Foresummer drought sensitivity of subalpine willow (*Salix*) shrublands: functional trait and topographic controls on spatial pattern

Adam Gentry

Rocky Mountain Biological Laboratory education program, final paper, summer 2026

Abstract

High-elevation ecosystems are warming faster than surrounding lowlands.n the Elk Mountain range of the Rocky Mountains, earlier snowmelt and a weakening summer monsoon are projected to lengthen the foresummer drought, the dry window between snowmelt and monsoon onset. Foresummer drought is known to control productivity in subalpine meadows, but the riparian willow (*Salix* spp.) shrublands that line valley bottoms have never been assessed directly: satellite studies are too coarse of scale and fine-scale mapping has been non-temporal. I combined a six-year record of drone multispectral and thermal imagery (2021–2026, 116 flights) over the Gothic townsite at the Rocky Mountain Biological Laboratory with field measurements of species identity and functional traits at 144 willow crowns to ask where willow sensitivity to foresummer drought is highest and whether topography, species, and traits explain that pattern. Drone and field measurements agreed well ($R^2 = 0.85$), and the six-year decline measured on hand-delineated crowns matched the automated segmentation ($r = 0.979$). Willow canopies declined dramatically since 2021: median canopy height fell 36% (81.7% of 27,991 segments declining), resulting in an implied leaf-area loss of ~20% (median LAI 2.72 → 2.17 $m^2 m^{-2}$). The decline correlated with neither same-year climate ($p = 0.85$) nor any measured terrain variable (spatially blocked $R^2 = 0.011$); only previous-year climate response was suggestive ($p = 0.082$). Greenness carried almost no structural information, but showed a pheneological response: Dry years had a shift in timing of green-up by roughly two weeks ($r = +0.95$) while season length was unaffected and varied between (97–102 days), integrated greenness stayed essentially the same across wet and dry years. Proportional decline differed by species, with taller wet-site species losing height up to seven times faster than short dry-site species (10.9 vs 1.5% per year, $p = 0.011$), while no measured trait predicted decline within species. These results help to identify which willow species are vulnerable as the foresummer lengthens. Because willow species account for a large portion of summer evapotranspiration and supply the habitat and food needed for bird and beaver species, a loss of one third canopy median height in 6 years is important to continue monitoring. Furthermore, this information can be used to inform future sampling, characterization, and modeling studies.

1. Introduction

Climate change poses a disproportionate threat to high-elevation ecosystems, where warming has occurred faster than in lower-elevation ecosystems (Mountain Research Initiative EDW Working Group, 2015; Nogués-Bravo et al., 2007). High-elevation ecosystems therefore offer a preview of plant

communities' reassembly and functional change under climatic conditions expected late this century (Dunne et al., 2003; Harte & Shaw, 1995). In temperate mountain systems this warming is especially important as snowmelt is the primary early-summer water source, and the timing of snowmelt and rain organizes plant performance.

The western United States has already lost 10–20% of maximum snow water equivalent since the 1980s, with up to a further ~60% loss projected within three decades (Fyfe et al., 2017), and models project a shift toward extreme low-snow years and earlier melt-driven runoff (Diffenbaugh et al., 2013). Additionally, the North American monsoon is projected to weaken (Pascale et al., 2017). Earlier snowmelt plus a weaker monsoon could lengthen the foresummer drought in the southern Rocky Mountains. Foresummer drought is the dry period between snowmelt-driven soil-moisture depletion and monsoon onset (typically early to mid-July).

Foresummer drought strongly influences subalpine plant performance. At the Rocky Mountain Biological Laboratory (RMBL), Sloat et al. (2015) combined 11 years of carbon-exchange monitoring with a watering manipulation and showed that the timing of growing-season water matters more than its amount: plants unwatered through June were stunted, reached critical predawn water potentials, and carried reduced net ecosystem productivity even after monsoon rains arrived. Long-term RMBL climate manipulation experiments have shown that experimental heating advanced snowmelt and dried summer soils, shifted biomass from forbs toward shrubs (Harte & Shaw, 1995), and after 23 years the unheated control plots had begun tracking the heated ones, with ambient snowmelt advance driving woody encroachment (Harte et al., 2015). Shifts in community traits are nonetheless not monolithic: the encroaching shrub in that experiment is drought-adapted sagebrush. How a wet-adapted shrub community responds to a drying foresummer is therefore an open question, and plant functional traits tied to the leaf economics spectrum (Reich, 2014; Wright et al., 2004) offer a framework for answering it.

Willows are an important and understudied community to ask that question. Despite roughly 75 years of research at RMBL, willows appear in the local literature almost exclusively as bird habitat andbeaver forage; no study has measured their drought response. What studies do exist suggest that Congeneric willows, willows of different species, sort along hydrologic microgradients by habitat affinity (Cottrell, 1996) and diverge in drought survival strategy, with dry-habitat species growing fast at high water-use efficiency while wet-habitat specialists tolerate less soil drying (Savage & Cavender-Bares, 2011; Savage et al., 2009). At biome scale, the climate sensitivity of shrub growth is itself spatially structured, greatest at moister sites and for taller shrubs (Myers-Smith et al., 2015). What research at RMBL has shown is that riparian *Salix* shrublands occupy the valley bottoms (Langenheim, 1962) and are quite dynamic in this landscape: a 65-year resurvey of Langenheim's transects found *Salix* increasing in upland herbaceous communities as species shifted upslope (Zorio et al., 2016), a similar shift is happening in tundra ecosystems in which expansion of *Salix* is a focal genus (Elmendorf et al., 2012; Myers-Smith et al., 2011). Willow ecology is also important hydrologically: modeling and flux towers in the western US have shown that riparian vegetation accounts for 20-33% of summer Evapotranspiration off rivers (Dahm et al., 2002). And two decades of experimental work in Yellowstone showed that water tables constrain willow height growth as strongly as browsing (Bilyeu et al., 2008; Johnston et al., 2011), with topographic wetness affecting stem growth as much as elk abundance (Marshall et al., 2014).

There is also a gap of scale. Wainwright et al. (2020) mapped foresummer drought sensitivity across the upper East River watershed as the regression slope of peak Landsat NDVI on June PDSI (1992–2010, 30

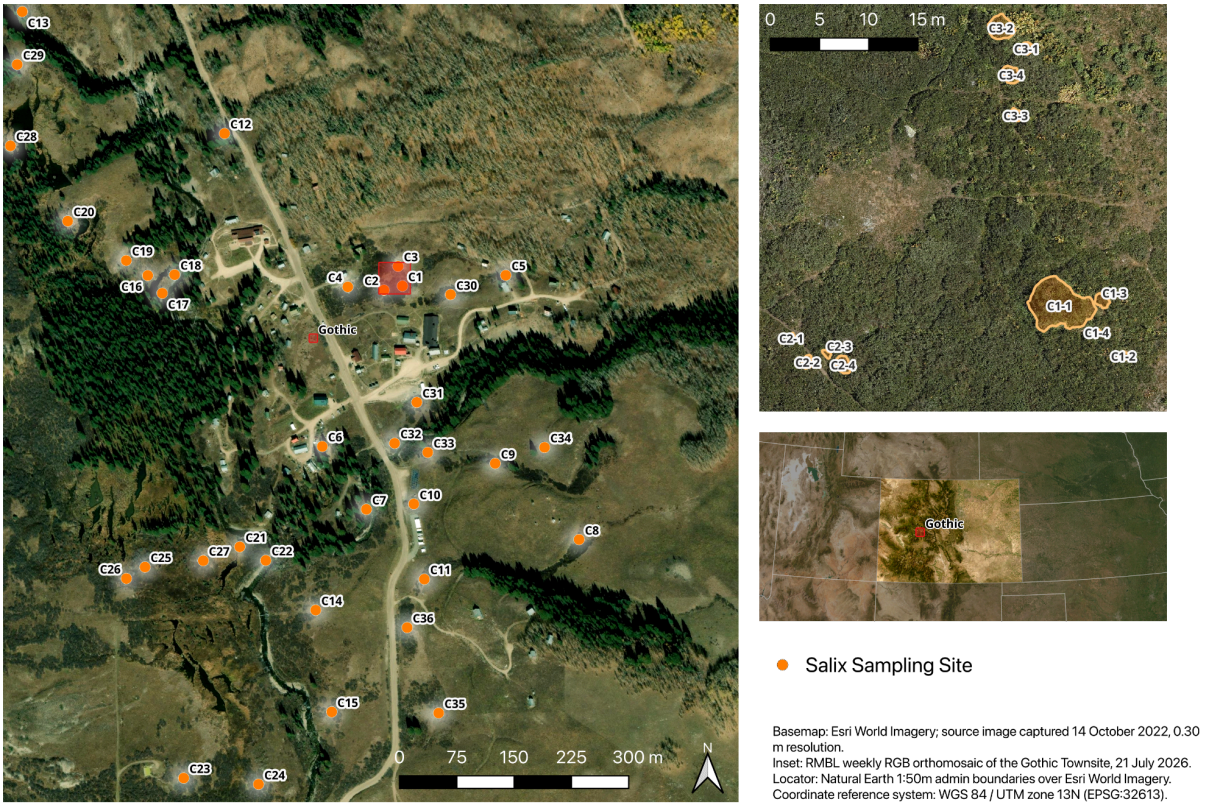
m), finding sensitivity controlled by plant type and elevation. Their classification, however, contains no willow class, and was at a coarse 30 m resolution. Falco et al. (2019) classified the East River's shrub and meadow communities at submeter resolution (shrub classification accuracy rising from 60% to 85% with LiDAR fusion) and showed that soil moisture, controlled by microtopography, drives where communities are spatially distributed. Their study however was only a static snapshot with no temporal dimension and no sensitivity metric. There is a substantial amount of information lost between the scales used by Falco et al. (2019) and Wainwright et al. (2020) but read together these two studies offer a framework for bridging plot-scale observations to larger community monitoring and read together present an unexplored opportunity for fine scale multitemporal drought sensitivity in willow shrublands. In this study I combined a six-year record of drone multispectral and thermal imagery (5cm scale, 2021–2026, 116 flights) over the Gothic townsite at the Rocky Mountain Biological Laboratory with field measurements of species identity and functional traits at 144 willow crowns. I ask: **(Q1)** Where in the Gothic townsite is *Salix* communities' sensitivity to foresummer drought highest over 2021–2026, and how has willow leaf area trended over that period? **(Q2)** How do topographic setting and remotely sensed canopy temperature predict drought sensitivity across the townsite, and do individuals' species identities and functional traits explain this sensitivity pattern?

2. Methods

2.1 Study area

The study area is a 900 × 1200 m rectangular region centered over the campus of the Rocky Mountain Biological Laboratory in Gothic, Colorado, USA (corners 38°57′55.1″N 106°59′44.2″W to 38°57′07.7″N 106°58′52.8″W). This is a subalpine site at ~2,900 m elevation in the upper East River valley and a heavily instrumented mountain "community testbed" (Hubbard et al., 2018). Valley bottoms support riparian shrub communities dominated by willow (*Salix* spp.) within a mix of herbaceous subalpine meadow, grassland, and Engelmann spruce, subalpine fir, and aspen forest; the vegetation in this basin has been described since Langenheim (1962).

Figure 1. The study area and field sampling design: the 900 × 1200 m Gothic townsite domain with cLHS-selected field sites spread across topographic wetness, snow persistence, and plant vigor.



2.2 UAS imagery

Imagery comes from the RMBL's uncrewed aircraft system (UAS) program: a MicaSense Altum v1 multispectral and thermal sensor (AgEagle Corp.) on a quadcopter yielding ~5 cm ground resolution for five multispectral bands and ~50 cm for radiometric surface temperature. Flights were conducted approximately weekly from the start of seasonal snowpack disappearance to the onset of permanent fall snow, within 3 hours of solar noon and in full-sun conditions where possible, on NNW–SSE flightlines at 28 m spacing progressing west to east (~75 min per flight, so western edges are imaged ~1 h before eastern edges). Orthomosaics were georeferenced in Agisoft Metashape (v1.7) using RTK GNSS. Radiometric calibration used the sensor's DLS2 irradiance sensor plus calibration-panel images before and after each flight, implementing the Micasense-reccomended calibration protocol formultispectral drone images. The analysis uses data collected from 116 flights: weekly coverage in 2022–2025 (27–30 flights per year, snowmelt to fall snow) plus single-date annual orthomosaics in 2021 (n = 2) and 2026 (n = 3, July 6–21). Canopy height models (CHMs; 5 cm, one per year, 2021–2026) were derived photogrammetrically with a DEM, using structure-from-motion. Band identity was verified against the physical constraint red < red-edge < NIR over vegetation (the data-platform documentation transposes NIR and red-edge relative to the files).

2.3 Terrain, snow, and climate layers

Topographic wetness is the RMBL SDP (Rocky Mountain Biological Laboratory Spatial Data Hub) *Multiscale Height-above-stream Wetness Index* (1 m) . This layer gives each $1m^2$ pixel a value based on its elevation above and distance from the nearest stream channel (Nobre et al., 2011). Snowmelt timing is the SDP *Spring Snowpack Persistence Mean Day of Year* product (27 m; 1993–2022). A value is given to each pixel based off of the mean day of year when the seasonal snowpack disappears and bare ground first appears, averaged over 1993–2022 and estimated from Landsat imagery (Gao et al., 2021). Additional terrain predictors (elevation, slope, eastness, northness, topographic position index (TPI), curvature) were computed from the DEM.

2.4 Field campaign

Field sites were selected by conditioned Latin hypercube sampling (cLHS; Minasny & McBratney, 2006) over three axes: topographic wetness, mean spring snowpack persistence, and plant vigor. Plant vigor was computed as the peak-season NDGI (described in §2.6) per willow segment from a 2025 orthomosaic. 40 sites were selected whose environmental spread mirrored the full study area. Only 36 sites were visited however an adequate amount of environmental variation was still achieved. In July 2026 I hand-delineated 144 willow crowns with a Trimble Catalyst DA2 GNSS (2 cm precision), tracing each crown as a polygon with a unique identifier and diagnostic photos. For each willow I recorded species identity (vouchers to the RMBL Herbarium) and measured maximum photosynthetic height. Additionally, I measured leaf water content (LWC) and leaf mass per area (LMA) from a stem sample making up 15-90 leaves. Leaf area index was measured with a LI-COR LAI-2200C plant canopy analyzer (one above-canopy and at least four below-canopy readings with the 90° cap, wand oriented north; k-record procedure under direct sun).

2.5 Analysis units

The analyses in this paper use the townsite willow at three different scales:144 field crowns, 27,991 willow segments and 900 25m gridded cells. Collection of field crowns information was described in (§2.4). To produce willow segments of the entire townsite I used a segment algorithm based on Dalponte and Coomes (2016). First a 5cm CHM (canopy height model) is used to find the maximum height per 5cm pixel and is then coarsened to 25cm CHM to lessen noise from individual branches or small clumps being mistaken as separate crown tops. Then, the algorithm checks each 25cm pixel that is at least 0.6m tall and checks that it is the highest within a circle centered on it and counts that pixel as a crown top if it is the tallest. From each top a crown is grown outward, adding pixels nearby until another crown is met or until the willow canopy stops. Polygons made up of more than half willow pixels, determined by a 5cm land classification map of the townsite, kept ending with 27,991 segments. The same segments created with the 2021 CHM are used in all six years so a polygon cannot grow or shrink. However, these segments would often conflict with the hand delineated field crowns with the median field crown containing 2 segments and the maximum containing 10 meaning many individual segments were multiple observations of the same plant. To mitigate this problem 900 25-m gridded cells were used. To create these a 25m grid was generated over the townsite and kept only the cells that had at least one willow segment ending up with 900. All other pixels that were not willow removed from the cell meaning the

cells value for any given value is the average of the willow segments inside the cell. How models on these cells are validated is described in (§2.6).

2.6 Responses and Statistics

Three response variables are measured with the units of (§2.5). Median canopy height is measured per segment every year. Decline is the slope of median height on year, then fit per segment for the overall townsite trend and decline map (§3.1), per field crown for validation and species comparisons where decline is expressed as a percent of height loss per year and per 25m cell for testing whether terrain explains where decline was fastest.

Seasonal greenness is computed per cell from each year's flights. Season greenness curves use NDGI (normalized difference greenness index) (Yang et al., 2019), an index designed to mitigate signal contamination from snow. For each cell, green up and senescence dates are determined where the curve crosses half of its seasonal amplitude each year. Seasonal amplitude is calculated by taking peak summer greenness value - lowest seasonal greenness value. Season length is the amount of days in between green up date and senescence date and seasonal integral is the area under the seasonal curve, or total seasonal greenness. In order to trace a greenness curve to generate these results I used the four fully flown seasons (2022-2025).

Field measured LAI was calibrated through regression against median crown height by and validated by leave-one-site-out (LOSO) cross-validation. LOSO refits the regression 36 times leaving a different site out each time and uses the regression to predict the left out sites crowns $r = 0.64$, RMSE = 0.77 $m^2\ m^{-2}$). Because sites share environmental conditions, removing an entire site from training data removes the ability for the model to predict LAI from the other willows at that site. Which, when applied across the entire townsite would be information the model would not have. The final regression, fit on all 144 crowns, is LAI = 1.182 + 1.206 × median crown height. For comparison, a random forest (20,000 trees) was fit and validated the same way. A random forest is a prediction model built from many decision trees: each tree predicts through a series of yes/no questions on the predictors e.g. (is height above 1.1 m? was snowmelt before day 150?) each tree is trained on a random subset of the data and the forest averages their predictions, so it can fit nonlinear relationships without assuming a straight line (Breiman, 2001). This is the random forest behind the “kitchen sink” model which was given all 672 structural, spectral, texture, phenology, and thermal predictors (§3.1).

Species classification. Species identity was extended from the 144 field crowns to the townsite with a random-forest classifier (1,000 trees) using class balanced sampling. Class balanced sampling is when each decision tree trains on equal numbers of crowns from each species so the model does not gain accuracy by favoring common species when guessing. The random forest used spectral, texture ,structure, and phenological predictors and was validated with LOSO cross validation. The model was fit on all 6 species and was refit on a pooled 4 species version which was then mapped. The pooled model assigns a label to every segment with a complete set of predictors, 27,918 of the 27,991; accuracies, the baseline comparison, and per-class recall are reported in (§3.1).

Spatial models used random forests (600 trees) validated by spatially blocked cross validation. Spatially blocked cross validation, the process of dividing the townsite into 250m tiles and withholds all the cells, if any, inside a tile, prevents spatial autocorrelation by having every held-out cell be predicted with data

elsewhere in the townsite. Spatial autocorrelation is when nearby cells or segments share similar terrain or willow species, so a model trained on non-spatially blocked units can seem to be predicting from terrain when it is really recognizing a cell's neighbors and their values. The 250m size comes from a variogram, which measures how far cells' resemblance extends. These 250m tiles are randomly held out by being shuffled into 5 random groups from which the model picks 1 group at a time to hold out. The model trains on the cells in the other four groups and predicts the cells in the held-out group, so every cell is predicted once by a model that never saw its tile. The shuffle is then redone and the validation rerun, 12 times in total; reported scores are the median of the 12 runs, with the range across runs shown as the whiskers.

Traits were related to proportional median canopy height decline with linear mixed models with site random intercepts, because crowns share site conditions (Table 1). LMA, LWC and initial plant height enter as two predictors, the species mean and each crown's deviation from that mean, so the between-species and within-species associations are estimated separately, with effects expressed per standard deviation of the trait. The within-species slope also uses bootstrapping to calculate a 95% confidence interval.

Canopy heights' relationship to climate was related with two permutations tests, one pairing each season's canopy with its own climate and one pairing it with the climate of the season before. Before the permutation tests canopy height and each index were detrended to remove any long term trend agreement. Climate or drought severity measured with seven different drought indices spanning different lengths of moisture history. Four are SPEI values (Vicente-Serrano et al., 2010). SPEI is precipitation minus evaporative demand, summed over a chosen window, the four different windows used here were: (May-June), 3 months (April-June), 6 months (January-June), and 12 months (the full year through June). The fifth is June PDSI, matching the foresummer drought variable of Wainwright et al. (2020). PDSI is a running water balance in which each month keeps roughly 90% of the previous month's value. This means PDSI is more representative of summarizing accumulated water, like winter snowpack, than SPEI. The sixth is foresummer water deficit: precipitation minus evaporative demand, summed over that same April 15 – June 15 window and standardized against the 1979–2026 record, so higher values are wetter. The seventh is a combined index of standardized winter precipitation and foresummer water deficit. All indices were computed from gridMET climate data (Abatzoglou, 2013).

Species proportional decline differences were tested with a permutation test. To remove any effect of species identity the 144 field crowns species labels were shuffled at random 999 times while keeping its original decline rate to remove any effect of species identity. The test only used species with 15 or more crowns, therefore excluding *S. glauca* and *S. planifolia*, to prevent the small number of samples adding noise.

Site selection used the clhs package (Roudier et al., 2012) and the drought indices were computed with the SPEI package (Beguería & Vicente-Serrano, 2023). The orthomosaic time series (§2.2) was accessed with the rSDP package (Breckheimer, 2026). Analyses were run in R 4.6.1 (R Core Team, 2026), with data.table (Barrett et al., 2026), terra (Hijmans et al., 2026), sf (Pebesma, 2018), and exactextractr (Baston, 2025) for data handling and extraction, ranger (Wright & Ziegler, 2017) for random forests, lme4 (Bates et al., 2015) with lmerTest (Kuznetsova et al., 2017) for the mixed models, and ggplot2 (Wickham, 2016) with patchwork (Pedersen, 2025) for figures.

3. Results

3.1 Field Validation

Drone and field measurements agreed well (Fig. 2). The CHMs maximum height per crown reproduced field measurements of maximum photosynthetic height well at $R^2 = 0.85$. While CHM derived measurements averaged 0.34m lower than field measurements this was expected as the thin uppermost leaders are often missed by imagery. Because the offset is consistent there is no problem with comparisons or trends.

To estimate LAI across the townsite field, measured LAI was calibrated with the canopy height data (§2.6). The model used LOSO and predicted LAI at (at $r = 0.64$ (RMSE = 0.77 $m^2\ m^{-2}$), and the kitchen sink random forest model using all 672 predictors improved on it only modestly ($R^2 = 0.447$ against 0.369), with the 648 non-height predictors alone predicting an (R^2 of 0.005) (Fig. 3). Across the 27,991 segments median LAI fell from 2.72 to 2.17$m^2\ m^{-2}$,a loss of 20.4% (Fig. 4). Because LAI over time was estimated from canopy median height alone, this is the height trend as LAI rather than a separate estimate of decline.

Species identity was predicted across the townsite with a random forest classifier trained on the 144 field identified crowns and validated with LOSO (§2.6). Over all six willow species identified at the 36 sites the classifier reached an accuracy of 0.57 against a basic chance of 0.17 and a baseline accuracy of 0.30, the score from guessing the most abundant species *S. drummondiana*. The three least abundant species *S. boothii*, *S. planifolia*, and *S. glauca* (accounting for 19, 9, and 8 crowns respectively) did not have enough identified crowns for the model to effectively learn their species information, with recalls of 0.16, 0.22 and 0.25 respectively. Therefore, they were pooled into an "other Salix" class and the model was refit on the new four class dataset model and reached an accuracy of 0.62 against the baseline accuracy of 0.30 showing a modest .05 improvement in accuracy (Fig. 5). The confusion matrix shows that *S. monticola* is most often misidentified as *S. drummondiana*, whereas S. wolfii is rarely misidentified (Fig. 5).Pooling also allows for the rarer willow species to remain on the map as an uncertain class rather than risk them being relabeled as an incorrect, more common species that they resemble. *S. wolfii* is abundant in the east of the townsite, *S. drummondiana* is distributed throughout with large patches in the southwest and west, *S. monticola* is evenly spread throughout and the pooled class is more abundant in the west of the townsite (Fig. 6). While reading the map it is important to consider each classes recall: *S. wolfii* (0.82), *S. drummondiana* (0.65), *S. monticola* (0.44) and pooled other salix (0.53). Because the map identifies each segment equally a 0.44 *S. monticola* segment is presented in the same way as a 0.82 *S. wolfii* segment which could lead to overconfidence if not read with recall values in mind.

Figure 2. Field validation of the canopy height model at the 144 field crowns, colored by species. Dashed line is 1:1 agreement, solid line shows the fitted regression for each CHM height statistic (max, p99, p95, median; bias, RMSE, and R^2 in each panel header).

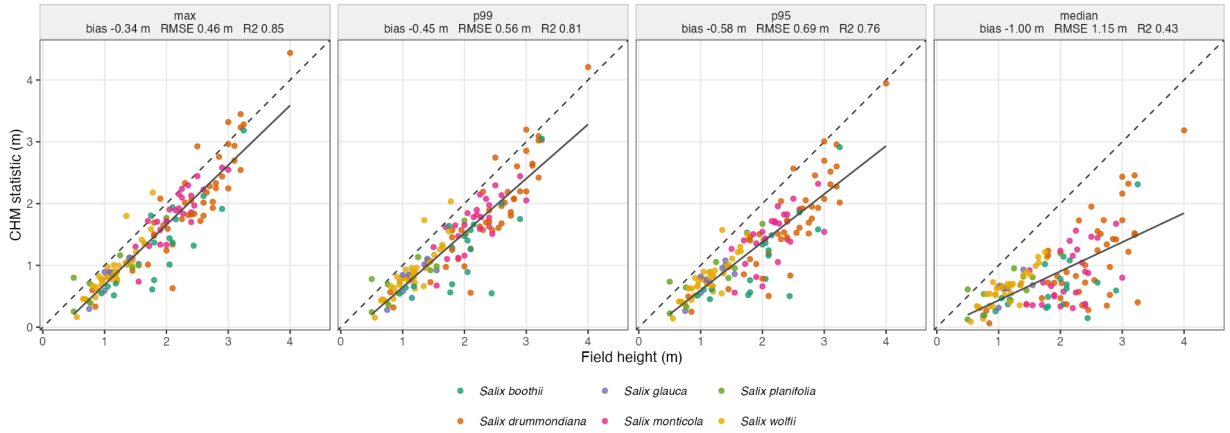


Figure 3. Prediction of field measured LAI at the 144 crowns: The kitchen sink model with all 672 predictors (left) against canopy median height only (right). Each point is predicted by withholding the entire cluster to prevent spatial autocorrelation (§2.6) from inflating the correlation.

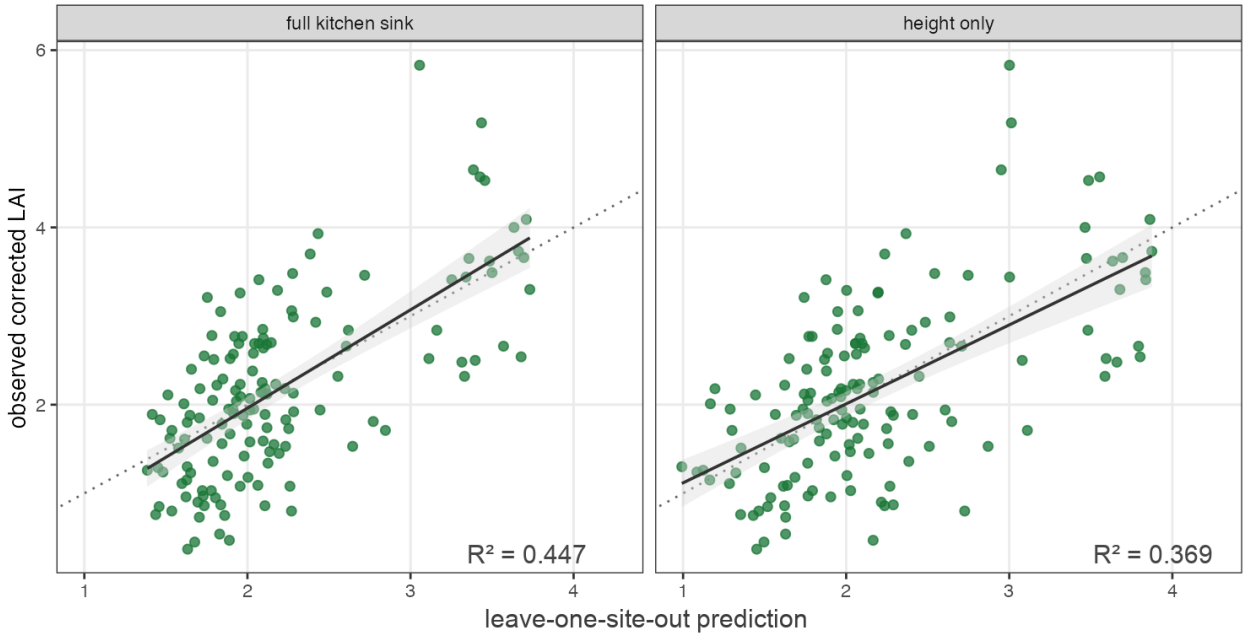


Figure 4. Leaf area implied by the height, 2021–2026: median LAI 2.72 → 2.17 m² m⁻² (−20.4%).

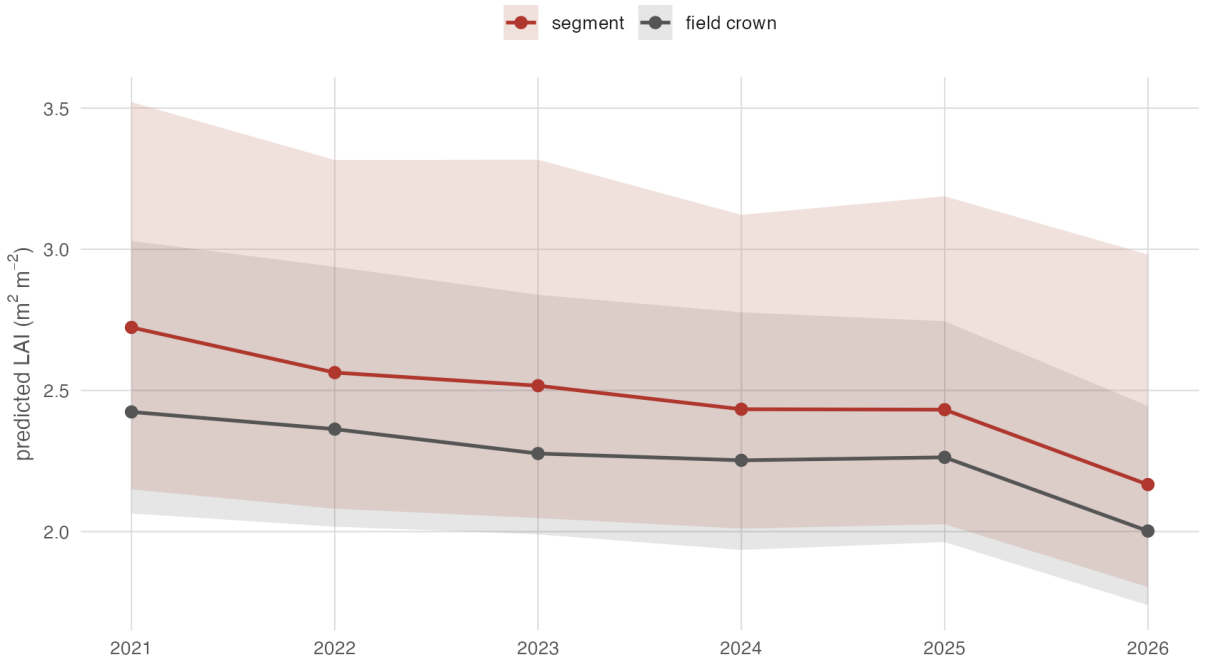


Figure 5. Confusion matrix for the merged four-class species classifier. rows are field-identified species, columns the model's predictions, per-class recall on the right.

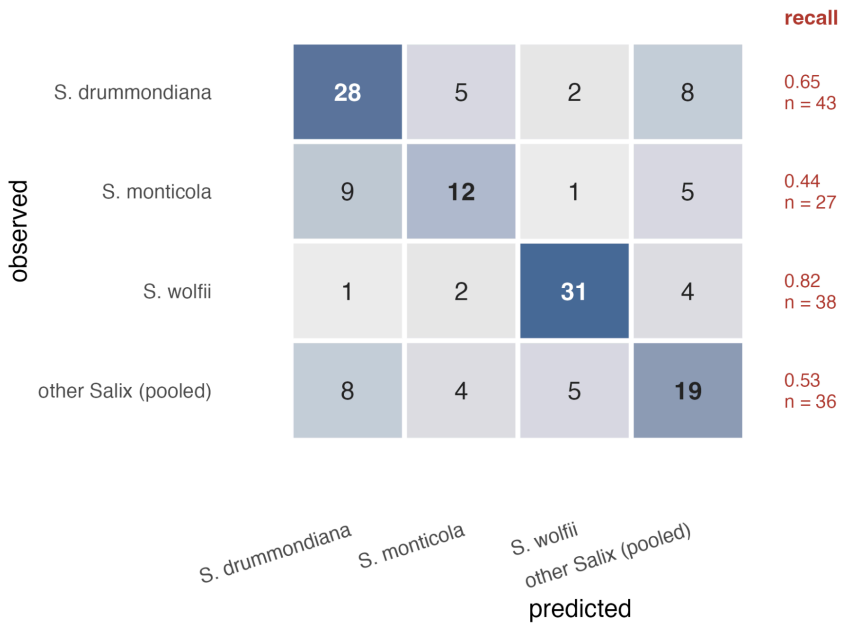
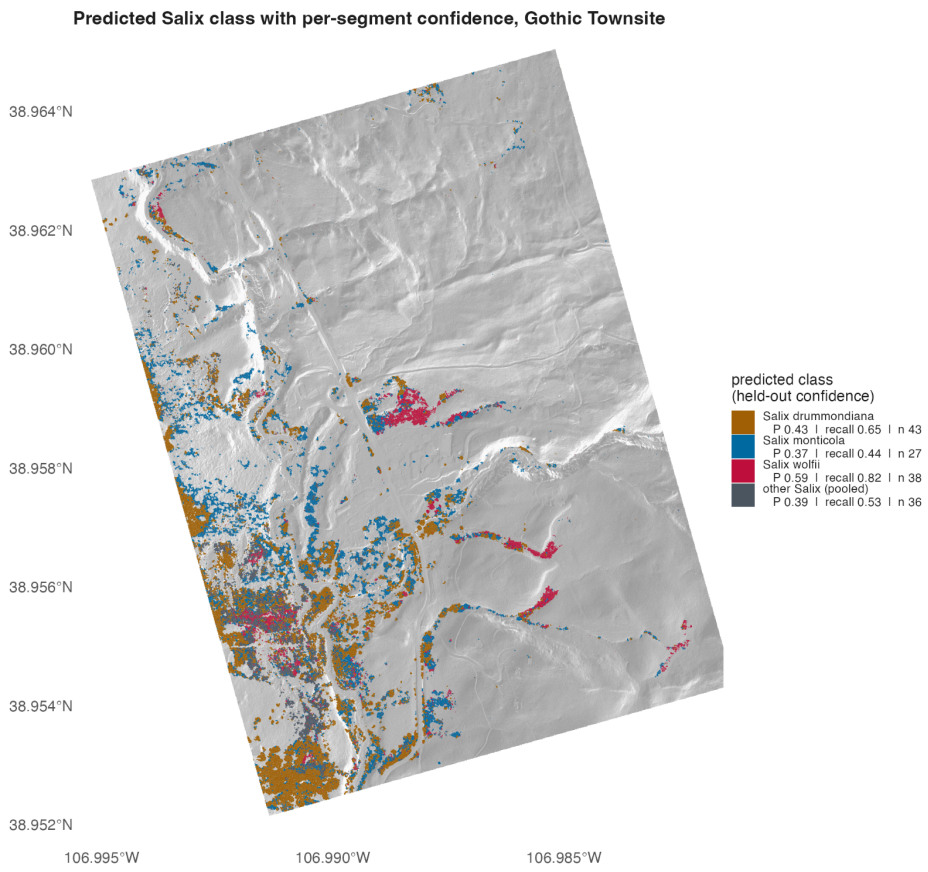


Figure 6. Predicted Salix classes across the townsite: all 27,918 willow segments labelled by the merged four-class random forest.Relief is a hillshade of the site DEM and carries no data.



3.2 Willow canopy decline

Willow Median Canopy Height fell 36.1% from a 2021 baseline across the 27,991 segments (Fig. 7A). The 144 hand-delineated field crowns agree with the decline falling 34.0% and followed the same year to year trend as the segments ($r = 0.979$) indicating representative sampling of decline (Fig 7B). However, canopy loss differed with measurement type: maximum height fell just 3.9% and 95th percentile fell 17.0%. The six-year decline measured on hand-delineated crowns matched the automated segmentation well ($r = 0.979$). To test whether terrain explains the variation, a random forest predicted each 25m cells canopy decline rate from terrain and snowmelt variables (§2.6). While decline was widespread in the townsite some areas declined faster than others, particularly in the southwest and west of the townsite (Fig. 8). With random 25m cells held out the models $R^2 = 0.108$ however, when using spatial blocking and holding out entire 250m blocks (§2.6) the models $R^2 = 0.011$ showing that the model was able to predict a 25m cells decline based off of the decline of its neighboring cells rather only the terrain and snowmelt variables. Adding thermal predictors in addition to the previous variables (§2.6) raised the spatially blocked $R^2 = 0.078$, showing canopy temperature predicts ~8% of cell to cell differences in decline. Interestingly, when thermal predictors alone were used the model scored worse than either terrain and snowmelt or terrain and snowmelt + thermal producing an $R^2 = -0.157$ (Fig. 9). Same year climate also did not correlate with the height decline ($p = 0.85$); however, previous years foresummer climate did show a suggestive correlation ($p = 0.082$) (Fig. 10), All 7 indices tested showed the same pattern (Fig. 11) of a wetter previous years climate corresponding to a taller than expected canopy the following year and

previous year climate correlating better than same year climate. The relationship was strongest measuring foresummer conditions (June SPEI-3, r = 0.96) and grew progressively weaker the wider the range of months included in the index (SPEI-12, r = 0.31).

Figure 7. Canopy decline and field validation: median height −36.1% (h_p95 −17.0%, h_max −3.9%) across 27,991 segments; hand-delineated field crowns −34.0%, concordance r = 0.979.

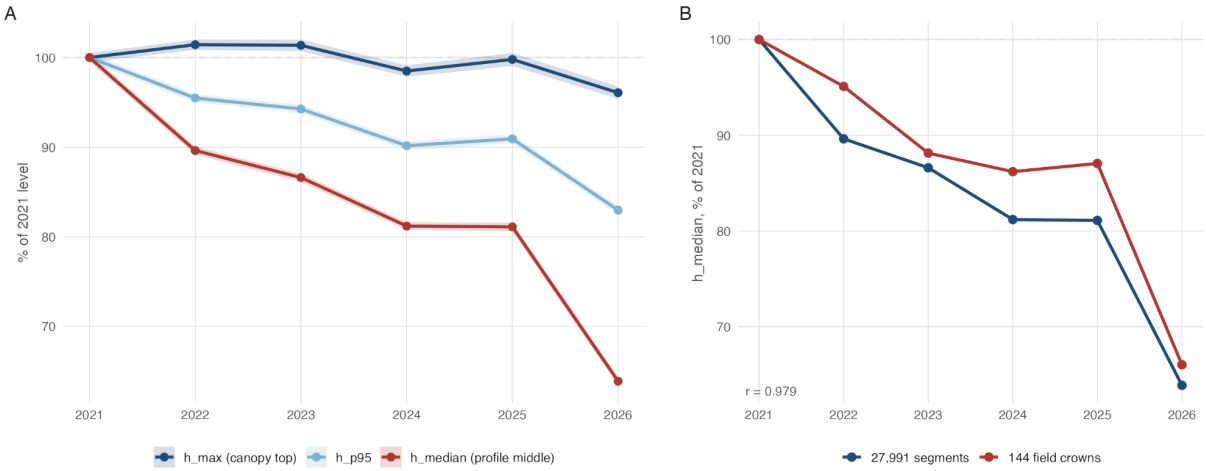


Figure 8. Decline rate in median canopy height per willow segment, 2021–2026, over hillshade.

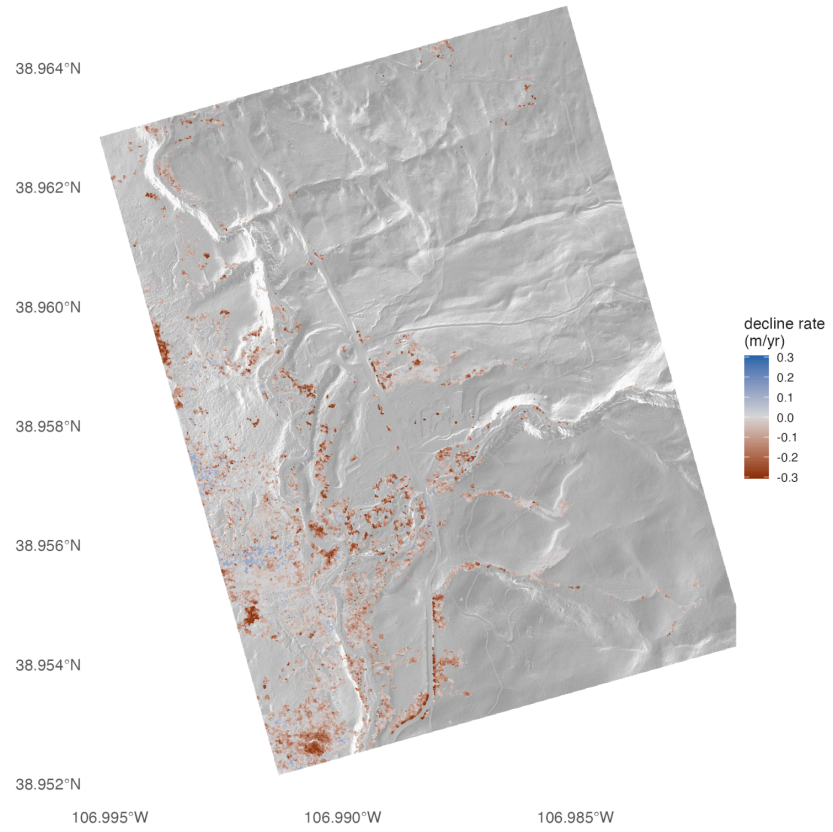


Figure 9. R^2 for predicting each 25 m cell's decline rate from terrain and snowmelt variables, using spatially blocked cross-validation. Whiskers span the 12 repeated fold assignments (§2.6)

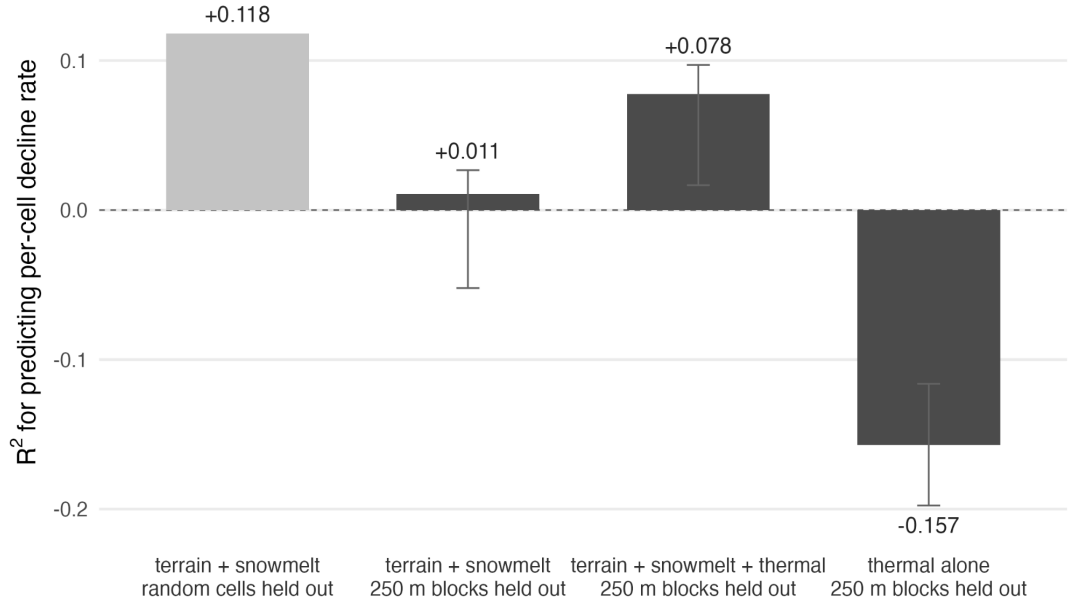


Figure 10. Detrended mean canopy height against detrended 3-month June SPEI, paired with previous-year (left) and same-year (right) climate.

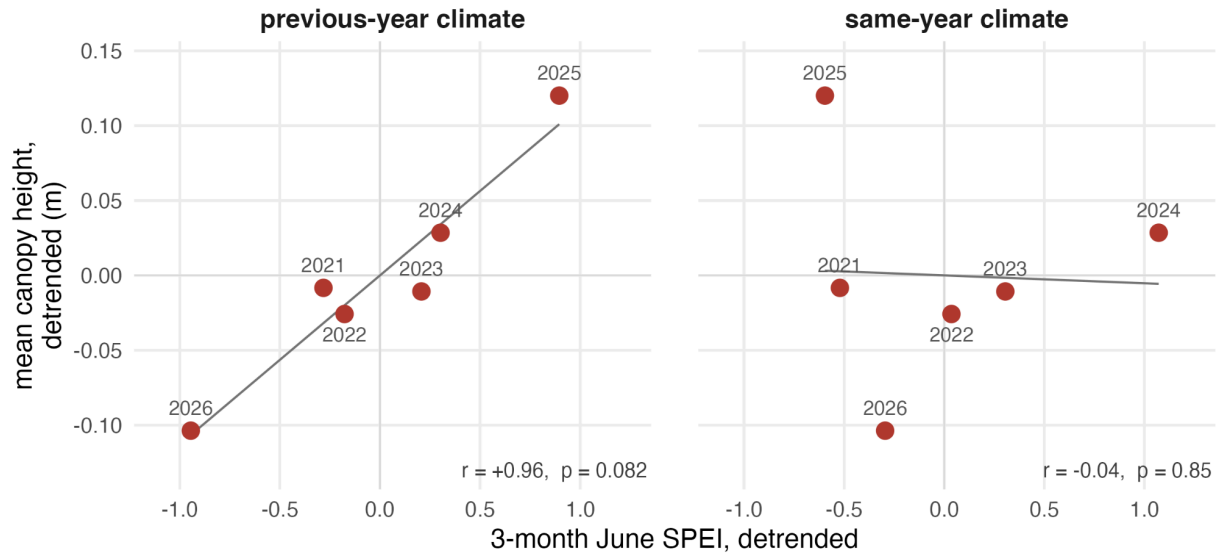
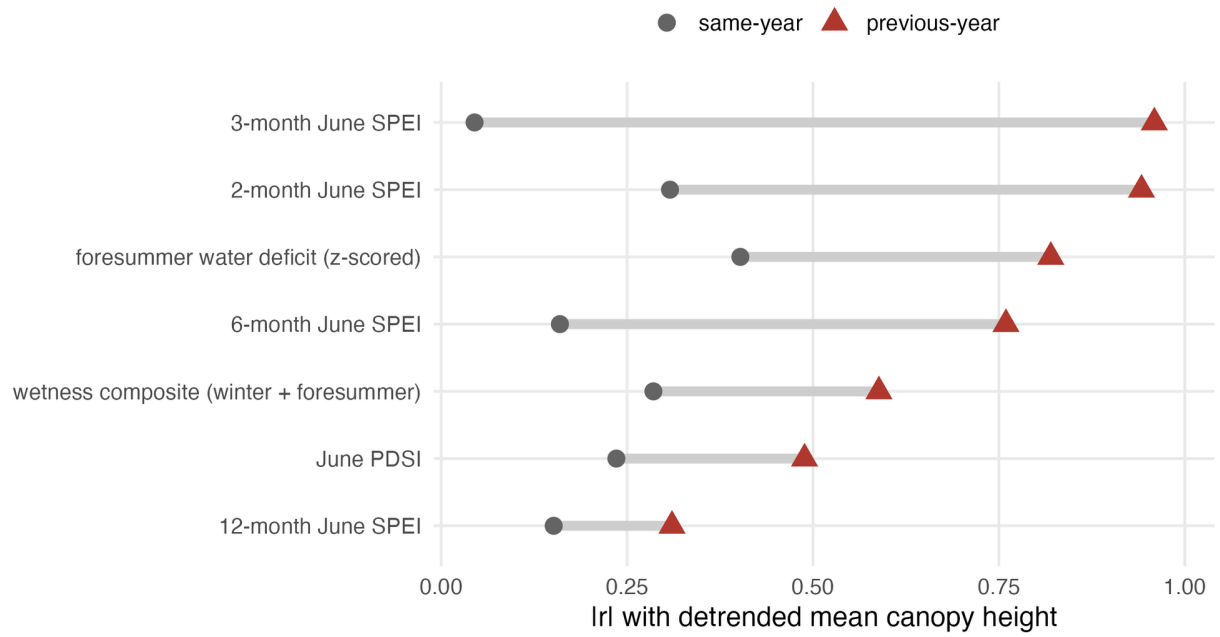


Figure 11.Correlation (|r|) between detrended mean canopy height and each of the seven drought indices (§2.3), paired with same-year and previous-year climate

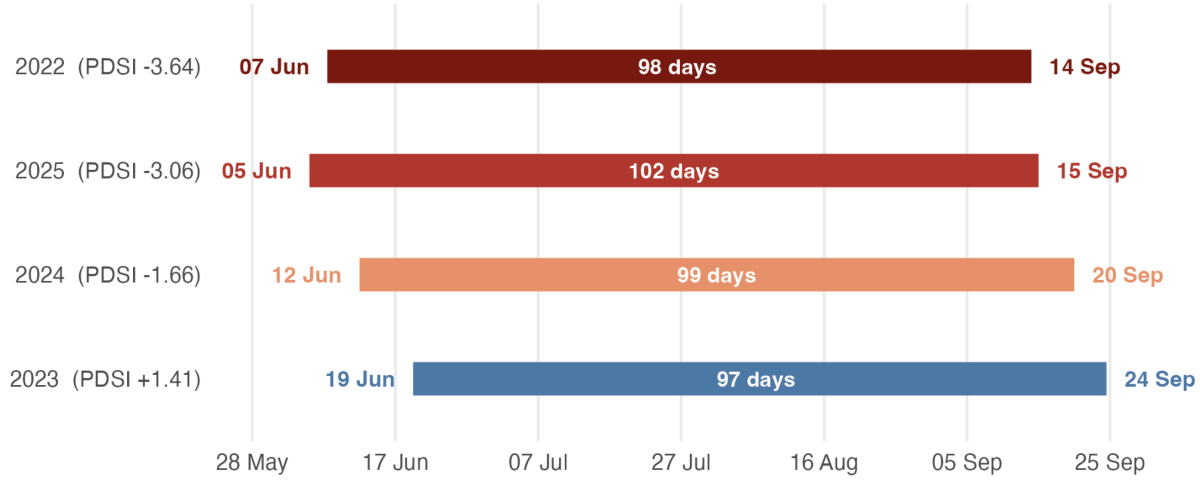


3.3 Drought creates a phase shift

Greenness was shown to add little information to the LAI model. The random forest model given all 672 structural, spectral, texture, phenology and thermal predictors improved prediction of field measured LAI

to (R^2 = 0.447) from canopy median height alone (R^2 = 0.369) (Fig. 3). While there was slim improvement, most of it came from the added height variables in the random first not the spectral, texture, phenology or thermal predictors. These predictors on their own only produced an (R^2 = 0.005). The LAI estimates across years therefore come from the height-only model (§2.6) rather than the kitchen sink model. Over the four fully sampled seasons (2022-2025), green up, determined by 50% of seasonal amplitude per segment, in dryer years happened on June 5 and June 7 compared to the wettest years green up date of June 19 (Fig. 12). The season started ~ two weeks earlier and senesced at a similarly fast rate September 14 and September 15 for the dryer years and September 24 for the wettest year (green up r = 0.95 with June PDSI, senescence r = 0.98). However, drought had no effect on the entire season's integral (r = -0.40).

Figure 12. Willow growing season per year from the NDGI curves: green-up date, season length, and senescence date, with years ordered by June PDSI.



3.4 Species differences and traits

Proportional decline differed by species (Fig. 13). Between the field identified crowns *S. monticola* lost 10.9% of its median canopy height per year, *S. boothii* 5.7%, *S. drummondiana* 4.7% while the shorter species lost proportionally less *S. planifolia* 3.5%, *S. wolfii* 1.5% and *S. glauca* 0.2%. To test whether differences this large could arise from chance the crowns species labels were shuffled at random while keeping their proportional decline rate, effectively removing any effect of species identity. The spread of difference between species was then recomputed 999 times and only 1.1% of shuffles produced differences as large as the observed differences (p = 0.011). Traits were related to decline between species but not within them (Table 1). Each trait was split into its species mean and each crowns standard deviation from that mean, modelled against proportional decline with site random intercepts (§2.6). Species means of LWC and initial plant height correlated with decline showing taller, wetter leaved plants declined proportionally faster. LMA did not predict between species decline. Within a given species none of the measured traits predicted which crowns would decline faster. Therefore, knowing a willows species

contains more valuable information about canopy height decline than knowing how an individual differs from those within its own species.

Figure 13. Median canopy height by field identified willow species, 2021-2026, (mean ± 1 SE per species).

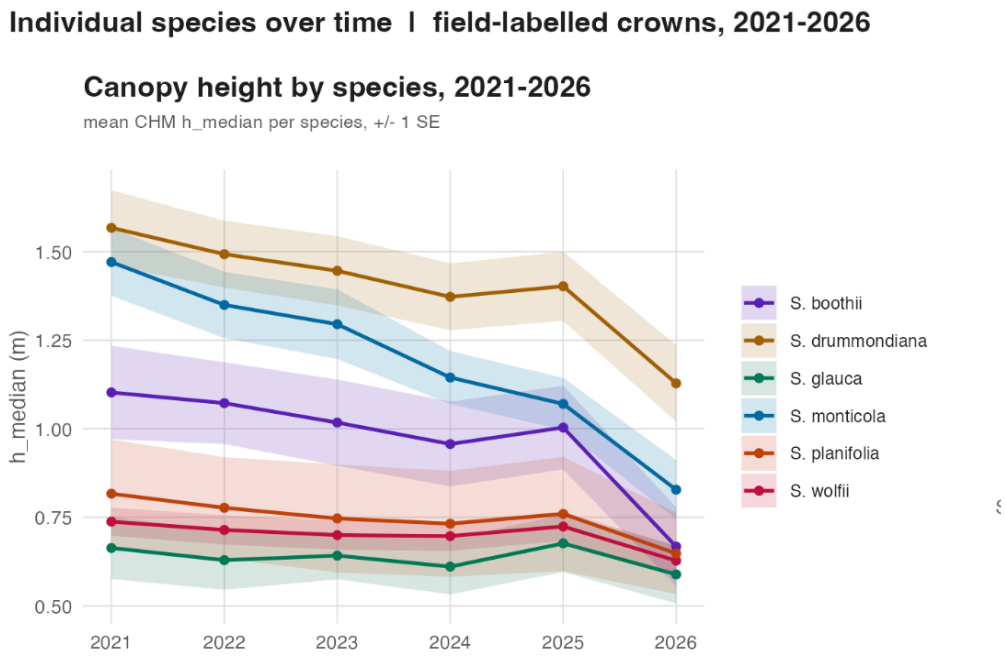


Table 1. Linear mixed models of proportional canopy decline per crown against each trait, with the trait split into its species mean (between-species) and each crown's deviation from its species mean (within-species). Models include site random intercepts; the within-species term carries a site-cluster bootstrap 95% interval (999 resamples). Effects are per standard deviation of the trait. Leaf traits were measured for 116 of the 144 crowns.

Trait	n	Between-species β (per SD)	Between-species p	Within-species β (per SD)	Bootstrap 95% CI	Within-species p
Leaf water content	116	−0.041	0.001	−0.010	−0.038 to +0.005	0.31
Leaf mass per area	116	−0.007	0.72	+0.008	−0.010 to +0.031	0.37
Initial height (2021)	144	−0.035	0.002	−0.003	−0.027 to +0.017	0.73

4. Discussion

4.1 Findings

Throughout the Gothic townsite willow canopies declined dramatically from 2021-2026: median canopy height fell 36% across the 27,991 willow segments resulting in an implied leaf-area loss of about 20%.

This result is replicated in the 144 field-delineated willow crowns declining by 34% and having an overall trend agreement of r = 0.979. Interestingly, maximum canopy height fell only 3.9% in the same timespan highlighting the importance of remote sensing as the sampling method as a survey built on only maximum photosynthetic height would have missed the structural trend. In Yellowstone resampling of young aspen previously shown to be recovering after wolf reintroduction found that the reported recovery was partly an effect of how the measurement was taken. Previous surveys targeted the tallest trees whereas the conflicting surveys targeted the entire stand (Brice et al., 2022). Decline was found to be structured by size and the largest decline was in the taller species; *S. monticola, S. drummondiana* and *S. boothii* declined proportionally more than the shorter *S. Wolfii, S. glauca*, and *S. planifolia*. This is consistent with previous findings showing increased summer temperature sensitivity at taller and wetter shrub sites (Myers-Smith et al., 2015). Individual species also showed differences as *S. monticola, S. drummondiana* started near the same median height in 2021 but monticola declined proportionally at a larger rate. A spatially blocked random forest model determined that observed decline did not correlate with any measured terrain variable, however when including thermal measurements R^2 improved to 0.078 showing canopy temperature accounted for ~ 8% of cell to cell difference in decline rate. Same-year climate was also shown not to correlate with decline, however, previous-years foresummer climate measured with SPEI-3 response showed a suggestive but non-significant correlation. A weaker than expected correlation with climate could point to a longer term trend driven by environmental conditions integrated over several years. This is consistent with willows preforming next season's leaves in the previous seasons buds (Diggle, 1997) and would explain why previous-year climate is suggestive of a correlation. Wainwright et al. (2020) showed that meadow ecosystems were the most drought sensitive in this watershed by analyzing single year greenness responses to drought stress. While riparian ecosystems were also responsive, although less so, willow shrubland was not in their analysis and it may be too simplistic of a framework for 6 years of data to answer for this community.

Greenness carried less structurally significant information than expected, adding little information to the LAI model. The “kitchen sink” model was given all 672 structural, texture, phenology, spectral and thermal measurements that were extracted from UAS imagery and lowered out-of-site RSME to 0.744 from canopy median height RSME of 0.795. A possible explanation for the minimal information carried spectrally is the amount of saturation in the vegetation indexes over the townsite willow. 94% of peak median NDVI values were over 0.85 on a scale of -1 to 1; NDVI is able to characterize vigor well for more sparse canopies but when canopies are dense the index is not as proficient at seeing fine scale change e.g. differences in LAI values 4 and 5 would carry less NDVI variation than LAI values of 1 and 2 (Tucker, 1979). Instead, greenness was proficient at characterizing the population level response of a phenological shift in drier years. In years with the strongest drought (as measured by June PDSI), the greenup date for willows was two weeks earlier than the wettest year while maintaining the same general length of season, between 97-102 days. Within the years of 2022-2025 where the seasonal integral curve was assessed dry years were earlier melt years and Dunne et al. (2003) showed that snowmelt date is the dominant control on seasonal timing in this area.

No measured trait predicted decline within species. While the taller and higher LWC willows did decline more this correlation did not translate to a withinspecies trend. The linear mixed model assessing LMA, LWC, and initial plant height found no reliable association between within species trait variation and decline. Read et al. (2017) and Siefert et al. (2015) reported similar findings that within species trait variation is much smaller than between species trait variation and that by adding within species trait

variation to trait based models it decreased their reliability. Maximum rooting depth, specific root length and other traits that are not leaf traits are all factors that affect drought stress but were unmeasured in this study. This showed that knowing a willow's species was therefore related to its traits and its amount of decline; however, species identity was only assessed for 144 individuals. This presented an opportunity to extend the species information I collected across the entire townsite with a classifier.

Most of the classifier's accuracy came from phenological predictors rather than the spectral, texture, and structural predictors. Previous research has shown that willows species differ in phenological timing of green-up and senescence more than general appearance through common garden experiments (Savage & Cavender-Bares, 2013), and has been shown to be an adequate predictor for species classification in previous remote sensing studies (Fassnacht et al., 2016).The resulting map allows species-level patterns to be known beyond the field sites, but its reliability differs by class and it should be used accordingly. S. wolfii can be mapped with relatively high confidence (recall 0.82), whereas S. monticola, the fastest-declining species, had the lowest recall among the named species (0.44). Additional field identifications as well as incorporating different datasets with more spectral information could improve classification and monitoring of species level decline.

4.2 Limitations

There are some limitations in this study that must be acknowledged. Most importantly is the short time frame. There is only six years of data for canopy structure and four fully flown seasons for greenness, which limited my ability to track cumulative or multi-year climate effects. Secondly, the lack of agreement between the drone-measured traits (spectra, texture, phenology, and thermal measurements) and field-measured LAI limited the ability to measure functional change over time. Finally, the small size of the study area , generally speaking, at just 900 m x 1200 m limits the broader application of our results to larger landscapes.

4.3 Next steps

Continued UAS flights over the gothic townsite would create a more robust dataset allowing for continued monitoring and analyses of the decline in canopy median height. Additionally, testing whether this decline of height in willow can extend beyond the townsite and could be investigated with the 2018 and 2025 CHESS imaging spectroscopy and LiDAR acquisitions of the East River watershed. Spectral information from these campaigns could also be used in the species classifier to investigate whether additional spectral information would increase the accuracy of the random forest model. In the townsite, soil moisture loggers could be placed throughout and could be paired with soil property analysis to create a more robust measure of water availability. Regarding the willows themselves, additional trait information could be taken such as Maximum rooting depth, specific root length and other traits that are not leaf traits. Dendrochronology, used extensively by Myers-Smith et al. (2015) in their investigation of tundra shrub growth and climate interactions, is the process of cutting a sample stem, taking a basal cross section and measuring their ring widths. Incorporating this within the townsite could illuminate and extend the growth record back in time allowing for more than 6 years of climate-decline analysis.

Acknowledgments

I thank the Breckheimer lab for the mentorship, the instrumentation and for hosting the RMBL Spatial Data Platform RMBL for GIS and Trimble GNSS access, and the RMBL education program. I also thank the National Science Foundation for the grant that allowed me to complete this work

Works Cited

Abatzoglou, J. T. (2013). Development of gridded surface meteorological data for ecological applications and modelling. *International Journal of Climatology, 33*(1), 121–131. https://doi.org/10.1002/joc.3413

Barrett, T., Dowle, M., Srinivasan, A., Gorecki, J., Chirico, M., Hocking, T., Schwendinger, B., & Krylov, I. (2026). *data.table: Extension of 'data.frame'* (R package version 1.18.4). https://doi.org/10.32614/CRAN.package.data.table

Baston, D. (2025). *exactextractr: Fast extraction from raster datasets using polygons* (R package version 0.10.1). https://doi.org/10.32614/CRAN.package.exactextractr

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beguería, S., & Vicente-Serrano, S. M. (2023). *SPEI: Calculation of the standardized precipitation-evapotranspiration index* (R package version 1.8.1). https://doi.org/10.32614/CRAN.package.SPEI

Bilyeu, D. M., Cooper, D. J., & Hobbs, N. T. (2008). Water tables constrain height recovery of willow on Yellowstone's northern range. *Ecological Applications, 18*(1), 80–92. https://doi.org/10.1890/07-0212.1

Breckheimer, I. (2026). *rSDP: Discover, query, and subset data from the RMBL Spatial Data Platform* (R package version 0.6). https://github.com/rmbl-sdp/rSDP

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brice, E. M., Larsen, E. J., & MacNulty, D. R. (2022). Sampling bias exaggerates a textbook example of a trophic cascade. *Ecology Letters, 25*(1), 177–188. https://doi.org/10.1111/ele.13915

Cottrell, T. R. (1996). Use of Plant Strategy Ordination, DCA and ANOVA to elucidate relationships among habitats of *Salix planifolia* and *Salix monticola*. *Journal of Vegetation Science, 7*(2), 237–246. https://doi.org/10.2307/3236324

Dahm, C. N., Cleverly, J. R., Allred Coonrod, J. E., Thibault, J. R., Mcdonnell, D. E., & Gilroy, D. J. (2002). Evapotranspiration at the land/water interface in a semi-arid drainage basin. *Freshwater Biology, 47*(4), 831–843. https://doi.org/10.1046/j.1365-2427.2002.00917.x

Dalponte, M., & Coomes, D. A. (2016). Tree-centric mapping of forest carbon density from airborne laser scanning and hyperspectral data. *Methods in Ecology and Evolution, 7*(10), 1236–1245. https://doi.org/10.1111/2041-210X.12575

Diffenbaugh, N. S., Scherer, M., & Ashfaq, M. (2013). Response of snow-dependent hydrologic extremes to continued global warming. *Nature Climate Change, 3*(4), 379–384. https://doi.org/10.1038/nclimate1732

Diggle, P. K. (1997). Extreme preformation in alpine *Polygonum viviparum*: An architectural and developmental analysis. *American Journal of Botany, 84*(2), 154–169. https://doi.org/10.2307/2446077

Dunne, J. A., Harte, J., & Taylor, K. J. (2003). Subalpine meadow flowering phenology responses to climate change: Integrating experimental and gradient methods. *Ecological Monographs, 73*(1), 69–86. https://doi.org/10.1890/0012-9615(2003)073[0069:SMFPRT]2.0.CO;2

Elmendorf, S. C., Henry, G. H. R., Hollister, R. D., et al. (2012). Plot-scale evidence of tundra vegetation change and links to recent summer warming. *Nature Climate Change, 2*, 453–457. https://doi.org/10.1038/nclimate1465

Falco, N., Wainwright, H., Dafflon, B., Léger, E., Peterson, J., Steltzer, H., Wilmer, C., Rowland, J. C., Williams, K. H., & Hubbard, S. S. (2019). Investigating microtopographic and soil controls on a mountainous meadow plant community using high-resolution remote sensing and surface geophysical data. *Journal of Geophysical Research: Biogeosciences, 124*(6), 1618–1636. https://doi.org/10.1029/2018JG004394

Fassnacht, F. E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L. T., Straub, C., & Ghosh, A. (2016). Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment, 186*, 64–87. https://doi.org/10.1016/j.rse.2016.08.013

Fyfe, J. C., Derksen, C., Mudryk, L., Flato, G. M., Santer, B. D., Swart, N. C., Molotch, N. P., Zhang, X., Wan, H., Arora, V. K., Scinocca, J., & Jiao, Y. (2017). Large near-term projected snowpack loss over the western United States. *Nature Communications, 8*, 14996. https://doi.org/10.1038/ncomms14996

Gao, X., Gray, J. M., & Reich, B. J. (2021). Long-term, medium spatial resolution annual land surface phenology with a Bayesian hierarchical model. *Remote Sensing of Environment, 261*, 112484. https://doi.org/10.1016/j.rse.2021.112484

Harte, J., & Shaw, R. (1995). Shifting dominance within a montane vegetation community: Results of a climate-warming experiment. *Science, 267*(5199), 876–880. https://doi.org/10.1126/science.267.5199.876

Harte, J., Saleska, S. R., & Levy, C. (2015). Convergent ecosystem responses to 23-year ambient and manipulated warming link advancing snowmelt and shrub encroachment to transient and long-term climate–soil carbon feedback. *Global Change Biology, 21*(6), 2349–2356. https://doi.org/10.1111/gcb.12831

Hijmans, R. J., Brown, A., & Barbosa, M. (2026). *terra: Spatial data analysis* (R package version 1.9-34). https://doi.org/10.32614/CRAN.package.terra

Hubbard, S. S., Williams, K. H., Agarwal, D., et al. (2018). The East River, Colorado, watershed: A mountainous community testbed for improving predictive understanding of multiscale

hydrological–biogeochemical dynamics. *Vadose Zone Journal, 17*, 180061. https://doi.org/10.2136/vzj2018.03.0061

Johnston, D. B., Cooper, D. J., & Hobbs, N. T. (2011). Relationships between groundwater use, water table, and recovery of willow on Yellowstone's northern range. *Ecosphere, 2*(2), art20. https://doi.org/10.1890/ES10-00150.1

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Langenheim, J. H. (1962). Vegetation and environmental patterns in the Crested Butte area, Gunnison County, Colorado. *Ecological Monographs, 32*(3), 249–285. https://doi.org/10.2307/1942400

Marshall, K. N., Cooper, D. J., & Hobbs, N. T. (2014). Interactions among herbivory, climate, topography and plant age shape riparian willow dynamics in northern Yellowstone National Park, USA. *Journal of Ecology, 102*, 667–677. https://doi.org/10.1111/1365-2745.12225

Minasny, B., & McBratney, A. B. (2006). A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computers & Geosciences, 32*(9), 1378–1388. https://doi.org/10.1016/j.cageo.2005.12.009

Mountain Research Initiative EDW Working Group. (2015). Elevation-dependent warming in mountain regions of the world. *Nature Climate Change, 5*(5), 424–430. https://doi.org/10.1038/nclimate2563

Myers-Smith, I. H., Elmendorf, S. C., Beck, P. S. A., et al. (2015). Climate sensitivity of shrub growth across the tundra biome. *Nature Climate Change, 5*, 887–891. https://doi.org/10.1038/nclimate2697

Myers-Smith, I. H., Forbes, B. C., Wilmking, M., et al. (2011). Shrub expansion in tundra ecosystems: Dynamics, impacts and research priorities. *Environmental Research Letters, 6*(4), 045509. https://doi.org/10.1088/1748-9326/6/4/045509

Nobre, A. D., Cuartas, L. A., Hodnett, M., Rennó, C. D., Rodrigues, G., Silveira, A., Waterloo, M., & Saleska, S. (2011). Height Above the Nearest Drainage – a hydrologically relevant new terrain model. *Journal of Hydrology, 404*(1–2), 13–29. https://doi.org/10.1016/j.jhydrol.2011.03.051

Nogués-Bravo, D., Araújo, M. B., Errea, M. P., & Martínez-Rica, J. P. (2007). Exposure of global mountain systems to climate warming during the 21st century. *Global Environmental Change, 17*(3–4), 420–428. https://doi.org/10.1016/j.gloenvcha.2006.11.007

Pascale, S., Boos, W. R., Bordoni, S., Delworth, T. L., Kapnick, S. B., Murakami, H., Vecchi, G. A., & Zhang, W. (2017). Weakening of the North American monsoon with global warming. *Nature Climate Change, 7*(11), 806–812. https://doi.org/10.1038/nclimate3412

Pebesma, E. (2018). Simple features for R: Standardized support for spatial vector data. *The R Journal, 10*(1), 439–446. https://doi.org/10.32614/RJ-2018-009

Pedersen, T. L. (2025). *patchwork: The composer of plots* (R package version 1.3.2). https://doi.org/10.32614/CRAN.package.patchwork

R Core Team. (2026). *R: A language and environment for statistical computing* (Version 4.6.1). R Foundation for Statistical Computing. https://www.R-project.org/

Read, Q. D., Henning, J. A., & Sanders, N. J. (2017). Intraspecific variation in traits reduces ability of trait-based models to predict community structure. *Journal of Vegetation Science, 28*, 1070–1081. https://doi.org/10.1111/jvs.12555

Reich, P. B. (2014). The world-wide 'fast–slow' plant economics spectrum: A traits manifesto. *Journal of Ecology, 102*(2), 275–301. https://doi.org/10.1111/1365-2745.12211

Roudier, P., Beaudette, D. E., & Hewitt, A. E. (2012). A conditioned Latin hypercube sampling algorithm incorporating operational constraints. In B. Minasny, B. P. Malone, & A. B. McBratney (Eds.), *Digital Soil Assessments and Beyond* (pp. 227–232). CRC Press. https://doi.org/10.1201/b12728-48

Savage, J. A., & Cavender-Bares, J. M. (2011). Contrasting drought survival strategies of sympatric willows (genus: *Salix*): Consequences for coexistence and habitat specialization. *Tree Physiology, 31*(6), 604–614. https://doi.org/10.1093/treephys/tpr056

Savage, J. A., & Cavender-Bares, J. (2013). Phenological cues drive an apparent trade-off between freezing tolerance and growth in the family Salicaceae. *Ecology, 94*(8), 1708–1717. https://doi.org/10.1890/12-1779.1

Savage, J. A., Cavender-Bares, J., & Verhoeven, A. (2009). Willow species (genus *Salix*) with contrasting habitat affinities differ in their photoprotective responses to water stress. *Functional Plant Biology, 36*(4), 300–309. https://doi.org/10.1071/FP08303

Siefert, A., Violle, C., Chalmandrier, L., et al. (2015). A global meta-analysis of the relative extent of intraspecific trait variation in plant communities. *Ecology Letters, 18*(12), 1406–1419. https://doi.org/10.1111/ele.12508

Sloat, L. L., Henderson, A. N., Lamanna, C., & Enquist, B. J. (2015). The effect of the foresummer drought on carbon exchange in subalpine meadows. *Ecosystems, 18*(3), 533–545. https://doi.org/10.1007/s10021-015-9845-1

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment, 8*(2), 127–150. https://doi.org/10.1016/0034-4257(79)90013-0

Vicente-Serrano, S. M., Beguería, S., & López-Moreno, J. I. (2010). A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index. *Journal of Climate, 23*(7), 1696–1718. https://doi.org/10.1175/2009JCLI2909.1

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., Steltzer, H., Williams, K. H., Hubbard, S. S., & Enquist, B. J. (2020). Satellite-derived foresummer drought sensitivity of plant productivity in Rocky Mountain headwater catchments: Spatial heterogeneity and geological-geomorphological control. *Environmental Research Letters, 15*(8), 084018. https://doi.org/10.1088/1748-9326/ab8fd0

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wright, I. J., Reich, P. B., Westoby, M., et al. (2004). The worldwide leaf economics spectrum. *Nature, 428*(6985), 821–827. https://doi.org/10.1038/nature02403

Wright, M. N., & Ziegler, A. (2017). ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software, 77*(1), 1–17. https://doi.org/10.18637/jss.v077.i01

Yang, W., Kobayashi, H., Wang, C., Shen, M., Chen, J., Matsushita, B., et al. (2019). A semi-analytical snow-free vegetation index for improving estimation of plant phenology in tundra and grassland ecosystems. *Remote Sensing of Environment, 228*, 31–44. https://doi.org/10.1016/j.rse.2019.03.028

Zorio, S. D., Williams, C. F., & Aho, K. A. (2016). Sixty-five years of change in montane plant communities in western Colorado, U.S.A. *Arctic, Antarctic, and Alpine Research, 48*(4), 703–722. https://doi.org/10.1657/AAAR0016-011