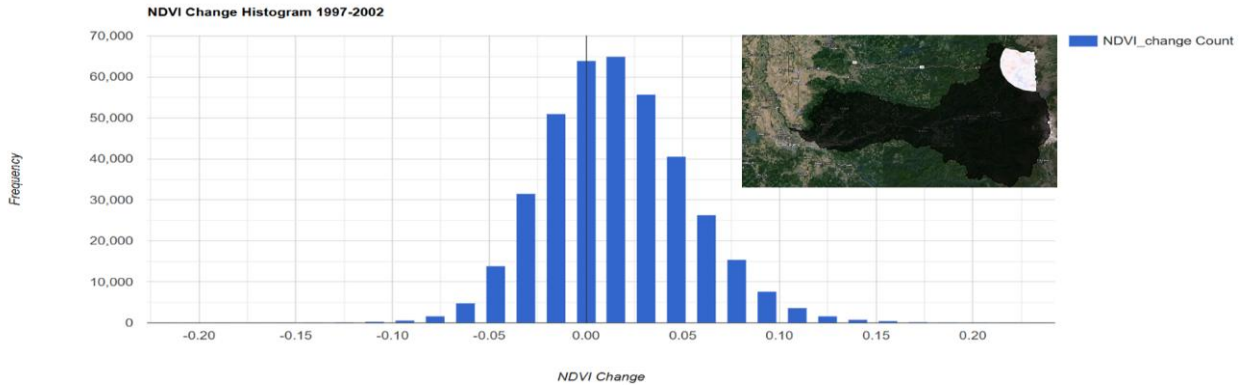

Fig. 4 *Histogram of land NDVI change from 1997 to 2002 over the region of interest, where a red color indicates a negative change and a blue color indicates a positive change.*

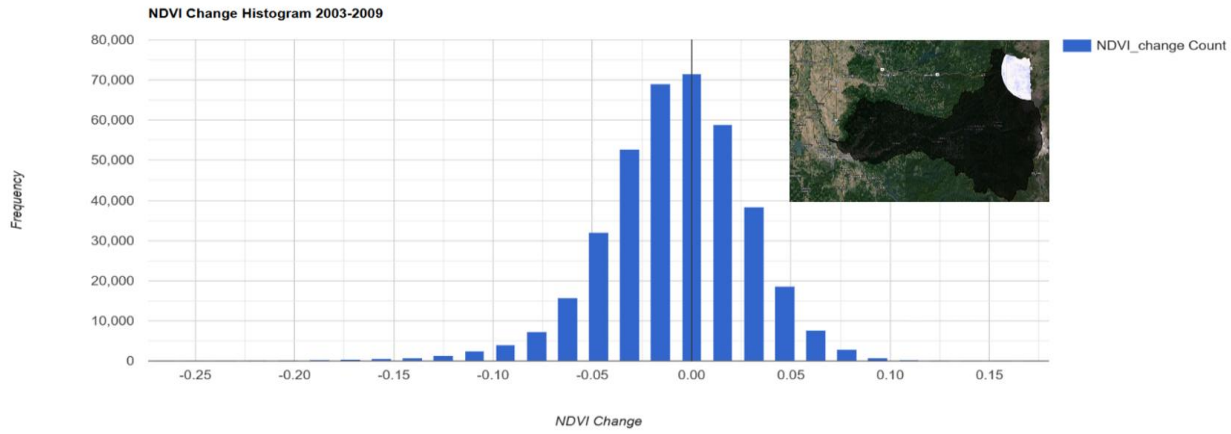

Fig. 5 *Histogram of land NDVI change from 1997 to 2002 over the region of interest, where a red color indicates a negative change and a blue color indicates a positive change.*

The coarse 1-km footprints from the FIRMS database blur out the exact outline of the B&B Complex Fire (see **Fig. 6**) as more fire detections straddled or spilt beyond the MTBS perimeter than initially expected. Still, the histogram of maximum $T_{(\beta)}$ from the buffered watershed exhibited a right-skew (see **Fig. 7**), with a distinct shoulder peaking around 375K. Setting a threshold at 300K cleanly separated that shoulder from the cooler majority, flagging pixels most likely to have experienced intense combustion while avoiding the need for another spatial cutoff. Under this threshold, 39.9% of the basin was classified under 'high-severity' burn, nearly double the ~20% burned-area fraction that Parks et al., (2019) linked to detectable streamflow increases. However, the low spatial resolution of this data relative to the study area indicated that this burn-extent percentage is inflated. To reduce this uncertainty, the finer-resolution CBI model was used to provide a clearer image of post-fire severity patterns.

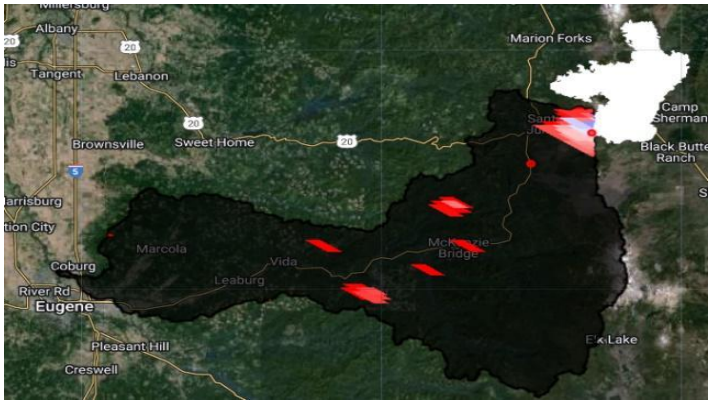

Fig. 6 *FIRMS maximum brightness temperature ($T_{(\beta)}$) from the B&B Complex fire, clipped to the study area watershed. The white polygon represents the actual fire extent.*

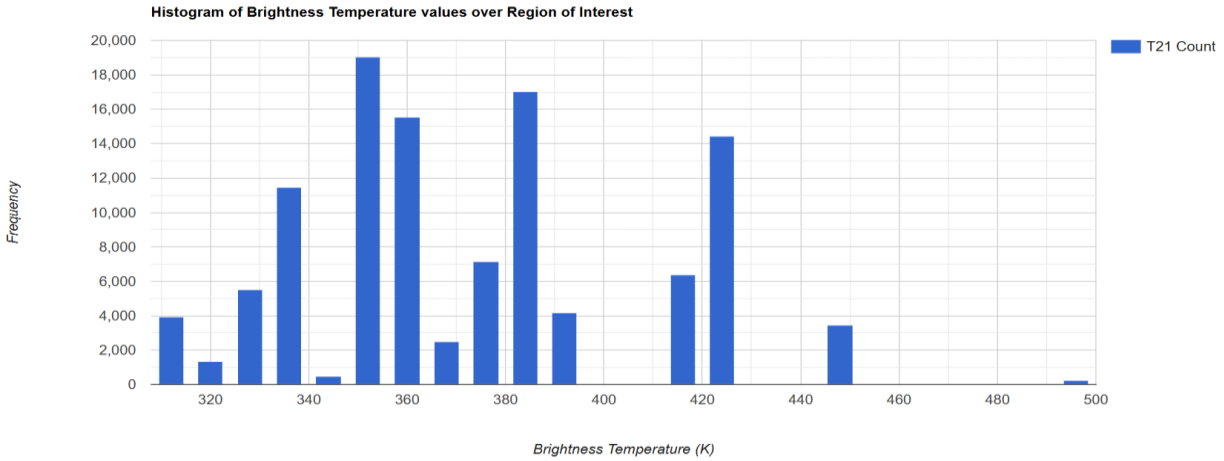

Fig. 7 *Histogram of maximum brightness temperature ($T_{(\beta)}$) values in the study area shown in* ***Fig. 6***

The Landsat-based CBI layer proved to be a more robust and meaningful way to understand and quantify burn severity over the study area. For the pixels within the region of interest, the mean bias-corrected CBI value was ~1.83 (averaging in unburned pockets inside the fire perimeter coded CBI = 0 but ignoring pixels outside the perimeter where CBI is undefined). Converting this continuous surface to categorized classes of burn severity was inherently subjective, yet commonly used thresholds (Parks et al., 2019) of CBI ≥ 0.99 (moderate-to-high) and CBI ≥ 1.99 (high, see **Fig. 8**) flag only about 9.6 % and 6.7 % of pixels, respectively. Those fractions were an order of magnitude smaller than the ~40% pixel share implied by the FIRMS data, confirming that MODIS resolution greatly exaggerated the burn extent footprint.

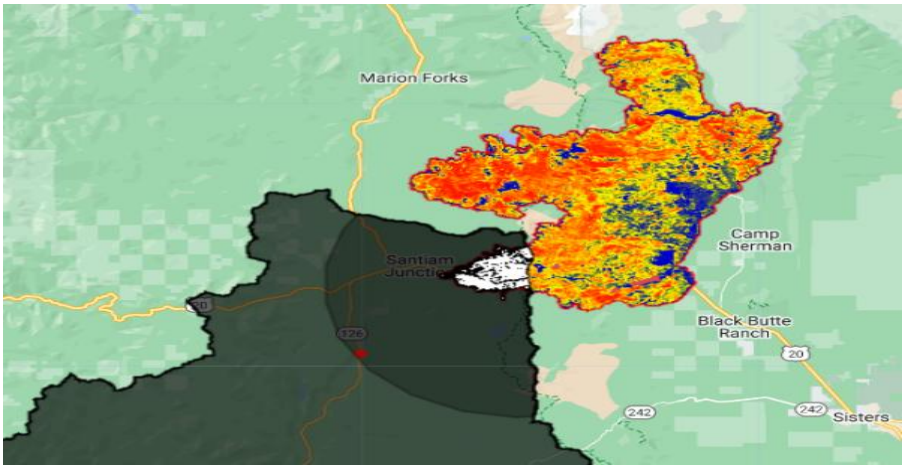

Fig. 8 *Bias corrected CBI clipped into the study area. The mask of CBI values ≥ 1.99 are shown in black within the buffer zone.*

This small severe-burn area is critical for interpreting the M^4 hindcast (**Fig. 2**). The Hogg Pass SNOTEL station, situated inside both the MODIS-based and Landsat-based high burn-severity pixel clusters, recorded the full post-fire snow response - greater accumulation due to canopy loss, lower albedo from ash presence, and hydrophobic soils routing a larger fraction of meltwater to channels to produce earlier, faster melt overall. However, roughly 90% of the watershed burned at low severity or not at all according to the CBI model, so the area-weighted snowpack and timing across the basin may have actually stayed closer to pre-fire norms. This mismatch likely created two compounding errors that can be attributed to the M^4 hindcast errors: First, it created a representativeness error by allowing the model to treat a localized disturbance as if it were basin-wide. Second, it created a structural error within M^4 since the model was calibrated before the wildfire disturbance event occurred. Because SWE and precipitation values were converted to streamflow values using coefficients learned under the intact-canopy conditions, the model underestimated the higher post-fire runoff efficiency (higher flow per unit SWE) and likely mistimed the peak flow. Together, these effects yielded forecasts that are biased, often too low relative to observed streamflow.

Conclusion & Future Research

This study used the 2003 B&B Complex Fire as a natural experiment to investigate the extent to which the Hogg Pass SNOTEL record represented watershed-scale snow-water equivalent after a wildfire disturbance. The M^4 model was trained on pre-fire SNOTEL data and compared its March predictions with observed April–July flows at the nearby McKenzie River outlet. The forecasts diverged sharply after water-year 2006, supporting our hypothesis that the Hogg Pass SNOTEL station, burned at moderate-to-high severity, no longer represented basin-wide snow-water equivalent (SWE). Yet, the fire severely burned only ~7-10% of the HUC-8 watershed buffer area versus ~40% suggested by the coarse MODIS maximum $T_{(\beta)}$ masks, so any hydrologic signal must be interpreted with caution. A larger burned fraction would likely have produced an even clearer bias in M^4 streamflow prediction.

Because M4's training window (1980-2002) predates the B&B Complex Fire, its six PCR-style models blended into a single ensemble mean are still keyed to a landscape with dense canopy, high-albedo snow, and freely draining soils. The baseline domain relationship between

March 1st SWE and April-July runoff was nearly linear, so the ensemble could map snowpack volume to streamflow with modest error. Post-fire, however, the burn scar flipped many of those physical controls - exposing snowpack likely absorbed more radiation, sublimation dropped, ash-darkened surfaces accelerated melt, and hydrophobic soils routed a larger fraction of meltwater straight to channels. Because none of those disturbance signals existed in the calibration data, every new forecast inherited the same structural bias - the ensemble simply has no precedent for a year in which reduced canopy or altered albedo decouple SWE from flow. Errors therefore compound rather than cancel, suggesting that predictions from the M^4 model are unlikely to self-correct when relationships between streamflow and nearby topography are disturbed. Three implications follow: First, point snow measurements inside the limited burn scar bias SWE-driven forecasts. The balancing effect that sensors in unburned parts of the same basin or that gridded snow products exhibit should therefore be investigated more thoroughly. Second, M^4 and similar data-driven models should be recalibrated after major disturbances, ideally with burn-severity metrics included as additional predictors so the algorithm can learn when a point record no longer reflects watershed conditions. Third, while FIRMS is valuable for rapid fire detection, hydrologic impact studies likely need higher-resolution, bias-corrected indices such as CBI to avoid overestimating the fraction of snowpack affected. The low severe-burn percentage revealed by CBI thus clarifies why Hogg Pass overstated post-fire snow gains and why M4 forecasts, trained on pre-fire relationships, diverged from reality.

More broadly, the case study underscores M^4 sensitivity to its input network. Reliance on a single, wildfire-affected SNOTEL site can exacerbate forecast error unless the model is recalibrated after the disturbance. Future applications should therefore target basins that (i) contain multiple SNOTEL stations spanning key elevation and aspect gradients, (ii) possess a stream-gauge at or near the outlet, and (iii) include reference (unburned) sites for normalizing hydro-climatic variability (Smoot & Gleeson 2021; Parks et al. 2019; Williams et al. 2022). With a richer observational backbone, M^4 can be retrained to recognize fire-altered snow dynamics rather than treating them as noise. The workflow developed here points the way for further investigation. Landsat-based bias-corrected CBI layers provide a higher resolution and more conservative burn-severity metric for smaller watershed scales. MODIS-derived maximum $T_{(\beta)}$ masks offered a quick initial scan and could be used in future studies to test broader-area M^4 recalibration or balancing effects from multiple SNOTEL stations. In short, the project

demonstrates both the promise and the pitfalls of machine-learning snow models in disturbance-prone watersheds and lays out a clear path for improving them.

Works Cited

Barnard, D. M., Green, T. R., Mankin, K. R., DeJonge, K. C., Rhoades, C. C., Kampf, S. K., Giovando, J., Wilkins, M. J., Mahood, A. L., Sears, M. G., Comas, L. H., Gleason, S. M., Zhang, H., Fassnacht, S. R., Harmel, R. D., & Altenhofen, J. (2023). Wildfire and climate change amplify knowledge gaps linking mountain source-water systems and agricultural water supply in the western United States. Agricultural Water Management, 286, 108377. https://doi.org/10.1016/j.agwat.2023.108377

Clark, M. P., Kavetski, D., & Fenicia, F. (2011). Pursuing the method of multiple working hypotheses for hydrological modeling. Water Resources Research, 47(9). https://doi.org/10.1029/2010WR009827

Deschamps-Berger, C., Gascoin, S., Shean, D., Besso, H., Guiot, A., & López-Moreno, J. I. (2023). Evaluation of snow depth retrievals from ICESat-2 using airborne laser-scanning data. The Cryosphere, 17(7), Article 7. https://doi.org/10.5194/tc-17-2779-2023

Fleming, S. W., Garen, D. C., Goodbody, A. G., McCarthy, C. S., & Landers, L. C. (2021). Assessing the new Natural Resources Conservation Service water supply forecast model for the American West: A challenging test of explainable, automated, ensemble artificial intelligence. Journal of Hydrology, 602(August), 126782. https://doi.org/10.1016/j.jhydrol.2021.126782

Fleming, S. W., & Goodbody, A. G. (2019). A machine learning metasystem for robust probabilistic nonlinear regression-based forecasting of seasonal water availability in the US West. IEEE Access, 7, 119943–119964. https://doi.org/10.1109/ACCESS.2019.2936989

Fleming, S. W., Zukiewicz, L., Strobel, M. L., Hofman, H., & Goodbody, A. G. (2023). SNOTEL, the Soil Climate Analysis Network, and water supply forecasting at the Natural Resources Conservation Service: Past, present, and future. JAWRA Journal of the American Water Resources Association, 59(4), 585–599. https://doi.org/10.1111/1752-1688.13104

Frigg, R., & Hartmann, S. (2020). Models in science. In E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy (Spring 2020). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2020/entries/models-science/

Giglio, L., Schroeder, W., & Justice, C. O. (2016). The Collection 6 MODIS active fire detection algorithm and fire products. Remote Sensing of Environment, 178, 31–41. https://doi.org/10.1016/j.rse.2016.02.054

Johnson, M., & Nolin, A. (2020). Forest disturbance and the impacts on maritime snow in the Oregon Cascades. https://westernsnowconference.org/node/1868

MacDonald, L. H., & Putz, J. (2011). Strength and persistence of fire-induced soil hydrophobicity under ponderosa and lodgepole pine. Retrieved from https://www2.nrel.colostate.edu/assets/nrel_files/labs/macdonald-lab/pubs/StrengthandPersistanceofFire-inducedSoilHydrophobicityunderPonderosaandLodgepolePine.pdf

MGHydro. (2025). Global Watersheds App [Web application]. Retrieved April 20, 2025, from https://mghydro.com/watersheds/

Natural Resources Conservation Service. (2023). M4 user manual: Multi-model machine-learning metasystem for water-supply forecasting. National Water and Climate Center. https://nrcs-nwcc.github.io/UserManual_NRCS_M4.html

Natural Resources Conservation Service. (2025). Air & Water Database Report Generator [Dataset]. U.S. Department of Agriculture. Retrieved April 20, 2025, from https://www.wcc.nrcs.usda.gov/reportgenerator/

Parks, S. A., Holsinger, L. M., Koontz, M. J., Collins, L., Whitman, E., Parisien, M.-A., Loehman, R. A., Barnes, J. L., Bourdon, J.-F., Boucher, J., Boucher, Y., Caprio, A. C., Collingwood, A., Hall, R. J., Park, J., Saperstein, L. B., Smetanka, C., Smith, R. J., & Soverel, N. (2019). Giving ecological meaning to satellite-derived fire severity metrics across North American forests. Remote Sensing, 11(14), Article 14. https://doi.org/10.3390/rs11141735

Pettorelli, N. (2013). The normalized difference vegetation index. Oxford University Press.

Roth, T. R., & Nolin, A. W. (2017). Forest impacts on snow accumulation and ablation across an elevation gradient in a temperate montane environment. Hydrology and Earth System Sciences, 21(11), 5427–5442. https://doi.org/10.5194/hess-21-5427-2017

Smoot, E. E., & Gleason, K. E. (2021). Forest fires reduce snow-water storage and advance the timing of snowmelt across the Western U.S. Water, 13(24), 3533. https://doi.org/10.3390/w13243533

U.S. Army Corps of Engineers, Bureau of Reclamation, & Bonneville Power Administration. (2024). 2025 water management plan – Draft 1. https://public.crohms.org/tmt/documents/wmp/2025/Draft_1_October_1/20241001_2025_WMP_Draft_1.pdf

U.S. Bureau of Reclamation. (2009). Teleconnections: Remote weather-condition relationships (Western Water & Power Solution Bulletin No. 25). Research and Development Office.

U.S. Geological Survey. (n.d.). National Water Information System data available on the Web (WaterData). Retrieved August 4, 2025, from https://waterdata.usgs.gov/nwis

U.S. Geological Survey. (2025). Data Basin: Hydrology and Watershed Layers [GIS data platform]. Retrieved April 20, 2025, from https://databasin.org/

U.S. Geological Survey. (2025). McKenzie River at Outlet of Clear Lake, OR (USGS 14159000) [Stream-gage record]. USGS Water Data for the Nation. Retrieved April 20, 2025, from https://waterdata.usgs.gov/

Williams, A. P., Livneh, B., McKinnon, K. A., Hansen, W. D., Mankin, J. S., Cook, B. I., Smerdon, J. E., Varuolo-Clarke, A. M., Bjarke, N. R., Juang, C. S., & Lettenmaier, D. P. (2022). Growing impact of wildfire on western US water supply. Proceedings of the National Academy of Sciences, 119(10), e2114069119. https://doi.org/10.1073/pnas.2114069119