

Supplementary methods and results

Seasonal Landsat summaries improve cross-fire prediction of post-fire conifer establishment

S1 Clarifying the reported estimands

The original manuscript combined three summaries: AP calculated after concatenating all repeat-level out-of-fold predictions, AP calculated after averaging those probabilities within each plot, and the mean of bootstrap AP differences. They are not generally equal. The revised primary point estimates use the second definition for repeated grouped validation and the single final prediction per plot for Locked13. Paired point estimates are directly calculated AP differences. Original percentile fire-bootstrap endpoints remain unchanged. No original prediction file was overwritten and no model was fitted during this revision.

The archived true-refit ablation bootstrap table already contains an observed_delta computed from mean repeated out-of-fold predictions. Figure 8 uses that field rather than combining pooled-repeat AP values with mean-OOF intervals. Complexity-control intervals use the matching observed_delta from their own separate fixed refits. The V4 and V5 All38 intervals remain attached to their original experiments and may differ slightly for repeated comparisons because of distinct archived Monte Carlo draws.

Table S1. Comparison of the two Pilot25 AP summaries.

Model	Set	Pooled repeat-level AP	Mean-OOF AP
Elastic Net	M0	0.671512	0.680268
Elastic Net	M1	0.688590	0.702698
Elastic Net	M2	0.641110	0.655215
HGB	M0	0.613919	0.634348
HGB	M1	0.657658	0.681367
HGB	M2	0.647796	0.664966

Pooled repeat-level AP treats the concatenated repeat predictions as observations; mean-OOF AP first averages predictions within plot. The main manuscript uses mean-OOF AP.

Table S2. Observed paired changes alongside bootstrap averages.

Cohort	Model	Contrast	Observed Δ AP	Bootstrap mean
Pilot25	Elastic Net	M1 – M0	+0.022430	+0.023137
Pilot25	Elastic Net	M2 – M1	-0.047483	-0.045783
Pilot25	Elastic Net	M2 – Prefire control	+0.003438	+0.005381
Pilot25	HGB	M1 – M0	+0.047019	+0.044989
Pilot25	HGB	M2 – M1	-0.016401	-0.014680
Pilot25	HGB	M2 – Prefire control	-0.017964	-0.016759
Locked13	Elastic Net	M1 – M0	+0.118034	+0.117473
Locked13	Elastic Net	M2 – M1	-0.011190	-0.011842
Locked13	Elastic Net	M2 – Prefire control	+0.034457	+0.035723
Locked13	HGB	M1 – M0	+0.103053	+0.096537
Locked13	HGB	M2 – M1	-0.021942	-0.019366
Locked13	HGB	M2 – Prefire control	-0.001127	-0.000276

Observed differences are calculated from matching pooled fire-weighted AP estimates; bootstrap averages are shown only for reconciliation with the archived summaries.

S2 Numerical safeguards and limits of legacy analyses

Before this revision, the All38 programme repaired 45 warning-bearing checkpoints using the original SAGA solver, parameter grid, seeds, samples and folds, with an iteration ceiling of 20,000 and tolerance of 0.002. All 45 selected the same best grid parameters; 255 already warning-free checkpoints were retained. The resulting 300-checkpoint archive contained no recorded ConvergenceWarnings. The original manuscript reports a maximum pooled-repeat AP shift of 0.00069 and a maximum point-increment shift of 0.00102 relative to the earlier archive. This is inherited numerical provenance, not a rerun conducted for the current manuscript.

Legacy proxy-ablation outputs and the empty/uncompleted TCN figure are not used as inferential evidence in the current figures. They remain only in the unchanged original archive for provenance. Twenty previously examined ecological interactions lacked the strict cross-cohort support criterion in both cohorts; they do not establish a mechanism and are not promoted to main-text findings.

S3 How to read the fixed transfer analyses

The contributor test removes all Pilot25 fires associated with the held-out contributor, including plots from other contributors sharing those fires. The label therefore identifies the exclusion rule, not a pure isolated collection of that contributor's plots. Contributor sets overlap and must not be treated as independent replicate studies. The five geographic blocks concern All38 instead. These protocols reuse earlier hyperparameter choices; fixed settings are not equivalent to fully nested selection. Full numerical outputs are retained in `source_data/contributor_transfer.csv` and `geographic_transfer.csv`.

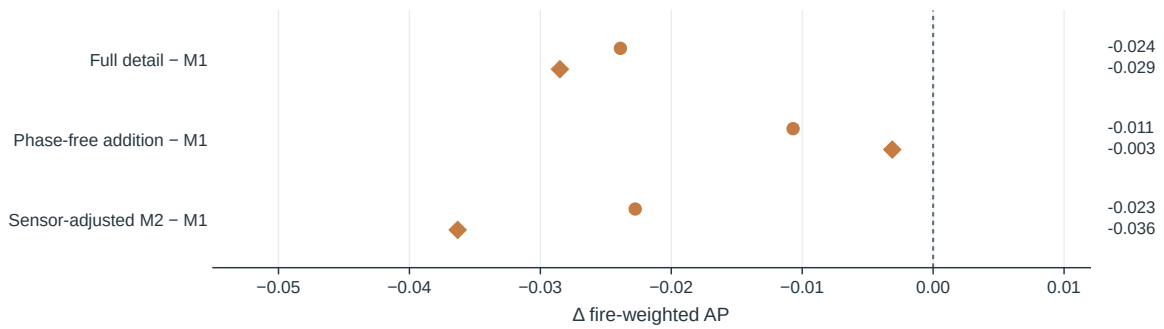
Table S5. Assignments used for the fixed geographic fire blocks.

Block	Fires	Fire identifiers
1	15	AMERICAN RIVER COMPLEX (GOVERNMENT) (2008); ANTELOPE COMPLEX (WHEELER) (2007); BASSETTS (2006); BTU LIGHTNING COMPLEX (LONG BRANCH-JACK) (2008); CANNON (2002); CUB COMPLEX (CUB) (2008); FRED'S (2004); MOONLIGHT (2007); PIGEON (2006); POWER (2004); RALSTON (2006); RUSH (2006); SIMS (2004); SPANISH (2003); SUMMIT COMPLEX (2003)
2	3	BULL COMPLEX (SUMMIT) (1996); HASH ROCK (2000); PORPHYRY SOUTH (2007)
3	9	BORREGO (2002); DOME (1996); LAKES (2002); MISSIONARY RIDGE (2002); PINE CANYON (2005); PONIL COMPLEX (2002); SPRING (2002); WALLOW (2011); WEST FORK (2008)
4	7	BUFFALO CREEK (1996); ELDORADO (2000); FOURMILE CANYON (2010); HAYMAN (2002); HIGH MEADOWS (2000); MASON (2005); OVERLAND (2003)
5	4	BATTLE CREEK (2002); JASPER (2000); RICCO (2005); ROGERS SHACK (2001)

Blocks are the frozen geographic holdout assignments over All38. Identifiers correspond to the original analysis-fire keys.

● Elastic Net ◆ Gradient boosting

a Bounded representation controls



b Complete source-transfer audit

	Elastic Net		Gradient boosting		
	Mc	M1	Mc	M1	Δ AP vs M0
AEB (n=2)	+0.036	+0.044	+0.032	+0.039	
HDS (n=4)	+0.069	+0.077	-0.011	-0.014	
JDC (n=3)	-0.055	-0.028	+0.041	+0.040	
JRO (n=3)	-0.036	-0.058	-0.043	-0.041	
KCR (n=4)	-0.058	-0.063	+0.015	-0.010	
KTD (n=2)	+0.043	+0.046	+0.063	+0.060	
LBH (n=2)	+0.050	-0.017	+0.076	+0.097	
MPS (n=2)	-0.089	-0.094	-0.118	-0.127	
MTR (n=5)	-0.039	-0.056	+0.039	+0.044	
PJF (n=8)	+0.016	-0.015	+0.075	+0.049	

Upper: point estimates. Lower: AP differences relative to M0. Source holdouts may overlap.

Figure S5. Bounded representation controls and the complete source-transfer audit. a, AP differences on one existing grouped partition: full M2 versus M1; a phase-invariant seasonal addition versus M1; and sensor-adjusted M2 versus identically adjusted M1. Circles identify Elastic Net and diamonds histogram gradient boosting. These fixed-partition point estimates have no supplied uncertainty intervals. Phase removal does not align observations to ecological time, and sensor-composition control is not radiometric harmonization. b, All ten Pilot25 contributor holdouts, including the four represented by two test fires. Cells display Mc–M0 and M1–M0 AP differences using a common zero-centered color scale. Parentheses give test-fire counts. Source test sets can overlap, and the settings were selected previously; the heatmap summarizes descriptive transfer diagnostics rather than independent external confirmations.

S4 Additional diagnostics separate from the primary experiment

The prefire-control diagnostic uses frozen Locked13 predictions for M0, the prefire-only control and M1, with 2,000 paired complete-fire bootstrap resamples and seed 20260906. AP is evaluated using fire-equal weights. For a resampled fire appearing multiple times, its plot weights are multiplied by the bootstrap multiplicity. The same resampled fire indices are used for the compared models. No test labels are used for model fitting, but the decision to examine this comparison is post hoc. Draw-level values are provided in placebo_bootstrap_draws.csv.

Table S3. Post hoc comparison with the prefire-only control.

Model	Contrast	Observed difference	95% fire-bootstrap interval
Elastic Net	M1 — Prefire control	+0.0456	[-0.0049, +0.1104]
Elastic Net	Prefire control — M0	+0.0724	[+0.0132, +0.1381]
HGB	M1 — Prefire control	+0.0208	[-0.0273, +0.0636]
HGB	Prefire control — M0	+0.0822	[+0.0024, +0.1547]

The prefire control and M1 use different feature inventories. Intervals resample complete fires and do not constitute a matched causal decomposition.

The absence-directed screening diagnostic ranks the lowest predicted establishment probabilities first within each test fire. Capacities are fixed at 10%, 20%, 30% and 50%, not selected to maximize an observed outcome. At each capacity, within-fire precision is averaged equally across fires. The bootstrap unit is the complete fire, with 2,000 resamples and seed 20260905. Marginal curve intervals and paired contrast intervals are generated in their respective passes through the same reproducible random-number stream. A difference between two marginal confidence limits is not used as a paired interval. Selected counts use ceiling rounding and can exceed the nominal fraction slightly.

Table S4. Paired changes in absence precision at 20% capacity.

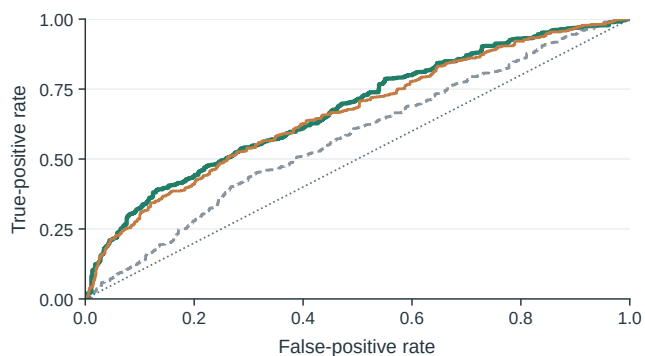
Model	Contrast	Observed difference	95% fire-bootstrap interval
Elastic Net	M1 — M0	+0.1069	[+0.0528, +0.1639]
Elastic Net	M2 — M1	-0.0062	[-0.0493, +0.0323]
HGB	M1 — M0	+0.0814	[+0.0217, +0.1477]
HGB	M2 — M1	+0.0083	[-0.0596, +0.0769]

Plots are ranked within fire by increasing predicted establishment. Precision is averaged equally across fires; intervals use paired whole-fire resampling.

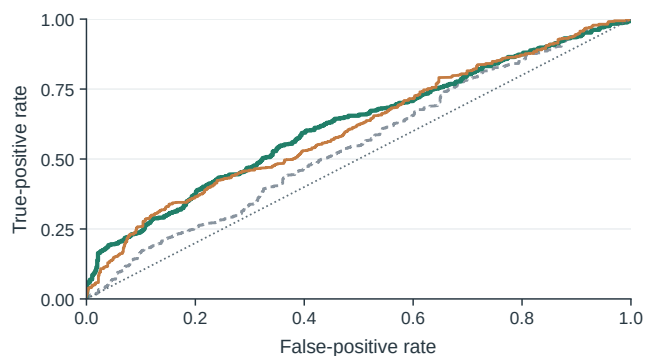
Reliability curves are descriptive ten-bin summaries of raw probabilities. Platt-calibrated Brier values use already archived Pilot25-trained transformations and do not fit on Locked13. Neither a universal probability threshold nor treatment benefit is estimated. Spectral trajectories are observed-group descriptive medians and interquartile ranges; plotting by later outcome is not part of model training and does not provide causal inference.

-- M0 — M1 — M2

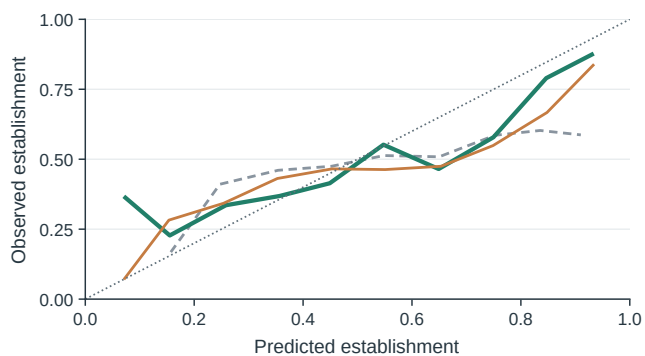
a Elastic Net · ROC



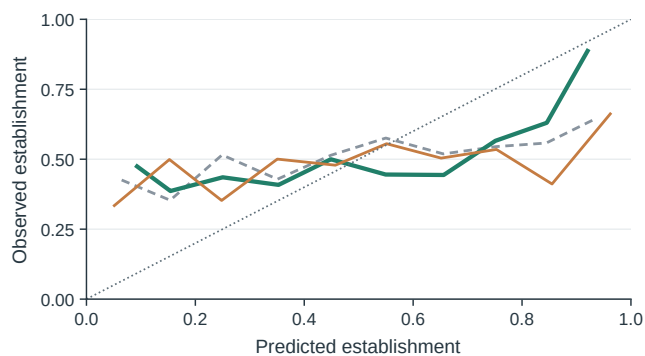
b Gradient boosting · ROC



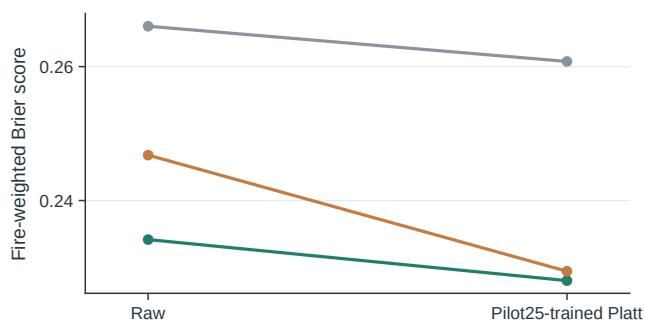
c Elastic Net · Raw calibration



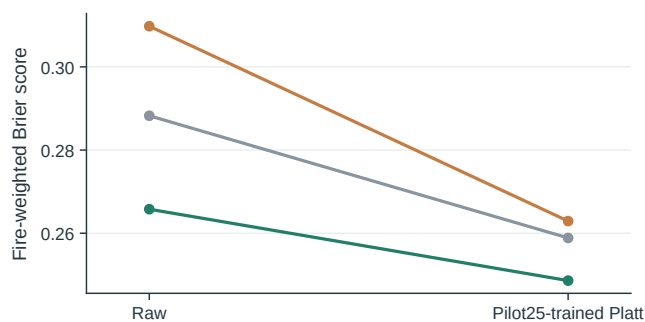
d Gradient boosting · Raw calibration



e Elastic Net · Calibration sensitivity



f Gradient boosting · Calibration sensitivity

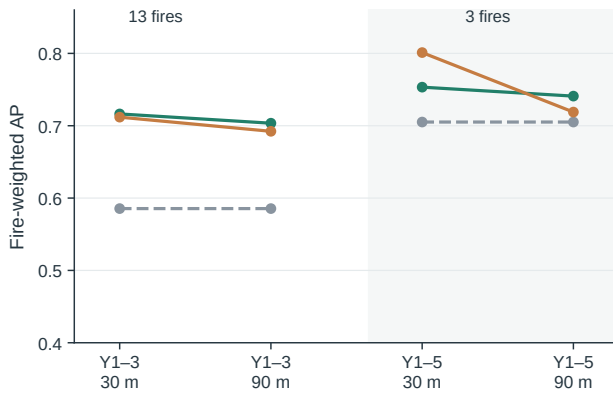


ROC and raw-calibration curves use Locked13 predictions. Platt parameters were learned on Pilot25.

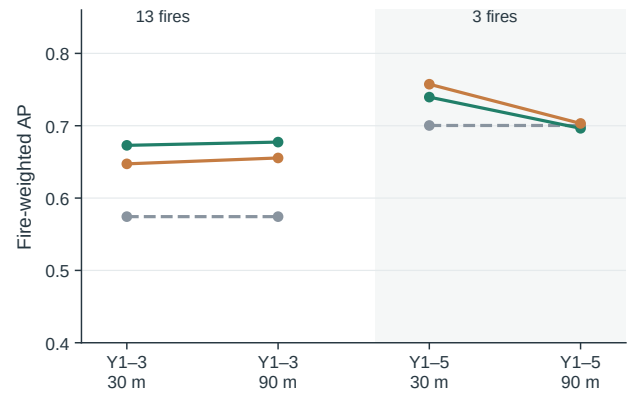
Figure S2. Discrimination, raw calibration and archived calibration sensitivity. a–b, Fire-weighted ROC curves for the original Locked13 predictions. c–d, Raw-probability calibration in ten fixed-width bins using fire-equal weights; the diagonal is perfect calibration. e–f, Fire-weighted Brier scores for the raw probabilities and the existing Pilot25-trained Platt transform applied to Locked13. Lower Brier scores indicate smaller probability error. The left column is Elastic Net and the right histogram gradient boosting. These descriptive curves and scores use archived predictions; the Platt transform was not refitted in this figure revision. They do not imply confidence intervals or calibration across unsampled populations.

-- M0 — M1 — M2

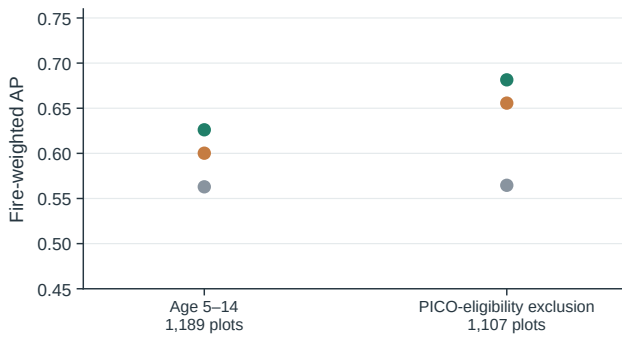
a Elastic Net · temporal / spatial support



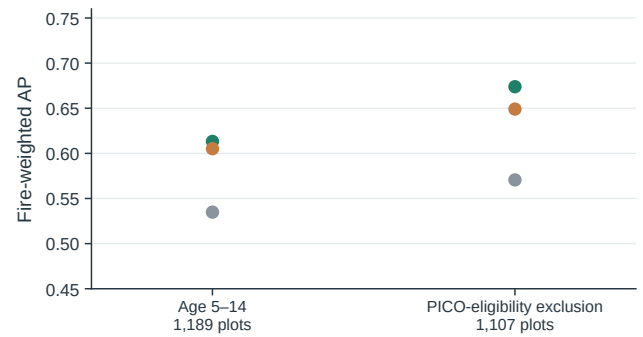
b Gradient boosting · temporal / spatial support



c Elastic Net · eligibility refits



d Gradient boosting · eligibility refits



Years 1-3: 13 test fires. Years 1-5: 3 test fires. Eligibility filters change the sample denominator.

Figure S3. Sensitivity to temporal windows, spatial support and outcome eligibility. a-b, Frozen Locked13 AP at 30 m and 90 m support for Years 1-3 and Years 1-5. Years 1-3 includes thirteen test fires, whereas Years 1-5 includes only three. Lines connect only the two spatial-support settings within the same window; the two window groups are deliberately disconnected because their denominators differ. c-d, Archived refits restricted to the 5-14-year survey-age overlap (1,189 plots) or excluding plots eligible for the source *Pinus contorta* model (1,107 plots). The latter flag denotes source-model eligibility, not proof that the species is absent. The two filters have different denominators and are plotted as discrete estimates rather than a temporal or paired trajectory.

S5 Evidence that remains descriptive or incomplete

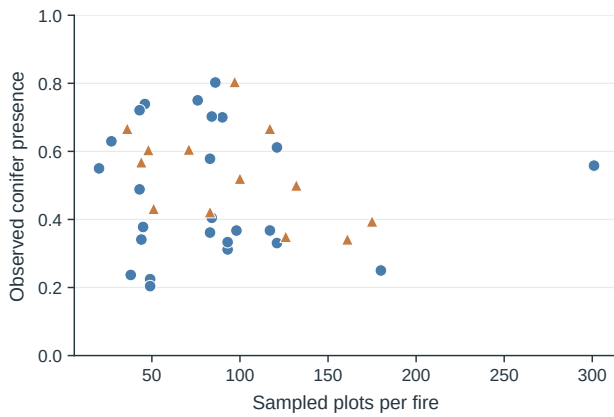
A matched information-source experiment should compare the same index families, feature definitions, support variables and tuning protocol using prefire context alone, post-fire additions alone, and their combination. A reduced-subset claim requires feature-family and hyperparameter selection inside each outer training fold or testing a completely fixed model in a new field programme. Prospective use requires timestamps for all predictors, an explicit year-3 decision point and a common prediction horizon. These tasks were not performed in this revision; no new independent cohort, species-level endpoint, stocking threshold or restoration-effect estimate is claimed.

S6 Archived inputs for deterministic figure checks

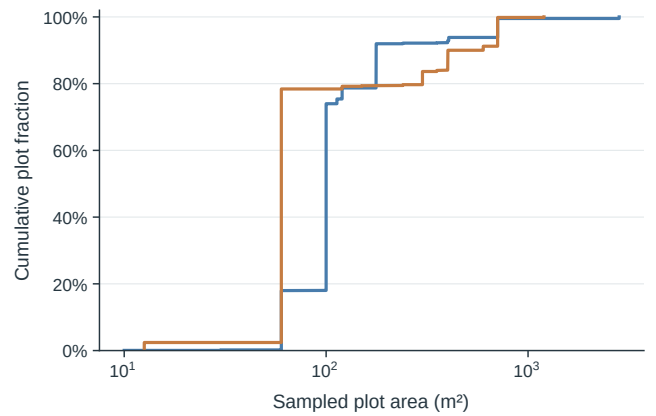
The expanded numerical figures use the archived source tables and five-repetition All38 probability snapshot. The derived tables contain unique per-plot probabilities; pooled AP, AUROC and Brier scores; per-fire metrics and contrasts; raw-probability reliability bins; Mc-M1 rank agreement; repetition-specific scores; observation support; and within-fire spectral comparisons. The All38 snapshot has five probabilities per plot for each model, not five independent plot observations. Their mean reproduces the archived mean-OOF probabilities within 2×10^{-16} . The current manuscript does not add fitted models, outcome-driven feature searches or resampling runs.

● Pilot25 · development ▲ Locked13 · held out

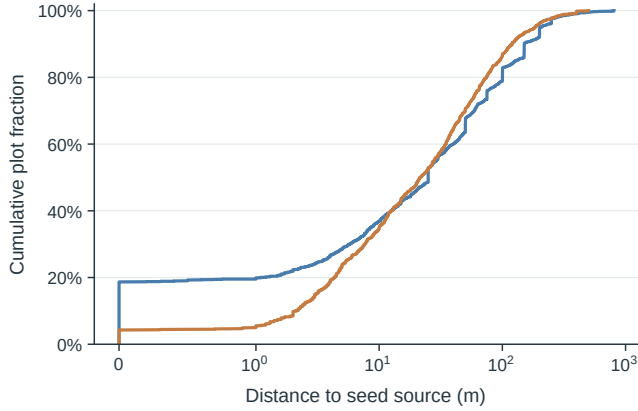
a Fire-level sample and outcome support



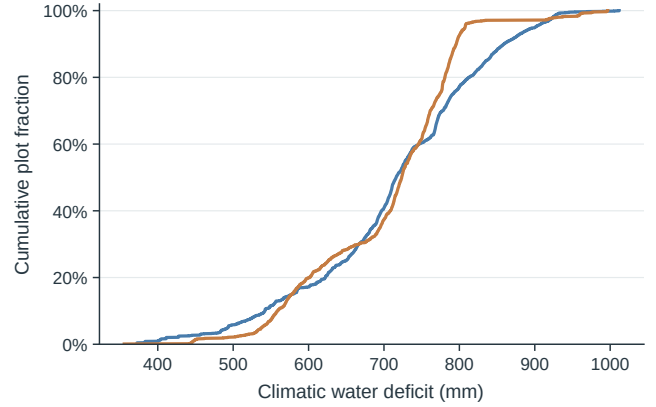
b Plot-area support



c Seed-source support



d Climatic support

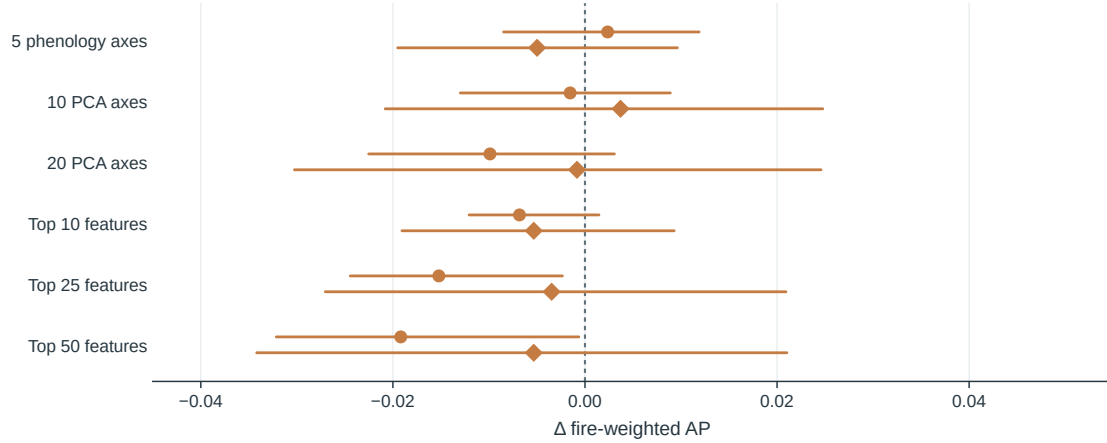


ECDFs represent observed field-plot distributions; symbols in a denote fire-level summaries.

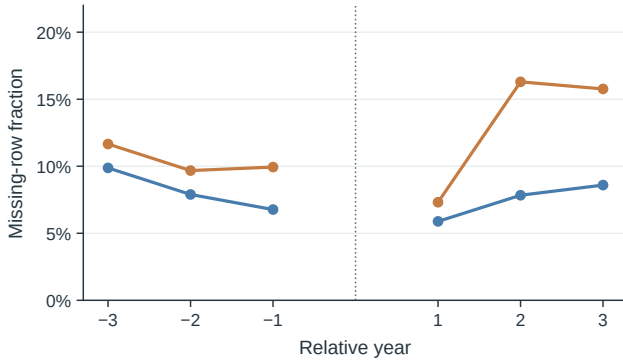
Figure S1. Additional sampling and covariate support. a, Fire-level sample sizes and observed conifer prevalence. b–d, Plot-level empirical cumulative distributions of sampled plot area, distance to a seed source and annual climatic water deficit in Pilot25 and Locked13. Plot area and distance use symmetric-log axes to retain small or zero values. Distributions represent the observed field compilation; no response matching or reweighting to a regional target population was performed.

● Elastic Net ◆ Gradient boosting ● Pilot25 · development ▲ Locked13 · held out

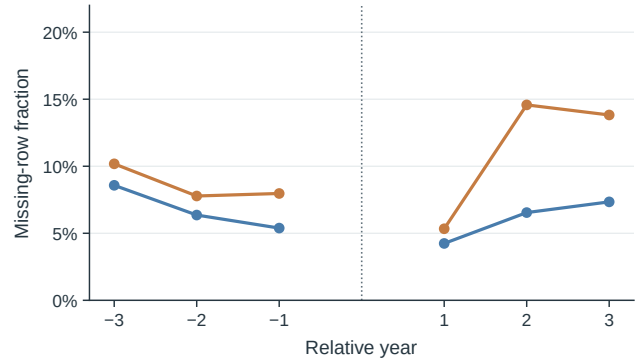
a Compressed seasonal detail versus the matched M1-only refit



b 30 m · missing time-series rows



c 90 m · missing time-series rows



Upper: observed differences and 95% fire-bootstrap intervals. Lower: row-weighted missingness; fire year excluded.

Figure S4. Low-dimensional seasonal additions and observation coverage. a, Corrected Pilot25 complexity-control refits: five phenology axes, 10 or 20 principal components, and the top 10, 25 or 50 seasonal predictors, each compared with its matched M1-only refit. Points are observed mean-out-of-fold AP differences and horizontal segments are the archived 95% fire-bootstrap intervals. Circles identify Elastic Net and diamonds histogram gradient boosting. b–c, Missing time-series rows divided by scheduled rows at 30 m and 90 m, restricted to common relative years -3 , -2 , -1 , $+1$, $+2$ and $+3$. Blue and orange distinguish Pilot25 and Locked13. The fire year has no eligible observations and is excluded. Row-weighted missingness is distinct from the fire-weighted prediction metric; lines only join observed summaries within the prefire or postfire period.

Figure 4 juxtaposes pooled AP, within-fire AP differences, prediction agreement and repetition-specific scores. These quantities answer different questions and are not combined into a single meta-analytic estimate. The plotted fire-level offsets only separate overlapping symbols. Hexagon shading represents the count of real plots in each bin. The complete mapping of the F01–F38 identifiers is supplied in Table S6 and in `source_data/derived/fire_key.csv` within the numerical figure package.

Table S6. Fire identifiers supporting the numerical panels.

Figure ID	Fire name / year	Cohort	Plots
F01	American River Complex 2008	Locked13	97
F02	Antelope Complex 2007	Locked13	126
F03	Bassetts 2006	Locked13	100
F04	Cub Complex 2008	Locked13	117
F05	Lakes 2002	Locked13	48
F06	Mason 2005	Locked13	36
F07	Moonlight 2007	Locked13	161
F08	Pigeon 2006	Locked13	83
F09	Porphyry South 2007	Locked13	71
F10	Power 2004	Locked13	175
F11	Rush 2006	Locked13	51
F12	Spanish 2003	Locked13	132
F13	Spring 2002	Locked13	44
F14	Battle Creek 2002	Pilot25	49
F15	Borrego 2002	Pilot25	46
F16	Btu Lightning Complex 2008	Pilot25	90
F17	Buffalo Creek 1996	Pilot25	117
F18	Bull Complex 1996	Pilot25	27
F19	Cannon 2002	Pilot25	180
F20	Dome 1996	Pilot25	43
F21	Eldorado 2000	Pilot25	49
F22	Fourmile Canyon 2010	Pilot25	20
F23	Freds 2004	Pilot25	121
F24	Hash Rock 2000	Pilot25	76
F25	Hayman 2002	Pilot25	301
F26	High Meadows 2000	Pilot25	84
F27	Jasper 2000	Pilot25	121
F28	Missionary Ridge 2002	Pilot25	83
F29	Overland 2003	Pilot25	93
F30	Pine Canyon 2005	Pilot25	43
F31	Ponil Complex 2002	Pilot25	83
F32	Ralston 2006	Pilot25	84
F33	Ricco 2005	Pilot25	45
F34	Rogers Shack 2001	Pilot25	93
F35	Sims 2004	Pilot25	86
F36	Summit Complex 2003	Pilot25	98
F37	Wallow 2011	Pilot25	44
F38	West Fork 2008	Pilot25	38

Fire names are display labels for the exact analysis-fire identifiers in the source table.

Figure 7 uses absence-oriented precision and recall. For each fire and capacity, recall is the fraction of all observed-absence plots captured in the selected subset. Averaging those values equally across fires retains the fire-level interpretation. No new intervals are attached to recall. The precision intervals remain the original, pointwise whole-fire bootstrap intervals; they are not simultaneous bands across all capacities.

Figure 9 calculates presence-minus-absence index contrasts within the same fire and year before aggregating across fires. Its direction counts are descriptive counts, not multiple-testing-adjusted findings. Recovery correlations use year +3 minus year +1 within each fire/outcome group, not a correlation between individual field trees. Neither the trajectories nor these correlations identify a conifer-specific causal pathway.

S7 Provenance of the conceptual illustration and inset

Figure 1 uses the author-supplied conceptual artwork generated with OpenAI image-2 (GPT Image 2; model gpt-image-2, OpenAI) as its pictorial substrate. The original raster is retained unchanged with a SHA-256 checksum. Only identified textual annotations and the empirical inset were corrected for the current manuscript; the forest scenes and hypothetical map tiles remain illustrations. The NIRv label, reduced-candidate composition and separate evidence roles of the original Locked13 test and exploratory All38 analysis were reconciled with the source tables. The inset was rebuilt as vector marks directly from the archived paired All38 contrasts, on a shared AP-difference scale from -0.06 to $+0.12$. Its 95% intervals are the original complete-fire bootstrap intervals; no image-generated statistical mark is treated as a quantitative source.

The original raster contains $1,644 \times 957$ pixels, corresponding to approximately 232 pixels per inch at a width of 180 mm. The exported high-resolution composite preserves editable numerical annotations, but exporting at a higher pixel density does not recover additional detail in the original illustration. The artwork tool and model were identified by the author as OpenAI image-2 (GPT Image 2; model gpt-image-2, OpenAI). This records the author-provided model identifier, not a model inference from the raster. The author remains responsible for publication rights and final scientific approval.

S8 Reproducibility tiers and release checks

The package separates integrity verification, deterministic figure regeneration and optional historical model refitting. The default verification checks SHA-256 hashes and the figure-number crosswalk. The numerical figure workflow regenerates derived tables and the eight data plates from frozen predictions and observations; the supplementary workflow retains the five data figures. Figures 2–9 map to the archived V9 plates, and Figure 1 retains the corrected V10 conceptual composite. The historical model-refitting workflow uses the original feature matrices, scripts and grouped partitions. It was not rerun for this typesetting release. Upstream Earth Engine extraction is separately documented and is not required for offline integrity checks or LaTeX compilation.

The full reproducibility archive retains all files from the supplied V7 full package in a historical directory, together with the V9 numerical-figure package and V8 supplementary-figure sources. The authoritative manuscript, captions and figure numbering are those of V11. Earlier manuscripts and templates are archival records, not alternative current submission files. V11 changes the image-generation disclosure and the placement of figures and tables; it does not change predictions, source-data values, model fits or resampling results.