remote sensing

Article

Complementing Topographic Wetness with Sentinel-2 Phenological Indices to Map Soil Moisture Gradients in a Snowmelt-Driven Mountain Floodplain

Edward Turk [1], Firstname Lastname [2] and Amalia Kokkinaki [2,*]

[1] Affiliation 1; turke6756@gmail.com
[2] Affiliation 2; e-mail@e-mail.com
* Correspondence: e-mail@e-mail.com; Tel.: (optional; include country code; Only one author should be designated as corresponding author)

Highlights

What are the main findings?

- Adding a single Sentinel-2 phenology metric (peak NDVI) to topographic wetness and plant community improves soil-moisture prediction at unsampled sites, with the added skill concentrated on high, steep terrain where topography alone explains little.
- Peak greenness ranks least important within a site but most important across sites, a within/across-site reversal showing it carries a transferable, not site-specific, moisture signal.

What are the implications of the main findings?

- Free, global Sentinel-2 imagery plus a lidar terrain index can map fine-scale (10 m) root-zone moisture across a snowmelt-driven basin, exposing wet pockets that point sensors and coarse satellites miss.
- The maps flag where topography is an incomplete indicator of moisture, guiding placement of field sensors for local moisture accumulation.
- More accurate moisture mapping can help identify biogeochemical hotspots and heterogeneities tied to greenhouse gas emissions in mountain floodplains.

Abstract

Soil moisture in mountain floodplains varies over a few meters, far finer than point sensors or coarse satellite images resolve; conventional optical and microwave methods sense only the top few centimeters, missing the root zone that governs plant function and biogeochemistry. Similarly, topography-derived moisture variations miss micro-heterogeneities in moisture related to intrinsic soil heterogeneity and/or the root zone. This has important implications for our ability to characterize local hotspots of biogeochemical reactions, like those related to the generation of methane. In this work, we read vegetation as a sensor of the subsurface, using Sentinel-2 land-surface phenology to infer root-zone moisture, and we test a phenology-conditioned moisture anomaly (PCMA) that reads the residual moisture signal after regressing late-season greenness on a pixel's own early-season phenology and vegetation class. Focusing our study on the upper East River watershed (Crested Butte, Colorado), we fit a pooled linear mixed model of volumetric water content on topographic wetness, plant community, and one Sentinel-2 phenology metric, calibrated on 2018 field measurements and evaluated by leave-one-site-out cross-validation across twelve sites. Our results show that peak NDVI carried the strongest transferable signal (out-of-site $R^2 = 0.34$, $\rho = 0.53$, RMSE = 7.2% VWC), out-predicting the PCMA residual. Peak NDVI ranked least important within sites (R^2 loss 16.3%), yet most important across sites (40.3%), and contributed most on high, steep terrain, where topographic wetness showed the least predictive value. Deeper soil moisture tracked snowmelt duration ($R^2 = 0.83$) and rate ($R^2 = 0.69$), not snowpack size ($R^2 = 0.14$). The framework maps fine-scale root-zone moisture from free inputs and flags where topography-only expectations do not suffice to capture moisture variability in mountain floodplains.

Keywords: soil moisture; land-surface phenology; Sentinel-2; NDVI; topographic wetness index; snowmelt; mountain floodplain; leave-one-site-out cross-validation

Academic Editor: Firstname Lastname

Received: date
Revised: date
Accepted: date
Published: date

1. Introduction

Soil moisture is a central variable of the terrestrial system, coupling the water, energy, and carbon cycles and exerting strong control on land-atmosphere exchange [1]. Much of that control operates below the surface, in the vadose zone: the unsaturated soil above the water table, where root-zone water availability governs both plant function and the soil's biogeochemical redox state, yet remains difficult to measure across the range of scales at which it varies [2].

The role of subsurface moisture in greenhouse-gas production from natural systems has more recently come into focus. Methane production tracks wetter conditions, but soils need not be fully saturated to emit it; even small, locally wet pockets can develop anoxic microsites within otherwise oxic soil [3]. Methanogens are ubiquitous in soils and persist even where it is dry, so the external environment, principally wetness, as well as the anoxia that follows, and the supply of labile carbon, are the decisive controls on methanogenesis [4]. Which locations hold those wet pockets therefore plays a significant role in determining where methane is produced [5], a dependence demonstrated in subalpine Colorado, where soil moisture alone flips the same landscape between wetland methane efflux and dry-soil uptake [6]. Predicting how and where these pockets generate methane thus depends on mapping subsurface moisture accurately and at high resolution.

Resolving moisture at that scale is difficult, because few landscapes are as heterogeneous as a mountain floodplain. Their mosaic of vegetation communities, microtopographic relief, shifting hydrological connectivity, and variable sediment compresses steep moisture gradients into very short distances [7], so that soil moisture can change sharply over just a few meters. Resolving that fine-scale variation would directly serve field research: it would let investigators place monitoring equipment where it is most representative, calibrate models against both the gradients and the hotspots they are meant to capture, and attribute biogeochemical fluxes to the conditions that drive them. This informed placement of sample collection is especially important given that a handful of samples must stand in for a landscape whose moisture varies over distances far shorter than the typical spacing between sampling points. The upper East River watershed near Crested Butte, Colorado, which includes the Slate River and was assembled as a snowmelt-driven testbed for exactly this class of hydrological and biogeochemical question, exemplifies the difficulty [8]. What is needed is a way to map subsurface moisture continuously, at a resolution fine enough to match the processes it governs, across terrain this variable.

There is a physical reason to expect vegetation to reflect these subsurface moisture gradients. Optical satellite and airborne instruments see the plant canopy, not necessarily the soil beneath it, yet plants are not detached from what lies below them. Their root systems couple them to the soil and bedrock, so a plant's water access, growth, and seasonal behavior are shaped by the moisture and substrate of its rooting zone. Cleverly et al. [9] showed this directly, finding that rooting-zone moisture governs the timing and magnitude of canopy growth across a season. More broadly, vegetation can be read as a surface expression of otherwise hidden critical-zone conditions, a premise advanced at this very watershed by Chadwick et al. [10]. A canopy growing over a persistently wet rooting zone should therefore behave differently across the season than one of the same kind growing over freely draining ground. One question remains open: whether that seasonal difference carries information about subsurface moisture specifically, beyond what vegetation class alone already reveals, and at a resolution useful for a basin-wide question.

Existing satellite and airborne soil-moisture methods share a limitation that matters for a basin-scale question: each senses only the near-surface, the top few centimeters, and each trades resolution against extent and cost while losing accuracy where the vegetation canopy interferes with the measurement [11,12]. Operational remote sensing spans ground, proximal, and satellite platforms, but for a question posed across an entire basin the satellite and airborne families of remote-sensing techniques are the ones of interest.

The most established method is passive microwave radiometry, which is what is most often meant by satellite soil moisture. Missions such as the Soil Moisture Active Passive (SMAP) satellite measure the land surface's natural microwave emission, which varies with soil water content, and retrieve moisture in the top 5 cm to an accuracy near 0.04 m^3/m^3, the benchmark for the field. They do so, however, at a footprint near 40 km, far too coarse to resolve a hillslope [13]. Active microwave sharpens that resolution. Sentinel-1 C-band synthetic-aperture radar reaches about 1 km in resolution by emitting its own pulse and reading the returned backscatter, but that backscatter is confounded by vegetation and surface roughness and performs poorly over the dense canopy of a meadow-to-conifer basin [12].

Optical and thermal feature-space methods push to finer resolution still, and onto freely available imagery. Older approaches read moisture from the relationship between land-surface temperature and a vegetation index, positioning each pixel within a triangular or trapezoidal feature space [14]. The Optical Trapezoid Model (OPTRAM) carries this idea onto optical reflectance: it places each pixel in a space of shortwave-infrared-transformed reflectance against greenness, locates it between a dry edge and a wet edge, the reflectance limits of the driest and wettest pixels in the scene, and reads surface moisture from that position, at 10 m resolution [15]; the approach has been validated across climates against ground networks at coarser MODIS resolution [11]. Greenness enters OPTRAM only as one axis of that space, used to position each pixel; the quantity the method estimates is still moisture at the soil surface, not derived by a property related to the plant itself.

These optical retrievals, whatever their resolution, capture only an instantaneous, surface-skin signal that degrades under dense canopy. Even the newest advance, uncrewed-aircraft L-band radiometry, reaches 3 to 50 m resolution but remains surface-only, costly, and limited to extents on the order of 10 km [16]. Across all of these methods the same gap remains: a near-surface measurement, delivered at a scale, extent, or fidelity that cannot answer a basin-wide, root-zone question.

Sentinel-2 optical imagery is the most promising asset of the set, being free, global, and acquired at 10 m, yet read in the conventional way it too sees only the surface. This study pursues a different reading of that same imagery. Rather than treating the canopy as an obstacle between the sensor and the soil surface, it treats the plants themselves as sensors of the subsurface, using their season-long behavior to infer moisture at depth in the root zone rather than at the surface. This returns to the main premise of this study: vegetation is a readable expression of the critical-zone conditions beneath it [10].

At the Crested Butte watershed, vegetation has already shown that it carries moisture information finer than land-cover class. On these hillslopes, Dafflon et al. [17] found that the season peak in Green Chromatic Coordinate (GCC) co-varied with soil electrical conductivity, the geophysical proxy for near-surface moisture, more tightly than with any other surface property they measured (r = 0.74). This is an in-situ, meter-scale correspondence of exactly the kind that coarse-footprint satellite retrievals, for all their 0.04 m^3/m^3 accuracy, cannot deliver at hillslope resolution. That demonstration, however, rested on a single hillslope across two years, leaving its generality untested over the broader landscape.

Working in the same watershed, Devadoss et al. [18] reduced multi-year NDVI time series to a few components and clustered them into distinct spatial zones whose moisture distributions differed significantly and aligned with the vegetation communities that occupy distinct topographic positions. Because such clustering aggregates sub-community variation into community-scale structure, resolving moisture beyond class identity depends on whether meaningful within-class variability exists. Wainwright et al. [19] showed that it does, finding the grassland class more internally heterogeneous in its drought response than any other vegetation type, with that heterogeneity governed by elevation, microtopography, and underlying geology.

Beyond this basin, groundwater-dependent-ecosystem studies likewise report that vegetation differs spectrally according to its access to subsurface water [20–22]. Together these studies support the hypothesis that vegetation encodes subsurface moisture. But clustering collapses that continuous signal into a handful of zones, discrete classes much like land-cover units, whereas the moisture question is ultimately a per-pixel (10 m) one, and no study in this system has yet tested whether a per-pixel spectral metric improves point predictions of moisture, expressed as volumetric water content (VWC), at sites not used to train moisture-predicting models.

In the Crested Butte basin, snowmelt gives reason to expect a plant's seasonal trajectory to function as a moisture record. Because the system is snow-dominated, we expect it to be the dynamics of melt, its timing, rate, and duration, rather than the size of the snowpack, that set the deep soil-moisture store the vegetation draws on. That store varies across the landscape at the scale of tens of meters, finer than any weather station can resolve. If a pixel's early-season growth integrates the consequences of that melt, it should carry a moisture signal that coarse climate or drought indices can only approximate.

Dafflon et al. [17] observed a consistent pattern in this watershed, where a deeper, slower-melting snowpack produced longer-lasting saturation and higher peak greenness. This reasoning motivates the study's central question: whether Sentinel-2 vegetation phenology, and peak greenness in particular, improves VWC prediction beyond topographic wetness and plant community, and where on the landscape it does so. A second motivation is to resolve local moisture accumulation even where bulk conditions appear unfavorable, the fine-scale wet pockets that landscape-average or topographic descriptors miss but that a per-pixel, vegetation-based map could expose.

In this study we utilize our own metrics for land-surface phenology, the seasonal trajectory of canopy greenness recorded at the pixel scale, rather than the life-cycle stages of an individual plant. We summarize each pixel's trajectory with a small, explicit set of metrics, namely June and September greenness, the timing and value of the seasonal peak, and the season's amplitude, all extracted from a per-pixel harmonic fit to the NDVI series.

We also evaluate one constructed predictor, built by inverting a method from land-degradation monitoring. Residual-trend methods (RESTREND) regress a vegetation index against an exogenous climate driver such as rainfall and read the residual, the greenness that climate cannot explain, as a sign of human disturbance [23,24]. Applied uniformly across a basin, climate data this coarse cannot resolve the pixel-by-pixel melt timing that governs root-zone moisture here. We therefore invert the logic. Rather than regressing greenness against an external climate predictor, we regress a pixel's late-season greenness against its own early-season phenology and vegetation class, and read the residual, the late-season greenness higher or lower than that trajectory and class would predict, as a moisture anomaly. We call this the phenology-conditioned moisture anomaly (PCMA) and treat it as one candidate among the individually derived land-surface phenology metrics.

Where OPTRAM reads an instantaneous signal from the soil surface, the approach described in this study is deliberately indirect; it uses the plants themselves as indicators, inferring root-zone moisture from how their greenness behaves across the whole season or specific months within the season. The remainder of the paper develops these predictors, fits and cross-validates a soil-moisture model over a network of field sites, compares it against a topography-and-community baseline, and maps where vegetation phenology adds skill across the watershed.

2. Materials and Methods

2.1. Study Area

The study area is the upper East River watershed in Gunnison County, Colorado, USA, a snowmelt-dominated montane-to-subalpine basin in the Elk Mountains near Crested Butte and the Rocky Mountain Biological Laboratory (RMBL) at Gothic (Figure 1). The landscape spans valley-bottom riparian corridors and wet meadows, dry montane meadows, mixed aspen and conifer forest, and high-elevation alpine and barren terrain, from approximately 2,600 to 4,190 m above sea level. The analysis footprint is centered near 38.92 degrees north, 107.00 degrees west (WGS 84 / UTM zone 13N, 326,670 E, 4,310,130 N).

Hydrologically, the basin is governed by a seasonal snowpack: deep winter accumulation, a spring-to-early-summer melt pulse, and progressive drying through the growing season [8,25]. Lateral subsurface flow concentrates moisture in valley bottoms [25,26], which topographic indices capture well, but not on hillslopes, where vegetation may instead integrate the unobserved subsurface signal. This contrast (moisture topographically organized in valley bottoms and steep dryer vegetation covered hillslopes) is what structures soil-moisture variability across the landscape, and it frames the study's question: whether remotely sensed vegetation phenology adds predictive skill for soil moisture beyond topographic wetness, and where on the landscape it does so.

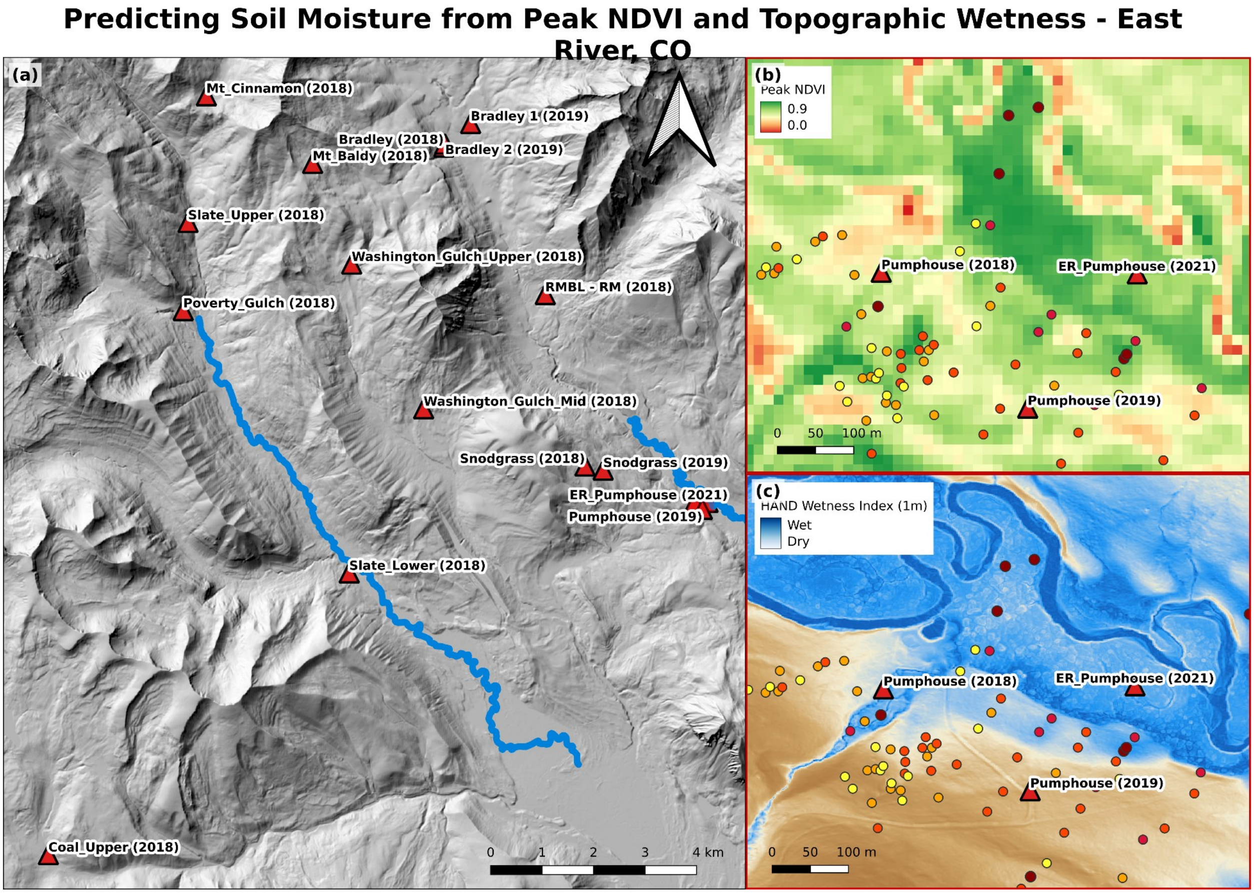

Figure 1. Study area and soil moisture sampling network in the upper East River watershed, Gunnison County, Colorado. (a) The main panel shows the 17 field sites overlaid on a hillshade, with the East River and Slate River. The two insets, zoomed to the Pump House sub-cluster, show the model predictor surfaces: (b) annual peak NDVI (top) and (c) the HAND topographical wetness index (bottom). Colored points are the field VWC measurements, shaded from yellow (dry) to red (wet).

2.2. Datasets

2.2.1. Field soil moisture measurements and site-polygon definitions

Direct measurements of ground volumetric water content (VWC) were obtained from three field campaigns archived at the U.S. DOE ESS-DIVE repository under the LBNL Watershed Function Science Focus Area, standardized into a single point table (VWC in percent; EPSG:32613):

i. TDR 2018: plot-level time-domain reflectometry (TDR), a ground-based probe method that infers volumetric water content from the soil's dielectric response, using a 20 cm probe with three replicates averaged per plot, collected 14 to 30 June 2018 concurrent with the NEON AOP flight; 336 rows [27].
ii. TDR 2019: spatial TDR survey (25 cm probe, recorded as 20 cm depth in the standardized table; RTK-GPS located), collected 2 to 7 July 2019; 101 rows [28].
iii. SWC 2021: monthly means of in-situ volumetric soil water content at the East River Lower Montane / Pump House cluster (TEROS12 / 5TE / Campbell CS655 sensors) at depths of 10, 20, 30, 35, 40, 50, 60, 106, and 115 cm, converted to percent; 487 rows [29].

The standardized table contains 447 rows/entries : 336 from the 2018 TDR campaign (12 sites), 101 from the 2019 TDR campaign (4 sites), and 10 from the 2021 ER-Pumphouse sensor array for the campaign-aligned June–July period at 20 cm depth.

Site identity was inherited from the source data: each archived measurement point already carried a site label, with 10 to 40 points per site. Around each measurement point, we placed a 50 m circular buffer and dissolved the buffers by (site, year, dataset), yielding 17 site-year polygons: 12 from the 2018 TDR campaign, 4 from the 2019 TDR survey, and 1 from the 2021 SWC cluster, the last assigned a single cluster label because its source contained only sensor identifiers.

The 50 m buffer radius was chosen a priori to approximate the spatial scale that a point sensor integrates, so the community context seen by a map pixel is comparable in scale to what a sensor at that location would represent. These site-year polygons provide the common spatial support for the site-level terrain and vegetation summaries computed in Section 2.3. The empirical justification for the 50 m support is reported in Section 3.4.3.

The two time-domain reflectometry campaigns were measured at a nominal depth of 20 cm, whereas the 2021 sensor record spans nine depths from 10 to 115 cm over twelve months. Because the campaigns were collected at different times of year and over different depth ranges, they were kept distinct in the analysis: the 2018 TDR campaign, a single mid-June campaign at a single 20 cm depth, provides the calibration dataset for the predictive model, while the 2019 TDR and 2021 SWC records provide additional site-year observations used in exploratory site-level comparisons. The 2021 cluster is summarized across its sensor depths and is therefore treated as a single exploratory site whose within-depth heterogeneity is not resolved.

2.2.2. Sentinel-2 Level-2A surface reflectance

Optical reflectance was obtained from the Copernicus Sentinel-2 Level-2A bottom-of-atmosphere product via the Google Earth Engine (GEE) collection COPERNICUS/S2_SR_HARMONIZED [30], over the four Sentinel-2 tiles for the 2016 to 2025 study period. The May–October window defines the growing-season extent analyzed each year, not the acquisition frequency: Sentinel-2 revisits the site far more often (sub-monthly), yielding the 42 to 54 usable scenes per growing season fit in Stage 1. Scene-days were screened for cloud cover, and a per-pixel Scene Classification Layer mask removed cloud, cloud shadow, and saturated or defective pixels. The Normalized Difference Vegetation Index (NDVI) was computed from the near-infrared (B8) and red (B4) 10 m bands as

$$\mathrm{NDVI} = (\mathrm{B8} - \mathrm{B4}) / (\mathrm{B8} + \mathrm{B4}) \quad (1)$$

then exported at 10 m, EPSG:32613, in 32-bit floating point. Cloud masking, across-tile compositing, and the NDVI computation were all performed server-side in Google Earth Engine. This yielded 419 single-scene NDVI rasters across 2016 to 2025.

The Sentinel-2 export used a single fixed bounding box of approximately 6,028 km² that spanned four Sentinel-2 tiles on the Military Grid Reference System grid (13SBC, 13SBD, 13SCC, 13SCD). This box represents only the extent from which raw scenes were downloaded and carries no analytical meaning.

2.2.3. Vegetation community classification

The vegetation community was derived from the Falco et al. 10 m classification of the East River basin (hereafter referred to as the NEON-campaign classification), produced from NEON Airborne Observation Platform (AOP) imaging spectroscopy acquired in 2018 and distributed as an open dataset [31]. The source legend includes 28 mapped land-cover classes (4 non-vegetation and 24 vegetation taxa) at 1 m resolution. For the per-pixel vegetation factor, we merged this legend into five classes: Aspen, Subalpine Conifer, Lodgepole Pine, Shrub Community, and Meadow Community.

The five class codes used to label these merged bins (41, 45, 47, 50, 60) are the NEON-campaign classification's original pixel codes, reused as bin identifiers. In the merge, code

45 (Subalpine Conifer) additionally includes Engelmann spruce, and codes 50 and 60 include the full shrub and meadow species sets, respectively, so the merged classes are broader than the NEON-campaign classification's single-code definitions. Because the five analysis classes are merged from the finer 1 m source legend, a 10 m pixel can straddle several types. A companion purity mask therefore restricts analysis to pixels in which a single merged class accounts for at least 60 % of the underlying 1 m source pixels; pixels more mixed than this are too ambiguous to label and are excluded. Two derived categorical labels are distinguished later in the Methods: a merged five-class per-pixel vegetation label (Veg_Type: 41, 45, 47, 50, 60), and a moisture-relevant five-class community label (Alpine/Barren, Dry Meadow, Forest, Riparian Shrub, Wet Meadow) derived from the NEON-campaign classification's fractional cover (Section 2.3.3). The first records species identity; the second instead names the community a pixel belongs to given its species mixture. The two are applied separately in the soil-moisture analyses (Section 2.3) to capture different controls on moisture.

The NEON-campaign classification footprint defined the analysis footprint: approximately 396 km² of classified vegetation pixels. Every raster product was masked to this footprint and analyzed at a 10 m ground sampling distance in WGS 84 / UTM zone 13N (EPSG:32613) unless stated otherwise.

2.2.4. Topographic wetness index and digital elevation model

Topographic wetness was represented by the Multiscale Height-above-stream Wetness Index for the Upper Gunnison Domain [32], a Height Above Nearest Drainage-based soil-moisture proxy. It was derived from a 1 m hydrologically corrected LiDAR digital elevation model (DEM) [33]. Following the Height Above Nearest Drainage concept [26], the index combines local slope with the vertical height above the nearest stream, averaged over four flow-accumulation thresholds, and rescales the result to [0, 1]. We refer to it throughout as the topographic wetness index (hereafter, the wetness index) and attribute its methodological lineage to Nobre et al. [26]; it is a rescaled wetness index, not the bare height-above-drainage surface. We resampled it from its native 1 m grid to the 10 m analysis grid by spatial averaging, so that each 10 m cell is the mean of the ~100 underlying 1 m cells.

Slope, elevation, and aspect used to characterize landscape gradients (Section 2.3.6) were derived from the same 1 m RMBL LiDAR DEM [33] and resampled to the 10 m analysis grid. The topographic wetness index and the terrain gradients therefore share a common 1 m DEM source but are distinct products: the wetness index is the published wetness-index product [32], and the slope, elevation, and aspect layers were computed directly from the 1 m RMBL LiDAR DEM [33] by us.

2.2.5. SNOTEL 380 snow and soil-moisture record

The snowmelt analysis (Section 2.3.7) used USDA NRCS SNOTEL station 380 (Butte, Colorado), drawing on daily snow water equivalent (SWE, in inches) and co-located soil-moisture probes at 2, 8, and 20 inch depths [34]. The analysis covered six water years, WY2020 to WY2025. WY2019 was excluded, because the station record begins on 1 January 2019 and thus lacks the October to December 2018 accumulation season (n = 6).

2.3. Methods

The analysis proceeds in a six-stage prediction pipeline, ordered from the simplest per-pixel summaries to the full spatial prediction, plus an independent premise analysis (Section 2.3.7) that motivates the approach. Each stage is introduced below by its purpose and then its method; what each stage found is reported in Section 3.

1) **Stage 1 - Harmonic curve fitting and phenology-metric extraction.** To compress each pixel's noisy, irregularly sampled NDVI time series into a few

comparable seasonal descriptors, we fit a per-pixel harmonic NDVI curve and read five phenology metrics from it.

2) **Stage 2 - Phenology-Conditioned Moisture Anomaly (PCMA) residual.** To test whether the late-season greenness a pixel reaches beyond what its early-season phenology and vegetation type would predict carries a moisture signal, we build the PCMA residual as a candidate spectral predictor. It is a self-contained branch of the workflow.
3) **Stage 3 - Community labeling and the community-context raster.** Because a soil-moisture model needs a moisture-relevant vegetation descriptor rather than a raw species map, we derive a per-site community label and a matching per-pixel community-context raster for prediction.
4) **Stage 4 - Per-site exploratory regressions.** Before pooling, we screen each site on its own to ask whether any spectral metric adds explanatory power to topographic wetness, the simplest form of the question. The screen is a simple per-site ordinary-least-squares regression of point VWC on each pixel-scale spectral predictor in turn, the harmonic NDVI phenology metrics of Stage 1 (Section 2.3.1) and the PCMA residual built from them; the vegetation-community label is not screened here because it is constant within a site and enters only the pooled between-site model.
5) **Stage 5 - Pooled mixed-model development and predictor comparison.** To learn a single transferable relationship, we fit a pooled linear mixed model to the 2018 field data and compare candidate spectral predictors by leave-one-site-out cross-validation and a drop-one-predictor analysis.
6) **Stage 6 - Spatial prediction and null-model comparison.** To produce maps and isolate the spectral term's contribution, we apply the selected model across the analysis footprint for each year and compare it against a null model with no spectral predictor.
7) **Premise analysis (Section 2.3.7) - Snowmelt and soil moisture.** To establish the mechanistic premise that motivates using phenology as a moisture proxy (that snowmelt dynamics, not snowpack size, set the deep soil-moisture store vegetation draws on), we relate the SNOTEL 380 snow record to co-located soil moisture. This analysis is independent of the prediction pipeline.

Stages 1–2 derive spectral predictors and Stage 3 the community predictor; Stages 4-6 form the deployed prediction chain; the snowmelt premise analysis (Section 2.3.7) sits outside this chain.

R notebooks were executed with R version 4.3.3, using lme4 version 1.1-35.1 for the mixed models, base lm() for fixed-effect regressions, and terra version 1.8.86 for raster operations. Python scripts ran under Python version 3.12.3, using NumPy version 2.4.4 for harmonic fitting and GDAL version 3.11.4 for raster input and output. Map layouts were exported from QGIS version 3.40.9. Sentinel-2 access used the Google Earth Engine Python API version 1.7.18.

2.3.1. Stage 1: Harmonic curve fitting and phenology-metric extraction

We summarized each pixel's seasonal NDVI time series (its temporal evolution) using a per-pixel harmonic fit. For each pixel and year, a third-order Fourier model was fit to NDVI as a function of day-of-year (denoted t):

$$\mathrm{NDVI}(t) = \beta_0 + \Sigma(k = 1..3)\ [\beta_{2k-1} \cos(k\omega t) + \beta_{2k} \sin(k\omega t)],\ \omega = 2\pi/365 \quad (2)$$

where β_0 is the annual mean and the remaining β terms are the harmonic coefficients. The fit compresses each pixel-year into seven coefficients that define a smooth daily curve. A fit required at least twelve valid cloud-free observations in the pixel-year. Because the 2016 and 2017 Sentinel-2 records are sparse and gap-prone (10 and 24 acquisition dates, respectively), too few pixels accumulated twelve valid cloud-free observations after masking in those years. Therefore, harmonic fitting was run only for the eight seasons from 2018 to 2025. Across these eight seasons, 385 NDVI scenes were fit in total, ranging from 42 to 54 images per season (date spans from early May to late October).

From each fitted curve, we extracted five NDVI phenology metrics over fixed day-of-year (DOY) windows, read from the fitted harmonic curve within the same calendar window each year rather than from raw scenes, so that differences in image acquisition timing between years do not bias the metrics:

- JunMean, the mean fitted NDVI over DOY 152 to 181 (June);
- Peak_Value, the maximum fitted NDVI over DOY 152 to 273 (June to September);
- Peak_DOY, the day-of-year of that maximum;
- SepMean, the mean fitted NDVI over DOY 244 to 273 (September);
- Amplitude, Peak_Value minus JunMean.

The above phenology metrics were retained only for pixel-years whose fit passed quality control, namely a coefficient of determination of at least 0.6 and a root-mean-square error no greater than 0.10, with a phenology-shape filter that discarded curves showing no growing-season rise (Amplitude not greater than zero). Metrics were written as rasters for 2018 to 2025 at 10 m resolution in EPSG:32613.

These phenology metrics serve two distinct roles in the analysis. First, they serve as inputs to the Phenology-Conditioned Moisture Anomaly product described in Stage 2 (Section 2.3.2). Second, each metric was evaluated independently as a candidate spectral predictor of soil moisture. In the per-site exploratory regressions (Stage 4), each metric was added in turn to a topographic wetness baseline, and in the 2018 pooled mixed model (Stage 5), each metric occupied a single rotating spectral slot alongside topography and community.

2.3.2. Stage 2: The Phenology-Conditioned Moisture Anomaly (PCMA) residual

The Phenology-Conditioned Moisture Anomaly (PCMA) residual is a metric developed as a candidate predictor of soil moisture. The PCMA approach treats the late-season NDVI level that cannot be explained by a pixel's early-season phenology and vegetation type as a candidate moisture-anomaly proxy. For each year from 2018 to 2025, an ordinary least-squares regression was fit to all quality-filtered pixels to predict the late season NDVI (i.e., SepMean) using the derived phenology metrics (Section 2.3.1) and the NEON-campaign merged 5-class raster at ≥ 60 % purity:

$$\text{SepMean} \sim \text{Amplitude} + \text{Peak\_DOY} + \text{JunMean} + \text{Veg\_Type} \quad (3)$$

Veg_Type is entered as a factor, providing a per-class intercept (as described in Section 2.2.3).

Table 1 distinguishes this Veg_Type factor from the community label used in the soil-moisture model (Section 2.3.3): two labels derived from the same NEON-campaign classification but never conflated.

Table 1. Veg_Type (PCMA regression) versus community (soil-moisture model): two distinct categorical labels derived from the same NEON-campaign classification.

community	Veg_Type	Attribute
Between-site fixed effect in the soil-moisture model (Sections 2.3.3 to 2.3.6)	Factor in the PCMA regression (Section 2.3.2)	Where used

5	5	Number of classes
Alpine/Barren, Dry Meadow, Forest, Riparian Shrub, Wet Meadow	Aspen (41), Subalpine Conifer (45), Lodgepole Pine (47), Shrub Community (50), Meadow Community (60)	Classes
Indicator-species fractions, then a priority cascade (Section 2.3.3, Table 2)	Merged from the NEON-campaign classification's 28-class legend by original pixel code	Derivation
Per site (50 m footprint) for fitting; per pixel for mapping	Per pixel (10 m)	Spatial support

Each annual fit used on the order of two million pixels. Peak_Value is deliberately excluded from Equation (3): because Amplitude = Peak_Value − JunMean by construction, entering Peak_Value alongside Amplitude and JunMean would make the predictors exactly collinear (rank-deficient). The per-pixel regression residual was taken as the PCMA moisture-anomaly product, producing annual PCMA residual rasters (2018 to 2025) from Equation (4).

$$\text{Residual} = \text{SepMean\_observed} - \text{SepMean\_predicted} \quad (4)$$

Two calculated layers were used for diagnostics: the multi-year mean residual, and a sign-consistency count that identifies pixels whose residual retains the same sign across the eight years. In addition, to assess whether the residual field carries structure rather than noise, we computed empirical semivariograms of the mean residual. These are diagnostics of the residual field's spatial structure used in the interpretation of the PCMA residual as a candidate predictor (Residual_PCMA) for the soil moisture predictive model (Stage 6).

2.3.3. Stage 3: Community labeling and the community-context raster

Community labels were assigned for each site to give the soil-moisture model a moisture-relevant vegetation descriptor rather than a raw species map. These three vegetation descriptors are not interchangeable and capture different controls: species identity is the raw classification label; community is a static, moisture-relevant grouping of species mixtures that varies between sites; and phenology is the within-season canopy trajectory that varies in time. Both community and phenology enter the pooled mixed model: phenology through the Peak_Value metric and community through the community-context raster. We constructed this label in two steps from the NEON-campaign classification. First, we condensed the classification's species into five indicator-species fractions: forest, non-vegetation, riparian willow (Salix), wet-meadow species, and dry-meadow species. The last two are each defined by four indicator species chosen as diagnostic of wet or dry soils rather than by the generic meadow class (Table 2). The same source product also supplies the merged five-class Veg_Type factor used in the PCMA regression (Section 2.3.2); the community label differs in targeting soil-moisture relevance rather than structural vegetation class (Table 1).

From each site-year polygon (Section 2.2.1) we summarized zonal terrain attributes (slope, elevation, aspect) from the aligned 10 m DEM derivatives for use as site-level predictors, together with the five vegetation-community fractions defined above.

Table 2. Indicator-species groups used to derive the community label. The forest and non-vegetation groups pool structural classes; the wet- and dry-meadow groups are each defined by four indicator species diagnostic of wet or dry soils. Thresholds are applied in the priority order shown.

Threshold (priority order)	NEON-campaign classes or indicator species	Community group
comm_Forest > 0.25	Aspen, subalpine fir, Engelmann	Forest

https://doi.org/10.3390/xxxxx

	spruce, lodgepole pine	
comm_NonVeg > 0.20	Water, snow, urban, barren	Alpine / Barren (non-vegetation)
comm_Salix > 0.20	Salix complex	Riparian shrub (willow)
comm_WetMdw > comm_DryMdw and > 0.05	Veratrum tenuipetalum, Ligusticum porteri, Caltha leptosepala, Corydalis caseana	Wet meadow
Dry Meadow (fallback)	Lupinus argenteus, Frasera speciosa, Wyethia amplexicaulis, Festuca thurberi	Dry meadow

Within each site-year polygon we computed these five fractions [35,36]; each fraction is that group's share of classified NEON-campaign pixels (comm_Forest, comm_NonVeg, comm_Salix, comm_WetMdw, comm_DryMdw). We then assigned one community per site from these fractions by a priority cascade, a rule that tests conditions in a fixed order and takes the first that is met, rather than a simple majority. The rules, in order, were:

- Forest if comm_Forest > 0.25;
- else Alpine/Barren if comm_NonVeg > 0.20;
- else Riparian Shrub if comm_Salix > 0.20;
- else Wet Meadow if comm_WetMdw > comm_DryMdw and comm_WetMdw > 0.05;
- else Dry Meadow (fallback).

Because a community is defined by its structurally dominant life form, not by whichever class covers the most ground [37], each rule is a minimum-presence threshold applied in priority order. The willow (0.20) and wet-meadow (0.05) cutoffs are deliberately low: willow grows in narrow corridors and wet meadows in small patches that a majority rule would erase. This community label is one of the fixed effects in the predictive soil-moisture model (Stages 4–6).

For spatial prediction, we applied the same community-label logic at every 10 m pixel: after aggregating the NEON-campaign classification to 10 m, we took a 50 m circular focal mean of the five fractions and applied the cascade in Table 2, producing a community-context raster used as a predictor in Stage 6.

2.3.4. Stage 4: Per-site exploratory regressions

Before fitting the pooled mixed model, we conducted an exploratory per-site screen. For each site, we fitted separate ordinary least-squares regressions of point VWC on topographic wetness plus one spectral candidate, where the candidate was drawn from the harmonic NDVI phenology metrics and the PCMA residual (itself a regression on those same phenology metrics). Vegetation community and the terrain metrics were not screened here; they were reserved for the pooled model (Section 2.3.5). This screen used data from all three field campaigns (Section 2.2.1).

2.3.5. Stage 5: Pooled mixed model development and predictor comparison

The primary objective of this research is the development of a predictive model for soil moisture. This is achieved by the construction of a pooled linear mixed model that uses metrics described in previous sections and is fitted to the 2018 TDR data only, the most complete single campaign, comprising 335 points across 12 sites. Restricting to 2018 avoids the cross-campaign month and depth mismatch (June 2018, July 2019, and multi-depth 2021). Each field point inherited its site's single community label (the per-site label from Stage 3), so the model is fit on one community value per site. The per-pixel community-context raster, the focal-mean (moving-window) version of that label, was not used in fitting; it was reserved for the spatial-prediction step (Stage 6), where a label is needed at every pixel.

Several model candidates were evaluated (lme4::lmer, REML = FALSE), each adding exactly one spectral term to a common base of topographic wetness, community, and a random site intercept. As examples of candidate model forms, Equations (5) and (6) show two such models, where the response VWC(%) is point volumetric water content.

(5) VWC(%) ~ Wetness_Index + community + Peak_Value + (1 | site_name)

(6) VWC(%) ~ Wetness_Index + community + Residual_PCMA + (1 | site_name)

Only a single spectral term was entered into each model candidate because including multiple phenology metrics together produced an exact rank deficiency (Amplitude = Peak_Value − JunMean by construction). By comparing different model candidates, we tested the value of different predictors. For example, comparing models (5) and (6), allows us to evaluate whether the more complex PCMA residual has better predictive skill than the simpler Peak Value.

This predictive skill was assessed by leave-one-site-out (LOSO) cross-validation (12 folds). Each site was held out in turn and its points predicted from the model fit on the other 11 sites, using the marginal, fixed-effects-only prediction. The per-site random intercept is omitted because a held-out site has no fitted intercept. The model includes a site-level random intercept, estimated during fitting to absorb between-site mean differences and to account for the non-independence of pixels within a site; all reported predictions are made at the population level (marginal), with the random effects set to zero, so the same fitted model transfers to unseen sites. Skill was summarized per held-out site by Spearman rank correlation, the out-of-sample coefficient of determination, and the root-mean-square error (RMSE).

We then ran a leave-one-variable-out (LOVO) analysis: each predictor was removed in turn, the model refit, and the resulting change in skill recorded, both in-sample (the model refit and scored on the full 2018 dataset, with no site held out) and under leave-one-site-out cross-validation. Because the predictors are correlated, the LOVO change represents each predictor's marginal contribution given the others, that is, the skill it adds that the remaining predictors cannot already recover, not an independent partition of variance. A predictor can therefore appear weak under LOVO simply because a correlated partner compensates when it is removed.

2.3.6. Stage 6: Spatial prediction analysis and null-model comparison

In Stage 6, the model selected from the Stage 5 candidate comparison (Equation 5) was applied across the entire study area for each year from 2018 to 2025. For each year, a predictor stack was assembled from the selected annual spectral raster, the static 10 m wetness index (Section 2.2.4), and the static community-context raster (Section 2.3.3). Predictions were generated using only the fixed effects, that is, with the site random intercept omitted, because a generic landscape pixel does not belong to any trained site.

To prevent extrapolation beyond the training range, pixels with Peak_Value outside [0, 1.1] were set to missing before prediction (the upper bound is set to 1.1 rather than 1.0 to retain pixels where the smoothed harmonic NDVI peak marginally exceeds unity), and any negative predicted VWC was clipped to zero. From the annual predictions, we derived a multi-year mean VWC raster and an inter-annual coefficient of variation (CV) raster (2018 to 2025). Because the topographic wetness and community rasters are static across years and only Peak_Value varies, the CV surface reflects spectral-transfer sensitivity, not leave-one-site-out skill or directly observed field VWC variability.

To isolate the spectral term’s contribution, we built a null model with the same base but no spectral predictor,

$$\text{VWC}(\%) \sim \text{Wetness\_Index} + \text{community} + (1 \mid \text{site\_name}) \quad (7)$$

giving a single static prediction surface. We compared full and null predictions over a common soil-valid mask (pixels at least half open water, permanent snow, or built surface were excluded; bare rock retained), reading both surfaces at the same random sample of pixels and summarizing the median predicted VWC (with interquartile ribbons) across elevation and slope bins. Because the full surface is a 2018 to 2025 mean while the null is a single static surface, and the highest bins hold few pixels, we read the comparison as descriptive rather than as a formal skill test.

2.3.7. Snowmelt–soil-moisture premise (SNOTEL 380, independent of the prediction pipeline)

To motivate why phenology should carry a moisture signal (testing whether snowmelt dynamics, rather than snowpack size, set the deep soil-moisture store that vegetation draws on), we characterized snow-to-soil-moisture relationships at SNOTEL 380, independently of the prediction pipeline. For each of the six water years, we derived snow metrics from the daily SWE record (peak SWE and its date, 1 April SWE, bare-ground date, melt duration, average and 7-day-maximum melt rate, and cumulative post-melt growing-degree-days) and paired each with the per-water-year soil-moisture response at 2, 8, and 20 inch depths (nearest valid reading within ±3 days). Given the small sample ($n = 6$) and the number of pairwise tests, the resulting ordinary least-squares relationships (Pearson r and p-values) are reported as exploratory.

3. Results

3.1. Snowmelt dynamics, not snowpack size, set deep soil moisture

We begin with the physical premise that motivates using vegetation phenology as a moisture proxy: the deep soil-moisture store that vegetation draws on is set by how the snowpack melts rather than by its size. Using the SNOTEL 380 record (Butte, CO) over six water years (WY2020–WY2025; WY2019 excluded for incomplete soil coverage), we related water-year snow metrics to soil moisture at the co-located pit.

Deep soil moisture (20 in) tracked melt duration closely but was essentially unrelated to peak snowpack (Figure 2). Melt duration explained 83% of the between-year variance in deep moisture ($r = +0.909$, $R^2 = 0.825$, $p = 0.012$), whereas peak SWE did not ($r = +0.375$, $R^2 = 0.141$, not significant). The same deep store also tracked melt intensity, measured as the 7-day maximum melt rate (Figure 2; $r = +0.831$, $R^2 = 0.691$, $p = 0.040$), whereas shallow moisture at 2 in was decoupled from melt intensity ($r = +0.336$, $R^2 = 0.113$, not significant). These results indicate that the rate and duration of melt, rather than the size of the snowpack, recharge the deep soil (Figure 2), and the resulting melt is detectable in deep stores (20 in) rather than in shallow surface-level soil moisture (Dafflon et al. [17]). Because deep soil moisture is set by melt timing rather than snowpack size, the season opens from a comparatively clean, spatially variable moisture initial condition that vegetation integrates early and carries forward, which is why end-of-season phenology, and its residual (Section 3.2), can serve as a moisture proxy.

These relationships are based on six water years and are presented as exploratory motivation, not as a calibrated snow-soil model.

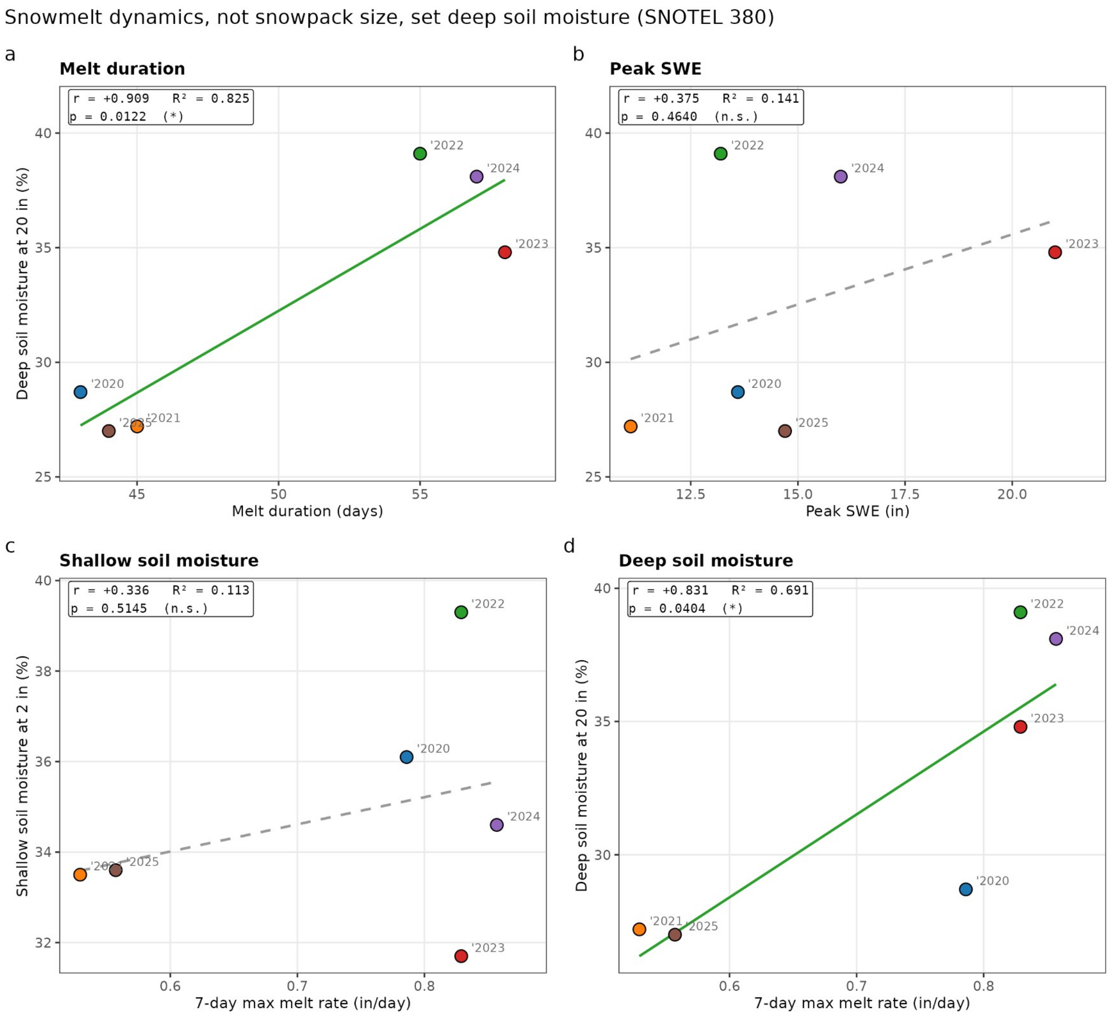

Figure 2. Snowmelt timing, not snowpack size, sets deep soil moisture at SNOTEL station 380 (water years 2020 to 2025, n = 6). Each point is one water year. (a) Deep soil moisture at 20 in versus melt duration. (b) Deep soil moisture versus peak SWE. (c) Shallow soil moisture at 2 in versus the 7 day maximum melt rate. (d) Deep soil moisture versus the 7 day maximum melt rate. Melt duration and intensity track deep recharge, while peak snowpack and shallow moisture do not.

3.2. The Phenology-Conditioned Moisture Anomaly (PCMA) residual product

The harmonic curve fitting (Stage 1) reduces each year's per-pixel NDVI curve to five metrics: June mean, September mean, peak value, peak timing, and the June-to-peak amplitude. PCMA asks how much of the late-season signal is predictable from early-season phenology and vegetation class, and treats the remainder as a moisture anomaly.

The PCMA regression (Equation (3)), which predicts late-season NDVI from each pixel's early-season phenology and vegetation class, fits well in every year. Fitting the per-year PCMA independently for each year, explained the large majority of the late-summer NDVI signal in every year from 2018 to 2025 (per-year adjusted R^2 0.585–0.833, mean 0.729). This predictability has a physical basis in the snowmelt record (Section 3.1): snowmelt sets the stage for early-season plant productivity, and that productivity helps determine where each pixel ends the growing season. Because the system is snowmelt-driven in this way (Figure 2; Dafflon et al.[17]), the season's growth trajectory, summarized by the PCMA phenology metrics, should itself carry a moisture signal, which helps explain why the PCMA regression attains high explanatory power and why its residuals merit interpretation. Because late-season greenness is this predictable from early-season phenology and vegetation class, we interpret the per-pixel residual (observed minus predicted September mean, Equation (4)) as a moisture-anomaly index: pixels greener or browner late in the season than their community and early-season trajectory would predict.

These anomalies are spatially organized and temporally persistent (Figure 3). The 2018–2025 mean residual map shows coherent wetter-than-expected and drier-than-expected patches rather than random noise, and a sign-consistency count, i.e., the number of years, of eight, in which a pixel retained the same residual sign, reaches at least 6 of 8 across 52% of pixels (Figure 3). The empirical semivariogram of the mean residual confirms this structure quantitatively: an exponential model fit gives a fitted range parameter of approximately 274 m (effective range ≈ 822 m) with roughly 65% of the variance structured rather than nugget (Figure 3). Together, the persistent residual sign, its multi-year stability, and this ~274 m structured range show the PCMA residual to be a spatially coherent, repeatable field rather than year-to-year noise, an interpretation we develop in the Discussion.

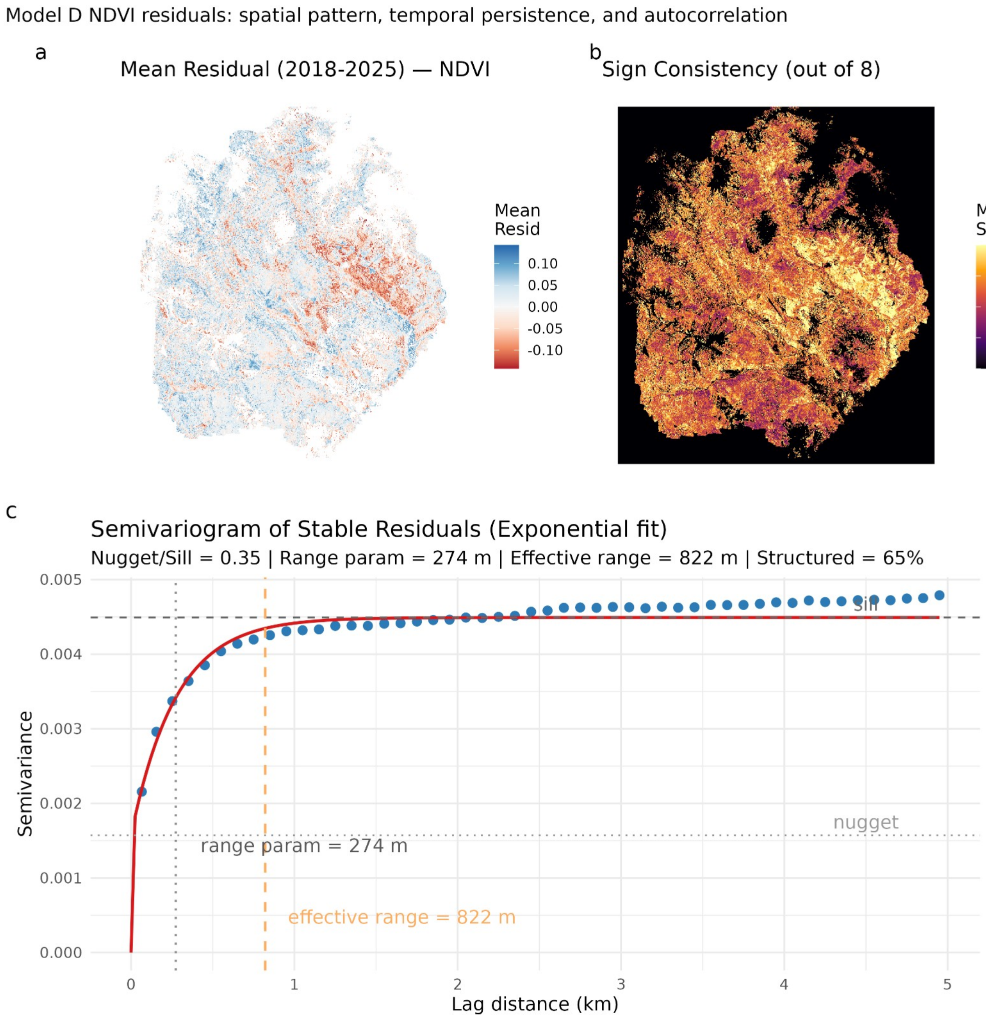

Figure 3. The phenology residual (PCMA) is spatially organized and persistent over 2018 to 2025. (a) Mean residual, with blue wetter than expected and red drier than expected. (b) Number of years out of eight in which each pixel kept the same residual sign. (c) Empirical semivariogram of the mean residual, showing spatial structure out to a range of about 274 m.

3.3. Where a spectral metric improves on topographic wetness alone

Before pooling the data, we asked, site by site, whether any spectral metric added explanatory power to topographic wetness for predicting point VWC. For each of 17 site-years, we compared a wetness-only fit with wetness plus that site-year's best spectral metric (Figure 4, with a 1:1 reference line). Adding a spectral metric helped most at sites where topographic wetness alone explained little: well-explained sites sat near the 1:1 line and moved little, while sites where the wetness-only model fit poorly rose furthest above it.

The sites that benefited most from adding spectral information did not belong to a single terrain class. The improvement tracked the weakness of the wetness-only baseline rather than specifically factors such as site slope or elevation, with no clean elevation or community pattern discernable from the site available. The gain was largest where topographic wetness alone explained little: at Bradley (2018, wet meadow), wetness-only skill of ϱ = 0.25 (adjusted R^2 = 0.20) rose to ϱ = 0.74 (adjusted R^2 = 0.73) once the PCMA residual was added, whereas already well-explained sites such as Slate Lower (2018, ϱ = 0.72 on wetness alone) gained almost nothing. The most helpful spectral metric was site-specific and showed no clear community pattern; no single metric was best everywhere: the PCMA residual was the most helpful single metric at 4 of the 17 site-years, an NDVI phenology metric at the other 13, which motivates testing whether a single metric can nonetheless carry a transferable signal once the data are pooled (Section 3.4).

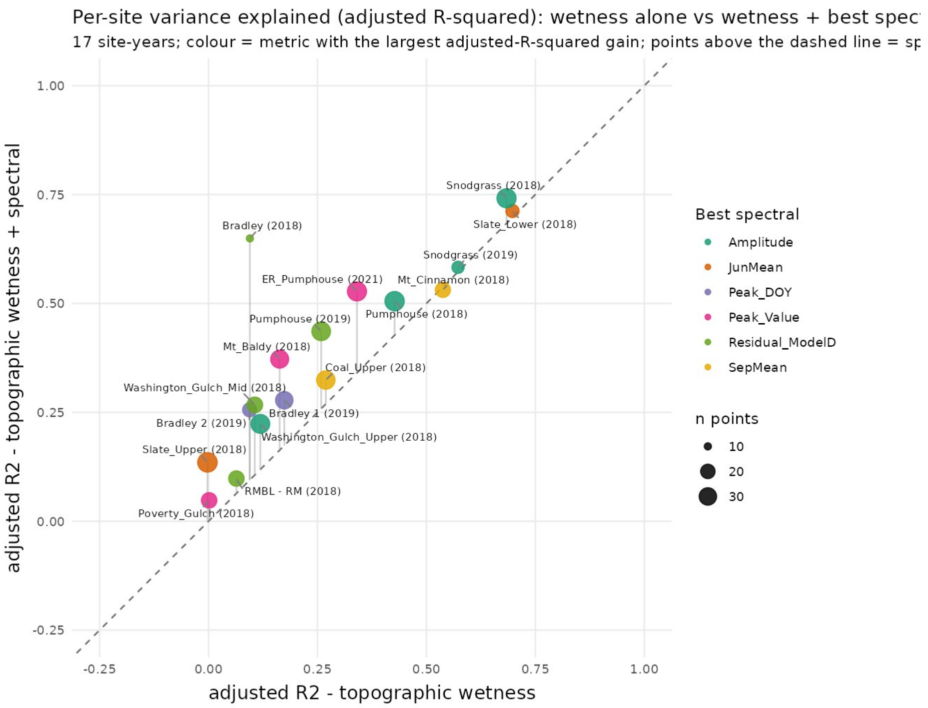

Figure 4. Per-site adjusted R-squared for a wetness-only fit (x axis) versus wetness plus the best NDVI metric (y axis) across 17 site-years, with a 1:1 line. Points above the line are sites where adding a spectral metric improved the fit, and the gain was largest where wetness alone explained little.

3.4. Pooled mixed-model skill and predictor transferability

3.4.1. Pooled model construction and community as a fixed effect

We fit a single pooled linear mixed model to the 2018 field data with a random intercept per site so that between-site offsets are not attributed to the fixed-effect predictors (see Equations 5 and 6). The Wetness Index is added as the baseline or null-model. The community fixed effect is motivated directly by the field data. Aggregating the 336 point observations to one mean per site (n = 12 sites across five communities), site-mean VWC ranged from 5.05% (Slate_Upper) to 16.49% (Mt_Baldy). The community label accounted for a sizeable share of this between-site spread descriptively (η^2 = 0.531), but with only 12 site means the difference was not statistically significant (one-way ANOVA, $F(4, 7) = 1.979$,

$p = 0.202$); we therefore retained community as a candidate fixed effect on mechanistic grounds rather than on a significance test. The effect size is large, but the test on 12 site means is underpowered, so this comparison cannot establish distinct moisture regimes on its own. We nonetheless carry community as a fixed effect because it is one of the three predictors we deploy and it improves fit (Section 3.4); this between-site check is independent of, and not a consequence of, the per-site phenology screen (Section 2.3.4).

3.4.2. Selecting the spectral predictor

Stages 2 and 3 of the analysis workflow derived several candidate spectral predictors from the fitted harmonic NDVI curves, along with the PCMA residual, which is itself constructed from those same phenology metrics via the community-phenology regression and retains only what the regression leaves unexplained (Equation (3), (4)). A central question of this study is which of these provide predictive skill to the soil moisture prediction model, alongside topographic wetness and community. We therefore entered the candidate spectral terms one at a time into the pooled model and scored each under leave-one-site-out (LOSO) cross-validation.

Peak Value was the strongest spectral candidate. It out-predicted the PCMA residual at held-out sites on every metric (LOSO R^2 0.340 vs 0.258; LOSO Spearman ρ 0.528 vs 0.464; RMSE 7.17 vs 7.68% VWC), and adding the PCMA residual on top of Peak Value did not improve out-of-site skill (Peak + Residual LOSO R^2 0.330, no gain over Peak Value alone).

Of the spectral metrics tested, Peak Value is therefore the best overall predictor of VWC in the pooled model. The simpler metric carries the transferable signal that the more elaborately derived residual does not improve on.

3.4.3. Leave-one-site-out prediction

We scored the selected model with Peak value as the spectral metric under LOSO across 335 complete-case observations from the 12 sites; held-out-site predictions used only the fixed effects (random intercept excluded), matching how the model is later applied to a map. Out-of-site skill was LOSO $R^2 = 0.340$, Spearman $\rho = 0.528$, and RMSE = 7.165% VWC (Figure 5). Because these metrics reflect performance at sites excluded from fitting, they represent the deployment-relevant skill on which the map products (Section 3.5) rest, though this remains cross-validation within a 12-site network rather than independent regional validation.

Figure 5 shows that the model preserves a useful out-of-site moisture ranking across held-out sites (Spearman $\rho = 0.528$), separating drier from wetter observations across communities. One systematic feature bounds this performance. The model underpredicts the wettest points: predicted VWC rarely exceeds about 24%, even where observed VWC is higher, so the highest-moisture observations are compressed downward, a ceiling the linear model cannot clear with these predictors. This effect narrows the predicted range relative to the observed range.

The 50 m site-footprint support used for the community label was checked for sensitivity to buffer size: rebuilding the label from raw 1 m NEON-campaign buffers at radii of 10, 25, and 50 m and comparing all variants by leave-one-site-out cross-validation of the 2018 pooled model, the 50 m site-footprint label achieved the highest out-of-sample skill (Spearman $\rho = 0.53$, versus 0.42–0.44 for the per-point variants) and was retained.

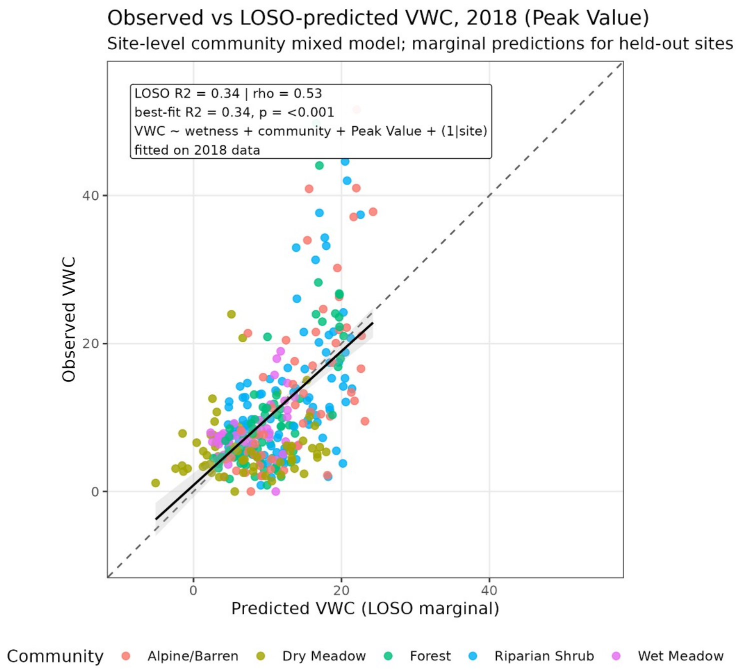

Figure 5. Leave-one-site-out observed versus predicted soil VWC for the 2018 pooled mixed model, colored by community (n = 335, 12 sites). Out-of-site skill was R-squared = 0.34, Spearman rho = 0.53, and RMSE = 7.17 percent VWC. The model pulls the wettest and driest values toward the middle of the range.

3.4.4. Drop-one predictor contributions

To quantify each predictor's contribution, we removed it, refit the model, and recorded the percentage of full-model skill lost, separately for in-sample and LOSO, and for two skill metrics (R^2 and Spearman ϱ). In-sample importance is measured by refitting and scoring on the same sites; leave-one-site-out (LOSO) importance is measured on held-out sites and is the relevant measure for deployment to unseen pixels (Section 3.4.3). We compare both because the expectation differs by regime: a predictor carrying a transferable signal should rank higher under LOSO than in-sample, whereas one fitting site-specific structure should do the reverse. Full-model references were in-sample R^2 = 0.415 and ϱ = 0.599, and LOSO R^2 = 0.340 and ϱ = 0.528 (n = 335). Under leave-one-site-out, Peak Value is the largest contributor on both metrics, with an R^2 loss of 40.3% (versus community 38.7% and topographic wetness 35.8%) and a ϱ loss of 25.1% (versus community 17.3% and topographic wetness 8.9%) (Figure 6). Looking at R^2, the Peak Value lead over community is narrow (40.3 vs 38.7%), but on Spearman ϱ the out-of-site advantage is clearer; for deployment to unseen pixels, this out-of-site skill is the relevant measure. We report the in-sample fit not as a deployment metric but as a diagnostic: Peak Value looking weak within a site yet strong across sites is direct evidence that it carries a transferable moisture signal rather than site-specific overfit.

Within site the ordering reverses. In-sample, Peak Value appears to be the least important: by in-sample R^2 loss it is the weakest of the three predictors (16.3%), behind topographic wetness (34.5%) and community (31.4%), and on in-sample ϱ it ranks below community (15.0% vs 19.6%). The spectral metric that adds least when a model is trained and scored on the same site's data (in-sample) thus adds most across sites. This within-site/across-site reversal (Figure 6), whose direction is the same whether skill is measured by R^2 or Spearman ϱ even as the predictor ranking itself flips between the in-sample and LOSO regimes, is the main field result.

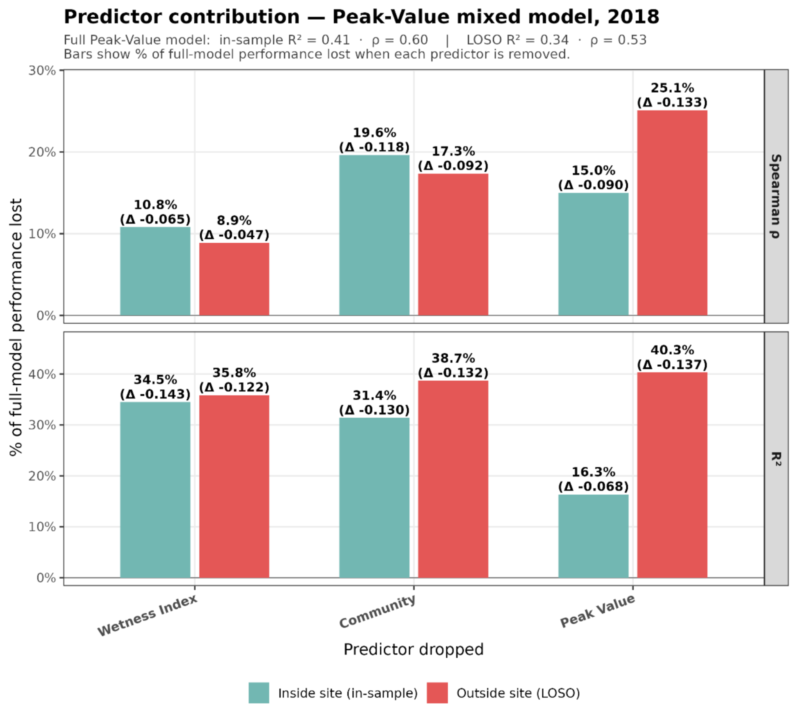

Figure 6. Drop-one predictor importance for the Peak Value mixed model, shown as the percent of model skill lost when each predictor is removed, both within sites and under leave-one-site-out. Peak Value matters least within a site but most when predicting at new sites.

3.5. Mapped VWC across the area of interest (AOI)

We applied the model fitted on the 2018 data across the Crested Butte study area for each year from 2018 to 2025 and summarized the eight annual 10 m predictions. The fitted peak-NDVI coefficient was +18.34% VWC per unit NDVI. The pixel-wise temporal mean (Figure 7) placed the wettest predictions in the valley bottoms and the driest on the hillslopes. Across 2,255,851 valid soil pixels (masked to the allowed Peak Value range $0 \leq \text{NDVI} \leq 1.10$), predicted mean VWC ranged from 0 to 29.06%, with a raster mean of 8.98%.

The lower tail of the map is floored at 0% by the clip-at-zero step, so the map should be read as a spatial ranking of wetter versus drier pixels in % VWC units rather than an exact VWC measurement at any single pixel.

Figure 7 maps the pixel-wise inter-annual coefficient of variation (temporal SD divided by temporal mean) for the same eight predictions. Its interpretation hinges on one point: at application, the topographic-wetness and community rasters are held fixed across years, so the only input that changes from year to year is the Sentinel-2 NDVI peak. The CV map therefore shows how sensitive the model's predictions are to year-to-year NDVI, serving as a spectral-transfer sensitivity surface. It is best read as that sensitivity diagnostic alone: model accuracy is reported separately as the LOSO skill in Section 3.4.3, and the surface is distinct from how much real field moisture varied.

Across all valid pixels, CV ranged from 1.9×10^{-6} to 2.83 (mean 0.144), with an average of 6.91 of 8 valid years per pixel. High CV marks low-mean or transitional pixels, where small absolute swings inflate the ratio. This effect is amplified at the low end by the clip-at-zero floor, which compresses the denominator.

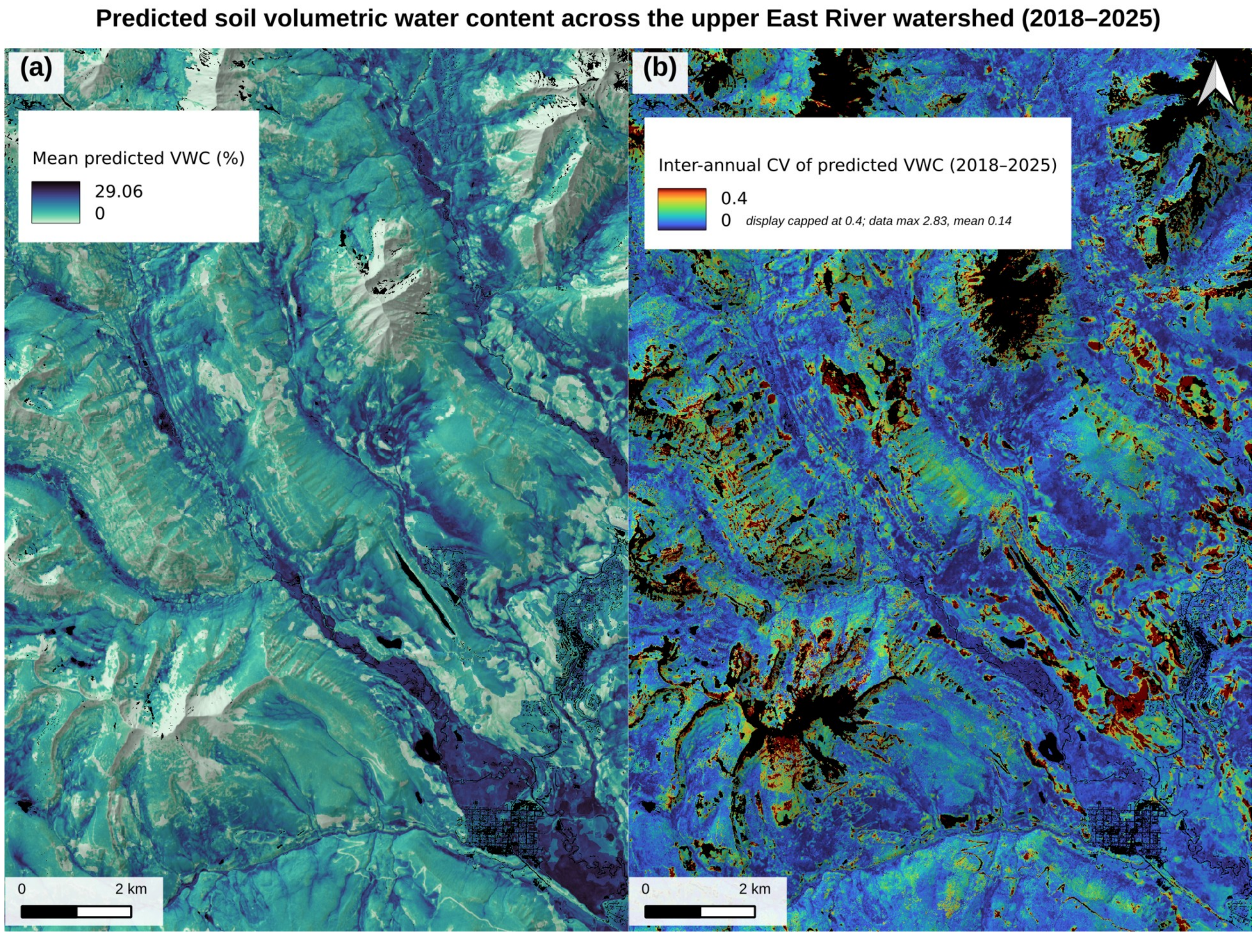

Figure 7. Predicted soil VWC across the study area, 2018 to 2025. (a) Mean predicted VWC, wettest in the valley bottoms and driest on the hillslopes (raster mean 8.98 percent). (b) Inter-annual coefficient of variation of the same predictions. Because wetness and community are held fixed across years, panel (b) reflects sensitivity to the yearly NDVI peak rather than model accuracy.

3.6. Spectral metric contribution along topographic gradients

Finally, we asked where the spectral term actually changes the prediction by comparing the full model (Peak Value + Wetness_Index + community, 2018–2025 mean; Equation (5)) with a null model that retains topographic wetness and community, but drops the spectral term (Wetness_Index + community, static 2018; Equation (7)). Whereas the per-site fits in Figure 4 each used that site's own best NDVI metric, the pooled model fixes the single metric that performs best across all sites, Peak Value; the contribution reported here is therefore the conservative, transfer-relevant counterpart of those per-site gains.

Summarizing both along elevation (25 m bins, approximately 2750–2975 m) and slope (5° bins, 0–55°) over the same 2,306 sampled soil-valid pixels (Figure 8), the full model sat above the null through the higher-elevation, moderate-slope terrain (at 2812.5 m, 8.63% full versus 4.78% null; at 22.5° slope, 8.79% versus 5.13%), while the two converged in the valley bottoms. The separation is largest on the steeper, higher terrain. In other words, at the pixel (10m) scale, topographic wetness explains much of the predicted moisture in the valley bottoms, while the spectral term extends the prediction to the higher, steeper terrain.

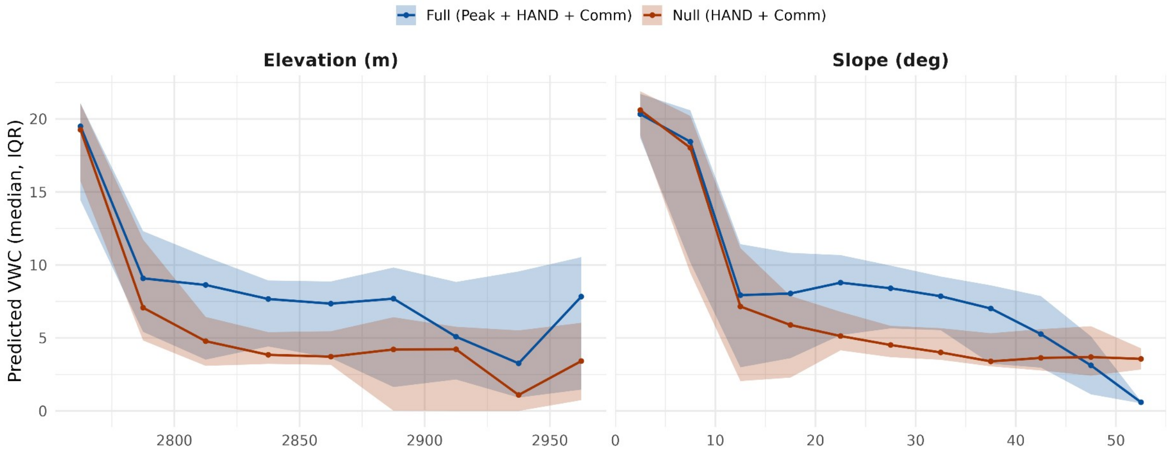

Figure 8. Median predicted VWC along elevation and slope gradients for the full model (Peak Value with wetness and community, blue) and a null model without the spectral term (red), with interquartile ribbons. The full model sits above the null on higher and steeper terrain, and the two converge in the valley bottoms.

4. Discussion

We set out to build a soil-moisture model from inexpensive, widely available inputs: topographic wetness, plant community, and a single Sentinel-2 derived phenology metric, fitted on a 2018 field campaign that provided direct soil moisture measurements and evaluated by leave-one-site-out cross-validation across twelve sites. The central result is an out-of-sample inversion. Under cross-validation Peak (NDVI) Value is the largest contributor to predictive skill (R^2 loss 40.3%, Spearman ϱ loss 25.1% when removed), yet within sites it ranks least important of the three predictors (in-sample R^2 loss 16.3%, behind topographic wetness at 34.5% and community at 31.4%) (Figure 6).

Peak Value was the best single spectral predictor of out-of-site VWC, the study's main finding. Of the phenology metrics tested (June mean, September mean, peak value, peak timing, amplitude), it outperformed every alternative at held-out sites (LOSO R^2 0.340 vs 0.258; ϱ 0.528 vs 0.464), and adding the PCMA residual provided no additional gain. The PCMA was a principled reorganization of those same metrics, conditioning late-season greenness on early-season phenology and vegetation class. While it carried relevant information, it finished as the runner-up to Peak Value. The comparison was made using pooled 2018 data, whereas the per-site screen found that the most helpful metric was site-specific (Figure 4), so Peak Value's primacy is a property of the signal that survives pooling, not a universal ranking. The residual is coherent only at the stand scale (fitted range about 274 m), so per pixel, it carries local noise, while the peak metric carries far less per pixel noise.

The reversal follows from how the model is fit and scored. The site random intercept absorbs each site's mean during fitting, so the fixed-effect slopes are estimated primarily from within-site variation; all reported predictions, in-sample and held-out alike, are made at the population level, with the random intercept set aside. Within a site the model was trained on, Peak Value is correlated with topographic wetness and community, so it adds little unique information and ranks last of the three predictors. At a held-out site, the coarse community class and the HAND-based wetness index reproduce only the site's approximate mean, while Peak Value (a continuous, canopy-integrated signal) carries the between-site moisture differences the others cannot recover, and so dominates transferable skill (LOSO R^2 = 0.340, ϱ = 0.528, RMSE = 7.17% VWC) ; Figure 5. A simple peak-greenness metric thus carries transferable soil-moisture information beyond topography and community, consistent with the critical-zone premise that the canopy integrates subsurface conditions [9,10,17].

These results support our hypothesis that phenology is an informative and necessary predictor of local (10m), deep (20cm) soil moisture. Deeper soil moisture, resulting from snowmelt dynamics that prolong melt rate and duration, regardless of the snowpack size, fuels early-season and peak productivity. As a result, September peak greenness integrates a deep-moisture state topographic wetness cannot fully encode. This is consistent with Dafflon et al. [17], who found that a slower-melting snowpack produced longer saturation and higher peak greenness here, and with the snow, plant, and soil co-dynamics reported by Devadoss et al. [18]. Our analysis of the snowmelt confirmed this to be the case at our study site. At SNOTEL 380, deep soil moisture (20 in) tracked melt duration ($R^2 = 0.825$, $p = 0.012$) and the 7-day maximum melt rate ($R^2 = 0.691$, $p = 0.040$) but not peak SWE ($R^2 = 0.141$, not significant), while shallow 2 in moisture was decoupled from melt intensity (Figure 2).

That the PCMA regression works at all is itself a result. Predicting late-season NDVI from early-season phenology and vegetation class alone explained the large majority of the late-summer signal in every year from 2018 to 2025 (adjusted R^2 0.585–0.833; Section 3.2). Such predictability is what the snowmelt premise anticipates: when meltwater rather than episodic summer rain sets the season's moisture, vegetation locks onto that initial condition early and its end-of-season state is largely determined from there. The PCMA residual isolates the part of the late-season canopy that this normal trajectory does not explain: the moisture anomaly. This is why a residual, rather than greenness itself, is the quantity of interest. As a point predictor the residual finished second to Peak Value, and we read that ordering as a scale effect rather than a failure of the premise: we hypothesize that the residual carries substantial pixel-level noise and retains temporally consistent spatial structure only at the coarser stand scale (fitted range ≈ 274 m), so that the phenological determinism that makes Peak Value transferable per pixel is expressed in the residual only once that stand-scale structure is resolved. We advance this as a hypothesis rather than a settled claim, to be tested against denser within-stand sampling.

The results presented here extend prior work in the East River watershed. Devadoss et al. [18] clustered NDVI series into moisture zones aligned with vegetation communities, and Wainwright et al. [19] showed that within-grassland spectral heterogeneity exceeded between-community heterogeneity and was explained by geology and topography. Both imply vegetation carries moisture information finer than land-cover class, but neither tested whether a spectral metric improves point VWC prediction at sites by the predictive model. We also reached a similar conclusion to Dafflon et al. [17] conclusion that peak greenness is the most informative single metric, but by an independent route, peak NDVI across twelve sites rather than peak GCC on one hillslope.

The spectral term contributes most where the topography-based wetness model is weakest. The wetness index follows the Height-Above-Nearest-Drainage construction [26], which penalizes slope, so it is inherently low on higher, steeper terrain; there the vegetation index carries complementary information, while in the valley bottoms, where wetness already explains most of the moisture, the two converge. The binned comparison shows exactly this: the full model performs better than the spectral-null on higher-elevation, moderate-slope terrain (8.6% vs 4.8% VWC at 2812 m) and performs similarly in the valley bottoms (Figure 8). This is the same pattern that the per-site screen showed, in which spectral gain was largest where wetness alone explained little (Figure 4).

In practice, the framework presented here flags where topography-only expectations are incomplete and where added sampling is warranted. It is broadly reproducible, since every input is free or common: Sentinel-2 for the community classification and the peak-greenness term, and a lidar DEM for the wetness model. Whether it transfers in skill remains to be tested, through regional validation beyond the twelve-site network, nonlinear or quantile models to clear the ceiling near 24% VWC, and a check of whether Peak Value's out-of-sample primacy holds in other snowmelt-driven montane basins.

Supplementary Materials: The following supporting information can be downloaded at: https://doi.org/10.5281/zenodo.21086627

Author Contributions: For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used "Conceptualization, X.X. and Y.Y.; methodology, X.X.; software, X.X.; validation, X.X., Y.Y. and Z.Z.; formal analysis, X.X.; investigation, X.X.; resources, X.X.; data curation, X.X.; writing—original draft preparation, X.X.; writing—review and editing, X.X.; visualization, X.X.; supervision, X.X.; project administration, X.X.; funding acquisition, Y.Y. All authors have read and agreed to the published version of the manuscript." Please turn to the CRediT taxonomy for the term explanation. Authorship must be limited to those who have contributed substantially to the work reported.

Funding: Please add: "This research received no external funding" or "This research was funded by NAME OF FUNDER, grant number XXX" and "The APC was funded by XXX". Check carefully that the details given are accurate and use the standard spelling of funding agency names at https://search.crossref.org/funding. Any errors may affect your future funding.

Data Availability Statement: Data Availability Statement: The analysis code, Jupyter notebooks, and derived figures supporting this study are openly available at https://github.com/turke6756/crested-butte-thesis (accessed on 30 June 2026) and permanently archived at https://doi.org/10.5281/zenodo.21086627. The underlying field, remote-sensing, and terrain datasets are third-party archives cited in Section 2.2: soil-moisture campaigns from the U.S. DOE ESS-DIVE repository (LBNL Watershed Function SFA), Sentinel-2 L2A imagery via Google Earth Engine, the Falco et al. vegetation classification, and the RMBL LiDAR-derived wetness index and DEM. Large raster and vector inputs are not redistributed in the code repository but are available from those cited sources.

AI Use Statement: During the preparation of this study, the author used Claude (Anthropic) to assist with the development and debugging of analysis scripts, including Python scripts for Sentinel-2 scene retrieval via Google Earth Engine and raster processing, and R code for harmonic phenology fitting, residual computation, and mixed-model analysis. The author reviewed and validated all generated code and takes full responsibility for the content of this publication.

Conflicts of Interest: Declare conflicts of interest or state "The authors declare no conflicts of interest." Authors must identify and declare any personal circumstances or interest that may be perceived as inappropriately influencing the representation or interpretation of reported research results. Any role of the funders in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; or in the decision to publish the results must be declared in this section. If there is no role, please state "The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results".

References

1. Seneviratne, S.I.; Corti, T.; Davin, E.L.; Hirschi, M.; Jaeger, E.B.; Lehner, I.; Orlowsky, B.; Teuling, A.J. Investigating Soil Moisture–Climate Interactions in a Changing Climate: A Review 2010, 99, 125.
2. Vereecken, H.; Huisman, J.A.; Bogena, H.; Vanderborght, J.; Vrugt, J.A.; Hopmans, J.W. On the Value of Soil Moisture Measurements in Vadose Zone Hydrology: A Review 2008, 44.
3. Keiluweit, M.; Wanzek, T.; Kleber, M.; Nico, P.; Fendorf, S. Anaerobic Microsites Have an Unaccounted Role in Soil Carbon Stabilization 2017, 8.
4. Angel, R.; Claus, P.; Conrad, R. Methanogenic Archaea Are Globally Ubiquitous in Aerated Soils and Become Active under Wet Anoxic Conditions 2011, 6, 847.
5. Kaiser, K.E.; Mcglynn, B.L.; Dore, J.E. Landscape Analysis of Soil Methane Flux across Complex Terrain 2018, 15, 3143.
6. Wickland, K.P.; Striegl, R.G.; Schmidt, S.K.; Mast, M.A. Methane Flux in Subalpine Wetland and Unsaturated Soils in the Southern Rocky Mountains 1999, 13, 101.
7. Hauer, F.R.; Locke, H.; Dreitz, V.J.; Hebblewhite, M.; Lowe, W.H.; Muhlfeld, C.C.; Nelson, C.R.; Proctor, M.F.; Rood, S.B. Gravel-Bed River Floodplains Are the Ecological Nexus of Glaciated Mountain Landscapes 2016, 2.
8. Hubbard, S.S.; Williams, K.H.; Agarwal, D.; Banfield, J.; Beller, H.; Bouskill, N.; Brodie, E.; Carroll, R.; Dafflon, B.; Dwivedi, D.; et al. The East River, Colorado, Watershed: A Mountainous Community Testbed for Improving Predictive Understanding of Multiscale Hydrological–Biogeochemical Dynamics 2018, 17, 1.
9. Cleverly, J.; Eamus, D.; Restrepo Coupe, N.; Chen, C.; Maes, W.; Li, L.; Faux, R.; Santini, N.S.; Rumman, R.; Yu, Q.; et al. Soil Moisture Controls on Phenology and Productivity in a Semi-Arid Critical Zone 2016, 568, 1227.
10. Chadwick, K.D.; Brodrick, P.G.; Grant, K.; Goulden, T.; Henderson, A.; Falco, N.; Wainwright, H.; Williams, K.H.; Bill, M.; Breckheimer, I.; et al. Integrating Airborne Remote Sensing and Field Campaigns for Ecology and Earth System Science 2020, 11, 1492.
11. Babaeian, E.; Sadeghi, M.; Franz, T.E.; Jones, S.; Tuller, M. Mapping Soil Moisture with the OPtical TRApezoid Model (OPTRAM) Based on Long-Term MODIS Observations 2018, 211, 425.
12. Entekhabi, D.; Njoku, E.G.; O'neill, P.E.; Kellogg, K.H.; Crow, W.T.; Edelstein, W.N.; Entin, J.K.; Goodman, S.D.; Jackson, T.J.; Johnson, J.; et al. The Soil Moisture Active Passive (SMAP) Mission 2010, 98, 704.
13. Bauer-Marschallinger, B.; Freeman, V.; Cao, S.; Paulik, C.; Schaufler, S.; Stachl, T.; Modanesi, S.; Massari, C.; Ciabatta, L.; Brocca, L.; et al. Toward Global Soil Moisture Monitoring With Sentinel-1: Harnessing Assets and Overcoming Obstacles 2019, 57, 520.
14. Zhang, D.; Zhou, G. Estimation of Soil Moisture from Optical and Thermal Remote Sensing: A Review 2016, 16.
15. Sadeghi, M.; Babaeian, E.; Tuller, M.; Jones, S.B. The Optical Trapezoid Model: A Novel Approach to Remote Sensing of Soil Moisture Applied to Sentinel-2 and Landsat-8 Observations 2017, 198, 52.
16. Stern, M.; Ferrell, R.; Flint, L.; Kozanitas, M.; Ackerly, D.; Elston, J.; Stachura, M.; Dai, E.; Thorne, J. Fine-Scale Surficial Soil Moisture Mapping Using UAS-Based L-Band Remote Sensing in a Mixed Oak-Grassland Landscape 2024, 5.
17. Dafflon, B.; Léger, E.; Falco, N.; Wainwright, H.M.; Peterson, J.; Chen, J.; Williams, K.H.; Hubbard, S.S. Advanced Monitoring of Soil-Vegetation Co-Dynamics Reveals the Successive Controls of Snowmelt on Soil Moisture and on Plant Seasonal Dynamics in a Mountainous Watershed 2023, 11.
18. Devadoss, J.; Falco, N.; Dafflon, B.; Wu, Y.; Franklin, M.; Hermes, A.; Hinckley, E.-L.S.; Wainwright, H. Remote Sensing-Informed Zonation for Understanding Snow, Plant and Soil Moisture Dynamics within a Mountain Ecosystem 2020, 12.
19. Wainwright, H.M.; Steefel, C.; Trutner, S.D.; Henderson, A.N.; Nikolopoulos, E.I.; Wilmer, C.F.; Chadwick, K.D.; Falco, N.; Schaettle, K.B.; Brown, J.B.; et al. Satellite-Derived Foresummer Drought Sensitivity of Plant Productivity in Rocky Mountain Headwater Catchments: Spatial Heterogeneity and Geological-Geomorphological Control 2020, 15.
20. Eamus, D.; Zolfaghar, S.; Villalobos-Vega, R.; Cleverly, J.; Huete, A. Groundwater-Dependent Ecosystems: Recent Insights from Satellite and Field-Based Studies 2015, 19, 4229.
21. Rohde, M.M.; Albano, C.M.; Huggins, X.; Klausmeyer, K.R.; Morton, C.; Sharman, A.; Zaveri, E.; Saito, L.; Freed, Z.; Howard, J.K.; et al. Groundwater-Dependent Ecosystem Map Exposes Global Dryland Protection Needs 2024, 632, 101.

22. Elmore, A.J.; Manning, S.J.; Mustard, J.F.; Craine, J.M. Decline in Alkali Meadow Vegetation Cover in California: The Effects of Groundwater Extraction and Drought 2006, 43, 770.
23. Evans, J.; Geerken, R. Discrimination between Climate and Human-Induced Dryland Degradation 2004, 57, 535.
24. Chen, H.; Liu, X.; Ding, C.; Huang, F. Phenology-Based Residual Trend Analysis of MODIS-NDVI Time Series for Assessing Human-Induced Land Degradation 2018, 18.
25. Carroll, R.W.H.; Deems, J.S.; Niswonger, R.; Schumer, R.; Williams, K.H. The Importance of Interflow to Groundwater Recharge in a Snowmelt-Dominated Headwater Basin 2019, 46, 5899.
26. Nobre, A.D.; Cuartas, L.A.; Hodnett, M.; Rennó, C.D.; Rodrigues, G.; Silveira, A.; Waterloo, M.; Saleska, S. Height Above the Nearest Drainage – a Hydrologically Relevant New Terrain Model 2011, 404, 13.
27. Falco, N.; Lamb, J.; Chen, J.; Tuvshintugs, O.; Balde, A.; Dafflon, B.; Wainwright, H.M. Geophysical Survey Associated with NEON AOP Survey, East River, CO 2018 Available online: https://doi.org/10.15485/3013006.
28. Falco, N.; Dafflon, B.; Devadoss, J.; Shirley, I.; Soom, F.; Uhlemann, S.; Wainwright, H. Time-Domain Reflectometer Survey across the East River Watershed, Colorado Available online: https://doi.org/10.15485/1648526.
29. Dafflon, B. Soil Moisture and Temperature from 2019 to 2024 along Northeast- and Southwest-Facing Hillslopes at the Lower Montane Site in the East River Watershed, Colorado Available online: https://doi.org/10.15485/2566877.
30. European, S.A. Copernicus Sentinel-2 MSI Level-2A Surface Reflectance, Harmonized Available online: https://developers.Google Scholar.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR_HARMONIZED.
31. Falco, N.; Wainwright, H.M.; Chadwick, K.D.; Dafflon, B.; Enquist, B.J.; Uhlemann, S.; Breckheimer, I.; Lamb, J.; Chen, J.; Tuvshintugs, O.; et al. Vegetation Classification Map and Covariates Associated with NEON AOP Survey, East River, CO 2018 Available online: https://doi.org/10.15485/1602034.
32. Breckheimer, I. 1 m Resolution Multiscale Height-above-Stream Wetness Index for the Upper Gunnison Domain Available online: https://www.rmbl.org/scientists/resources/data-catalog/data-catalog-entry/?catalog-id=102.
33. Breckheimer, I. 1 m Resolution Digital Elevation Model for the Upper Gunnison Domain Derived from 2015 and 2019 LiDAR Data Available online: https://www.rmbl.org/scientists/resources/data-catalog/data-catalog-entry/?catalog-id=83.
34. Natural Resources, C.S. Butte SNOTEL Station (Station 380), Colorado: Snow Water Equivalent and Soil Moisture Available online: https://wcc.sc.egov.usda.gov/nwcc/site?sitenum=380.
35. Dufrêne, M.; Legendre, P. SPECIES ASSEMBLAGES AND INDICATOR SPECIES:THE NEED FOR A FLEXIBLE ASYMMETRICAL APPROACH 1997, 67, 345.
36. Sharma, R.C. Dominant Species-Physiognomy-Ecological (DSPE) System for the Classification of Plant Ecological Communities from Remote Sensing Images 2022, 3, 323.
37. Kuchler-AnalyzingPhysiognomyStructure-1966.