

## ARTICLE

## Vegetation Ecology

# New methods provide a 300-year perspective on modern area burned in two wilderness areas of the southwest United States

C. A. Farris<sup>1</sup>  | E. Q. Margolis<sup>2</sup>  | J. M. Iniguez<sup>3</sup> | D. A. Falk<sup>4,5</sup>  |  
K. G. Gerow<sup>6</sup> | C. H. Baisan<sup>4</sup> | C. D. Allen<sup>7</sup> | T. W. Swetnam<sup>4</sup>

<sup>1</sup>National Park Service, Division of Fire and Aviation Management, Klamath Falls, Oregon, USA

<sup>2</sup>U.S. Geological Survey, Fort Collins Science Center, New Mexico Landscapes Field Station, Santa Fe, New Mexico, USA

<sup>3</sup>U.S. Forest Service, Rocky Mountain Research Station, Flagstaff, Arizona, USA

<sup>4</sup>Laboratory of Tree-Ring Research, University of Arizona, Tucson, Arizona, USA

<sup>5</sup>School of Natural Resources and the Environment, University of Arizona, Tucson, Arizona, USA

<sup>6</sup>Department of Mathematics and Statistics, University of Wyoming, Laramie, Wyoming, USA

<sup>7</sup>Department of Geography and Environmental Studies, University of New Mexico, Albuquerque, New Mexico, USA

**Correspondence**

E. Q. Margolis

Email: [emargolis@usgs.gov](mailto:emargolis@usgs.gov)

**Funding information**

National Park Service; U.S. Geological Survey

**Handling Editor:** Franco Biondi

**Abstract**

Climate change, expanding human ignitions, and increased fuels from fire exclusion are driving increases in area burned and fire severity in dry conifer forests of the western United States. Increasing area burned is occurring against the backdrop of a large fire deficit caused by over a century of fire exclusion. A key land management question is whether historically frequent fire regimes can be restored. Accurate estimates of historical annual area burned (prior to circa 1900) are necessary to evaluate modern area burned (after circa 1900), but are difficult to derive, and have rarely been calibrated or validated against modern fires, leaving their accuracy uncertain. We developed new methods to use tree-ring fire scars to reconstruct historical annual area burned and compare it to modern annual area burned. We focused on two southwestern US wilderness areas—Saguaro National Park (SAGU) and the Gila Wilderness (GILA)—that have a long history of using prescribed and managed fires. The abundant modern low- and moderate-severity fires allowed us to (1) calibrate and validate the fire-scar models against mapped fires to derive the first uncertainty estimates of reconstructed annual area burned and (2) test whether active fire management can help restore annual area burned to historical levels. A multi-model ensemble consisting of 10 individual member models accurately estimated area burned of mapped modern fires with no consistent biases. Each member model had distinct strengths and assumptions

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Ecosphere* published by Wiley Periodicals LLC on behalf of The Ecological Society of America. This article has been contributed to by U.S. Government employees and their work is in the public domain in the USA.

that made them suitable for specific applications (e.g., the synchrony model is easily applied, and Thiessen polygons provide spatially explicit area burned estimates). The accurate reconstruction of modern area burned from relatively sparse fire-scar data at GILA suggests that dense grids may not be necessary for accurate reconstructions. Our findings reveal that despite the near absence of fire in the early 20th century, both annual and 20-year sums of area burned in recent decades are back within historical levels at GILA, and trending toward historical levels at SAGU. These results demonstrate that fire management can help restore the historically prevalent, ecologically important process of widespread, frequent, low-to-moderate-severity fire in dry conifer forests.

#### KEYWORDS

area burned, fire, fire scar, Gila Wilderness, model, Saguaro National Park, tree ring

## INTRODUCTION

Area burned is increasing across the western United States, driven by climate change and human land use (Abatzoglou & Williams, 2016; Balch et al., 2017; Littell et al., 2009; Westerling et al., 2006). Fire severity is also increasing in dry conifer forests across the region (Parks & Abatzoglou, 2020; Singleton et al., 2019), driven in part by increased fuel loads from more than a century of fire exclusion (Covington & Moore, 1994; Hagmann et al., 2021). These fires have adversely affected human life and property (Burke et al., 2021), ecosystems (Coop et al., 2020), and ecosystem services, such as declining surface water supplies (Williams et al., 2022), posing major challenges for management. The increase in fire is occurring in the context of a modern fire deficit (Marlon et al., 2012; McClure et al., 2024; Parks et al., 2025), due to the disruption of fire regimes circa 1900 across the West (Margolis et al., 2022; Swetnam et al., 2016). In the 1970s, fire management in some wilderness areas of the western United States (e.g., the Gila Wilderness in New Mexico) began to shift toward using prescribed burns and managed lightning ignitions to restore low-severity fire as a keystone ecological process in dry conifer forests (van Wagtendonk, 2007). These areas have some of the highest rates of burning in recent decades in the western United States with predominantly low and moderate fire severity (Farris, 2009; Hunter et al., 2014; McClure et al., 2024), which raises the important question of whether historical (pre-1900) annual area burned has been restored.

Paleoecological records provide centennial to multi-millennial context for variability in ecological and climatic processes, and their relationships (Palli et al., 2023; Swetnam & Anderson, 2008). Pollen and charcoal sediments record multi-millennial

dynamics of fire–vegetation–climate–human interactions (Marlon et al., 2012; Roos et al., 2021) and were key in documenting a West-wide post-1900 fire deficit that is anomalous on multi-millennial timescales (Marlon et al., 2012). Tree-ring fire-scar reconstructions have revealed that the lack of fire in most dry conifer forests of the western United States during the 20th century is anomalous over the last 400–500 years (Johnston et al., 2016; Margolis et al., 2022), particularly in the southwestern United States (Swetnam & Baisan, 1996). The widespread collapse of low-severity frequent fire regimes due to human land use changes has dramatically altered forest structure and composition, and increased fire severity (Allen et al., 2002; Fulé et al., 1997; Hagmann et al., 2021; McClure et al., 2024). In many cases, these changes have led to fire-driven type conversion of forests to non-forest vegetation (Coop et al., 2020; Falk et al., 2022; Guiterman et al., 2022). Tree-ring reconstructions can also provide context and targets for restoring fire regime components, including fire frequency, severity, and seasonality (Margolis et al., 2022). However, accurate estimates of historical annual area burned, an important component of fire regimes that aligns well with management targets, are more difficult to derive, and existing estimates of historical area burned (e.g., Roos & Swetnam, 2012; Safford et al., 2022; Swetnam et al., 2011) have not been calibrated or validated using modern fire atlas data and therefore lack measures of uncertainty.

Quantifying historical annual area burned is useful to evaluate restoration effectiveness in landscapes where frequent modern (post-fire exclusion) low- and moderate-severity fire has been reintroduced. Understanding modern and future area burned trends requires an objective reference point, which can be provided by anchoring them to historical fire regimes. This reference becomes increasingly

valuable as efforts to restore fire regimes and reduce fuels through prescribed or lightning fires expand (Hessburg et al., 2021). Concepts and metrics informed by fire history, such as fire regime departure or condition class (Rollins, 2009), have provided important guidance for fuels and fire management (Swetnam et al., 1999). Accurate estimates of the historical range of variability in area burned—including annual and multi-decadal means, ranges, and variance—can serve as important benchmarks for comparing different fire management periods, or when evaluating modern or future fire management targets.

Area burned is one of multiple components of a fire regime, and when possible, should be interpreted in tandem with other metrics such as fire severity (see [Discussion](#)). Area burned is critical for understanding the extent and feedbacks of landscape-scale fire regimes over time. Moreover, annual area burned is one of the longest consistently recorded metrics of modern fire regimes, and centuries-long tree-ring reconstructions can offer unique insights into changes over long timescales. Despite its importance, the current lack of historical reference information for area burned limits our ability to compare modern annual area burned against the full range of historical variability. We aim to address this knowledge gap to better understand how different management strategies have influenced fire activity, as has been done with fire severity and frequency (e.g., McClure et al., 2024).

Some of the most frequently burned modern landscapes in the western United States are in the southwestern United States, due to remote locations and the early implementation of prescribed fire and managed lightning fires (Hunter et al., 2014). The Saguaro Wilderness in Saguaro National Park (SAGU), and the Gila Wilderness in the Gila National Forest (GILA) both began implementing “prescribed natural fire” programs in the 1970s. Both landscapes have forests that have burned 5–10 times in the last century, with many trees and forest stands surviving these fires (Farris et al., 2010; Hunter et al., 2014; McClure et al., 2024). These programs were part of a broader fire management emphasis to restore fire regimes in multiple locations in the western United States, including Yosemite and Sequoia National Parks in California and the U.S. Forest Service Selway Bitterroot Wilderness in Montana (van Wagtenonk, 2007). Comparisons of these modern frequent fire regimes with historical burning provide important opportunities to evaluate the influence of changing fire management through time (Collins & Stephens, 2007). These landscapes with frequent, modern, low-severity fire are also valuable because they provide rare, long records of mapped modern fires that can be used to calibrate and validate tree-ring methods to reconstruct annual area burned and quantify estimation uncertainty.

Multiple methods exist for reconstructing area burned from tree rings through direct or indirect inference. Indirect estimates of area burned often rely on rates of burning over time, derived from tree-ring reconstructions of fire frequency and then extrapolated across vegetation types. Using this approach, Mallek et al. (2013) and Haugo et al. (2019) concluded that modern area burned (mid-1980s through the mid-2000s) in six forest types in Oregon and California, was well below historical levels. Similar methods were used to determine that the “record” 2020 fires in California were near historical average area burned (Safford et al., 2022), but that the proportion of high-severity fire exceeded historical conditions (Williams et al., 2023).

Direct inferences of historical area burned can be made from tree-ring-based fire history specific to the study area, including tree-ring fire scars, or forest ages in high-severity fire regimes (Agee, 1993; Farris et al., 2010; Taylor & Skinner, 1998). The ratio or “synchrony” method is a commonly used direct inference, which estimates annual area burned by multiplying the size of the study area by the proportion of recording tree-ring fire-scar sites or trees burned in a year (Morrison & Swanson, 1990; Taylor & Skinner, 1998). Romme and Despain (1989) used the ages of forest stands that were assumed to have regenerated following high-severity fire to reconstruct area burned in Yellowstone and concluded that the 1988 fires were not anomalous in size when evaluated in a multi-century context. Additional variations on direct methods include spatial interpolation of historical fire occurrence from tree-ring fire-scar point records (Hessl et al., 2007; Swetnam et al., 2011).

Another approach models area burned using centuries-long, climate-sensitive tree-ring width chronologies through both direct and indirect inference. This method combines relationships between climate and area burned, as derived from modern fire atlas data or the percentage of sites scarred from tree-ring fire-scar chronologies (Girardin, 2007; Roos & Swetnam, 2012; Westerling & Swetnam, 2003). It assumes that annual to decadal moisture variability, inferred from tree growth (ring-width indices), is the primary driver of area burned and may provide the best estimate of potential area burned. An advantage of this approach is the ability to extend area-burned reconstructions to periods beyond the fire-scar record. However, fire is a contingent process, and interannual climate conditions alone are insufficient for predicting when widespread, landscape-scale fires will occur. Climate variability alone also fails to capture bottom-up factors such as human and nonhuman ignition patterns, short-term fire weather, or fuel availability (e.g., Dewar et al., 2021; Parks et al., 2015).

These existing methods for reconstructing area burned from tree rings offer important context for

modern rates of burning, but many lack a spatial component and they are rarely calibrated and validated with modern fire maps, leading to unknown accuracy. Farris (2009) and Farris et al. (2010) demonstrated that tree-ring fire-scar networks can accurately reconstruct spatial patterns of mapped modern fires and their fire regime parameters in dry conifer forests. In this paper, we build on those efforts to provide a historical perspective of modern area burned by presenting a continuous tree-ring fire-scar time series of annual area burned (in hectares) in two southwestern US wilderness areas with relatively abundant modern fires: Saguaro National Park and the Gila Wilderness. We aimed to determine whether modern annual area burned falls within the historical range of variability. Our specific research objectives were to: (1) develop new methodologies for creating and assessing the performance of 11 tree-ring fire-scar models for area-burned estimation, including a multi-model ensemble; (2) produce validated, multi-century time series of annual and multi-decadal area burned; and (3) compare modern and historical area burned in two wilderness areas across three different fire management periods. To achieve these objectives, we applied four time-series analysis techniques to fire-scar area-burned reconstructions for the first time: (1) a multi-model ensemble; (2) quantification of estimation uncertainty in paleofire records; (3) time-varying predictor subset calibration to address

potential spatiotemporal variability from sample attrition in tree-ring records; and (4) empirical validation metrics to assess multiple aspects of model performance.

## MATERIALS AND METHODS

### Study area

The study areas are in the Rincon and Mogollon Mountains in the southwestern United States. The SAGU site is situated in the Rincon Mountains within Saguaro National Park, east of Tucson, Arizona, USA (Figure 1). It encompasses the coniferous forest belt on Mica Mountain covering an area of 2780 ha (elevation 1800–2640 m). The GILA site is in the Mogollon Mountains of southwestern New Mexico, encompassing approximately 8186 ha of coniferous forests (elevation 2000–3150 m) within the White Creek Watershed in the Gila National Forest.

Both sites are dominated by ponderosa pine (*Pinus ponderosa* P. & C. Lawson), but also include southwestern white pine (*Pinus strobiformis* Engelm.), Gambel oak (*Quercus gambelii* Nutt.), white fir (*Abies concolor* [Gord. & Glend.] Lindl. ex Hildebr), and Douglas-fir (*Pseudotsuga menziesii* [Mirbel] Franco). We masked the GILA study area to include only ponderosa pine and



**FIGURE 1** Location of study areas in the Saguaro Wilderness Area in Saguaro National Park (SAGU), southern Arizona, United States, and the Gila Wilderness (GILA) in southern New Mexico, United States. Green shaded areas represent the geographic range of coniferous forests. *Source:* Esri, DeLorme, HERE, TomTom, Intermap, increment P Corp., GEBCO, USGS, FAO, NPS, NRCAN, GeoBase, IGN, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), swisstopo, MapmyIndia, and the GIS User Community (accessed October 21, 2024).

mixed-conifer forests to be consistent with the locations of the tree-ring fire-scar samples. This excluded 11% of the watershed that was piñon-juniper and spruce-fir forest.

Average annual precipitation at GILA ranges from 785 mm at the highest elevations to 466 mm at lower elevations (1991–2020; PRISM, 2022). At SAGU, average annual precipitation varies between 680 and 630 mm between higher and lower elevations (1991–2020; PRISM, 2022). The seasonal distribution of precipitation at both sites is bimodal: approximately half falls as rain between May and September, peaking in July and August during the summer monsoon season (1991–2020; PRISM, 2022), and the remainder falls as rain or snow between October and March, peaking in December and January. Fire season in both study areas typically occurs between April and September. Maximum area burned peaks during June, whereas the maximum number of ignitions is in early July, coincident with the monsoon and peak lightning occurrence season (Baisan & Swetnam, 1990; Crimmins & Comrie, 2004). Lightning fires are common during the monsoon season in July and August, but rarely become widespread before being extinguished by rain, reflecting a unique climatological fire regime of the region (Arizpe et al., 2020; Margolis et al., 2017).

## Fire management history

Hunter et al. (2014) identified three fire management periods common to both study areas that reflect major human-caused influences on the fire regime: pre-suppression, fire suppression, and managed fire. We modified that classification slightly by replacing fire “suppression” with fire “exclusion.” This distinction includes the indirect exclusion of fires by land use and cultural shifts years before organized fire suppression began in the early 1900s, which better aligns with observed declines in fire at SAGU and GILA in the late-19th and early-20th centuries. Due to their remote locations, both study areas had multiple fires in the mid-1900s; therefore, there were more fires during the “fire exclusion” period than in most other areas in the southwestern United States. We define each period as follows:

1. *Pre-exclusion period* (prior to 1894 at SAGU; prior to 1905 at GILA): During this period, frequent fires generally spread largely unobstructed. Ignitions from lightning and Indigenous fire use were abundant, with the primary drivers of total area burned being favorable climate conditions and available fine fuels (Allen, 2002).
2. *Fire exclusion period* (1894–1971 at SAGU; 1905–1974 at GILA): This period marked the onset of

(a) disruptions to widespread burning due to heavy grazing that removed fine fuels, and the displacement of Indigenous populations (Swetnam & Baisan, 2003) and (b) direct fire suppression efforts by federal land management agencies, beginning around 1910, aimed at minimizing the total area burned (Hunter et al., 2014).

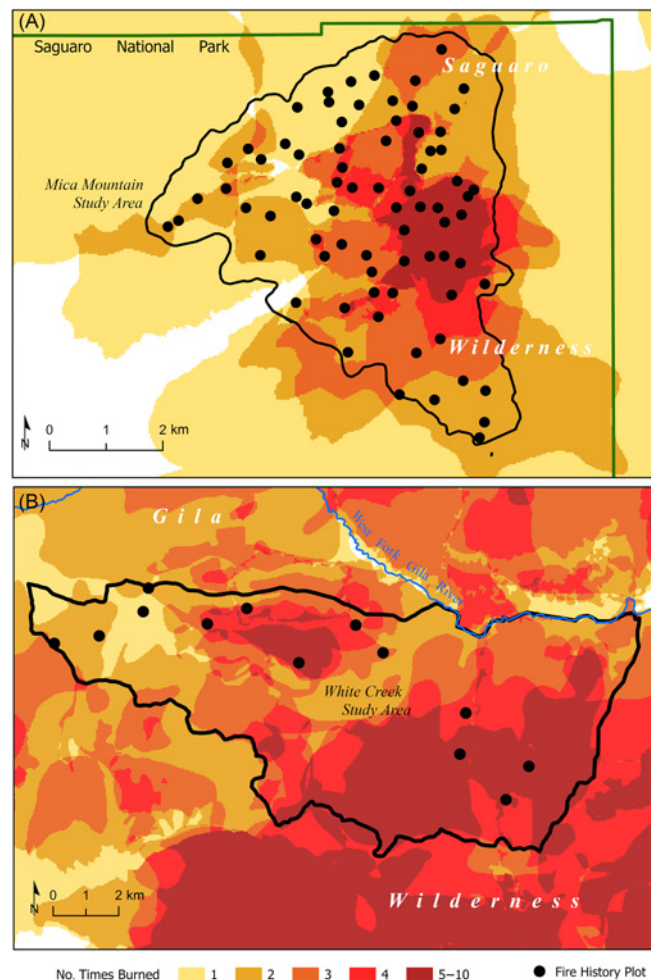
3. *“Managed fire” period* (1972 to “present” at SAGU; 1975 to “present” at GILA, where present was 2016 for our study): During this period, fire managers implemented prescribed burns and utilized lightning ignitions to increase area burned, aiming to restore fire as an ecological process. The extent of these “managed fires” varied over time and space as management programs evolved and experience increased (see Hunter et al., 2014). At SAGU, more than half of the acreage burned between 1993 and 2010 resulted from management-ignited prescribed fires, often in areas that had previously burned. At GILA, all fires during the managed fire era were lightning-ignited and also included many reburns.

We hereafter refer to the fire exclusion and managed fire periods collectively as the “modern era” (i.e., after fire exclusion began in 1894 at SAGU and 1904 at GILA). The pre-exclusion period will be referred to synonymously as the “historical era.”

## Fire data

### Tree-ring fire-scar data

We assembled all available tree-ring fire-scar data to obtain the most comprehensive, cross-dated fire-scar plot network in each study area (Figure 2). At SAGU, we used a systematically sampled fire-scar network consisting of  $n = 60$ , ~1 ha plots (Farris et al., 2010), three additional plots (<1 ha each) collected opportunistically by the National Park Service (NPS) and University of Arizona staff before prescribed burns (archived at the Laboratory of Tree-Ring Research), and a targeted fire-scar dataset consisting of  $n = 12$  plots (~1–2.5 ha each; Baisan & Swetnam, 1990). In all three collections, fire scars were collected from multiple trees within a small area, and all fire years were composited into a single master chronology (i.e., “composite plots”; sensu Dieterich, 1980, Farris et al., 2010). Our resulting gridded fire-scar network consists of  $n = 75$ , covering both major vegetation types proportionately (ponderosa pine = 89% of the forested portion of the study area and 87% of the plots; mixed conifer = 11% of the study area and 13% of plots). Importantly, Farris et al. (2010) collected samples from



**FIGURE 2** Modern fire frequency derived from fire atlas maps and the locations of tree-ring fire-scar sampling plots in (A) the Saguaro Wilderness study area (SAGU) atop Mica Mountain (2780 ha) in southern Arizona and (B) the Gila Wilderness study area (GILA) in the White Creek Watershed (8186 ha) in southern New Mexico. Fire atlas data cover the period 1937–2016 for Saguaro (maximum number of times burned = 10) and 1909–2016 for Gila (maximum number of times burned = 8).

both live trees and dead remnant wood in 2002, so overall sample depth for this network is robust through 2001. Samples collected by Baisan and Swetnam (1990) and NPS were restricted to remnant wood.

At GILA, we used fire-scar networks from two studies in the White Creek Watershed. Swetnam and Dieterich (1985) sampled fire scars in 1980 from multiple small stands to characterize fire history across the eastern portion of the study area (Figure 2B). Eight discrete areas were sampled covering approximately 0.75–2 ha each. For the purpose of this study, we considered those collection sites to be eight individual plots because they were sufficiently dispersed and consisted of multiple fire-scarred specimens collected within small, discrete areas (consistent with the composite plot data structure

at SAGU). Similarly, Abolt (1997) sampled fire scars at five discrete sites in the western portion of the GILA study area in ponderosa pine and mixed conifer forests. Each collection site was treated as a separate plot for this study. The combined GILA dataset consists of  $n = 13$  composite plots. They are proportionately representative of the two major forest types (ponderosa pine = 76% of the forested portion of the study area and 69% of plots; mixed conifer = 24% of study area and 31% of plots). Sample depth decreased sharply at GILA after 1980 because that is when the Swetnam and Dieterich (1985) sites were sampled, and Abolt (1997) sampled primarily remnant wood.

## Fire atlas data

The NPS developed a comprehensive fire atlas database of fire perimeters at SAGU dating back to 1937, originally compiled by Swantek (1997), updated through 2013 by Hunter et al. (2014) and through 2016 for this study (Figure 2A). This is considered to be a very complete and accurate fire atlas because of intensive mapping efforts for a relatively small area, including early survey-based ground mapping for major fires in the 1940s and 1950s (see Farris et al., 2010). The period of overlap between the fire atlas and fire-scar record at SAGU was 1937–2001. This period was used to calibrate and validate the fire-scar models of annual area burned versus mapped fires and is referred to hereafter as the calibration and validation period at SAGU.

For GILA, we used a long-term fire atlas of fire perimeters dating back to 1909, compiled originally by Rollins et al. (2001), and updated through 2013 by Hunter et al. (2014), and through 2016 for this study (Figure 2B). The period of overlap of empirical fire-scar records and fire atlas maps at GILA is 1909–1980 because tree-ring fire scars were collected from live trees in 1980. Due to the fact that relatively few fires occurred during the period of overlap, we extended the end year of the calibration and validation period at GILA to 2016 (see methods below).

## Area-burned reconstruction

### Ensemble modeling approach

To address our study objectives, we sought not only to develop the most accurate overall estimate of area burned for a given year but also to develop a robust framework for characterizing estimation uncertainty and variability over time using an ensemble modeling approach. Our

multi-model ensemble consists of 10 individual models (members) that vary in complexity and the assumptions used to infer annual area burned from fire-scar networks (Table 1). These 10 models include five core models and five “filtered” variants of the core models. The mean and prediction ranges of the member models are expected to provide a more robust estimation of area burned and its uncertainty over time than any individual member alone, due to offsetting errors and biases (Weigel et al., 2008).

## Member model definitions

### *SYNC (fire-scar synchrony model)*

Fire-scar synchrony is defined as the proportion of actively recording fire-scar plots that detected a fire scar in a given year (Farris et al., 2010):

$$P_i = \frac{n_i^d}{n_i^a}, \quad (1)$$

where  $n_i^d$  is the number of plots that detected a fire scar in year  $i$ ;  $i = 1, 2, \dots, I$ , and  $n_i^a$  is the number of plots that are actively recording in year  $i$ . A plot is considered “actively recording” each year if (1) there was at least one living tree to record a fire and (2) the tree(s) were previously scarred, making them susceptible to subsequent scarring (Romme, 1980; Stephens et al., 2010).

The quantitative empirical relationship between annual area-burned ( $A_i$ ) and fire-scar synchrony ( $P_i$ ) is rarely establishable for contemporary fire history studies, as most study areas have not burned enough times since the onset of fire exclusion. The SYNC model assumes that for a network of fire-scar samples within a defined study area ( $S$ , in hectares), the proportion of the study area burned ( $A_i S^{-1}$ ) is equivalent to  $P_i$ . Conceptually, the SYNC model represents a special case of regression through the origin ( $Y = \beta X + \epsilon$ ), where  $A_i S^{-1} \cong 0$  when  $P_i = 0$ , and  $\beta$  is assumed to be reasonably approximated by a blind imputation of 1.0 (i.e., isometric scaling with a constant but undefined variance):

$$\hat{A}_i S^{-1} = (1.0)(P_i). \quad (2)$$

This can be simplified to the more commonly used form found in Taylor and Skinner (1998) to estimate annual area burned in hectares:

$$\hat{A}_i = (P_i)(S). \quad (3)$$

The model assumes that  $\beta = 1.0$  for all years in a chronology. Estimation error is presumably assumed to be

constant, though it has never been explicitly specified or calculated in practice. The model does not account for temporal variations in  $\beta$  due to factors such as the loss of tree-ring samples or shifts in fire-size distributions over time, and therefore requires no ancillary data to fit this assumed coefficient.

### *Ratio of means regression*

If the fire-scar record overlaps with mapped fire perimeters, the actual relationship between  $A_i$  and  $P_i$  can be estimated empirically through calibration (i.e., fitting the model coefficient using independent reference data). When assumptions of linearity and a 0-origin still hold (which can now be tested), and when variance in  $A_i$  increases with higher values of  $P_i$  (typical of fire-scar data; see Farris et al., 2013), the ratio of means (RoM) is a robust and efficient estimator compared to ordinary least squares methods (Farris et al., 2013; Schaeffer et al., 2006).

The RoM regression model is of the form,  $\mu_{Y|X} = \beta X$ , where  $\beta = \hat{\rho}_{\text{RoM}}$ , the fitted RoM estimator. It is calculated as:

$$\hat{\rho}_{\text{RoM}} = \frac{\sum_{i=1}^n A_{ic} S^{-1}}{\sum_{i=1}^n P_{ic}} = \frac{\bar{A}_c S^{-1}}{\bar{P}_c}, \quad (4)$$

where  $A_{ic}$  is observed (mapped) area burned and  $P_{ic}$  is fire-scar synchrony for year  $i$  within calibration period  $c$ . The second expression represents the simplified form of the estimator. The value of  $\hat{\rho}_{\text{RoM}}$  will be  $<1.0$  if mean fire-scar synchrony during the calibration period is greater than the mean observed proportion of the study area burned;  $>1.0$  if mean fire-scar synchrony is lower than the mean observed proportion of area-burned; and 1.0 if there was a perfect 1:1 agreement (i.e., the assumed coefficient of the SYNC model). CIs can be estimated using a bias-corrected and accelerated (BCa) bootstrapping approach (see Gerow et al., 2021).

Annual area burned in hectares for historical years in which only  $P_i$  is known (independent of the calibration set) is calculated as:

$$\hat{A}_i = (\hat{\rho}_{\text{RoM}})(P_i)(S). \quad (5)$$

Conveniently, the RoM model is an area-weighted estimate because years with larger fires contribute more to RoM relative to small fire years (consistent with area-burned-based fire history metrics such as filtered Mean Fire Return Intervals). Unlike the SYNC model, which has a constant coefficient over time, different values of  $\hat{\rho}_{\text{RoM}}$  can be fitted for any period in the tree-ring record containing different sets of samples using a time-varying predictor subset calibration (see section below).

**TABLE 1** Models used to reconstruct area burned from tree-ring fire scars.

| Model code | Model name                 | Description                                                                                                                                                           |
|------------|----------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| SYNC       | Fire-scar synchrony        | Assumes an ideal 1:1 relationship for the ratio of annual area burned and fire-scar synchrony.                                                                        |
| SYNCF      | SYNC filtered              | Excludes fire years when only a single plot recorded a fire.                                                                                                          |
| RoM        | Ratio of means regression  | Area-weighted ratio regression model fitted to the observed ratio of <i>total</i> area burned to <i>total</i> fire-scar synchrony during the full calibration period. |
| RoMF       | RoM filtered               | Excludes fire years when only a single plot recorded a fire.                                                                                                          |
| MoR        | Mean of ratios             | Unweighted ratio estimate fitted to the observed ratio of <i>annual</i> area burned to <i>annual</i> fire-scar synchrony during the calibration period.               |
| MoRF       | MoR filtered               | Excludes fire years when only a single plot recorded a fire.                                                                                                          |
| TP         | Thiessen polygon           | Spatially explicit interpolation that generates a polygon around each plot, encompassing all areas closer to that plot than any other plot.                           |
| TPF        | TP filtered                | Excludes fire years when only a single plot recorded a fire.                                                                                                          |
| TPA        | Thiessen polygons adjusted | Rule-based algorithm applied to the TP model to adjust for the scarring record of surrounding polygons.                                                               |
| TPAF       | TPA filtered               | Excludes fire years when only a single plot recorded a fire.                                                                                                          |
| ENS        | Ensemble model             | The mean of all 10 member model estimates, representing our best fire-scar estimate of area burned.                                                                   |

### Mean of ratios

The mean of ratios (MoR) model is similar in form and assumptions to the RoM model, but the MoR estimator ( $\hat{\rho}_{\text{MoR}}$ ) represents the unweighted average of the annual ratios of fire-scar synchrony to observed area burned during the calibration period (unlike the RoM, which sums the annual ratios for the entire calibration period) (Rao 2002). The estimator is calculated as:

$$\hat{\rho}_{\text{MoR}} = \frac{1}{n} \sum_{i=1}^n \left( \frac{A_{ic} S^{-1}}{P_{ic}} \right). \quad (6)$$

Values of  $\hat{\rho}_{\text{MoR}}$  have the same range and interpretation as  $\hat{\rho}_{\text{RoM}}$ . Similarly, a separate estimator can be fitted to empirical data for different periods in a chronology. For years where only  $P_i$  is known, annual area burned in hectares is estimated as:

$$\hat{A}_i = (\hat{\rho}_{\text{MoR}})(P_i)(S). \quad (7)$$

Years with small fires may be mapped less accurately or have lower fire-scar detection probabilities than larger fire years (Farris, 2009; Farris et al., 2010), so equal weighting of years may seem unwarranted. However, this limitation is partially offset in practice because MoR can be calculated only for fire years in which  $A_{ic}$  and  $P_{ic}$  are both  $>0$  (typically, larger fire years). Moreover, the fitted

MoR model is useful for bracketing estimation variability and uncertainty within the ensemble framework, especially if the relationship between  $A_{ic}$  and  $P_{ic}$  is skewed for one or more large fire years.

### Thiessen Polygons

The first three core models rely on general statistical relationships between  $A_i$  and  $P_i$ . An alternative approach utilizes a spatially explicit model that considers the specific location of plots that recorded fires. Thiessen polygon tessellation has proven to be an accurate and efficient binary spatial interpolator for this purpose at SAGU (Farris et al., 2010). Given a set of plots  $\{p_1, p_2, \dots, p_n\}$ , the Thiessen polygon region surrounding a specific “target plot” in year  $i$  (denoted  $R_{pi}$ ) is defined as the area (in hectares) encompassing all locations closer to the target plot than any other plot (Burrough & McDonnell, 1998).

$\hat{A}_i$  can then be derived from the conditional summation of polygon regions, as:

$$\hat{A}_i = \sum_{p=1}^n R_{pi} d_{pi}, \quad (8)$$

where the conditional variable  $d_{pi} = \{1 \text{ if a fire was detected at plot } p, 0 \text{ otherwise}\}$ . The size and spatial extent of each polygon determine the relative weighting of individual plots.

A new polygon set must be generated each year as available plots are added or lost in the chronology, providing a built-in form of calibration over time through the readjustment of Thiessen polygon configurations and resultant plot weightings. The sizes and shapes of Thiessen polygons (TP) are determined solely by the distance and angle between available plots (not the location of observed fires), meaning independent fire atlas maps are not required for parameterization. Variance is challenging to estimate and has not been reported in previous TP-based fire-scar studies, but the relative size of polygons can serve as a proxy for spatial uncertainty, as lower plot densities lead to larger polygons.

#### *Thiessen polygon adjusted*

Farris et al. (2010) observed that burned plots that did not record a fire scar (i.e., Type II errors) were most common when all surrounding plots/polygons recorded the fire year. To reduce potential underestimation of annual area burned, we applied a rule-based algorithm from Farris et al. (2010) to reassign plots without scars to “burned” status when they were surrounded by scarred plots (i.e.,  $d_p = 0 \rightarrow d_p = 1$  in Equation 8). This method corresponded better to the largest mapped fires at SAGU than the basic TP model. For historical years, there is no way to know for certain whether the absence of a fire scar indicates that the associated TP region was completely unburned, partially burned, or completely burned. However, the Thiessen polygon adjusted (TPA) model provides a useful complementary estimate to help bracket the range of annual area-burned estimates, and to better characterize uncertainty within the ensemble framework (particularly for large fire years).

#### *F (filtered models)*

In a typical fire-scar chronology, there are many years where a fire scar is detected at only a single plot. These could represent small “spot fires” that did not spread beyond the tree or plot, or fires that spread beyond the plot but failed to burn or scar trees in adjacent plots. We observed both scenarios during recent fire events at each study area. The default for the five core models is to always convert single-plot fire scars to an annual area-burned estimate. Over long time periods, however, this might result in cumulative overestimation of total area burned for a given period if single-plot fires were (1) common and (2) consisted mostly of small spot fires. Overestimation would be more likely when the sample size is lowest, such as earlier periods in a chronology because the area contribution of a single-plot fire scar would be disproportionately high (e.g., produce a relatively high value of  $P_i$  or a large Thiessen polygon region).

The filtered variant of the five core models (denoted with an F at the end of the model code) addresses this potential overestimation by assuming that *all* single-plot fires are spot fires with 0 ha burned (i.e., a filter of a minimum of two plots scarred):

$$\hat{A}_{i|d} = \begin{cases} \hat{A}_i, & n_i^d \geq 2 \\ 0, & n_i^d < 2 \end{cases} \quad (9)$$

This assumption is not always accurate; however, the five *F*-variant models offer a conservative lower bound that is useful for characterizing long-term estimation ranges and uncertainty. This correction could be especially valuable when comparing total area burned over extended time periods, where small errors could accumulate.

### Ensemble fire-scar model

The ensemble fire-scar model (ENS) represents our primary estimate of annual area burned. The ENS solution for a given year, denoted as  $\bar{A}_i$ , is calculated as the unweighted mean of all 10 member models:

$$\bar{A}_i = \frac{1}{M} \sum_m \hat{A}_{im}, \quad (10)$$

where  $\hat{A}_{im}$  is the annual area-burned estimate from member model  $m$ ;  $m = 1, 2, \dots, M$  for year  $i$ ;  $i = 1, 2, \dots, I$ . CIs for any individual year can be derived using either parametric methods or bootstrapping. However, to assess estimation uncertainty over multiple years within a time series, the distribution of interannual predictions from the member models provides a more effective approach (see time series section below).

### Time-varying predictor subset calibration

A key challenge in calibrating multi-century models using paleo tree-ring indices (e.g., paleoclimate or paleofire) is that calibration data are usually restricted to recent periods with the largest sample sizes (i.e.,  $n_i^a$ ), which may be less representative of tree-ring signals for earlier periods in the chronology (Farris et al., 2013; Meko, 1997). Specifically, we are concerned that the relationship between fire-scar synchrony and annual area burned may change over time due to factors such as differential sample attrition, plot clustering, or shifts in fire-climate regimes (Abatzoglou & Williams, 2016).

Thus, a global model coefficient fitted only to recent data may inadequately represent earlier years of the chronology. To address this, we applied a time-varying predictor subset approach from dendroclimatology (see Meko, 1997) to fit ratio estimators (i.e.,  $\hat{\rho}_{\text{ROM}}$  and  $\hat{\rho}_{\text{MoR}}$ ) separately for each period in the chronology, using only the samples that would have been available during that period. This approach is not used for the SYNC and SYNCF models, which have a constant coefficient of 1.0 across all years and sample sizes.

To implement this approach, we divided the fire-scar record into a series of distinct temporal periods called epochs  $\{E_1, \dots, E_n\}$ , defined as periods containing the same membership of actively recording plots (Reed & Johnson, 2004). We then calculated fire-scar synchrony separately for each epoch using only plots that were actively recording during that period (i.e.,  $P_{\{E_1, \dots, E_n\}}$ , which is substituted for  $P_{ic}$  in Equation 4). This ensures that for every year in the chronology that we estimate area burned, the model coefficient was fit to mapped fires using the same set of plots available during that time.

A limitation of this approach, highlighted by Meko (1997) in the context of dendroclimatic reconstructions, is that some plots from historical epochs lack trees still alive to record climate—or mapped fires in our case—during the modern calibration era. For example, of the 40 plots recording fires at SAGU in 1733, only 30–35 contained trees still alive to record mapped fire years throughout the 64-year calibration period (which spanned five epochs; Appendix S1: Table S1).

For plots lacking live trees capable of recording a mapped fire during the calibration period, we assigned the scarring outcome based on (1) the plot's location inside or outside of a modern fire perimeter, (2) Type I and Type II detection error rates derived from empirical fire-scar data at SAGU (1% and 17%, respectively; see Farris et al., 2010), and (3) a random error algorithm. We first assigned a random value from 0 to 1.0 to each plot. We then assigned each plot a status of “scarred” if it met either of the following conditions: (1) it fell within a mapped fire perimeter and had a uniform random number greater than 0.17 or (2) it fell outside the fire perimeter and had a random number less than 0.01. Plots that did not meet these criteria were assigned a status of “unscarred.” This approach generates a distribution of “scarred” plots that aligns with the empirical observations at SAGU while incorporating randomized error to account for the fact that not all burned plots will record a given fire and, a small percentage of fire perimeter boundaries are inaccurately mapped.

Time-varying predictor subset calibration preserves the epoch-specific relationship between available samples and annual area burned, producing a more representative

coefficient than if those plots were not included. At SAGU, a high proportion of plots recorded fires across both historical and modern epochs (80%–91%), so modeling scarring rates was necessary for only a small number of plots (Appendix S1: Table S1). In contrast, the proportion of historical plots at GILA still actively recording in the calibration period ranged from 0% to 39% among epochs (Appendix S1: Table S2). This discrepancy reflects the reliance on remnant wood sampling at GILA, and the fact that 85% of the total modern-era area burned occurred after fire scars were sampled at GILA in 1980.

## Area-burned time series

We developed two time series for each study site to assess differences in area burned over time at two temporal resolutions. We constructed an annual time series by plotting the area-burned estimate from the ENS fire-scar model and the fire atlas maps. Years in which fire atlas and fire-scar records overlap will show both estimates. This resulted in a continuous area-burned time series for every year between 1700 and 2016 at SAGU and 1733 and 2016 at GILA.

Annual area-burned fire-scar time series often include many years with no recorded fire, which can make it challenging to identify multi-year trends. To address this, we developed a moving window time series to better highlight multi-year trends. We chose 20-year windows because that period is sufficiently long to typically encompass at least one major fire year, but sufficiently short to be sensitive to short-term management or climate-driven fluctuations in area burned (Swetnam & Betancourt, 1998). Another advantage of a moving window time series is that it provides a robust framework for quantifying temporal trends in estimation uncertainty that may not be readily apparent in annual resolution data, including cumulative or compounding uncertainty. This is particularly useful for our objective of evaluating modern-era area burned in a historical context.

For the moving window time series, let  $j$  be the target year for which a 20-year sum will be calculated, where the period includes year  $j$  and the preceding 19 years (e.g., for  $j = 1900$ , the 20-year window would be 1881–1900). The estimated total area burned for each 20-year window is denoted as  $\hat{A}_j^{20}$  and calculated as:

$$\hat{A}_j^{20} = \sum_{i=j-19}^j \bar{\hat{A}}_i, \quad (11)$$

where the summation term  $\bar{\hat{A}}_i$  is the ENS fire-scar estimate of annual area burned (Equation 10).

## Uncertainty estimates

We calculated two composite measures of uncertainty from the distribution of member model estimates. The first is combined model spread, defined as the moving window estimation range of all 10 member models (i.e., the highest and the lowest 20-year average model estimate for every window start year). Combined model spread is a broader, more conservative measure of uncertainty than the CI of the ENS model solution, which will always be contained within the combined model spread. Let  $l_j^{20}$  and  $u_j^{20}$  represent the lower and upper bounds of combined model spread for moving window start year  $j$ . Model spread is calculated as:

$$\begin{aligned} l_j^{20} &= \min(\hat{W}_m^{20}; m = 1, 2, \dots, M); \\ u_j^{20} &= \max(\hat{W}_m^{20}; m = 1, 2, \dots, M). \end{aligned} \quad (12)$$

where  $\hat{W}_m^{20}$  represents the 20-year moving window estimate of area burned for each member model  $m$ :

$$\hat{W}_m^{20} = \sum_{i=j-19}^j \hat{A}_{im}. \quad (13)$$

We developed an even broader, more conservative measure of time series uncertainty for this study, termed blended model uncertainty. This measure assumes that, for each individual year, the lowest and highest area-burned estimates are always the most accurate. Blended model uncertainty is the resultant artificial annual times series constructed from those low and high values, where  $L_j^{20}$  and  $U_j^{20}$  denote the lower and upper bounds for a given window start year  $j$ . The bounds are calculated as:

$$L_j^{20} = \sum_{i=j-19}^j L_i; \quad U_j^{20} = \sum_{i=j-19}^j U_i, \quad (14)$$

where  $L_i$  and  $U_i$  represent the lowest and highest annual area-burned estimates of the 10 member models, defined as:

$$\begin{aligned} L_i &= \min(\hat{A}_m; m = 1, 2, \dots, M); \\ U_i &= \max(\hat{A}_m; m = 1, 2, \dots, M). \end{aligned} \quad (15)$$

## Model validation

### Empirical validation (SAGU)

We assessed four measures of model accuracy by comparing predicted (fire-scar models) versus independently

observed (fire atlas maps) area-burned estimates. At SAGU, there were enough mapped fires of varying sizes during the validation period (1937–2001) for a completely independent empirical validation. This was not the case at GILA, where fire-scar data were collected prior to most modern fires. Therefore, we used an alternative method that assumes that plots record fire scars at an empirically derived detection rate below 1.0 (see “Fire detection probability for model assessment at GILA” below).

### Annual area-burned estimation

In an ideal scenario, the relationship between mapped and predicted annual area burned would be linear with a slope of 1.0, an intercept of 0.0, an  $r^2$  of 1.0, and prediction intervals (PIs) that are parallel to and encompass an idealized 1:1 line. To test this, we fit a linear regression model between the empirical fire atlas and the ENS fire-scar model area-burned estimates and used bootstrapping (1000 iterations) to determine if the slope, intercept,  $r^2$ , and  $PI_{95\%}$  differed significantly from theoretical expectations. We plotted area-burned estimates for individual member models to show the full spread of model estimates.

Because fire atlas data were also used to empirically fit the RoM and MoR models in this study (including the two  $F$ -variants of those models), we used a leave-one-out cross-validation method to ensure independence in the validation (Meko, 1997). Fire atlas data were not used directly to estimate annual area burned by the other six member models, so they are considered independent *ipso facto*.

### Total period area-burned estimation

A common application of fire history reconstructions (including this study) is to compare total area burned between long, multi-decadal time periods. Even relatively small annual estimation errors may become important if they propagate over several decades (e.g., predicting areas for “spot fires” over many years). To test the accuracy of total area-burned estimates across the entire validation periods in each study area, we used Percent Bias (Moriassi et al., 2015):

$$\%BIAS = \frac{\sum_{i=1}^n (\hat{A}_i - A_i)}{\sum_{i=1}^n A_i} \times 100. \quad (16)$$

where  $\hat{A}_i$  is the modeled area burned and  $A_i$  is the observed area burned in year  $i$  during the validation period.  $BIAS > 0.0$  indicates that the fire-scar model overestimated total area burned for the period by a certain percentage;  $\%BIAS < 0.0$  indicates that the model underestimated total area burned for the period; and  $\%BIAS = 0.0$  indicates a perfect agreement between

modeled and observed. We tested whether BIAS differed significantly from 0 using a BCa bootstrapping procedure (Efron, 1987). We also plotted the individual %BIAS and  $CI_{95\%}$  for all member models to show the full spread and variability of model estimates in relation to the ENS model and mapped fires. We used a leave-one-out cross-validation method for the RoM and MoR models (and filtered variants) to ensure independence (Meko, 1997).

### Temporal model consistency

Model consistency over time is an important component of paleofire reconstructions because changes in cumulative rates of burning may signal either a shift in fire regimes or inconsistencies in the model (i.e., temporal changes in the underlying relationship between fire-scar synchrony and annual area burned). We used Double Mass Analysis plots (Searcy & Hardison, 1960) to determine if the relationship between predicted and observed area burned was constant over the entire validation period. We plotted the cumulative fraction of the study area burned for each year of the ENS fire-scar model against mapped fires and used bootstrapping (1000 iterations) to generate 95% CIs. A plot with constant slope and a 95% CI encompassing the 1:1 line would indicate model consistency, whereas statistically significant deviations from the 1:1 line (referred to as breakpoints) would suggest systematic changes in underlying model performance (Searcy & Hardison, 1960).

### Empirical sample size sensitivity

We evaluated the influence of sample size variation at SAGU on the two primary tree-ring fire-scar indices used in this study: fire-scar synchrony and Thiessen polygon regions (average size). These indices are the basis of all 10 member models, as  $P_i$  is a term in all six statistical models, and TP are the basis of the four spatial models. Thus, any systematic sample size bias in member models would be evident in these two raw metrics. We limited this analysis to SAGU, which had sufficient sample size and temporal depth (spanning many epochs) in the validation period to produce rigorous empirical estimates.

Using random resampling without replacement from the full empirical validation dataset, we produced 1000 combinations of fire-scar plots for five sample sizes spanning the range of epochs at SAGU:  $n = 20, 30, 40, 50$ , and  $60$  (maximum validation-period sample size was  $n = 68$ ). We used sampling without replacement rather than bootstrapping because (a) it simulates actual field sampling scenarios where there would not be duplicate plots and (b) the spatially explicit TP cannot be generated from duplicate plot selections. For each sample size, we calculated mean fire-scar synchrony and Thiessen polygon area of scarred plots, and asymmetric  $CI_{95\%}$  using a

percentile rank of the bootstrapped distribution. Relative percent bias was calculated by normalizing the reduced sample estimates to the full dataset:

$$r\%BIAS = \frac{\sum_{i=1}^n (\hat{A}_{sub} - \hat{A}_{full})}{\sum_{i=1}^n \hat{A}_{full}} \times 100. \quad (17)$$

where  $\hat{A}_{full}$  is the area-burned estimate from the full dataset ( $n = 68$ ) and  $\hat{A}_{sub}$  is the mean area-burned estimate from the resampled subset of plots for a given sample size. An  $r\%BIAS$  of 0 would indicate no relative bias; an  $r\%BIAS > 1.0$  indicates a potential overestimation bias; an  $r\%BIAS$  significantly  $< 1.0$  indicates potential underestimation bias. The width of the 95% CI error bars reflects the level of uncertainty in the estimation.

### Fire detection probability for model assessment at GILA

At GILA, the number and size of fires during the validation period that had corresponding tree-ring fire-scar data (1909–1980) were insufficient to conduct a statistically valid and independent empirical validation, as was done at SAGU. Most major fires, including 85% of the total area burned in the GILA fire atlas, occurred after fire-scar data were collected in 1980. Therefore, we extended the GILA fire-scar record from 1980 to 2016 using the same conditional probability methods described above in “Time-Varying Predictor Subset Calibration,” which assumes that plots within fire perimeters would have recorded fire at similar empirical rates observed at SAGU.

Using this extended fire-scar chronology, we applied our multi-model ensemble methodology to reconstruct annual area burned. We then compared the resulting plot-based estimates to the observed area burned from fire atlas maps, employing the first three validation metrics established for SAGU. In our statistical analyses, we combined the empirical and modeled data but used distinct symbology to differentiate between data points in our figures. A sample size sensitivity analysis was not conducted at GILA, as the full range of sample sizes is only 9–13 plots.

Our validation assessment at GILA measures how representative the fire-scar model estimates are based on actual plot locations, assuming that they would have recorded modern fires after 1980 with empirical fire-scar detection probabilities. This hybrid empirical-modeled validation approach provides a baseline accuracy of the GILA fire-scar models. While empirical results may vary slightly depending on how many of the fires were detected by fire scars in a given year, this approach

provides a reasonable assessment of model performance based on expected fire detection rates observed in similar ecosystems at SAGU.

## Comparison of modern and historical area burned metrics

We conducted two-sample *t* tests (Zar, 2010) to determine if there were significant differences in mean annual area burned between the three fire management periods at each study area. We also performed a series of two-sample randomization tests with 1000 simulations to validate that skewness of the data did not influence the *t* results (Edgington & Onghena, 2007). We used  $\alpha = 0.05$  for all analyses.

To evaluate modern-era 20-year area burned totals in a historical context, we used the maximum 20-year total as a reference to compare with the pre-exclusion range of variability. We calculated the exceedance probability (which we express as a percentage) that the 20-year maximum area burned during both the managed fire period (MAN) and the fire exclusion period (EXC) exceeded the 20-year ENS estimates from the pre-exclusion period (PRE). Specifically, we calculated  $\Pr(\max(\text{MAN}) > \min(\text{PRE}))$ , and  $\Pr(\max(\text{EXC}) > \min(\text{PRE}))$ , where  $\Pr$  is the exceedance probability. We repeated these exceedance calculations using the lower and upper combined model spread and blended model uncertainty to account for estimation uncertainty around the ENS mean. Only 20-year windows entirely within a given management period were included for the exceedance calculations (e.g., the fire exclusion period at SAGU began in 1894, so the first 20-year window for this distribution would be  $j = 1913$ ). Wilson score CIs (Wallis, 2013) were calculated for all exceedance probabilities.

Lastly, we compared the relative distributions of 20-year area burned across the three fire management periods in each study area. To quantify the similarity between distributions, we calculated Szymkiewicz–Simpson coefficients (also known as the overlap coefficient), which measure the overlap by dividing the number of elements in the intersection by the number of elements in the smaller set (Franco-Pereira et al., 2021). CIs were generated using the parametric bootstrapping procedure described in Franco-Pereira et al. (2021).

## RESULTS

### Model validation

#### Annual area-burned estimation

We found strong linear relationships between fire atlas maps and fire-scar estimates of annual area burned at

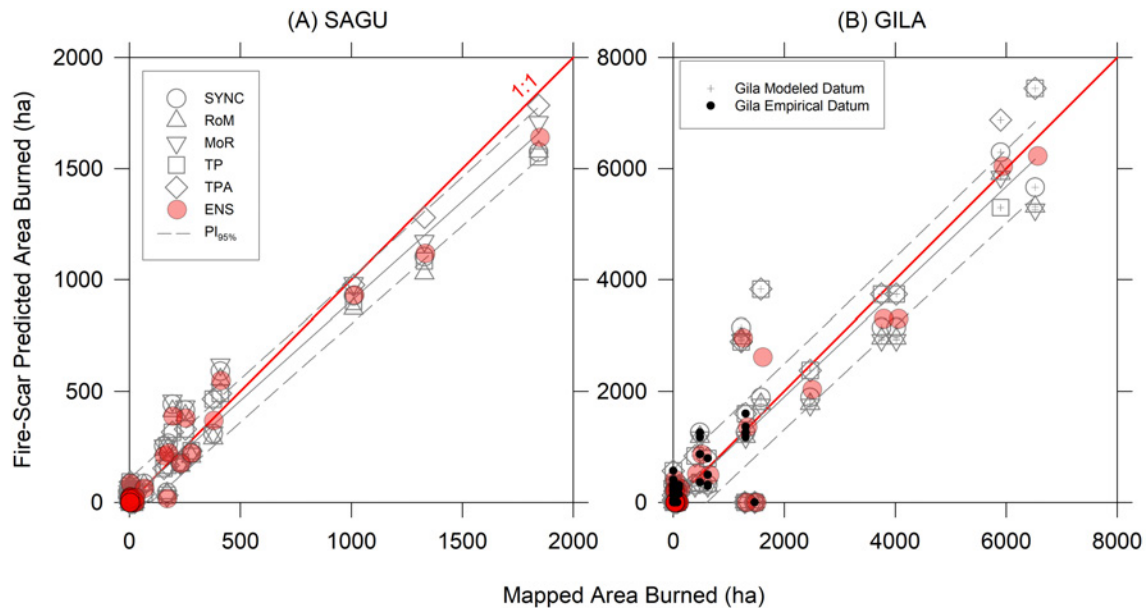
both study areas (Figure 3). At SAGU, the regression of ENS versus fire atlas had a slope of 0.89 (95% CI = 0.84–1.03) and an intercept of 10.9 (95% CI = –1.01–25.8), neither of which differed significantly from hypothetical expectations of 1.0 and 0.0, respectively. The  $PI_{95\%}$  encompassed the 1:1 line for years at SAGU except for the largest fire year in 1943, which was still within 70 ha of the upper bound. This suggests a potential slight underestimation of study-area-wide fire events >2000 ha. The regression  $r^2$  was 0.97. At GILA, the regression slope was 0.95 (95% CI = 0.84–1.06) and the intercept was 12.49 (95% CI = –23.55 to 51.94), neither of which differed significantly from 1.0 or 0.0. The  $PI_{95\%}$  encompassed the 1:1 line for the entire data range. The regression  $r^2$  was 0.93. Most individual member models also performed well, with slopes that did not differ significantly from 1.0. Exceptions included the RoM, RoMF, TP, and TPF models at SAGU, which had upper CIs slightly <1.0. Coefficients and constants for all individual member models are reported in Appendix S1: Table S3.

### Validation period total-area-burned estimation

Fire-scar estimation of the total area burned for the validation period did not differ significantly from the fire atlas data for the ENS or any member model at SAGU or GILA (Figure 4). Consistent with expectations, the ENS model was optimal at both study areas (i.e., combination of low %BIAS and low estimation variability encompassing 0.0). At SAGU, the empirical ENS model had a %BIAS of –2.9%, with individual member models ranging from –12.0% (TPF) to 6.3% (SYNC). At GILA, %BIAS for the ENS model was –1.1% and the individual model estimates ranged from –18.8% (ROMF) to 20.0% (TPA). The *F*-variant models had slightly lower and more symmetrical CIs than the unfiltered variants in all cases due to the elimination of “spot fire” years. Although uncertainty and CI bands of member models were generally wider at GILA than SAGU due to considerably lower sample densities, the differences were less pronounced for the ENS model solutions.

### Temporal model consistency

The relationship between predicted and observed area burned was consistent through time for the ENS model at both study sites (Figure 5). Trajectory breakpoints associated with differential accuracy of major fire events were minor and CIs did not differ significantly from the 1:1 reference line at any point during the validation period in either study area. Furthermore, the trajectory of breakpoints associated with differential accuracy of major



**FIGURE 3** Relationship between estimated area burned by fire-scar models and independently mapped fire perimeters (fire atlas data) for individual fire years in (A) the Saguaro Wilderness (SAGU, 1937–2001) in southern Arizona and (B) the Gila Wilderness (GILA, 1909–2016) in southern New Mexico. The red diagonal line represents an ideal 1:1 relationship. Regression and 95% prediction intervals were calculated for the ensemble fire-scar model (ENS). Filtered and unfiltered estimates differ only for the smallest fires (<200 ha) that were detected by a single plot. GILA empirical datum derived from overlap between fire-scar data and fire atlas (<1980) and GILA model datum from modeled scarring rates (>1980) when fire-scar data were absent. ENS, ensemble; MoR, mean of ratios; RoM, ratio of means; SYNC, synchrony; TP, Thiessen polygon; TPA, Thiessen polygon adjusted.

fire events was distributed evenly above and below the reference line, an indication of randomly distributed error.

### Sample size sensitivity

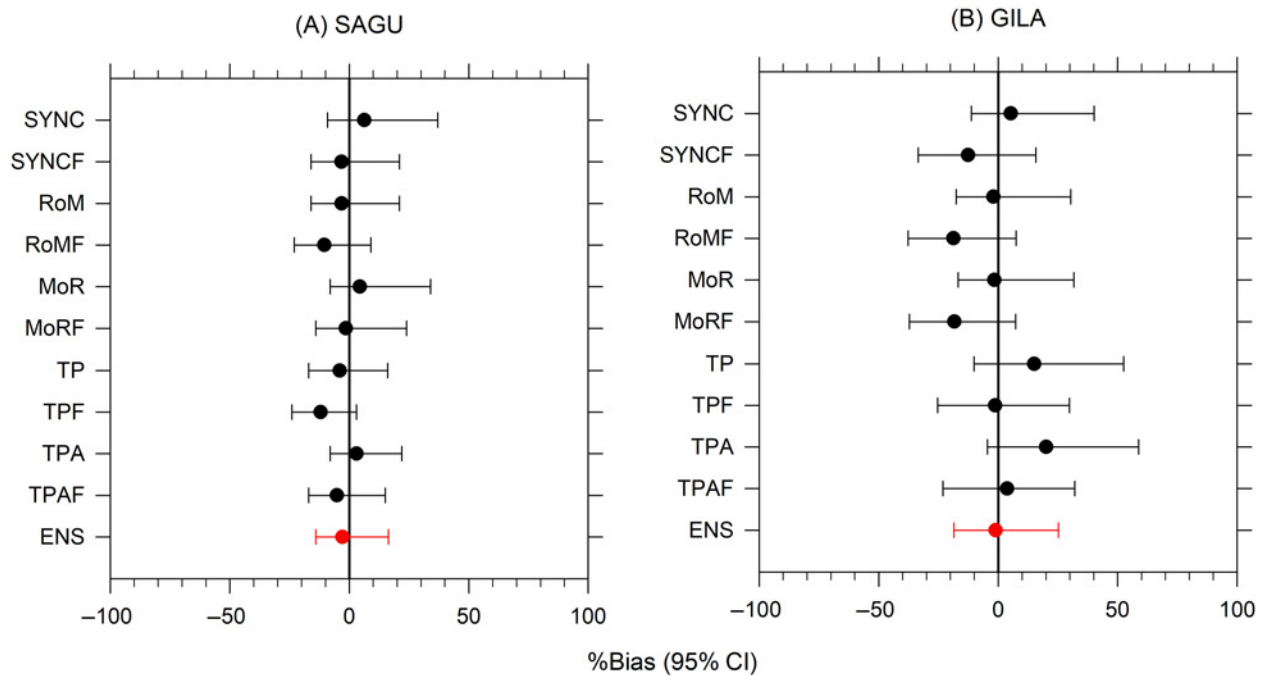
Sample size reductions did not result in significant relative estimation biases for empirical fire-scar synchrony or Thiessen Polygon areas at SAGU (Figure 6), indicating that area-burned reconstructions were largely robust across the full range of chronology sample sizes. As expected, however, estimation variability increased considerably with decreasing sample sizes (that would be typical of earlier epochs). There was a slight trend toward relative underestimation of synchrony and annual area burned for the filtered models with decreasing sample size, even though  $r\%BIAS$  was not significantly less than 0.0 at any sample size. This is consistent with expectations because the likelihood of fires scarring multiple plots should decrease with increasingly lower sample densities.

### Area-burned time series

The fire-scar time series at SAGU covered the period from 1700 to 2001 (Figure 7) and included 40 distinct epochs.

Sample sizes ranged from a minimum of 25 recording plots in epoch 40 (1700–1702) to a maximum of 73 plots in epoch 9 (1908–1920), which then declined to 62 plots by epoch 1 (1985–2001) (Appendix S1: Table S1). Estimated area burned at SAGU across the entire time series from 1700 to 2016—including both fire-scar and fire atlas estimates—varied widely. Annual area burned ranged from 0 to 2632 ha (average = 216 ha; 8% of the study area), while 20-year area burned ranged from 411 to 7661 ha (average = 4650 ha) (Table 2, Figure 7). Large fire years at SAGU were relatively frequent: There were 29 years in which at least 25% of the study area burned and 19 years in which at least 50% burned, occurring on average every 10.9 and 16.7 years, respectively.

At GILA, where sample size was considerably smaller, the fire-scar time series covered the period from 1733 to 1980 (Figure 8). Sample depth declined sharply before 1733, and again after 1980 (the year most samples were collected) (Appendix S1: Table S2). The GILA series included six epochs, with a minimum of nine recording plots in epoch 6 (1733–1779) and a maximum of 13 in epoch 2 (1847–1922) (Appendix S1: Table S2). Estimated area burned across the entire time series from 1733 to 2016 varied widely. Annual area burned ranged from 0 to 8179 ha (average = 714 ha; 9% of the study area), and 20-year area burned ranged from 49 to 29,436 ha



**FIGURE 4** Percent bias of the total area-burned estimates for the ensemble fire-scar model (ENS, red circle) and 10 individual member models (black circles) at Saguaro (SAGU; 1937–2001) in southern Arizona and the Gila Wilderness (GILA; 1909–2016) in southern New Mexico. A bias of 0.0 indicates perfect agreement between fire-scar models and mapped area burned; values  $>0.0$  indicate that a fire-scar model overpredicted area-burned compared to fire atlas maps; values  $<0.0$  indicate that a fire-scar model underpredicted area-burned compared to fire atlas maps. ENS, ensemble; MoR, mean of ratios; RoM, ratio of means; SYNC, synchrony; TP, Thiessen polygon; TPA, Thiessen polygon adjusted. “F” models are filtered to exclude fires that burned only one plot.

(average = 14,742 ha) (Table 2). Large fire years were also frequent at GILA: There were 42 years in which at least 25% of the study area burned and 19 years in which at least 50% burned, occurring on average every 6.8 and 15.0 years, respectively.

Both measures of fire-scar estimation uncertainty for the moving window time series were generally wider in the earlier years of the chronologies at both sites (Figures 7 and 8). This reflects typical patterns of sample attrition consistent with the loss of older wood and scars over time (Appendix S1: Tables S1 and S2), which were more pronounced at GILA due to lower overall sample sizes. At SAGU, the combined model spread bands for 20-year periods during the pre-exclusion period (1700 to 2001) ranged from a high of 2599 ha in 1729 to a low of 221 ha in 1906. At GILA, they ranged from a high of 8614 ha in 1767 to a low of 1610 ha in 1904.

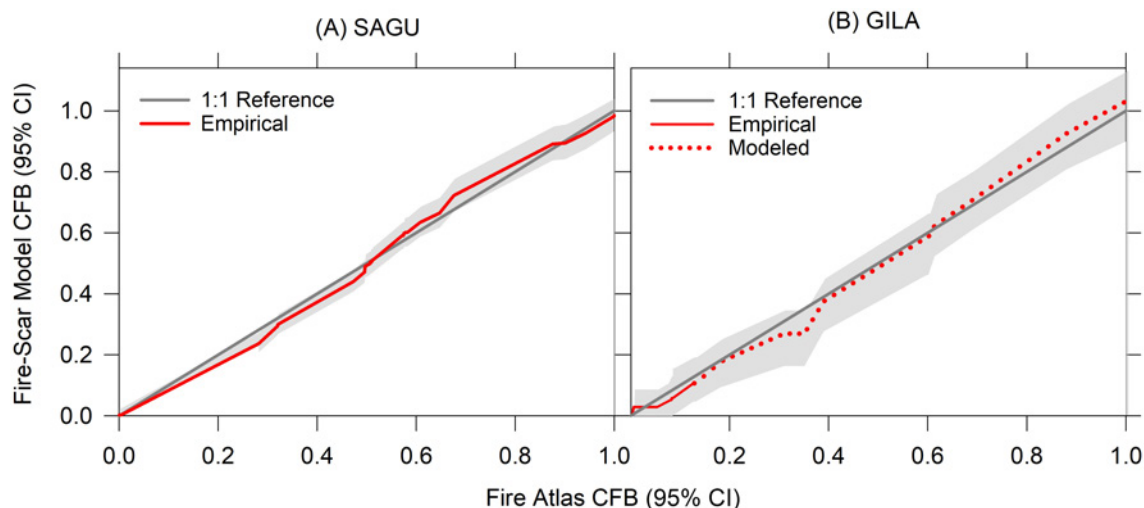
### Comparison of modern and historical area burned metrics

Mean annual area burned at SAGU during the pre-exclusion period was  $301 \pm 605$  ha (mean  $\pm$  SD), which was significantly higher than the fire-exclusion

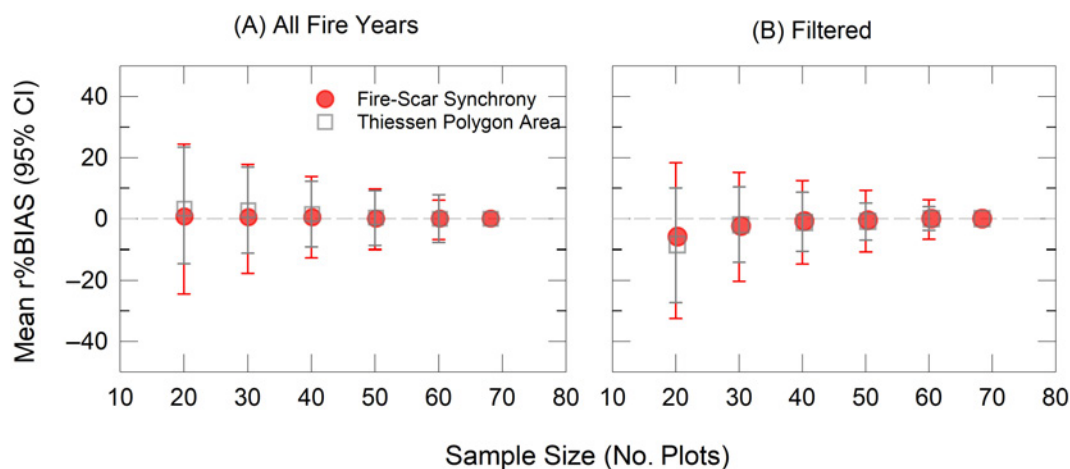
period ( $64 \pm 209$  ha) and the managed fire period ( $112 \pm 233$  ha; Tables 2 and 3). Although the mean annual area burned increased in the managed fire period relative to the fire exclusion period, the difference was not statistically significant (Table 3), largely due to two of the largest 20th-century fires occurring in 1943 and 1954, despite suppression efforts (Figure 7A).

At GILA, the mean annual area burned of  $995 \pm 1894$  ha during the pre-exclusion period was significantly higher than the fire exclusion period ( $42 \pm 197$  ha; Tables 2 and 3). This was followed by a significant increase in burning during the managed fire period ( $684 \pm 1574$  ha). Most notably, we found no significant difference in mean annual area burned between the pre-exclusion period and the managed fire period at GILA (Table 3). Interannual variability in burning was high during the managed fire period, starting with minimal burning early in the period followed by a large increase beginning in the early 1990s, which continued through 2016 (Figure 8).

At SAGU, modern-era 20-year area-burned totals were generally lower than historical levels, although there was some overlap (Table 2, Figures 7B and 9). The modern-era maximum 20-year area burned was 3753 ha during the managed fire period in 2004, compared to an average of



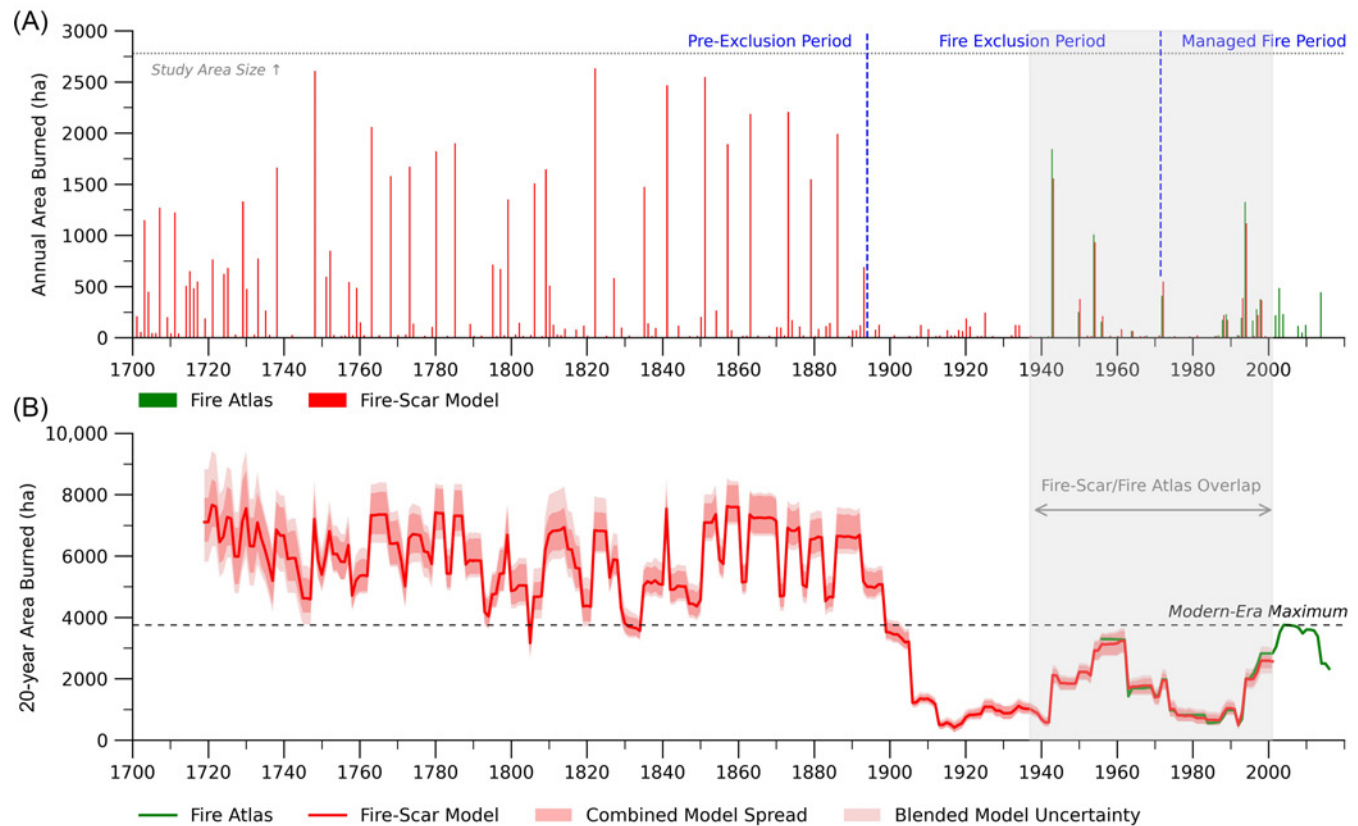
**FIGURE 5** Double-mass analysis curves for cumulative fraction burned (CFB) for the ensemble fire-scar model (red lines) compared to an idealized 1:1 reference (dark gray); SAGU, Saguaro Wilderness, southern Arizona; GILA, Gila Wilderness, southern New Mexico. At GILA, the double mass line is shown as solid red for the period of empirical overlap between the fire-scar and fire atlas record ( $\leq 1980$ ), and as dotted for the period of modeled scarring when the two records do not overlap ( $> 1980$ ; see [Materials and methods](#)). Gray shading indicates 95% confidence bands generated with a bootstrapping procedure (1000 iterations).



**FIGURE 6** Mean relative percent bias (r%BIAS) showing the influence of empirical sample size variation on estimates of fire-scar synchrony (SYNC model) and Thiessen polygons (TP) in the Saguaro Wilderness in southern Arizona for (A) all fire years, and (B) filtered fire years that burned two or more plots. Maximum sample size was  $n = 68$  plots for the calibration period epochs (1937–2001). A value of 0.0 (dashed reference line) indicates that subsample estimates did not vary from the full sample estimate; values  $< 0.0$  indicate incomplete detection of fire area due to smaller simulated sample sizes. Mean bias was not significantly different from 0% bias for any sample size.

5984 ha during the pre-exclusion period. This modern-era maximum exceeded 2.9% of the pre-exclusion 20-year windows based on the ENS model estimate, 4.0% and 0.6% using the low and high combined model spread estimates, respectively, and 6.3% and 0% using the low and high blended model uncertainty estimates, respectively (Table 4, Figure 7B). Maximum 20-year area burned during the fire exclusion period at SAGU was 3297 ha in 1956, exceeding 0.6% of the pre-exclusion 20-year windows based on the ENS model estimate, 0.6% and 0% of using

the low and high combined model spread estimates, respectively, and 1.1% and 0% using the low and high blended model uncertainty estimates, respectively (Table 4). The Szymkiewicz–Simpson coefficient indicated minimal distribution overlap for 20-year area burned totals in the pre-exclusion period vs. the fire exclusion period (0.03, 95% CI = 0–0.05) and vs. the managed fire period (0.03, 95% CI = 0.01–0.06), but showed moderate overlap between the fire exclusion and managed fire periods (0.43, 95% CI = 0.23–0.59) (Figure 9).



**FIGURE 7** Time series of area burned for the Saguaro Wilderness (SAGU) in southern Arizona based on ensemble fire-scar reconstruction (1700–2001) and fire atlas data (1937–2016). (A) Fire atlas map and fire-scar model estimates of annual area burned (ensemble mean). (B) Total area burned for 20-year moving windows. The value for each year represents the sum of area burned during the prior 20-year period, inclusive. The dark red line is the 10-model prediction mean; the red shading shows combined model spread. The gray shading is blended model uncertainty, which represents the time series that would occur if the single lowest and highest member model estimate for *every* given year were used to calculate the running sum. The horizontal dashed line is the maximum modern-era (1894–2016) 20-year area burned.

At GILA, recent 20-year area burned values were similar to historical levels (Figures 8 and 9). The maximum 20-year area burned during the managed fire period was 21,932 ha in 2012, compared to an average of 19,477 ha during the pre-exclusion period (Table 2, Figure 8B). This managed fire period maximum exceeded 65.4% of the pre-exclusion 20-year windows based on the ENS model estimates, respectively, and 86.3% and 49.7% of pre-exclusion windows using the low and high combined model spread estimates, respectively, and 92.2% and 34.6% of pre-exclusion windows using the low and high blended model uncertainty estimates, respectively (Table 4, Figure 8B). The maximum 20-year area burned during the fire exclusion period at GILA was 2692 ha in 1956, which exceeded 0% of the pre-exclusion and managed fire 20-year windows based on the ENS estimate and all measures of uncertainty (Table 4). The Szymkiewicz–Simpson coefficient was 0 for all comparisons involving the fire exclusion period, but moderate overlap was found between the pre-exclusion

and managed fire period distributions (0.39, 95% CI = 0.24–0.54) (Figure 9).

## DISCUSSION

Annual area burned is a key consideration in the ecology and management of ecosystems. While tree rings have been used to quantify some properties of historical fire regimes (fire frequency, severity, seasonality, and fire climatology) in many forest ecosystems (Falk et al., 2011; Margolis et al., 2022), estimates of pre-fire exclusion area burned are rare and often lack metrics that quantify accuracy and prediction uncertainty (see Farris, 2009; Safford et al., 2022; Taylor & Skinner, 1998). In this study, we present the first comprehensive tree-ring fire-scar reconstruction combining a multi-model ensemble framework, time-varying predictor subset calibration, and multiple measures of uncertainty and validation to reconstruct landscape-scale annual area burned in

**TABLE 2** Summary statistics<sup>a</sup> for annual and 20-year running totals of area burned for the Saguaro Wilderness (SAGU) in southern Arizona and the Gila Wilderness (GILA) in southern New Mexico, for the full record and three fire management periods.

| Time period <sup>b</sup> | Annual area burned |      |     |     |      | 20-year area burned |      |        |      |        |
|--------------------------|--------------------|------|-----|-----|------|---------------------|------|--------|------|--------|
|                          | AVG                | SD   | IQR | MIN | MAX  | AVG                 | SD   | IQR    | MIN  | MAX    |
| SAGU                     |                    |      |     |     |      |                     |      |        |      |        |
| All years                | 216                | 504  | 121 | 0   | 2632 | 4650                | 2195 | 3844   | 411  | 7661   |
| Pre-exclusion            | 301                | 605  | 167 | 0   | 2632 | 5984                | 1076 | 1867   | 3165 | 7661   |
| Fire exclusion           | 64                 | 209  | 24  | 0   | 1556 | 1518                | 859  | 1162   | 411  | 3236   |
| Managed fire             | 112                | 233  | 168 | 0   | 1328 | 2778                | 955  | 1235   | 493  | 3753   |
| Modern era               | 81                 | 219  | 64  | 0   | 1556 | 1904                | 1059 | 1884   | 411  | 3753   |
| GILA                     |                    |      |     |     |      |                     |      |        |      |        |
| All years                | 714                | 1642 | 251 | 0   | 8179 | 14,742              | 8613 | 13,504 | 49   | 29,436 |
| Pre-exclusion            | 995                | 1894 | 834 | 0   | 8179 | 19,477              | 4745 | 5863   | 7794 | 29,436 |
| Fire exclusion           | 42                 | 197  | 0   | 0   | 1467 | 1191                | 1053 | 2292   | 49   | 2692   |
| Managed fire             | 684                | 1574 | 70  | 0   | 6522 | 13,296              | 4497 | 7570   | 7053 | 21,932 |
| Modern era               | 283                | 1018 | 4   | 0   | 6522 | 4953                | 6219 | 8101   | 49   | 21,932 |

<sup>a</sup>AVG, average; IQR, interquartile range.

<sup>b</sup>SAGU: All years 1700–2016; pre-exclusion 1700–1893; fire exclusion 1894–1971; managed fire 1972–2016; modern era 1894–2016. GILA: all years 1733–2016; pre-exclusion 1733–1904; fire exclusion 1905–1974; managed fire 1975–2016; modern era 1905–2016.

absolute terms (hectares). Our findings show that multiple models can accurately estimate area burned when compared to independently mapped perimeters of modern fires. Applying these methods to two wilderness landscapes in the southwestern United States with a long history of active fire restoration, we found that area burned is within historical levels at GILA and is approaching historical levels at SAGU.

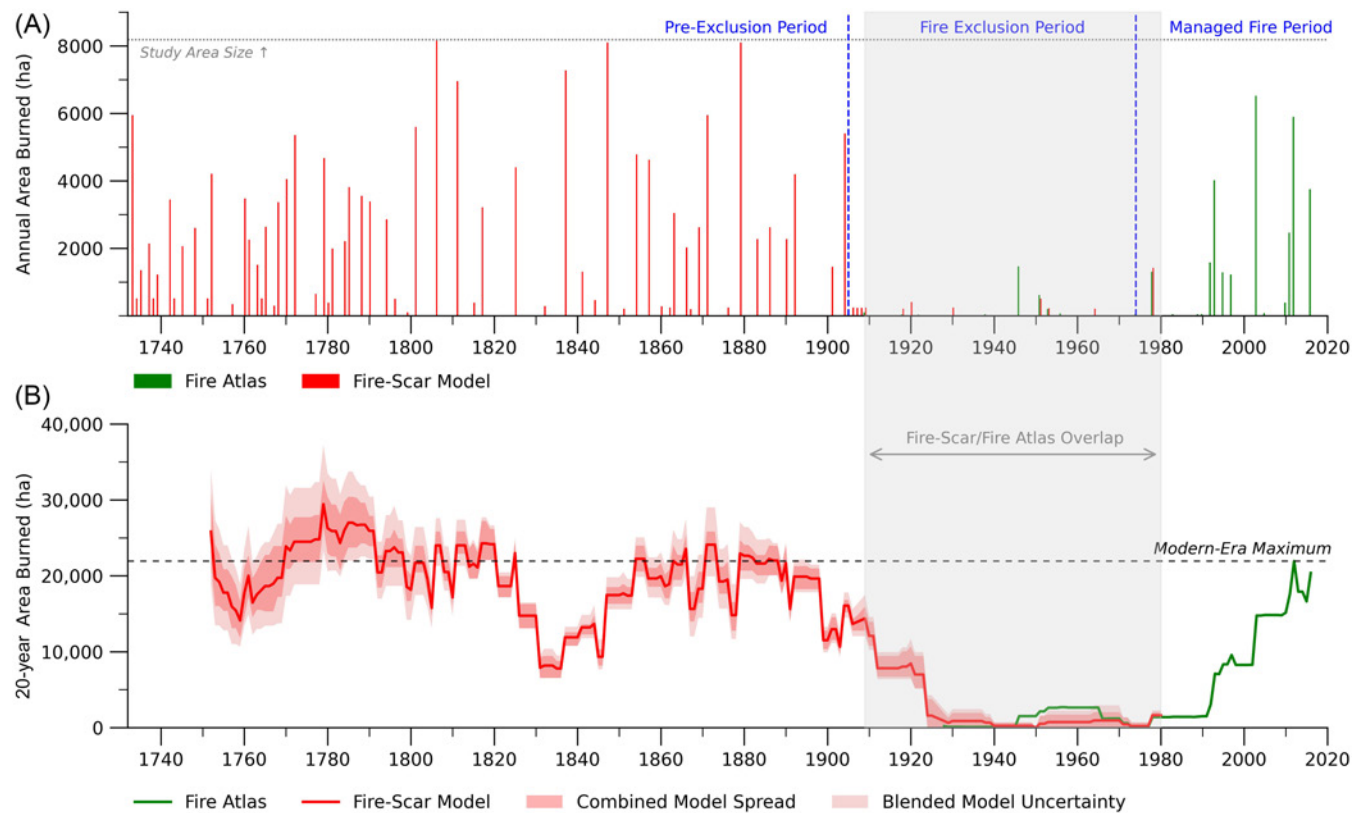
## Area-burned reconstruction models and methods

The multi-model ensemble produced accurate tree-ring reconstructions of both annual and total area burned compared to mapped fire perimeters in dry conifer forest landscapes that have experienced some of the highest modern-era rates of burning in the western United States. The ENS model estimates were stable over time and robust across various sample sizes, providing confidence that variation in area burned in each time series reflects actual changes in the fire regime rather than model error. Although each individual member model produced statistically valid area-burned estimates compared to modern fire atlas data, the multi-model ensemble (ENS) solution achieved the best balance of high accuracy and low variance across both study areas (Figure 4). This aligns with expectations and findings from studies of ensemble modeling in climatology and other fields (Hagedorn et al., 2005; Parsons et al., 2021). Our study supports using ensemble

modeling for tree-ring fire history studies to leverage the diverse assumptions and strengths of different fire-scar models, thereby improving accuracy and reducing uncertainty (Weigel et al., 2008).

Estimating area burned using tree-ring fire-scar records presents several significant challenges: (1) sample size and spatial distribution vary over time due to attrition (Farris et al., 2013), (2) fire-scar formation and retention fluctuate both within and between individual fires due to differences in fire intensity, tree condition and wound susceptibility (Stephens et al., 2010), and (3) the fire frequency–area distribution changes with climate and land-use patterns over time, affecting the proportion of fire-scar sampling sites that experience fire (Swetnam et al., 2016). Given these challenges, it is unlikely that any single model will be universally superior across all situations over space and time, so it is useful to integrate results from a range of models.

Our findings support this assertion in two key ways. First, each core model produced the most accurate annual area-burned estimate at least once during the validation period, highlighting differential performance in a dynamic environment (see Figure 3). Second, the top-performing member models varied between study areas. For example, the TPA model was the third most accurate for total burned area at SAGU but the least accurate at GILA. In contrast, the TPF model ranked second in accuracy at GILA (behind ENS) but was the least accurate at SAGU (Figure 4). These differences are not a limitation in this study; rather, our unweighted



**FIGURE 8** Time series of area burned for the Gila Wilderness (GILA) in southern New Mexico based on the ensemble fire-scar reconstruction (1733–1980) and fire atlas data (1909–2016). (A) Fire atlas and fire-scar model estimates of annual area burned (ensemble mean). (B) Total area burned for 20-year moving windows. The value for each year represents the sum of area burned during the prior 20-year period, inclusive. The dark red line is the 10-model prediction mean; the red shading shows combined model spread. The gray shading is blended model uncertainty, which represents the time series that would occur if the single lowest and highest member model estimate for *every* given year were used to calculate the running sum. The horizontal dashed line is the maximum modern-era (1905–2016) 20-year area burned.

**TABLE 3** Two-sample *t* test and randomization (Rand) test results for comparison of mean annual area burned between fire management periods for the Saguaro Wilderness (SAGU) in southern Arizona and the Gila Wilderness (GILA) in southern New Mexico.

| Management periods <sup>b</sup> | SAGU <sup>a</sup> |          |                            | GILA <sup>a</sup> |          |                            |
|---------------------------------|-------------------|----------|----------------------------|-------------------|----------|----------------------------|
|                                 | <i>t</i> (df)     | <i>p</i> | Rand <i>p</i> <sup>c</sup> | <i>t</i> (df)     | <i>p</i> | Rand <i>p</i> <sup>c</sup> |
| Pre-exclusion versus exclusion  | 3.376 (270)       | 0.001    | <0.001                     | 4.194 (240)       | <0.001   | <0.001                     |
| Managed versus exclusion        | 1.166 (121)       | 0.246    | 0.262                      | 3.375 (110)       | 0.001    | <0.001                     |
| Pre-exclusion versus managed    | 2.061 (237)       | 0.039    | 0.035                      | 0.985 (212)       | 0.326    | 0.280                      |

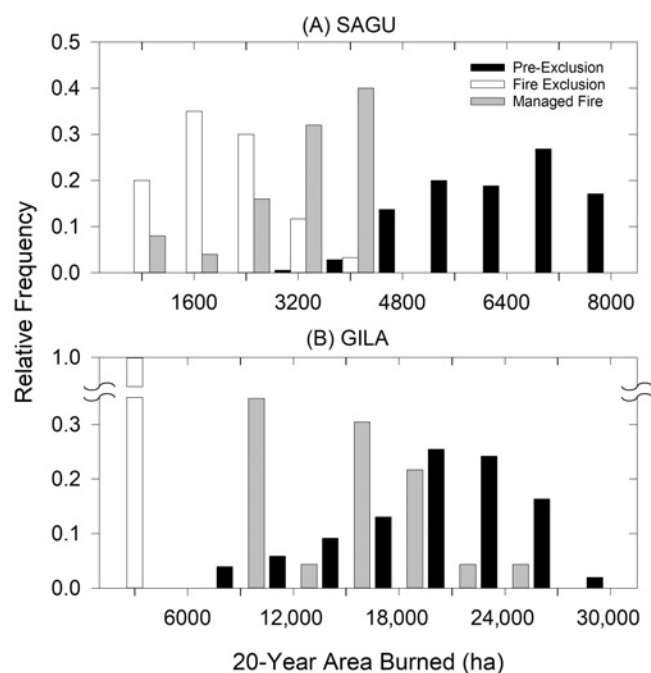
<sup>a</sup>SAGU: the fire-scar ensemble model was used for pre-exclusion and fire exclusion periods and fire atlas data were used for the managed fire period. GILA: the fire-scar model was used for the pre-exclusion period and the first 4 years of the fire exclusion period and fire atlas data were used for the rest of the fire exclusion period and the managed fire period.

<sup>b</sup>SAGU: pre-exclusion 1700–1893; fire exclusion 1894–1971; managed fire 1972–2016. GILA: pre-exclusion 1733–1904; fire exclusion 1905–1974; managed fire 1975–2016.

<sup>c</sup>The average *p* of 10 randomization tests using 1000 simulations for each test, for a total of 10,000 simulations.

ensemble model provided a flexible framework that minimized bias by avoiding overreliance on any single member model. It also leveraged the spatiotemporal variability in member model performance to capture the range of uncertainty. For example, while the TPA

overestimated some large fire years at GILA (though results were not statistically significant), it compensated for underestimations by other models, ultimately leading to an ENS estimate closer to the reference (Figure 3B).



**FIGURE 9** Relative frequency distributions of 20-year total area burned moving windows in the (A) Saguaro Wilderness (SAGU) in southern Arizona and (B) Gila Wilderness (GILA) in southern New Mexico during three fire management periods: pre-exclusion, fire exclusion, and managed fire. SAGU: pre-exclusion 1700–1893; fire exclusion 1894–1971; managed fire 1972–2016. GILA: pre-exclusion 1733–1904; fire exclusion 1905–1974; managed fire 1975–2016.

Ensemble averages are often weighted based on the observed accuracy of member models (Knutti, 2010). However, unweighted ensembles can perform competitively across a range of conditions due to challenges in optimal weighting and error propagation (Bukovsky et al., 2019). For paleo fire-scar data, we argue that unweighted ensembles are suitable because relative member model performance varies over time and between study areas depending on site-specific differences in sample attrition and quality, and fire regimes (Farris, 2009). Even within a single study area, the most accurate member model in 1780 may not be the most accurate in 1980. Several other factors favor unweighted models: fire scars are imperfect recorders over time (Farris et al., 2010), so anomalies from short validation periods may propagate to older (or future) periods; fire atlas records are more uncertain compared to instrumental records like temperature and rainfall (Rollins et al., 2001); opportunities for modern validation—and thus, weighting—are rare; individual validation metrics may fail to capture all aspects of model performance; and fire-scar data are used primarily to reconstruct past fire occurrence rather than to predict future fires, thereby negating a primary advantage of weighting (but see Marlon et al., 2012, Roos et al., 2021).

Assessing empirical accuracy to weight models would require longer fire atlas records and a consensus on best metric(s) of accuracy.

Our goal in this study was not to identify the single best member model but rather to draw from all models to generate robust comparisons of management periods. However, given the wide range of assumptions and plot weightings used by each model, it is notable that no member models differed significantly from the fire atlas data (even the SYNC model). In an ideal scenario where fires are widespread relative to the study area and samples are densely and evenly distributed, we would expect all models to perform similarly well because fires would be detected in proportion to their area. As real-world conditions deviate from these assumptions, the performance of individual member models will vary depending on how well their underlying assumptions match the new conditions. Significant clustering of fires and/or plots should presumably favor spatially explicit and/or calibrated statistical models that adjust the plot weightings based on ancillary information. While calibrated and uncalibrated model performance did not differ significantly for the full modern-era validation period (but variance was lower for calibrated estimates), differences between ratio estimators were more pronounced during earlier periods in each time series when both sample size and annual area burned were more variable. This is consistent with findings from paleoclimate reconstructions that found ensemble models reduced error most in areas with the fewest observations (Parsons et al., 2021).

The strong performance of the SYNC model across both the dense, systematic fire-scar network at SAGU and the sparse, clustered network at GILA suggests that this model may be more robust than previously recognized. This is significant because its simplicity and lack of calibration make it potentially widely applicable. The robustness of SYNC is further supported by sample size simulations, which show stable fire-scar synchrony estimates even with smaller sample sizes (Figure 6). Despite these observations, SYNC estimates can exhibit relatively high and skewed variance (a common issue with any ratio-based estimate). It is therefore important to carefully consider both the strengths and potential biases of the SYNC model in relation to the sampling network and modeling objective. In fact, we recommend this consideration be applied to any fire-scar model or research objective, including the ensemble solution. Additional case studies evaluating model performance under a wider range of conditions would be valuable, particularly given the potential for broad application of some models.

We found no evidence that fire-scar models in this study systematically over- or underpredict area burned. Estimates from member models were distributed

**TABLE 4** Exceedance probabilities showing the percentage of times the maximum 20-year area-burned window during the fire exclusion period (max(EXC)) and managed fire period (max(MAN)) exceeded a 20-year window in the pre-exclusion period distribution (PRE) in the Saguaro Wilderness (SAGU) in southern Arizona and the Gila Wilderness (GILA) in southern New Mexico.

| Pre-exclusion estimate | SAGU <sup>a</sup>          |                         | GILA <sup>b</sup>       |                         |
|------------------------|----------------------------|-------------------------|-------------------------|-------------------------|
|                        | Pr(max(MAN) > min(PRE))    | Pr(max(EXC) > min(PRE)) | Pr(max(MAN) > min(PRE)) | Pr(max(EXC) > min(PRE)) |
| ENS estimate           | 2.9 (1.2–6.5) <sup>c</sup> | 0.6 (0.1–3.2)           | 65.4 (57.5–72.4)        | 0                       |
| CMS <sup>d</sup> low   | 4.0 (2.0–8.0)              | 0.6 (0.1–3.2)           | 86.3 (79.9–90.8)        | 0                       |
| CMS <sup>d</sup> upper | 0.6 (0.1–3.2)              | 0                       | 49.7 (41.9–57.5)        | 0                       |
| BMU <sup>e</sup> low   | 6.3 (3.6–10.9)             | 1.1 (0.3–4.1)           | 92.2 (86.8–99.9)        | 0                       |
| BMU <sup>e</sup> upper | 0                          | 0                       | 34.6 (27.6–42.5)        | 0                       |

Note: Pre-exclusion values are based on the fire-scar ensemble (ENS) model and include the associated measures of model uncertainty.

<sup>a</sup>MAN, managed fire period (1972–2016); PRE, pre-exclusion period (1700–1893); EXC, fire exclusion period (1894–1971).

<sup>b</sup>MAN, managed fire period (1975–2016); PRE, pre-exclusion period (1733–1904); EXC, fire exclusion period (1905–1974).

<sup>c</sup>Wilson score intervals (95%) with continuity adjustment in parentheses.

<sup>d</sup>CMS, combined model spread.

<sup>e</sup>BMU, blended model uncertainty.

relatively evenly in both directions for annual burned (measured by %BIAS and variance) and total area burned (indicated by slope and regression residuals). This is contrary to concerns that fire scars might systematically overestimate landscape-scale area burned by disproportionately emphasizing small fire years or favoring areas that burn more frequently than the landscape as a whole (Minnich et al., 2000; Baker & Ehle, 2001). The fire-scar network at GILA, consisting of 13 opportunistically sampled tree clusters with extensive unsampled areas, represents a prototypical scenario where overestimation bias would be expected. We instead found that many member models slightly underestimated some of the largest fire years. For example, two major fires that covered over 34% of the study area were undetected entirely because they occurred between plots, and some of the most frequently burned areas were located between plots. Additionally, even the largest modern-era fire year (~80% of the study area) overlapped disproportionately fewer plots than expected (62%). Notably, the slight underestimation of the largest fire years was not restricted to GILA; the ENS model slightly underestimated seven of the eight largest fires that burned >25% of both study areas, despite SAGU having a denser, systematic sampling network. Thus, in many cases, our fire-scar models provided slightly conservative estimates of area burned during historically large fire years.

This study is the first to apply a time-varying predictor subset approach for area burned reconstructions from fire-scar data to address the sample attrition and potential recency bias highlighted by Meko (1997). This method is commonly used in dendrochronology reconstructions of climate, streamflow, or drought indices

where lengthy modern instrumental data are readily available and consistent through time (Meko, 1997; Cook et al., 2004). However, using fire atlas maps as “instrumental data” for calibration presents numerous challenges in many study areas, including: (1) modern fires rarely overlap temporally with older fire-scar collections (e.g., GILA) or are entirely absent due to fire exclusion; (2) even when they do, there are often too few modern fires for a statistically robust calibration and validation (especially overlapping fires) (McClure et al., 2024), (3) increasingly severe modern fires may kill trees that historically recorded and retained fire scars (McClure et al., 2024), and (4) completeness and precision of modern fire maps vary over time and space, especially for older fires that pre-date satellite-era data (Rollins et al., 2001). These challenges are further compounded in some wilderness areas where restrictions limit fire-scar sampling to dead remnant wood.

SAGU provided a rare and ideal case study setting for rigorous empirical calibration and validation of fire-scar models. At GILA, where high modern fire occurrence is exceptional for dry conifer forests in the western United States, minimal overlap with fire-scar records required assumptions about scarring probabilities for model calibration and validation (Meko, 1997). The methods we presented, including modeling modern fire-scar occurrence with known plot-level scarring probabilities, can be applied broadly to areas lacking overlap of empirical fire scars and fire atlas data. When local scarring probabilities are unavailable, a sensitivity analysis using a range of probabilities (e.g., 0.7–1.0) within our randomization framework can assess how well the fire-scar network can potentially reconstruct fire

perimeters. Even if calibrated models are not used, a time-varying predictor approach would be useful to (1) identify periods in the time series when fire-scar estimates might be most sensitive and (2) guide model interpretation and selection. Our methodology produced plot-level scarring outcomes having a specified target probability, while also incorporating random error to account for uncertainties in scar formation and retention. Further research on scarring probabilities across landscapes will improve fire-scar reconstructions and reduce uncertainties.

A significant advancement of our research is the development of multiple measures of uncertainty for our fire-scar area-burned time series. An advantage of ensemble modeling is that it provides a robust distribution of estimates to quantify estimation uncertainty. While combined spread is a common measure of uncertainty in ensemble modeling, our novel approach to create a blended uncertainty time series metric provides an even broader and more dynamic measure of interannual estimation uncertainty. Temporal patterns of model agreement not only inform confidence in predictions but also may highlight key aspects of the fire regime and sampling regimes that warrant further investigation. For example, higher uncertainty in the earliest periods of each time series was expected due to declining sample sizes. However, the pattern of more frequent but smaller fires in the early 1700s differs qualitatively from the slightly less frequent but more widespread fires of the 1800s. This difference could result in greater variation among member models even if sample sizes were consistently high. Additionally, the symmetry of the distribution of member models around the ENS mean provides further insights into modeling behavior, such as model consistency and the identification of potential outliers.

While our study primarily focused on the uncertainty of ENS model estimates using the distribution of member model predictions, there are additional opportunities to calculate variance for individual member models depending on research objectives. For example, methods are available for deriving CIs for RoM and MoR estimators using bootstrapping or parametric techniques (Gerow et al., 2021; Schaeffer et al., 2006). Although CIs for SYNC model estimates have not yet been reported in fire-scar studies, binomial distribution methods for proportions could potentially address this shortcoming (Brown et al., 2002). Standard methods for calculating and mapping variance for binary Thiessen polygon interpolations are lacking (unlike those available for Kriging), but jackknife and resampling strategies—based on sample size or scarring probabilities—offer promising avenues for future area-burned research.

Filtering in fire-scar reconstructions is typically used to calculate mean fire intervals for fire years above a minimum relative “size” (e.g., years when >10% or >25% of samples burned). For our purpose of estimating absolute area burned, the primary function of filtering was to reduce cumulative Type I error over long periods. However, filtering may also increase Type II error if sample size is low because it becomes more likely that larger fires will be recorded by only one plot (see Figure 6). While it is impossible to know whether single-plot fires in any given year were truly small or burned larger areas undetected by neighboring plots, our modeling approach hedges for both outcomes. First, both filtered and unfiltered scenarios are incorporated directly into the combined and blended model spread metrics. Second, the ENS prediction minimizes long-term error by averaging all estimates. This approach acknowledges that while individual years may have lower or higher area burned, averaging across years brings us closer to the true value in the long term.

Reconstructing area burned from tree-ring fire scars represents a key frontier in fire history research (Harley et al., 2018; Margolis et al., 2022). While some models and methods in this study are not feasible in regions lacking modern fires for calibration and validation, several aspects remain applicable. For example, our results show that the uncalibrated SYNC model alone can offer a basic, yet useful estimate of area burned. A more robust approach could involve supplementing the SYNC model estimate with a basic spatial interpolation method, such as the TP model used in this study. Neither model requires modern fire maps for calibration, and together they should offer a more robust estimate in landscapes where clustering of fires and/or plots could distort the output of a single model. Including a spatially explicit model would also shed light on the spatial processes shaping fire-prone landscapes. Adding filtered variants could further strengthen inferences by creating a simple four-model average, with the model spread providing a range of uncertainty likely to bracket the actual values. One challenge of retroactively applying these models to existing collections is the need to clearly define an area of inference to minimize boundary error.

## A multi-century perspective of changing fire regimes

After average annual area burned at GILA fell to just 64 ha/year in the fire exclusion period (a 94% decline from 995 ha/year in the pre-exclusion period), active fire management and restoration strategies in recent decades have helped restore area burned to within the historical

range of variability (Table 2, Figure 8). All fires during the managed fire period at GILA were lightning-ignited, indicating that human ignitions are not required to reach high pre-exclusion levels of area burned. Although the peak in modern area burned was heavily influenced by the 2003 Dry Lake Complex (37,600 ha) and 2012 Whitewater Baldy megafire (120,534 ha), which affected 80% and 73% of the study area, respectively, there were multiple pre-exclusion years where all plots burned (Figure 8), indicating that fires of this size at GILA were not unusual historically. Area burned increased substantially beginning in the 1990s following a series of successfully managed lightning fires, with further acceleration in the 2000s due to more managed fires and a series of large wildfires (Figure 8) (Hunter et al., 2014). This aligns with a broader increase in fire activity across the southwestern United States (Singleton et al., 2019), linked to higher temperatures and aridity (Mueller et al., 2020). Therefore, the recent increase in area burned at GILA likely reflects a combination of aggressive fire restoration and climatic changes, though we did not quantify their relative contributions.

Area burned in the dry conifer forests of SAGU has also increased in recent decades, similar to GILA, with a number of 20-year periods reaching historical levels (Table 3, Figures 7B and 9). However, overall area burned remains slightly below historical levels when considering the entire managed fire period (Table 2). Area burned during the managed fire period increased substantially beginning in the 1990s, despite the managed fire program at SAGU starting two decades earlier. This lag was likely due to a combination of management actions—such as the onset of the management-ignited prescribed burning program in 1993 and increased program capacity to manage lightning fires—along with increasing drought severity that contributed to an increase in wildfire activity (Hunter et al., 2014). Notably, the area with the highest fire frequency in either study area (11 overlapping fires since 1925; Figure 2A) consisted entirely of lightning-ignited fires, indicating that human ignitions are not required to restore historical levels of burning at SAGU.

Over the past two decades, no individual fire at SAGU has burned more than half of the study area, and the 20-year area burned has slightly decreased from its peak in the early 2000s. Despite ongoing climate pressures, the lack of a sustained increase in area burned suggests that local factors and management actions are effectively dampening these trends. The proximity of SAGU to Tucson, AZ—a metropolitan area of over one million people—limits the feasibility of managing lightning ignitions over large areas due to concerns about infrastructure risk and smoke. As a result, fires are more

often suppressed compared to GILA, and managed fires tend to be smaller than their natural potential. Additionally, fires igniting outside the study area and spreading into it are constrained by development, the small surrounding landscape, and the proximity to Sonoran Desert habitats with limited fuel loads. To address these challenges, SAGU has relied more on management-ignited prescribed burns to reintroduce fire into unburned areas with high fuel loads and sustain fire occurrence in previously burned areas. Combined with opportunistic management of lightning fires under favorable conditions, these efforts have reduced landscape-scale fuel loads and expanded options for controlling wildfire size and extent, even under severe conditions. This suggests that a management signature at SAGU is evident despite increasing aridity.

The recent acceleration in area burned is higher at GILA relative to SAGU, but uncertainty exists regarding whether this trend will continue or stabilize due to fuel limitation feedbacks. This increase is likely influenced by rising aridity in recent decades, similar to regional trends (Mueller et al., 2020), and projections indicate that climate change will further drive increases in area burned across the western United States (Abatzoglou et al., 2021). At GILA, fuel availability may be the primary constraint on continued further increases in area burned (e.g., Hurteau et al., 2019; McKenzie & Littell, 2017). Studies in the Gila and Aldo Leopold Wilderness Areas indicate that reduced fuels following fires can limit subsequent fire spread for approximately five to 20 years, depending on climate conditions (Parks et al., 2015). Additionally, increasing aridity may reduce the production of grass and herbaceous fuels, which are primary drivers of fire spread in dry conifer forests of the southwestern United States. Consequently, GILA could be one of the first landscapes to test whether area burned will continue to rise with increasing aridity—potentially surpassing pre-exclusion levels driven by a warming climate—or stabilize due to fuel limitations.

Area burned is a critical fire regime component, offering important insights into the extent and rate of fire as an ecological process across landscapes. Other fire regime components—such as fire severity—are also important for comparing modern and pre-exclusion fire regimes (McClure et al., 2024). Evolutionary adaptations of plants are strongly influenced by fire severity (Keeley et al., 2011). However, fire regime components are often inter-related. For example, increasing modern annual area burned is often associated with extreme single-day fire spread events (Coop et al., 2022), especially in long-unburned areas with high fuel loads. Most of the pre-exclusion area burned at both study sites, and in dry conifer forests across the southwestern United States, was

low- and moderate-severity fire. This is evidenced by the long-lived fire-scarred trees that survived numerous fires over centuries at our study sites, and at hundreds of other sites across the region (McClure et al., 2024; Swetnam & Baisan, 1996). Additionally, there is a lack of evidence for large (>100 ha) pre-exclusion high-severity patches in dry conifer forests of the region (Fulé et al., 2014; Hagmann et al., 2021; O'Connor et al., 2014). Changes in the rate and extent of area burned are important ecologically, as more frequent burning historically is often inversely correlated with lower fire severity. When possible, changes in annual area burned should be considered in the context of increasing modern fire severity in dry conifer forests of the region (McClure et al., 2024).

Our study areas are relatively unique, in that their occurrence of frequent, landscape-scale contemporary fires had largely beneficial effects. Despite large fire-free intervals in the modern era at GILA and across parts of SAGU, the first major fires during the managed fire period burned primarily with low-to-moderate severity (Farris et al., 2010; Hunter et al., 2014; Parks et al., 2023). This is significant because studies indicate that re-burns in dry conifer forests are more likely to stay low severity if the initial post-exclusion fire also burned at low or moderate severity (Harris & Taylor, 2017). This pattern was observed in SAGU where the largest modern-era fire in 1943, which burned under relatively cool conditions and only 50 years after the last pre-exclusion period fire in 1893, resulted in mostly low-to-moderate severity burning (based on early aerial imagery). In another part of the study area, low-severity burning resumed in 1925 after just a 32-year fire-free interval, which constrained the amount of fuel that could accumulate. At GILA, fire managers were able to manage a number of lightning fires in the early 1990s under moderate weather conditions to limit intensity and head fire development, which likely mitigated subsequent fire behavior under more severe conditions. For example, the 35,950 ha Miller Fire in 2011 and the 38,336 ha Johnson Fire in 2021 both burned during severe regional droughts but had limited high-severity areas where previous fires had occurred, with most high-severity burning in previously unburned areas (Hunter et al., 2014; Parks et al., 2023). More research is needed to better understand these complex interactions in these study areas.

## Implications for fire and land management

Fire exclusion over the last century has negatively impacted many forest ecosystems. Recent increases in wildfire activity in dry conifer forests across the southwestern United States are linked to higher fire severity

compared to historical fire regimes, although the total area burned remains below historical levels (McClure et al., 2024). Given the high fuel loads resulting from a century of fire exclusion and increasing aridity driven by climate change, it is likely that most southwestern US forests will experience wildfires within the next 50 years (Dillon et al., 2023; Stavros et al., 2014). Failure to reintroduce fire and reduce fuels may increase the risk of these forests burning under extreme fire weather conditions, during which suppression efforts are often ineffective (Young et al., 2020). Fires under such conditions are more likely to be of high severity, potentially leading to forest loss (Coop et al., 2020; Guiterman et al., 2022), as well as increased post-fire flooding and debris flows (Lopez et al., 2024) and reduced carbon storage along with infrastructure damage (Singleton et al., 2021).

There is broad consensus on the need to restore low-to-moderate-severity fire for ecosystem resilience, but uncertainty has remained about whether fire management can achieve the historically high levels of burning documented in the tree-ring records. Our multi-century area-burned reconstructions demonstrate that fire management can restore landscape-scale burning to historical levels, providing crucial context for evaluating restoration benchmarks, ecological outcomes, and future trends. While achieving historical levels may not be feasible in all landscapes, significant benefits can still be realized with burning levels below historical thresholds. For example, areas of SAGU and GILA that have burned just two to three times at low-to-moderate severity have reduced fuel loads and created open stands dominated by old-growth trees, thereby enhancing resistance and resilience to future fires and climate change (Hunter et al., 2014; van Mantgem et al., 2016, 2021). These initial burns may also enable future burns to be more cost-effective, produce less smoke, cover larger areas, and reduce firefighter risk by lowering fire severity and expanding strategic and operational options (Harris et al., 2024; Hunter et al., 2014).

While GILA and SAGU are notable in the western United States for achieving extraordinary levels of burning relative to historical levels, other wilderness areas have also made significant progress in restoring fire regimes. These include portions of Yosemite National Park in California and the Selway-Bitterroot Wilderness in Montana and Idaho (Berkey et al., 2021; van Wagtendonk, 2007). These areas began using prescribed burns and lightning ignitions in the late 1960s and early 1970s, demonstrating how sustained fire management efforts can restore fire regimes and enhance ecosystem resilience in dry conifer forests (Parks et al., 2023). Although wilderness areas face fewer constraints than other lands, lessons from frequent burning in dry forests within these areas can inform fire restoration efforts elsewhere. Fire

restoration is a long-term process, and the beneficial impacts of initial low-to-moderate-severity treatments can persist long term when followed by subsequent treatments. Initial fuels treatments—whether planned or unplanned—are a critical first step to increase operational opportunities on the landscape, enabling more effective and safer management of future fires (Harris et al., 2024).

## CONCLUSIONS

This study demonstrates that networks of tree-ring dated fire-scar chronologies can accurately reconstruct annual area burned in dry conifer forests without systematic bias. Applying ensemble and time-varying calibration methods to other landscapes could improve reconstruction accuracy and uncertainty estimation. However, estimates from individual models, including simple uncalibrated models like SYNC, can also yield reasonable reconstructions. Integrating statistical and spatially explicit models (e.g., TP) may further offset bias and reveal spatial properties of self-limiting fire behaviors over time, offering valuable insights for fire management strategies. The newly established North American Tree-Ring Fire-Scar Network (Margolis et al., 2022) offers the potential to scale historical area-burned estimates across the dry conifer forests of the southwestern United States and other regions with tree-ring fire-scar networks, potentially allowing for the reconstruction of area burned for forests in much of North America. These data are critical for understanding the landscape ecology of fire regimes, informing management practices, and improving forecasts of future fire regimes in the context of climate change.

Our methods enabled the first multi-century assessment of whether absolute area burned has been restored to pre-exclusion levels in two wilderness areas with active fire management programs: GILA and SAGU. These programs are notable for restoring such high amounts of burning to forests historically dominated by frequent, low-severity fire regimes. Our findings indicate that area burned has been restored to within historic levels at GILA and to near historic levels at SAGU compared to the pre-exclusion range of variability (Hunter et al., 2014). To our knowledge, no other ponderosa pine-dominated landscapes in the United States have experienced such a high frequency of overlapping fires. This information offers managers valuable context for assessing fire management effectiveness in these landscapes and beyond.

While pre-exclusion area-burned benchmarks are remarkably high and may be unachievable in many landscapes, we hypothesize that restoring low- and moderate-severity fires

even to levels well below historical rates can still enhance resilience in dense, fire-suppressed dry conifer forests now experiencing higher rates of uncharacteristically severe fires. Determining “how much fire” is enough to impart resilience requires more research to quantify fuel loading and forest structure thresholds that will support low-to-moderate fire severity under increasingly severe fire weather conditions driven by climate change.

## ACKNOWLEDGMENTS

Support for analysis and writing was provided by the U.S. Geological Survey Ecosystems Land Change Science Program, and the National Park Service Division of Fire and Aviation Management. Michael Crimmins and two anonymous reviewers provided helpful comments.

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

## DATA AVAILABILITY STATEMENT

The tree-ring fire-scar data (Margolis & Guiterman, 2021) are available from the International Multiproxy Paleofire Database (IMPD): <https://doi.org/10.25921/pef0-zz47>. The modern fire perimeters and study area boundaries (Farris et al., 2025) are available from Dryad: <https://doi.org/10.5061/dryad.k6djh9wh2>.

## ORCID

C. A. Farris  <https://orcid.org/0000-0001-9748-4432>

E. Q. Margolis  <https://orcid.org/0000-0002-0595-9005>

D. A. Falk  <https://orcid.org/0000-0003-3873-722X>

## REFERENCES

- Abatzoglou, J. T., D. S. Battisti, A. P. Williams, W. D. Hansen, B. J. Harvey, and C. A. Kolden. 2021. “Projected Increases in Western US Forest Fire despite Growing Fuel Constraints.” *Communications Earth & Environment* 2: 227.
- Abatzoglou, J. T., and A. P. Williams. 2016. “Impact of Anthropogenic Climate Change on Wildfire across Western US Forests.” *Proceedings of the National Academy of Sciences of the United States of America* 113(11): 770.
- Abolt, R. A. 1997. *Fire Histories of Upper Elevation Forests in the Gila Wilderness, New Mexico via Fire Scar Analysis and Stand Age Structure Analysis*. Tucson, AZ: University of Arizona.
- Agee, J. K. 1993. *Methods for Fire Ecology* 75–112. Washington, DC: Island Press.
- Allen, C. D. 2002. In *Lots of Lightning and Plenty of People: An Ecological History of Fire in the Upland Southwest*, edited by T. R. Vale, 143–193. Covelo, CA: Island Press.
- Allen, C. D., M. Savage, D. A. Falk, K. F. Suckling, T. W. Swetnam, T. Schulke, P. B. Stacey, P. Morgan, M. Hoffman, and J. T. Klingel. 2002. “Ecological Restoration of Southwestern Ponderosa Pine Ecosystems: A Broad Perspective.” *Ecological Applications* 12: 1418–33.

- Arizpe, A. H., D. A. Falk, C. A. Woodhouse, and T. W. Swetnam. 2020. "Widespread Fire Years in the US–Mexico Sky Islands Are Contingent on both Winter and Monsoon Precipitation." *International Journal of Wildland Fire* 29: 1072–87.
- Baisan, C. H., and T. W. Swetnam. 1990. "Fire History on a Desert Mountain Range: Rincon Mountain Wilderness, Arizona, U.S. A." *Canadian Journal of Forest Research* 20: 1559–69.
- Baker, W. L., and D. Ehle. 2001. "Uncertainty in Surface-Fire History: The Case of Ponderosa Pine Forests in the Western United States." *Canadian Journal of Forest Research* 31: 1205–26.
- Balch, J. K., B. A. Bradley, J. T. Abatzoglou, R. C. Nagy, E. J. Fusco, and A. L. Mahood. 2017. "Human-Started Wildfires Expand the Fire Niche across the United States." *Proceedings of the National Academy of Sciences of the United States of America* 114: 2946–51.
- Berkey, J. K., C. Miller, and A. J. Larson. 2021. "A History of Wilderness Fire Management in the Northern Rockies." Rocky Mountain Research Station, Fort Collins, CO.
- Brown, L. D., T. T. Cai, and A. DasGupta. 2002. "Confidence Intervals for a Binomial Proportion and Asymptotic Expansions." *The Annals of Statistics* 30: 160–201.
- Bukovsky, M. S., J. A. Thompson, and L. O. Mearns. 2019. "Weighting a Regional Climate Model Ensemble: Does It Make a Difference? Can It Make a Difference?" *Climate Research* 77: 23–43.
- Burrough, P. A., and R. A. McDonnell. 1998. *Principles of Geographical Information Systems*. New York, NY: Oxford University Press.
- Burke, M., A. Driscoll, S. Heft-Neal, J. Xue, J. Burney, and M. Wara. 2021. "The Changing Risk and Burden of Wildfire in the United States." *Proceedings of the National Academy of Sciences of the United States of America* 118: e2011048118.
- Collins, B. M., and S. L. Stephens. 2007. "Fire Scarring Patterns in Sierra Nevada Wilderness Areas Burned by Multiple Wildland Fire Use Fires." *Fire Ecology* 3: 53–67.
- Cook, E. R., C. A. Woodhouse, and C. M. Eakin. 2004. "Long-Term Aridity Changes in the Western United States." *Science* 306: 1015–18.
- Coop, J. D., S. A. Parks, C. S. Stevens-Rumann, S. D. Causbay, P. E. Higuera, M. D. Hurteau, A. Tepley, et al. 2020. "Wildfire-Driven Forest Conversion in Western North American Landscapes." *Bioscience* 70: 659–673.
- Coop, J. D., S. A. Parks, C. S. Stevens-Rumann, S. M. Ritter, and C. M. Hoffman. 2022. "Extreme Fire Spread Events and Area Burned under Recent and Future Climate in the Western USA." *Global Ecology and Biogeography* 31: 1949–59.
- Covington, W. W., and M. M. Moore. 1994. "Southwestern Ponderosa Forest Structure." *Journal of Forestry* 92: 39–47.
- Crimmins, M. A., and A. C. Comrie. 2004. "Interactions between Antecedent Climate and Wildfire Variability across South-Eastern Arizona." *International Journal of Wildland Fire* 13: 455–466.
- Dewar, J. J., D. A. Falk, T. W. Swetnam, C. H. Baisan, C. D. Allen, R. R. Parmenter, E. Q. Margolis, and E. J. Taylor. 2021. "Valleys of Fire: Historical Fire Regimes of Forest-Grassland Ecotones across the Montane Landscape of the Valles Caldera National Preserve, New Mexico, USA." *Landscape Ecology* 36: 331–352.
- Dillon, G. K., J. H. Scott, M. R. Jaffe, J. H. Olszewski, K. C. Vogler, M. A. Finney, K. C. Short, et al. 2023. *Spatial Datasets of Probabilistic Wildfire Risk Components for the United States (270m)*. Fort Collins, CO: U.S. Department of Agriculture.
- Dieterich, J. H. 1980. "The Composite Fire Interval — A Tool for More Accurate Interpretation of Fire History." In *Proceedings of the Fire History Workshop, 20–24 October 1980, Tucson, AZ. USDA Forest Service General Technical Report RM-81*, edited by M. A. Stokes and J. H. Dieterich, 8–14. Fort Collins, CO: Rocky Mountain Forest and Range Experiment Station.
- Edgington, E., and P. Onghena. 2007. *Randomization Tests*, 4th ed. New York: Chapman and Hall/CRC.
- Efron, B. 1987. "Better Bootstrap Confidence Intervals." *Journal of the American Statistical Association* 82(397): 171–185. <https://doi.org/10.1080/01621459.1987.10478410>.
- Falk, D. A., E. K. Heyerdahl, P. M. Brown, C. Farris, P. Z. Fule, D. McKenzie, T. W. Swetnam, A. H. Taylor, and M. L. Van Horne. 2011. "Multi-Scale Controls of Historical Forest-Fire Regimes: New Insights from Fire-Scar Networks." *Frontiers in Ecology and the Environment* 9: 446–454.
- Falk, D. A., P. J. van Mantgem, J. E. Keeley, R. M. Gregg, C. H. Guiterman, A. J. Tepley, D. Jn Young, and L. A. Marshall. 2022. "Mechanisms of Forest Resilience." *Forest Ecology and Management* 512: 120129.
- Farris, C., E. Margolis, and C. D. Allen. 2025. "Data from: New Methods Provide a 300-Year Perspective on Modern Area Burned in Two Wilderness Areas of the Southwest United States." Dataset. Dryad. <https://doi.org/10.5061/dryad.k6djh9wh2>
- Farris, C. A. 2009. "Spatial and Temporal Validation of Fire Histories Reconstructed from Fire Scars." Ph.D. Dissertation. University of Arizona, Tucson, Arizona.
- Farris, C. A., C. H. Baisan, D. A. Falk, M. L. Van Horne, P. Z. Fulé, and T. W. Swetnam. 2013. "A Comparison of Targeted and Systematic Fire-Scar Sampling for Estimating Historical Fire Frequency in South-Western Ponderosa Pine Forests." *International Journal of Wildland Fire* 22: 1021–33.
- Farris, C. A., C. H. Baisan, D. A. Falk, S. R. Yool, and T. W. Swetnam. 2010. "Spatial and Temporal Corroboration of a Fire-Scar-Based Fire History in a Frequently Burned Ponderosa Pine Forest." *Ecological Applications* 20: 1598–1614.
- Franco-Pereira, A. M., C. T. Nakas, B. Reiser, and M. C. Pardo. 2021. "Inference on the Overlap Coefficient: The Binormal Approach and Alternatives." *Statistical Methods in Medical Research* 30: 2672–84.
- Fulé, P. Z., W. W. Covington, and M. M. Moore. 1997. "Determining Reference Conditions for Ecosystem Management of Southwestern Ponderosa Pine Forests." *Ecological Applications* 7: 895–908.
- Fulé, P. Z., T. W. Swetnam, P. M. Brown, D. A. Falk, D. L. Peterson, C. D. Allen, G. H. Aplet, et al. 2014. "Unsupported Inferences of High-Severity Fire in Historical Dry Forests of the Western United States: Response to Williams and Baker." *Global Ecology and Biogeography* 23: 825–830.
- Gerow, K., D. R. Stewart, and C. Farris. 2021. "Relative Inference for Paired Data: More than Meets the Eye." *The Bulletin of the Ecological Society of America* 103: e01960.
- Girardin, M. P. 2007. "Interannual to Decadal Changes in Area Burned in Canada from 1781 to 1982 and the Relationship to Northern Hemisphere Land Temperatures." *Global Ecology and Biogeography* 16: 557–566.

- Guiterman, C. H., R. M. Gregg, L. A. E. Marshall, J. J. Beckman, P. van Mantgem, D. A. Falk, J. Keeley, et al. 2022. "Vegetation Type Conversion in the US Southwest: Frontline Observations and Management Responses." *Fire Ecology* 18: 1–16.
- Hagedorn, R., T. N. Palmer, and F. Vitart. 2005. The rationale behind the success of multi-model ensembles in seasonal forecasting — I. *Basic concept. Tellus A: Dynamic Meteorology and Oceanography* 57: 219–233. [doi: [10.1111/j.1600-0870.2005.00103.x](https://doi.org/10.1111/j.1600-0870.2005.00103.x)]
- Hagmann, R. K., P. F. Hessburg, S. J. Prichard, N. A. Povak, P. M. Brown, P. Z. Fule, R. E. Keane, et al. 2021. "Evidence for Widespread Changes in the Structure, Composition, and Fire Regimes of Western North American Forests." *Ecological Applications* 31: e02431.
- Harley, G. L., C. H. Baisan, P. M. Brown, D. A. Falk, W. T. Flatley, H. D. Grissino-Mayer, A. Hessler, et al. 2018. "Advancing Dendrochronological Studies of Fire in the United States." *Fire* 1: 1–6.
- Harris, L., and A. H. Taylor. 2017. "Previous Burns and Topography Limit and Reinforce Fire Severity in a Large Wildfire." *Ecosphere* 8: e02019.
- Harris, L. B., C. A. Farris, D. Niziolek, and A. H. Taylor. 2024. "Severity of a Megafire Reduced by Interactions of Wildland Fire Suppression Operations and Previous Burns." *Environmental Research Letters* 19: 114070.
- Haugo, R. D., B. S. Kellogg, C. A. Cansler, C. A. Kolden, K. B. Kemp, J. C. Robertson, K. L. Metlen, N. M. Vaillant, and C. M. Restaino. 2019. "The Missing Fire: Quantifying Human Exclusion of Wildfire in Pacific Northwest Forests, USA." *Ecosphere* 10: e02702.
- Hessburg, P. F., S. J. Prichard, R. K. Hagmann, N. A. Povak, and F. K. Lake. 2021. "Wildfire and Climate Change Adaptation of Western North American Forests: A Case for Intentional Management." *Ecological Applications* 31: e02432.
- Hessler, A., J. Miller, J. Kernan, D. Keenum, and D. McKenzie. 2007. "Mapping Paleo-Fire Boundaries from Binary Point Data: Comparing Interpolation Methods." *The Professional Geographer* 59: 87–104.
- Hunter, M. E., J. M. Iniguez, and C. A. Farris. 2014. "Historical and Current Fire Management Practices in Two Wilderness Areas in the Southwestern United States: The Saguaro Wilderness Area and the Gila-Aldo Leopold Wilderness Complex." U.S. Forest Service, Rocky Mountain Research Station, Fort Collins, CO. 38 p.
- Hurteau, M. D., S. Liang, A. L. Westerling, and C. Wiedinmyer. 2019. "Vegetation-Fire Feedback Reduces Projected Area Burned under Climate Change." *Scientific Reports* 9: 2838.
- Johnston, J. D., J. D. Bailey, and C. J. Dunn. 2016. "Influence of Fire Disturbance and Biophysical Heterogeneity on Pre-Settlement Ponderosa Pine and Mixed Conifer Forests." *Ecosphere* 7: e01581.
- Keeley, J. E., J. G. Pausas, P. W. Rundel, W. J. Bond, and R. A. Bradstock. 2011. "Fire as an Evolutionary Pressure Shaping Plant Traits." *Trends in Plant Science* 16: 406–411.
- Knutti, R., R. Furrer, C. Tebaldi, J. Cermak, and G. A. Meehl. 2010. "Challenges in Combining Projections from Multiple Climate Models." *Journal of Climate* 23(10): 2739–58. <https://doi.org/10.1175/2009JCLI3361.1>.
- Littell, J. S., D. McKenzie, D. L. Peterson, and A. L. Westerling. 2009. "Climate and Wildfire Area Burned in Western U. S. Ecoregions, 1916–2003." *Ecological Applications* 19: 1003–21.
- Lopez, M., E. Margolis, A. Tillery, S. Bassett, and A. Hook. 2024. "Pre-Fire Assessment of Post-Fire Debris-Flow Hazards in the Santa Fe Municipal Watershed." *International Journal of Wildland Fire* 33: WF23065.
- Mallek, C., H. Safford, J. Viers, and J. Miller. 2013. "Modern Departures in Fire Severity and Area Vary by Forest Type, Sierra Nevada and Southern Cascades, California, USA." *Ecosphere* 4: 1–28.
- Margolis, E. Q., and C. H. Guiterman. 2021. "NOAA/WDS Paleoclimatology – North American Tree-Ring Fire Scar Synthesis (NAFSS)." NOAA National Centers for Environmental Information. <https://doi.org/10.25921/pef0-zz47>.
- Margolis, E. Q., C. H. Guiterman, R. D. Chavardès, J. D. Coop, K. Copes-Gerbitz, D. A. Dawe, D. A. Falk, et al. 2022. "The North American Tree-Ring Fire-Scar Network." *Ecosphere* 13: e4159.
- Margolis, E. Q., C. A. Woodhouse, and T. W. Swetnam. 2017. "Drought, Multi-Seasonal Climate, and Wildfire in Northern New Mexico." *Climatic Change* 142: 433–446.
- Marlon, J. R., P. J. Bartlein, D. G. Gavin, C. J. Long, R. Anderson, C. E. Briles, K. J. Brown, et al. 2012. "Long-Term Perspective on Wildfires in the Western USA." *Proceedings of the National Academy of Sciences of the United States of America* 109: E535–E543.
- McClure, E. J., J. D. Coop, C. H. Guiterman, E. Q. Margolis, and S. A. Parks. 2024. "Contemporary Fires Are less Frequent but more Severe in Dry Conifer Forests of the Southwestern United States." *Communications Earth & Environment* 5: 581.
- McKenzie, D., and J. S. Littell. 2017. "Climate Change and the Eco-Hydrology of Fire: Will Area Burned Increase in a Warming Western USA?" *Ecological Applications* 27: 26–36.
- Meko, D. 1997. "Dendroclimatic Reconstruction with Time Varying Predictor Subsets of Tree Indices." *Journal of Climate* 10: 687–696.
- Minnich, R. A., M. G. Barbour, J. H. Burk, and J. Sosa-Ramirez. 2000. "Californian Mixed-Conifer Forests under Unmanaged Fire Regimes in the Sierra San Pedro Martir, Baja California, Mexico." *Journal of Biogeography* 27: 105–129.
- Moriasi, D. N., M. W. Gitau, N. Pai, and P. Daggupati. 2015. "Hydrologic and Water Quality Models: Performance Measures and Evaluation Criteria." *Transactions of the ASABE* 58(6): 1763–85. <https://doi.org/10.13031/trans.58.10715>.
- Morrison, P. H., and F. J. Swanson. 1990. "Fire History and Pattern in a Cascade Range Landscape." USDA Forest Service, Portland, OR.
- Mueller, S. E., A. E. Thode, E. Q. Margolis, L. L. Yocom, J. D. Young, and J. M. Iniguez. 2020. "Climate Relationships with Increasing Wildfire in the Southwestern US from 1984 to 2015." *Forest Ecology and Management* 460: 117861.
- O'Connor, C. D., D. A. Falk, A. M. Lynch, and T. W. Swetnam. 2014. "Fire Severity, Size, and Climate Associations Diverge from Historical Precedent along an Ecological Gradient in the Pinaleno Mountains, Arizona, USA." *Forest Ecology and Management* 329: 264–278.
- Palli, J., S. A. Mensing, E. M. Schoolman, F. Solano, and G. Piovesan. 2023. "Historical Ecology Identifies Long-Term Rewilding Strategy for Conserving Mediterranean Mountain Forests in South Italy." *Ecological Applications* 33: e2758.
- Parks, S. A., and J. T. Abatzoglou. 2020. "Warmer and Drier Fire Seasons Contribute to Increases in Area Burned at High Severity in Western US Forests from 1985 to 2017." *Geophysical Research Letters* 47: e2020GL089858.

- Parks, S. A., C. H. Guiterman, E. Q. Margolis, M. Lonergan, E. Whitman, J. T. Abatzoglou, D. A. Falk, et al. 2025. "A Fire Deficit Persists across Diverse North American Forests despite Recent Increases in Area Burned." *Nature Communications* 16: 1493.
- Parks, S. A., L. M. Holsinger, K. Blankenship, G. K. Dillon, S. A. Goeking, and R. Swaty. 2023. "Contemporary Wildfires Are more Severe Compared to the Historical Reference Period in Western US Dry Conifer Forests." *Forest Ecology and Management* 544: 121232.
- Parks, S. A., L. M. Holsinger, C. Miller, and C. R. Nelson. 2015. "Wildland Fire as a Self-Regulating Mechanism: The Role of Previous Burns and Weather in Limiting Fire Progression." *Ecological Applications* 25: 1478–92.
- Parsons, L. A., D. E. Amrhein, S. C. Sanchez, R. Tardif, M. K. Brennan, and G. J. Hakim. 2021. "Do Multi-Model Ensembles Improve Reconstruction Skill in Paleoclimate Data Assimilation?" *Earth and Space Science* 8: e2020EA001467.
- PRISM Climate Group, Oregon State University. 2022. "PRISM." <http://prism.oregonstate.edu>
- Rao, T. J. 2002. "Mean of Ratios or Ratio of Means or both?" *Journal of Statistical Planning and Inference* 102: 129–138.
- Reed, W. J., and E. A. Johnson. 2004. "Statistical Methods for Estimating Historical Fire Frequency from Multiple Fire-Scar Data." *Canadian Journal of Forest Research* 34: 2306–13.
- Rollins, M. G. 2009. "LANDFIRE: A Nationally Consistent Vegetation, Wildland Fire, and Fuel Assessment." *International Journal of Wildland Fire* 18: 235–249.
- Rollins, M. G., T. W. Swetnam, and P. Morgan. 2001. "Evaluating a Century of Fire Patterns in Two Rocky Mountain Wilderness Areas Using Digital Fire Atlases." *Canadian Journal of Forest Research* 31: 2107–23.
- Romme, W. H. 1980. "Fire History Terminology: Report of the Ad Hoc Committee." 135–137.
- Romme, W. H., and D. G. Despain. 1989. "Historical-Perspective on the Yellowstone Fires of 1988." *Bioscience* 39: 695–99.
- Roos, C. I., and T. W. Swetnam. 2012. "A 1416-Year Reconstruction of Annual, Multidecadal, and Centennial Variability in Area Burned for Ponderosa Pine Forests of the Southern Colorado Plateau Region, Southwest USA." *Holocene* 22: 281–290.
- Roos, C. I., T. W. Swetnam, T. J. Ferguson, M. J. Liebmann, R. A. Loehman, J. R. Welch, E. Q. Margolis, et al. 2021. "Native American Fire Management at an Ancient Wildland-Urban Interface in the Southwest United States." *Proceedings of the National Academy of Sciences of the United States of America* 118: e2018733118.
- Safford, H. D., A. K. Paulson, Z. L. Steel, D. J. N. Young, R. B. Wayman, and M. Varner. 2022. "The 2020 California Fire Season: A Year like no Other, a Return to the Past or a Harbinger of the Future?" *Global Ecology and Biogeography* 31: 2005–25.
- Schaeffer, R. L., W. Mendenhall, and L. Ott. 2006. *Elementary Survey Sampling*, 6th ed. Belmont, CA: Thompson Brooks/Cole.
- Searcy, J. K., and C. H. Hardison. 1960. "Double-Mass Curves." U. S. Geological Survey Report, Washington D.C. 66.
- Singleton, M. P., A. E. Thode, A. J. Sánchez Meador, and J. M. Iniguez. 2019. "Increasing Trends in High-Severity Fire in the Southwestern USA from 1984 to 2015." *Forest Ecology and Management* 433: 709–719.
- Singleton, M. P., A. E. Thode, A. J. Sánchez Meador, J. M. Iniguez, and J. T. Stevens. 2021. "Management Strategy Influences Landscape Patterns of High-Severity Burn Patches in the Southwestern United States." *Landscape Ecology* 36: 3429–49.
- Stavros, E. N., J. Abatzoglou, N. K. Larkin, D. McKenzie, and E. A. Steel. 2014. "Climate and Very Large Wildland Fires in the Contiguous Western USA." *International Journal of Wildland Fire* 23: 899–914.
- Stephens, S. L., D. L. Fry, B. M. Collins, C. N. Skinner, E. Franco-Vizcaino, and T. J. Freed. 2010. "Fire-Scar Formation in Jeffrey Pine-Mixed Conifer Forests in the Sierra San Pedro Martir, Mexico." *Canadian Journal of Forest Research* 40: 1497–1505.
- Swantek, P. 1997. "Development of a Database to Analyze Historical Fire Activity in Southern Arizona: A Prototype from Saguaro National Park." M.S. Thesis. University of Arizona Tucson, AZ.
- Swetnam, T. L., D. A. Falk, A. Hessler, and C. Farris. 2011. "Reconstructing Landscape Patterns of Historic Fires and Fire Regimes." In *The Landscape Ecology of Fire*, edited by D. McKenzie, C. Miller, and D. A. Falk. Dordrecht, Netherlands: Springer.
- Swetnam, T. W., C. D. Allen, and J. L. Betancourt. 1999. "Applied Historical Ecology: Using the Past to Manage for the Future." *Ecological Applications* 9: 1189–1206.
- Swetnam, T. W., and R. Anderson. 2008. "Fire Climatology in the Western United States: Introduction to Special Issue." *International Journal of Wildland Fire* 17: 1–7.
- Swetnam, T. W., and C. H. Baisan. 1996. "Historical Fire Regime Patterns in the Southwestern United States since AD 1700." In *Fire Effects in Southwestern Forests, Proceedings of the Second La Mesa Fire Symposium, March 29–31, 1994, Los Alamos, New Mexico. RM-GTR-286, 11-32*, edited by C. D. Allen. Fort Collins, CO: U.S. Department of Agriculture, Rocky Mountain Forest and Range Experiment Station.
- Swetnam, T. W., and C. H. Baisan. 2003. In *Tree-Ring Reconstructions of Fire and Climate History in the Sierra Nevada and Southwestern United States*, edited by T. T. Veblen, B. H. Baker, G. Montenegro, and T. W. Swetnam, 158–195. New York, NY: Springer.
- Swetnam, T. W., and J. L. Betancourt. 1998. "Mesoscale Disturbance and Ecological Response to Decadal Climatic Variability in the American Southwest." *Journal of Climate* 11: 3128–47.
- Swetnam, T. W., and J. H. Dieterich. 1985. "Fire History of Ponderosa Pine Forests in the Gila Wilderness, New Mexico." 390–397. USDA Forest Service INT-182, Missoula, MT, Nov 15-18, 1983.
- Swetnam, T. W., J. Farella, C. I. Roos, M. J. Liebmann, D. A. Falk, and C. D. Allen. 2016. "Multi-Scale Perspectives of Fire, Climate and Humans in Western North America and the Jemez Mountains, USA." *Philosophical Transactions of the Royal Society B: Biological Sciences* 371: 20150168.
- Taylor, A. H., and C. N. Skinner. 1998. "Fire History and Landscape Dynamics in a Late-Successional Reserve, Klamath Mountains, California, USA." *Forest Ecology and Management* 111: 285–301.
- van Mantgem, P. J., A. C. Caprio, N. L. Stephenson, and A. J. Das. 2016. "Does Prescribed Fire Promote Resistance to Drought in

- Low Elevation Forests of the Sierra Nevada, California, USA?" *Fire Ecology* 12: 13–25.
- van Mantgem, P. J., A. C. Caprio, N. L. Stephenson, and A. J. Das. 2021. "Forest Resistance to Extended Drought Enhanced by Prescribed Fire in Low Elevation Forests of the Sierra Nevada." *Forests* 12: 1248.
- van Wagtenonk, J. W. 2007. "The History and Evolution of Wildland Fire Use." *Fire Ecology* 3: 3–17.
- Wallis, S. 2013. "Binomial Confidence Intervals and Contingency Tests: Mathematical Fundamentals and the Evaluation of Alternative Methods." *Journal of Quantitative Linguistics* 20: 178–208.
- Weigel, A. P., M. A. Liniger, and C. Appenzeller. 2008. "Can Multi-Model Combination Really Enhance the Prediction Skill of Probabilistic Ensemble Forecasts?" *Quarterly Journal of the Royal Meteorological Society* 134: 241–260.
- Westerling, A. L., H. G. Hidalgo, D. R. Cayan, and T. W. Swetnam. 2006. "Warming and Earlier Spring Increase Western US Forest Wildfire Activity." *Science* 313: 940–43.
- Westerling, A. L., and T. W. Swetnam. 2003. "Interannual to Decadal Drought and Wildfire in the Western United States." *Eos* 84: 545–560.
- Williams, A. P., B. Livneh, K. A. McKinnon, W. D. Hansen, J. S. Mankin, B. I. Cook, J. E. Smerdon, et al. 2022. "Growing Impact of Wildfire on Western US Water Supply." *Proceedings of the National Academy of Sciences of the United States of America* 119: e2114069119.
- Williams, J. N., H. D. Safford, N. Enstice, Z. L. Steel, and A. K. Paulson. 2023. "High-Severity Burned Area and Proportion Exceed Historic Conditions in Sierra Nevada, California, and Adjacent Ranges." *Ecosphere* 14: e4397.
- Young, J. D., A. M. Evans, J. M. Iniguez, A. Thode, M. D. Meyer, S. J. Hedwall, S. McCaffrey, P. Shin, and C.-H. Huang. 2020. "Effects of Policy Change on Wildland Fire Management Strategies: Evidence for a Paradigm Shift in the Western US?" *International Journal of Wildland Fire* 29: 857–877.
- Zar, J. H. 2010. *Biostatistical Analysis*, 5th ed. Upper Saddle River, NJ: Prentice Hall.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Farris, C. A., E. Q. Margolis, J. M. Iniguez, D. A. Falk, K. G. Gerow, C. H. Baisan, C. D. Allen, and T. W. Swetnam. 2026. "New Methods Provide a 300-Year Perspective on Modern Area Burned in Two Wilderness Areas of the Southwest United States." *Ecosphere* 17(2): e70471. <https://doi.org/10.1002/ecs2.70471>