Evaluating critiques of evidence of historically heterogeneous structure and mixed-severity fires across dry-forest landscapes of the western USA

William L. Baker
Emeritus Professor
University of Wyoming
Laramie, WY 82071
Email: bakerwl@uwyo.edu,

Chad T. Hanson
Earth Island Institute
2150 Allston Way, Suite #460
Berkeley, CA 94704
Email: cthanson1@gmail.com

Mark A. Williams
P. O. Box 271135
Salt Lake City, UT 84127
markalanwilliams@hotmail.com

Dominick A. DellaSala
Wild Heritage, a project of Earth Island Institute
2150 Allston Way, Suite #460
Berkeley, CA 94704
Email: dominick@wild-heritage.org

Keywords: ponderosa pine, mixed conifer, dry forests, historical range of variability, ecological restoration, fire, forest structure, forest management

Abstract. The structure and role of fire in historical dry forests, ponderosa pine (*Pinus ponderosa*) and dry mixed-conifer forests, of the western USA, have been debated for 25 years, leaving two theories. The first, that these forests were relatively uniform, low in tree density and dominated by low- to moderate-severity fires was recently reviewed, including a critique of opposing evidence. The second, that these forests historically had heterogeneous structure and a mixture of fire severities, has had several published reviews. Here, as authors in part of the second theory, we critically examined evidence in the first theory's new review, which presented 37 critiques of the second theory. We examined evidence for and against each critique, including evidence presented or omitted. We found that a large body of published evidence against the first theory and supporting the second theory, presented in 10 published rebuttals and 25 other published papers, by us and other scientists, was omitted and not reviewed. We reviewed omitted evidence here. Omitted evidence was extensive, and included direct observations by early scientists, maps in early forest atlases, early newspaper accounts and photographs, early aerial photographs, seven paleo-charcoal reconstructions, ≥18 tree-ring reconstructions, eight land-survey reconstructions, and an analysis of forest-inventory age data. This large body of omitted published research provides compelling evidence supporting the second theory, that historical dry forests were heterogeneous in structure and had a mixture of fire severities, including high-severity fire. The first theory is rejected by this large body of omitted evidence.

Introduction

Sound evidence about the historical structure of natural vegetation and processes affecting this vegetation is important in understanding and managing ecosystems. By historical, here we mean prior to the expansion of industrial development and displacement of Indians. Biological diversity has embedded genetic composition from longterm response to historical variability in ecosystems where organisms live. Evidence about the historical range of variability (HRV) of ecosystems thus provides an essential frame of reference for restoring and managing ecosystems to maintain biological diversity and ecosystem services (Landres et al. 1999). Reconstructing the past is difficult, and has implications for public interests, so contrasting theories may be debated.

Historical forest structure and fire in dry forests in the western USA have been debated for the last 25 years (e.g., Covington and Moore 1994, Shinneman and Baker 1997). Dry forests include ponderosa pine (*Pinus ponderosa*) forests and dry mixed conifer forests (ponderosa with several associated trees). A major cause of debate is that all sources of historical evidence for large land areas have limitations (e.g., Williams and Baker 2010, 2011). These sources include (1) reconstructions from tree-ring data, early land surveys, and paleo-charcoal deposits and (2) early records from newspaper accounts, inventories, scientific reports, forest atlases, oblique photographs, and aerial photographs. Larger reviews of these sources about historical forest structure and fire in dry forests include Baker and Ehle (2001, 2003), Odion et al. (2014), Baker and Williams (2015, 2018), Hanson et al. (2015), Baker (2017a), Baker and Hanson (2017), and Hagmann et al. (2021). There are also many published local studies (e.g., Hessburg et al. 2007).

Given this large body of research, evidence for historically heterogeneous dry forests and mixed-severity fire has a sufficiently compelling evidence basis to qualify as an established theory. However, since the 1990s, there has also been evidence in support of the theory that historical dry forests were more uniform, low-density forests with predominantly low- to moderate-severity fire (e.g., Covington and Moore 1994). A recent review (Hagmann et al. 2021, "H et al." hereafter) for the first time synthesized evidence against the heterogeneous, mixed-severity theory, and in support of this more uniform, low- to moderate-severity theory.

Our purpose here is to present a critical review of the diverse sources of evidence, all in one place, that makes H et al. a logical focus. We use H et al.'s structure and refer to, and critique H et al. extensively, as it provides a sensible framework for this critical review–it will be easier for readers to compare the evidence and arguments if we follow H et al.'s structure, including tables they used to summarize evidence. However, we review relevant evidence, whether included or not included in H et al., as the overall goal is to address each of the critiques brought together in H et al. In so doing, this paper offers an updated review of evidence relevant to the theory that historical dry forests were heterogeneous in forest structure and shaped by mixed-severity fires.

We wish the reader to know that we found that H et al. omitted major bodies of published evidence that do not support their theory. H et al. said our publications misrepresented the state of the science, but did not claim we omitted a large body of evidence. Here we refute that our publications misrepresent the state of the science, and show that it is H et al. that did this by omitting a large body of evidence that does not support their theory. The following sections and tables point out evidence omitted by H et al. about each topic. Our conclusions summarize omissions. Unfortunately, H et al.'s omissions became a theme, because they are so significant.

H et al. presented about 37 published critiques in a section on "Evaluating evidence of lack of change" divided into "Misrepresented historical forest conditions" and "Misrepresented fire regimes." We also divided our text here into: (A) historical forest density and (B) historical fire rates and severity in dry forests. To facilitate comparison of evidence in H et al. and evidence reviewed here, we replicated Tables 3-6 in H et al. and added evidence, that H et al. omitted, into a new column in each of four tables here summarizing evidence and sources of critiques.

A. *H et al.'s "Misrepresented historical forest conditions" section omitted key evidence*

H et al. began with a critique of a new method to reconstruct historical tree density from original land-survey records (Williams and Baker 2011; "WB method" hereafter). This method has been used across >11 large landscapes in >1.9 million ha of dry forests (Baker and Williams 2018), and provided substantial new evidence that dry-forest landscapes were heterogeneous in structure (e.g., tree density, basal area) and had mixed-severity fires. H et al. thus may critique this method, because they support the first theory, not the second theory.

With reconstruction methods (e.g., WB method), there is a need to evaluate evidence about the development of the method and validations against independent modern and historical sources. Validations are inherently multi-proxy evidence, which H et al cited as most valuable. H et al. did not gather and evaluate all this available evidence, instead they omitted evidence in rebuttals and publications that does not support their theory. Here we present, defend, and discuss all available evidence regarding the WB method: (a) evidence about the development of the method relative to other methods, (b) evidence from critiques, (c) evidence from rebuttals of critiques, (d) evidence from modern and historical validations, including multi-proxy evidence, and (e) independent evidence, from other dry forests, that they were historically dense. To summarize the implications of the evidence that H et al. omitted, we replicated their Table 3 in our Table 1, and added a column that shows how H et al.'s conclusions were incorrect, because they omitted evidence in published rebuttals and other publications.

A1. Evidence about the development of the WB method relative to earlier methods

H et al. suggested Cogbill et al. (2018) is the correct analysis to use in evaluating the WB method, implying this method was not derived and tested properly: "...valid methods exist for deriving estimates from spatial point patterns, such as GLO bearing trees" (p. 15). However, Cogbill et al. only tested old existing point-pattern measures, with no test of the WB method at all, and they did no testing in western dry forests, only moister forests in the Midwest. They showed that old point-pattern measures typically have low accuracy, are biased, and require large sample sizes. These limitations, which Williams and Baker also studied and reported, were part of what spurred development of improved design-based estimators, including Voronoi-based estimators, that are more robust to a wide range of spatial patterns (Delincé 1986). Following Delincé, Williams and Baker (2011) explicitly improved on old methods by developing and validating Voronoi-based estimators for use in western dry forests. For comparison, Williams and Baker also tested common point-pattern measures in modern validations; they did generally perform poorly, were biased, and required larger sample sizes, as Cogbill et al. found. Williams and Baker (2010, 2011) had already shown, by the time of Cogbill et al., that their WB method was well derived, statistically sound (Delincé 1986), and overcame limitations of methods reviewed by Cogbill et al. Neither Cogbill et al. nor H et al. explained these motivations, advances, and tests of the WB method.

H et al. also incorrectly implied that the WB method can only provide an accurate estimate over a very large land area, but this is a known limitation of earlier methods, not a limitation of the WB method or an inherent property of land-survey data. H et al. incorrectly said: "...the extremely low sampling density of this national land survey limits reliable estimates to the average forest density for a large area" (p. 16). H et al. listed some accuracies for large land areas (3,000+ ha), but these are only from using the old, inaccurate, biased point-pattern methods that require pooling data across large land areas (Cogbill et al. 2018). Using the WB method, modern and historical validations (details below) showed that sample areas of ~518 ha in dry forests provide tree-density estimates with weighted mean errors of 19.3%. The WB method had already been well validated (Williams and Baker 2010, 2011) as an advance over earlier methods, that previously provided just one estimate for very large land areas (Cogbill et al. 2018). Our conclusion from this updated body of evidence is that Cogbill et al. did not test the WB method, was about midwestern forests, and had no relevance to the WB method or its findings. This validation evidence, omitted by H et al., does not support H et al's theory.

A2. Evidence from critiques in H et al.'s "misrepresented historical forest conditions" section

Another of H et al.'s arguments in their "Misrepresented historical forest conditions" section is that papers that used the WB method "have suggested that densities and fire severities of dry forests were higher and more variable than previously thought (Table 3)..." (p. 16), implying these estimates are erroneous and too high.

H et al.'s evidence (their Table 3) that tree-density estimates from the WB method are too high rests largely on their own published comments on the WB method and other publications that commented on, but did not test the WB method: (1) evidence from simulation modeling that the WB method leads to large overestimation errors (Levine et al. 2017), and evidence from a local empirical validation test against permanent plots that reported overestimation by the WB method (Levine et al. 2019), (2) findings of lower tree density from application of old point-pattern methods (Johnston et al. 2018, Knight et al. 2020), (3) findings of lower tree density from early timber inventories (Hagmann et al. 2013, 2014, 2017, 2018, 2019; Stephens et al. 2015, 2018), (4) findings of lower tree density from comparisons of tree-ring reconstructions in Colorado (Battaglia et al. 2018) and Oregon (Johnston et al. 2018) with land-survey reconstructions using the WB method, and (5) a mistaken entry in their Table 3 that has nothing to do with tree density; it is all about fire (Hanson and Odion 2016a, Collins et al. 2016). H et al. also expressed concern about comparisons of tree density from small plots with WB-method reconstructions for ~518 ha areas, and the inability of the WB method to reconstruct historical evidence at finer scales. They summarized evidence the WB method overestimates tree density as: "Density estimates based on Williams and Baker (2011) methods are also inconsistent with tree-ring reconstructions and early 20th-century timber inventory records for areas where the data overlap..." (p. 16), and "Dendrochronological reconstructions and early timber inventories demonstrate consistency with each other and with other independent sources" (p. 16). We address these criticisms next.

A3. Evidence from four published rebuttals of these critiques, all omitted by H et al.

H et al. did not cite or discuss evidence in four published rebuttals of their comments on articles that used the WB method (Table 1). Only their comments alone were the basis for arguments and evidence presented in their Table 3 and section on "Misrepresented historical forest conditions." H et al.'s evidence, that estimates from using the WB method with land surveys are too high, was refuted in these four omitted published rebuttals, discussed next.

(A3a). Levine et al. simulation modeling fatally flawed, as shown by key omitted rebuttal

Evidence from simulation modeling that argued the WB method overestimated tree density (Levine et al. 2017, 2019) actually showed the WB method works well. Levine et al. (2017) first incorrectly coded the WB method, a fatal error that invalidated this study, as shown in the rebuttal omitted by H et al. (Baker and Williams 2018). Levine et al. (2019) next used revised code in permanent plots and again reported overestimation by the WB method (Levine et al.

2019). However, another omitted rebuttal (Baker and Williams 2019) showed Levine et al. (2019) this time used incorrect equations. For their three sample sites, using their own coding of the WB method, when correct equations were used, relative mean errors were only 6.2%, 7.0%, and 25.9%, well within expected accuracy for the WB method (Williams and Baker 2011). Levine et al. (2017, 2019) are listed incorrectly in H et al.'s Table 3 and the text as evidence the WB method is wrong, but both Levine et al. (2017, 2019) are fatally flawed by use of incorrect code and equations. Omitted rebuttals (Baker and Williams 2018, 2019) showed that the WB method worked correctly and accurately even in highly altered modern forests in tiny plots, well outside their historical landscape-scale design, evidence of robust validity. H et al. omitted this key evidence, that does not support their theory.

(A3b). Old point-pattern methods, with lower accuracy and bias, not relevant to the WB method

H et al. said two studies, that used land-survey data, showed tree densities from the WB method are too low. However, their findings of low tree density were from application of old point-pattern methods (Johnston et al. 2018, Knight et al. 2020). These methods have no relevance to the WB method, since neither Johnston et al. nor Knight et al. actually used or tested the WB method at their sites. The methods they instead used have well-known low accuracy and documented underestimation bias (Williams and Baker 2011, Cogbill et al. 2018). Neither Johnston et al. nor Knight et al. expressed awareness of this significant limitation of the methods they chose to use. These two studies thus have no basis for claiming anything about the WB method. Johnston et al.'s critique also implied that a very large scale-mismatch, comparing their findings to Williams and Baker's, is valid, without reviewing its limitations, discussed next.

(A3c). H et al. showed a double standard, not objectivity, on scale mismatches

Critiques in the past, including several by these same authors, used a double standard on scale mismatches (Baker et al. 2018, Baker and Williams 2019), as they do here again. H et al. were concerned about mismatches in spatial scale in comparisons between a ~518 ha reconstruction polygon and a tree-ring reconstruction. Of course, this is not ideal, but it is also inherent in tree-ring reconstructions that their small plots produce scale mismatches with other historical sources. A limitation of tree-ring reconstructions is their often small spatial extent.

If H et al. were concerned about scale mis-matches, why did they not cite, mention, and review evidence from the most closely scale-matched validations of the WB method, which are the modern validations done in three states (Williams and Baker 2011)? These validations compared tree-density estimates from land-survey section-corner data and from small plots placed over these same section corners (Baker and Williams 2018). These closely scale-matched comparisons showed the WB method has high accuracy (details below). H et al.'s omission of these closely scale-matched validations showed lack of objectivity about evidence that tested and validated the WB method.

Although H et al. critiqued scale mismatches, they employed much larger scale mis-matches as evidence against the WB method. Battaglia et al. (2018) and Johnston et al. (2018) were presented in H et al.'s Table 3 as showing the WB method overestimates tree density. Battaglia et al.'s study area is ~30 times the Williams and Baker study area in the Front Range (Williams and Baker 2012a), and Battaglia et al. did not report estimates for just our study-area portion, so this is a very large scale mismatch. At most, only 6 of their 28 sampling points (21%) might occur within the Williams and Baker study area. Why did they not compare just these plots to our data, if they were seeking to objectively evaluate their own work? Johnston et al. (2018) compared their tree-ring reconstructions in five small plots with the Williams and Baker (2012a) overall estimates for their entire Blue Mountains study area. As was explained in the Baker and Williams (2019 Appendix S1) rebuttal, that H et al. omitted: "...Johnston et al. sampled and summarized Blue Mountains forests from only five clustered points covering the equivalent of perhaps 4 six-corner GLO pools, while our study sampled and summarized over a much larger area including over 500 six-corner GLO pools. Johnston et al. cannot validly infer from a small, nonrandom sample to the entire Blue Mountains landscape..." These are two examples of the double standard that H et al. used, but neither comparison they made is valid, because of large scale mismatches.

One small source (e.g., Johnston et al. 2018--Blue Mountains) or even a few sources within a large reconstruction area does not provide a valid comparison, particularly if its estimate is within the reconstructed historical range of variation. Land-survey reconstructions using the WB method show that variability was large across historical dry-forest landscapes (Williams and Baker 2013). H et al. cited Johnston et al. (2018) as evidence the WB method overestimates tree density. However, the Blue Mountains reconstruction (Williams and Baker 2012a) showed a mean of 167.3 trees/ha (median 146 trees/ha) and standard deviation of 89.8 trees/ha, so Johnston et al.'s weighted mean estimate of 112 trees/ha (Baker and Williams 2018) is well within the historical range of variability for Blue Mountain forests, even though their estimate is not from a statistically valid sample. If Johnston et al. had randomly selected their study sites and directly compared them to the same locations using the WB method, as in validations of other reconstructions (Baker and Williams 2018), the numbers would likely have been within the range of expected errors (Williams and Baker (2011).

We think that when comparing other sources, at finer spatial scales, to overall study-area estimates, it is only valid to do "general cross-validation" (Baker and Williams 2018) with findings from multiple sites in a land-survey study area. The two largest general cross-validations are: (1) in California's western Sierra, where Baker and Williams (2019) compared means, quartiles, and confidence intervals from 30 independent historical estimates of tree density with similar data from the Baker (2014) land-survey reconstruction. They found overlapping 95% confidence intervals for historical mean tree density (independent=257 trees/ha, land-surveys=293 trees/ha), similarity in distributions, and 14% relative error if independent estimates are considered the truth, and (2) on Arizona's Mogollon Plateau (mean study area estimate was 141.5 trees/ha versus the mean from eight tree-ring reconstructions of 122.0 trees/ha, a relative error of 16.0%, assuming tree-ring reconstructions represent truth (Baker and Williams 2018 Appendix Table S9). This is compelling multi-proxy evidence, omitted by H et al., that the WB method accurately reconstructs historical tree density across large landscapes.

Thus, the fuller set of evidence reviewed here shows scale mismatches to be inherent limitations of comparisons with some methods of reconstruction (e.g., tree-ring reconstructions), H et al. criticized validations for scale mismatches, but then used much larger scale mismatches to support their own arguments, evidence of their use of a double standard. When appropriate general cross-validations with multiple sites in land-survey study areas are evaluated, they show compelling multi-proxy evidence the WB method accurately reconstructs historical tree density across large landscapes. This evidence, omitted by H et al., does not support their theory.

(A3d). Agreement that early two-chain timber inventories underestimate and need correction

H et al. implied tree-density estimates from the WB method are too high. H et al. in their Table 3 said "...early timber inventory records and tree-ring reconstructions for the same study areas documented substantially lower tree densities than those estimated using Williams and

Baker (2011) methods," implying that estimates from the WB method are in error. This conclusion is incorrect, based on evidence in the original paper (Baker and Hanson 2017), that H et al. omitted, and evidence in the rebuttal (Baker et al. 2018) that H et al. also omitted, evidence that does not support H et al.'s theory.

Early timber inventories using two-chain wide strips failed early in modern evaluations and tests and later also in historical validations that found similar errors (Baker and Hanson 2017). These inventories required visual estimation over too large a distance (40 m) to be accurate, and were reported in the early-1900s to have large underestimation errors and require correction multipliers of about 2.0-2.5 (Baker and Hanson 2017). Even one-chain-wide inventories, with estimation over shorter distances (20 m), had errors of 21-25% in the earliest modern validation against plot data (Candy 1927). By the early 1930s, early timber inventories had been widely disparaged by agencies as not authentic data, and were abandoned for better methods, including plot samples (Baker and Hanson 2017). Large underestimation bias by early timber-inventory estimates can also be seen in other validations: (1) in comparing mean tree density from three early timber-inventory estimates (48 trees/ha) versus 19 estimates from independent sources (254 trees/ha) in the California-Western Sierra and (2) in comparing two early timber-inventory estimates (67 trees/ha) versus estimates from four other independent sources (218 trees/ha) in the Oregon-E. Cascades (Baker and Williams 2018 Appendix Table S9). Nonetheless, Hagmann et al. (2018) commented, regarding tree density, that early timber inventories had double-checking, comparisons did not consider differences in scale, minimum diameters, or natural variability, placement of inventories was not biased, and their cross-validations are valid.

However, Baker et al.'s (2018) rebuttal of Hagmann et al.'s (2018) comment, which H et al. omitted, confirmed that Hagmann et al. (2018) actually did not contest Baker and Hanson's

(2017) central findings about these early timber inventories: (1) "early timber inventory data, particularly from two-chain-wide transects, were documented between 1911 and 1916 to underestimate and be unreliable and were abandoned and replaced by more accurate methods by the 1930s..." (p. 2), (2) "...comparisons between timber inventory estimates and other sources...showed that it is timber inventory estimates, not other sources, that underestimate and need correction." (p. 3), (3) "...one-chain-wide inventories, if all available data are used, could be fairly accurate, but further validation is needed..." (p. 3), (4) quantitative estimates of immature conifer density and non-conifer trees "were not included in Stephens et al. (2015)" (p. 3), and, if included, historical tree density "...was ~17 times higher than the 25 trees/ha reported in ponderosa pine, and ~7 times higher than the 75 trees/ha reported in mixed-conifer forests...by Stephens et al. (2015)" (p. 3). This evidence does not support H et al.'s theory.

Regarding other points made by Hagmann et al. (2018): (1) the rebuttal (Baker et al. 2018) showed that early inventory "quality control records" were not accuracy tests and did not correct erroneous estimates, (2) the rebuttal agreed we had overestimated time available (more likely 15-30 min) for tallying trees in a transect, (3) the rebuttal updated Baker and Hanson's (2017) Table 1 to address concern about matching tree species, sizes, and time periods, and (4) the rebuttal found needed correction multipliers for early timber-inventory estimates were then 1.6-2.3, not 1.6-3.2, still large errors showing the need for large correction multiplication of early timber-inventory tree-density estimates, which we did here in Table 2.

Although Hagmann et al. (2018) did not dispute the central findings of Baker and Hanson (2017), that early timber-inventory data substantially underestimate tree density, and still need to use 1.6-2.3 correction multipliers before reporting tree-density estimates, they omitted any mention of our rebuttal (Baker et al. 2018) and did not do the necessary correction in this H et al.

paper. H et al. (their Table 3) still claimed Baker and Hanson (2017) is among several papers where "Fundamental errors compromise conclusions, including...(2) incorrect assumptions about the methodological accuracy of early timber inventories" (H et al. Table 3). We repeat that Hagmann et al. (2018) did not dispute the large inaccuracy of early timber inventory estimates of tree density. Moreover, Baker and Hanson (2017) and Baker et al. (2018) did not at all discuss "assumptions" about the accuracy of early timber inventories, as H et al. put it, they instead presented evidence, including documents, agency reports, and field tests, that showed early timber inventories have low accuracy and need correction multipliers of 1.6-2.3 to estimate tree density. These are documented failures, not "assumptions" as H et al. characterized them, that led to the abandonment of early timber inventories by the 1930s.

The papers that used two-chain-wide early timber inventories to estimate tree density and did not use correction multipliers, so their conclusions are invalid, are in Table 2. Shown are the missing corrected estimates using 1.6-2.3 correction multipliers, and also corrections for missing non-coniferous trees and small trees in one case. What emerges from this evidence, after these corrections, is that the forests that received timber inventories often had historical tree-density estimates that were near the first quartile to median tree density reconstructed from land-survey data for these areas, thus are within the estimated historical range of variability for tree density, but have lower density (Table 1). We made the case (Baker and Hanson 2017, Baker et al. 2018) that areas that received timber inventories likely had concentrations of large trees that typically are less dense than in younger forests with smaller trees. Thus, the full set of available evidence, reviewed again here, shows that early records and reports had documented that timber inventories underestimate, correction multipliers of 1.6-2.3 needed to be applied, and, when applied, these estimates are congruent with those from other historical older forests with large trees. H et al.

omitted evidence in the original paper (Baker and Hanson 2017) and the published rebuttal (Baker et al. 2018), that does not support their theory.

(A3e). The fourth entry in H et al. Table 3 is mis-placed, as Collins et al. (2016) is not about tree density or forest density and did not belong in this table, but instead in their Table 5. However, this is another case where H et al. cited their own comment (Collins et al. 2016), but omitted the rebuttal of this comment by Hanson and Odion (2016b). Hanson and Odion showed that: (1) Collins et al. said maps were wrong and therefore the interpretation, that forests had burned at high severity, was wrong, but Collins et al. just missed that areas that were forested by 1992, having recovered from early high-severity fires, had burned again, after the early high-severity fires, and (2) Collins et al. had omitted including essential 1911 field survey notes that directly described these high-severity fires. Both errors show that Collins et al.'s critiques were incorrect, and Hanson and Odion (2016a) remains valid, evidence that does not support H et al's theory.

A4. Omitted multi-proxy evidence of high accuracy from modern and historical validations

H et al. did not mention or review substantial published evidence on the accuracy and lack of bias of the WB method from both modern and historical validations (Williams and Baker 2010, 2011, 2012a, 2012b, 2014, Baker and Williams 2018), as noted above. These validations included considerable multi-proxy agreement, something H et al. had highlighted as strong evidence, but they did not review or report the validations, or the abundant multi-proxy evidence in them. We have to again update their incorrect summary from them omitting all this evidence.

In modern forests, H et al. omitted evidence that the WB method's Voronoi estimators and nine other existing estimators of tree density from land-survey data were tested and compared in field validations at 499 section corners in dry forests in three states (Williams and Baker 2011, Baker and Williams 2018 Appendix Table S1). The latest summary showed a weighted mean error of 19.3% relative to plot estimates (Baker and Williams 2018 Appendix Table S1). Nearly all other estimators, except the two new Voronoi estimators, including some tested by Cogbill et al. (2018), were significantly biased and underestimated modern tree density (Williams and Baker 2011). The WB method's Voronoi estimators are validated as the most accurate, unbiased estimators of tree density for use with land-surveys in modern dry forests in the western USA. H et al. omitted all of this evidence that the WB method is very well validated in modern forests.

In historical forests, H et al. also did not cite or review published evidence (Baker and Williams 2018 Appendix Table S4) that the WB method is quite accurate in reconstructing historical tree density, based on specific and general cross-validations with multiple sources, that also show high multi-proxy agreement. Specific cross-validations compare tree density from the six-corner reconstruction polygon that intersects an alternative source location with tree density at this source. Specific cross-validations at 18 source locations in Arizona, California, and Oregon had relative mean errors of 10.4-11.2% (Baker and Williams 2018 Appendix S4), much better than the 19.3% from modern validations. Relative mean errors were 9.6-10.7% in comparison with 12 tree-ring reconstructions, 10.0% in comparison with two early one-chain-wide timber-inventories, and 13.1% in comparison with four early permanent plots or other non-timber inventories. The WB method cross-validated well against multi-proxy historical sources, evidence that H et al. said they especially valued, but H et al. still omitted all this evidence.

General cross-validations compared sets of mean tree densities from independent historical studies (imprecisely located so cannot be overlaid) in or near reconstruction areas with tree-density reconstructions using the WB method for that area. For example, 19 tree-ring reconstructions across Arizona's Mogollon Plateau had a mean of 122 trees/ha, whereas the land-survey reconstruction from the WB method had a mean of 141.5 trees/ha, a relative error of 16.0% (Baker and Williams 2018 Appendix Table S9). A recent compilation of 15 tree-ring reconstructions, early inventories, and land-survey reconstructions for dry mixed conifer in the Southwest found a mean of 144.5 trees/ha, close to the WB-method estimate for mixed conifer on the Mogollon Plateau of 144.3 trees/ha (Wasserman et al. 2019). Others with smaller sample sizes include Oregon's Blue Mountains (4 early inventories) with a relative error of 27.8%, Oregon's Eastern Cascades (2 early inventories, 2 tree-ring reconstructions) with a relative error of 14.2%, and California's western Sierra (18 early inventories, 1 tree-ring reconstruction) with a relative error of 6.0%. This corrected full dataset shows that H et al.'s implication, that the WB method overestimates historical tree density, is incorrect, since the method showed relative errors of only 6-28% in validations across large land areas, which is supported by multi-proxy evidence and independent compilations (e.g., Wasserman et al. 2019). H et al. omitted all of this large body of validation evidence. Amy Waltz, an author of H et al., published evidence the WB method works well (Wasserman et al. 2019), then omitted any mention of that evidence in H et al. But, then, H et al. omitted all of this evidence, from extensive cross-validations, that the WB method is well validated and its reconstructions are sound. Evidence from these reconstructions does not support H et al.'s theory.

A5. Independent evidence from other dry forests that they were historically highly heterogeneous in tree-density and included substantial dense areas

Baker et al. (2007) reviewed evidence from 20 tree-ring reconstructions, forest-reserve reports, and other early scientific reports that dry forests in four Rocky Mountain states had highly variable tree densities, ranging from 17 to 19,760 trees/ha. Baker (2012 Appendix Table A1) published nine quotes from early forest-reserve reports and other early scientific reports that historical dry forests in the eastern Cascades of Oregon varied in historical tree-density, including some dense forests. Similarly, Baker (2014 Appendix A) published 47 quotes from early forest-reserve reports and other scientific reports documenting that Sierran mixed-conifer forests in California were highly variable in density, but typically dense. Also, Baker and Williams (2019) published evidence from 30 independent early estimates of historical tree density in Sierran mixed-conifer forests in California that had a mean of 257 trees/ha and a standard deviation of 100 trees/ha, showing that these historical forests were highly variable in tree density and generally dense. H et al. omitted all of this independent, multi-proxy evidence from more than half of the 11 western states that historical dry forests varied in density, and included substantial areas that were dense. This is an omission by H et al. of a large body of independent evidence, which they said they especially valued, that does not support their theory that historical dry forests were generally low in tree density and rather uniform in density.

<u>A. Conclusions–Abundant evidence the WB method accurately reconstructs forest density</u>

We showed here that what H et al. (p. 16) called "multiple weaknesses" and "...demonstrated methodological biases and errors" regarding land-survey reconstructions of historical tree density using the WB method had already been shown, in original papers and in rebuttals that H et al. omitted, to be invalid critiques. H et al. could have presented the evidence in original papers and in omitted rebuttals, then offered new counter-evidence, but they did not. H et al. simply summarized their previous comments, then omitted all evidence in published rebuttals of these comments and nearly all evidence in original papers. As a result, H et al.'s review is very incorrect regarding historical tree density in western USA dry forests. The WB method of reconstructing historical tree density had been validated to accurately estimate historical tree density by many closely scale-matched modern validations at section corners, and through many specific and general cross-validations with independent multi-proxy evidence. The reconstructions were validated by substantial independent, multi-proxy historical evidence. Independent sources (not land-survey reconstructions) in more than half of the 11 western states agreed that historical dry forests were highly variable in tree density and included dense forests. H et al. omitted all of this evidence, that does not support their theory.

B. H et al.'s "Misrepresented fire regimes" section omitted more evidence

It is basic to science, and objectivity in general, that available evidence both for and against a hypothesis or theory must be cited and evaluated, including both critiques and corresponding rebuttals of critiques. H et al. began this section with an incorrect summary of publications cited in their Tables 4-6: "Counter-evidence publications have also posited that the high-severity component of contemporary wildfires is consistent with historical fire regimes." Reconstructions using the WB method did find evidence of historical high-severity fire but did not report "consistency" with modern high-severity fire. What was found was that the proportion of high-severity effects on historical landscapes was higher than previously thought. Thus, some modern wildfires considered abnormal, are likely well within the historical range of variability.

B1. H et al. Table 4 omitted/mis-interpreted evidence on historical rate of low-severity fire

Evidence in H et al.'s Table 4 "Counter-premise" list mentions some concerns about past methods of estimating rates of historical low-severity fires. H et al. said "Counter-evidence" publications showed that historical rates of low-severity fires were not as frequent (short) as reported using “composite fire interval” (CFI) methods. Yes, this began with Baker and Ehle (2001, 2003), who critiqued the theoretical basis of CFI and ITFI for estimating the essential fire rate-parameters of fire rotation (FR) and population mean fire interval (PMFI), that they showed to be equivalent estimators of historical fire rates across landscapes. They theorized that the true fire rate, PMFI/FR, may lie between a CFI estimate, that is too short, and an ITFI estimate, that is too long. Baker and Ehle hypothesized and presented evidence that omission of origin-to-scar intervals, inclusion of small fires, targeted sampling, and known decline in mean CFI as samples increase, could together explain CFI estimates that are too short. H et al. cited studies in their Table 4 that presented evidence defending against these concerns with CFI estimates (e.g., Van Horne and Fulé 2006, Collins and Stephens 2007, Brown et al. 2008, Stephens et al. 2010), but these studies did not analyze why CFI estimates are too short relative to the PMFI/FR.

H et al.’s Table 4 omitted citing and reviewing the much larger body of evidence in Baker (2017a S1 Text), where there is detailed analysis, using 342 fire-history sampling sites, of all known hypotheses that could explain why CFI and ITFI estimates of PMFI/FR are inaccurate and biased toward intervals that, this study discovered, are both too short. These explanations included: (1) overcompensation from the compositing process, (2) destruction of long fire intervals by compositing, (3) insufficient CFI restriction rules, (4) censoring causing loss of long fire intervals, (5) targeted sampling also causing loss of long fire intervals, and (6) unstudied fire-severity inflating low-severity fire rates, because some of the fires likely were not low severity.

Even more important is that H et al.’s Table 4 column “Implications of evaluation” omitted extensive new evidence about how much CFI and ITFI underestimate PMFI/FR, and how they now can both be corrected to accurately estimate PMFI/FR (Baker 2017a). Baker used a 96-case calibration and analysis dataset from 44 fire-history studies where both CFI and/or ITFI were calculated, or could be calculated, and could be compared with estimated PMFI/FR. CFI measures all produced estimates that were too short (biases of 38-72%) and were quite inaccurate (errors of 43-70%) in estimating PMFI/FR. ITFI measures also produced estimates that were too short, but less so (biases of 3-28%) and were also less inaccurate (errors of 16-33%). Most important, linear regression showed that historical PMFI/FR could be very accurately estimated from Weibull mean ITFI (RMSE = 7.52, R^2_{adj} = 0.972) and quite accurately (R^2_{adj} > 0.900) from eight other CFI/ITFI measures. These linear regressions: (1) showed that all the CFI and ITFI measures and methods produced historical estimates that were too short, and (2) enabled correction of all CFI/ITFI estimates of historical PMFI/FR at 342 sites across the western USA.

Fortunately, a new landscape-scale method has been developed and validated for directly estimating PMFI/FR using random or systematic plots in which all scarred trees are sampled, fire years are cross-dated, and individual fire years are reconstructed spatially and used to estimate PMFI/FR (Farris et al. 2010, Dugan and Baker 2015). Baker (2017a) was able to find and use 24 of these fire-year reconstructions, showing that the fire-year reconstruction method is being widely used. This method does not require further use of inaccurate CFI or ITFI estimates, thus earlier debates over compositing, targeted sampling etc., that were the focus of H et al.'s comments, are no longer of much interest, since the science has moved on beyond those debates.

Plot methods can still be used, but have lower accuracy than these newer landscape methods, and require pooling over several plots, limiting their value. CFI and the all-tree-fire-interval (ATFI) plot methods (Kou and Baker 2006a, b) were tested in a modern and historical validation at Grand Canyon (Dugan and Baker 2014) that H et al. did not present or review. In these tests, ATFI outperformed all CFI measures. ATFI was always correct in modern tests at the plot scale and CFI mostly failed. In historical tests, ATFI had mean relative error of 14.3% and the best traditional CFI measure, scar-to-scar 25% filtered CFI, had mean relative error of 35.3%. ATFI was thus superior to all other plot-scale methods. ATFI at the plot scale can possibly achieve errors < 26.6%, but errors < 20% require at least four plots over 600-1000 ha (Dugan and Baker 2014). H et al.'s discussion of their Table 4 claimed that "Additionally, as acknowledged by Kou and Baker (2006: Accessory Publication), ATFI will always be much longer than any MFI..." (p. 20). H et al. thought this was a failing of the ATFI method, but this is actually because CFI's are always erroneously too short (Baker 2017a), and ATFI is longer and thus more correct. H et al. did not understand the ATFI method, and their critique is uninformed and incorrect.

Regression-corrected CFI/ITFI plot estimates and landscape-scale PMFI/FR estimates (n = 342) for western USA dry forests are available together in Baker (2017a). These show that frequent low-severity fire was historically much less prevalent than suggested incorrectly by the old CFI/ITFI methods that H et al. cited. H et al. defended old, out-of-date, inaccurate and biased methods of reconstructing historical rates of fire, without reviewing published evidence that these old CFI/ITFI measures and small-plot methods have been replaced with newer, more accurate PMFI/FR measures and spatial reconstruction methods, and the old CFI/ITFI estimates have been corrected to PMFI/FR estimates in Baker (2017a). H et al. omitted the large, significant body of evidence in Baker (2017a), that does not support their theory that low-severity fire dominated and was frequent in all dry forests.

B2. H et al. Table 5 omitted/mis-interpreted evidence about historical fire severity

H et al.'s theory is that low-severity fire with a little moderate-severity fire historically dominated dry forests. Our theory is that a mixture of fire severities occurred historically in all dry forests, with more low-severity fire in lower, drier settings and more high-severity fire in upper, moister settings. H et al.'s low-severity fire theory, however, is based on false and omitted evidence, covered in the following sections: (B2a) incorrect interpretation of fire scars and age-structure omits historical severe fires, (B2b) incorrect implication historical forests did not have high-severity fires, based on tree-ring reconstruction of fire in old growth, which typically lacked high-severity fire for centuries, (B2c) critiques of land-survey reconstructions of historical high-severity fires in dry forests, that were refuted, are repeated without reviewing the refutations, reporting only one side of the evidence, (B2d) use of early timber-inventories that found mostly low-severity fires, but from omitting key documents that showed evidence of high-severity fires, (B2e) omission of early forest-reserve reports, other scientific reports, and photographs, including their own publication, that found evidence of severe fires in historical dry forests, (B2f) omission of tree-ring reconstructions, including their own, that found evidence of severe fires in historical dry forests, (B2g) omission of 7 paleo-charcoal and 8 land-survey reconstructions that found evidence of severe fires at similar rates in historical dry forests, (B2h) omission of published validations of WB-method fire-severity reconstructions against independent multi-proxy sources in both modern and historical settings, (B2i) omission of Odion et al. (2016) that showed FIA data can still reconstruct fire severity, and (B2j) omission of rebuttal and new evidence of historically large high-severity fire patches.

(B2a). Incorrect interpretation of fire scars and age-structure omits role of historical severe fires

In H et al. Table 5, the citation of Brown (2006) as a counter to Shinneman and Baker (1997) repeats an incorrect interpretation of evidence by Brown. Prior to Brown (2006), Brown et al. (1999) studied fire history in the Colorado Front Range. Trees that died about the time of a dated fire and trees that regenerated in a pulse after a dated fire were interpreted as strong evidence of high-severity fire. Brown et al. also accepted that a pulse of trees established after a dated fire may also indicate a high-severity fire. However, Brown and Wu (2005) found the same evidence, but interpreted tree-regeneration pulses as having an unknowable disturbance cause and instead regional climate forcing: "...cohort structure is uncoupled from any single mortality event and instead appears to be the result of broader scale climate forcing of fire timing that resulted in successful recruitment episodes" (p. 3036). The flaw in this interpretation is that disturbance history and climate history are confounded; to determine the effect of one variable, the other must be controlled, which Brown and Wu did not do. It is not possible to validly conclude climate forcing was the cause, without showing fire was not the cause of tree-regeneration pulses.

Brown (2006 Figure 3) showed the same set of evidence, that should have led to recognition of confounding and possible interpretation as high-severity fire (Brown et al. 1999), but Brown instead said: "Abundant synchronous tree recruitment affected by optimal climate forcing is probably the reason for extensive stands of even-aged forests in the Black Hills, rather than widespread crown fires..." (p. 2507). However, Brown provided no explanation for how trees present before this period were all killed, so that regenerating stands became even-aged. If prior trees had not been mostly killed prior to a pulse of tree regeneration, resulting stands would not have been even-aged, but instead multi-aged. Again, the more likely explanation, that moderate- to high-severity fires produced the evidence presented in Brown (2006 Figure 3) was never analyzed. Failure to exclude a confounded variable, fire, before assuming climate-forcing as the cause, has been a repeated error in inference (e.g., O'Connor et al. 2017).

This climate-forcing theory of tree-recruitment pulses of Brown and others, was not supported in a key test. In Dugan and Baker (2015), these authors directly tested whether fires, fire-quiescent periods, droughts, or pluvials, in some combination or permutation, had separate or combined influences on the occurrence of historical tree-recruitment pulses in ponderosa pine forests in Grand Canyon National Park, Arizona. The conclusion was: "Permutation analysis showed that mortality-inducing influences of fire and drought played the primary role in initiating pulses as they occurred first for 90% of pulses, significantly more than expected...drought was the most important single initiator...as the first influence for 65% of pulses. Mixed-severity fire was the initial influence for 30% of fires...none of the 20 pulses had a pluvial influence alone" (p. 704). It remains essential to test for effects of canopy-opening disturbances before assuming that moist periods trigger these pulses; this test showed moist periods do not trigger pulses without a canopy-opening event, such as a moderate- to high-severity fire, drought, or possibly a beetle outbreak (not reconstructed). The climate-forcing conclusions of the Brown studies (e.g., Brown and Wu 2005, Brown 2006) are invalid, because no evidence was analyzed to exclude the possibility that severe fires were the cause of pulses.

(B2b). Incorrect implication historical forests did not have high-severity fires, based on tree-ring reconstruction of fire in old growth, which typically lacked high-severity fire for centuries.

Tree-ring reconstructions of fire history have commonly been biased against the detection of historical moderate- to high-severity fires. In a revealing moment, Grissino-Mayer (1995) said of volcanic landscapes in New Mexico: "We found no fire-scarred samples on the kipukas in the northern and eastern portions of the malpais, and found few samples in the southern portions. These areas contained ponderosa forests that appeared younger than elsewhere, perhaps due to more recent, intense stand-replacing fires..." (p. 136). This study did no analysis of fire-severity or fire frequency overall, instead excluded areas with possible evidence of high-severity fires and focused on older forests with abundant scars and lower-severity fires. This, of course, is biased sampling. Conclusions about historical fire-severity, in general, from biased sampling cannot be validly extrapolated to other areas. Yet, this is not unusual for fire-history studies in dry forests. Baker (2017a) found 32% of 342 fire-history sites explicitly targeted plots in old forests with concentrations of fire scars, where moderate- to high-severity fire likely had not occurred for long periods. Moreover, 74% of fire-history sites did not include any analysis of fire severity, and just assumed historical fires were low severity. In contrast, where fire severity was studied, some mixed- and high-severity fire was usually found, showing the low-severity bias in most studies.

Brown (2006), which is cited in H et al.'s Table 5 as countering Shinneman and Baker's (1997) finding of historically severe fires in the Black Hills, was similarly conducted in mostly old growth, where the probability of finding high-severity fires is very low (Baker 2017a), so it is not surprising that Brown (2006) found little evidence of historical high-severity fire. Merschel et al. (2014), similarly, intentionally sampled in "areas of older forest" (p. 1673), but nonetheless claimed: "The ubiquitous presence of large, multi-aged ponderosa pine at all sites, regardless of environmental setting, suggests historical fires were frequent and predominantly low severity..." Thus, most previous fire-history studies, including those cited by H et al. in their Table 5 (Brown 2006, Merschel et al. 2014), do not provide valid inference about historical fire severity across larger landscapes, as they are not random samples, they are mostly from rarer old-growth forests that inherently lacked moderate- to high-severity fires for long periods (Baker 2017a).

(B2c). Critiques of reconstructions of historical high-severity fires in dry forests, that were refuted, are repeated without reviewing the refutations, reporting only one side of the evidence.

Fulé et al. (2014) critiqued Williams and Baker (2012a) and received 95 citations by 9-29-

2021 (Google Scholar). Williams and Baker (2014) responded with a forceful refutation that received only 23 citations. Stevens et al. (2016) critiqued Odion et al. (2014) and received 50 citations. Odion et al. (2016) responded with a detailed refutation that received only 11 citations. Levine et al. (2017) critiqued Williams and Baker (2012a) and received 38 citations. Baker and Williams (2018) responded with a detailed refutation that received only 13 citations. Similarly, Levine et al. (2019) critiqued Baker and Williams (2018) and received 8 citations. Baker and Williams (2019) responded with a detailed refutation, and received only 1 citation. These data suggest many scientists are not reporting and weighing the evidence equally, but simply endorsing critiques, without examining and citing published rebuttals. These are also cases of omission of evidence, but by a broader part of the scientific community.

(B2d). Use of early timber-inventories that found mostly low-severity fires, but from omitting key documents that showed evidence of high-severity fires

H et al., in their Table 5, cited Hagmann et al. (2018) as evidence ostensibly rebutting Baker and Hanson (2017) regarding their findings of historical high-severity fire occurrence in ponderosa pine and mixed-conifer forests of the Sierra Nevada and Oregon. H et al., however, omitted the evidence in Baker et al. (2018), which rebutted Hagmann et al. (2018). Baker et al. (2018) explained that Hagmann et al. (2018) actually did not challenge or dispute the abundant evidence of historical high-severity fire presented in Baker and Hanson (2017). This evidence included: (a) extensive U.S. Forest Service field notes and maps documenting the occurrence of high-severity fire, and young, naturally-regenerating conifer forests following severe fire, from forest surveys circa 1911 in two different areas of the Sierra Nevada, and (b) explicit notes and observations from three different U.S. Forest Service reports, circa 1904-1912, regarding small and large high-severity fire patches, and naturally-regenerating conifer forest following severe fire. H et al. thus again omitted available evidence that does not support their theory.

(B2e). Omission of early forest-reserve reports, other scientific reports, and photographs, including their own publication, that found evidence of severe fires in historical dry forests

Authors of H et al. previously omitted or overlooked abundant evidence of historically severe fires in dry forests. Fulé et al. (2014), which included eight authors of H et al., incorrectly said: "W&B also fail to acknowledge the lack of contemporary evidence for large, patch-size crown fires in low- and mid-elevation dry forest landscapes, such as primary observation or photographic documentation in the 19th and early 20th centuries. The lack of direct documentary evidence of extensive crown fire in ponderosa pine forests in particular has been noted and reported repeatedly by ecologists and land-use historians for nearly 90 years..." (p. 826). This was incorrect, since Williams and Baker (2012a), which they were critiquing, had actually summarized direct independent evidence of high-severity fires in their study areas in AZ, CO, and OR (Williams and Baker 2012a, Appendix S1). This evidence included early journal articles from the turn of the century, forest-reserve reports by government scientists, analysis of early aerial photographs, tree-ring and fire-scar studies, and paleo-charcoal reconstructions.

Another author of H et al., Paul Hessburg, published early aerial photographic evidence of historically severe fires in >300,000 ha of dry northwestern forests (Hessburg et al. 2007), but H et al. remarkably omitted any review of the extensive evidence in this publication.

Yet another author of H et al, A. G. Merschel of Merschel et al. (2014) thought "the wave of tree establishment that began in ~1900...was likely caused by a variety of factors, including changes in fire regimes, selective tree harvesting, and domestic livestock grazing" (p. 1684) but rejected Baker's (2012) finding that late-1800s moderate- to high-severity fires led to this wave, by explaining: "it would require moderate- to high-severity fires occurring over an immense area...before 1900. Such fires are not recorded in written archives or tree-ring records from the region." However, Baker (2012 Supplemental Materials Appendix A) contained evidence from the written archives in early forest-reserve reports and other scientific reports of very extensive high-severity fires in the late-1800s in and near Merschel et al.'s study area that Merschel et al. did not report or review, nor was this evidence reported by H et al.

A large body of independent evidence, discussed in other sections, was also omitted by H et al. Baker et al. (2007) published 43 quotes from ca 1900 forest-reserve reports from throughout the Rocky Mountains that showed a diversity of historical fire severities, including abundant evidence of moderate- and high-severity fires. Baker (2009) published six early photographs of the aftermath of severe fires in dry forests in the Rocky Mountains. Baker (2014 Appendix A) published 208 quotes from early forest-reserve reports and other early scientific reports that documented historical moderate- to high-severity fires in Sierran mixed-conifer forests. Baker (2017b, 2018, 2020), documented that large late-1800s moderate- to high-severity fires occurred in dry forests on the Uncompahgre Plateau and in the San Juan Mountains, Colorado, based on forest-atlases, land-survey records, early photographs, early scientific publications, and other early records, including newspaper reports. All of this evidence, much of it independent and multi-proxy, which H et al. said was especially valuable, was omitted by H et al.

The repeated idea that there are no independent records of historically severe fires in dry forests is incorrect. These records have been available since the 1990s, and even more widely published in reviews (e.g., Odion et al. 2014) and other papers cited above since 2014. Eight authors of H et al. since 2014 in their published papers omitted this large body of evidence, and now H et al. again omitted all of this evidence, that does not support their theory.

(B2f). Omission of ≥18 tree-ring reconstructions, including their own, that found evidence of severe fires in historical dry forests.

H et al. did not cite or review that there have been ≥18 tree-ring reconstructions that found evidence of moderate- to high-severity fires in historical dry forests. Many of these were reported in Odion et al. (2014), including six published studies from the southern Cascades and Sierra in California, one from southern British Columbia, 10 from the Rocky Mountains, and two from the Southwest. Others include Wu (1999) and Tepley and Veblen (2015) in the San Juan Mountains. Remarkably again, H et al. did not cite or review Brown et al. (1999) from the Colorado Front Range, by an author of H et al., which documents severe fires in dry forests. The idea there are no independent tree-ring reconstructions of historical severe fires in dry forests has been incorrect for about two decades, and again is incorrect. H et al. omitted all of this evidence, including their own study, that does not support their theory.

(B2g). Omission of 7 paleo-charcoal and 8 land-survey reconstructions that found evidence of severe fires at similar rates in historical dry forests

H et al. did not cite or review that there have been seven paleo-charcoal studies that found evidence of severe fires in the last 500-600 years in dry forests (cited in Table 1 in Baker 2015a). These include Long et al. (2011) from the Eastern Cascades, Oregon (estimated fire rotation = 333 years), Fitch (2013) from northern New Mexico (~500 years), Pierce and Meyer (2008) and Pierce et al. (2004) from central Idaho (154-286 years, mean = 220 years), Jenkins et al. (2011) from northern Arizona (250 years), Bigio (2013) from southwestern Colorado (> 471 years), and

Colombaroli and Gavin (2010) from southern Oregon (500 years). The overall estimated high-severity fire rotation from these studies (Baker 2015a) had a mean of ~379 years, and a range of 154-500 years. The mean is 515 years, and the range 217-849 years from eight land-survey reconstructions (Baker 2015a). Both sources, which are independent of each other, document and validate each other in showing that infrequent high-severity fires occurred historically in dry forests. H et al. omitted all of this evidence, that does not support their theory.

(B2h). Omission of published validations of WB-method fire-severity reconstructions against independent multi-proxy sources in both modern and historical settings

Williams and Baker (2012a) calibrated and then validated their fire-severity reconstruction method using information directly from tree-ring reconstructions or direct measurements from historical forest plots where fire severity was assessed. Methods were directly calibrated using 55 estimates from areas where low-severity fire was dominant and from nine areas where mixed- or high-severity fire was dominant. The calibrated definitions and methods correctly predicted fire severity at all of the low-severity sites and all but one of the higher-severity locations, which was incorrectly assigned low severity as the high-severity event occurred 300 years ago.

For historical validations, Baker and Williams (2018) reported: "For historical fire severity, 10 specific cross-validations in six study areas in four states had high mean accuracy of 89.1-90.1%, based on PSC..." (p. 288), with the individual cross-validations in their Appendix S1 Table S7. Also, they reported: "There is substantial corroborating evidence that moderate/mixed-to-high-severity fires occurred and were extensive in some areas, based on evidence for five study areas in four states...These include 99 quotes from early forest-reserve and other reports, four tree-ring reconstructions, two paleo studies, and two using early photographs." This evidence was presented in detail in their Appendix S1 Tables S1 and S11.

Also, Williams and Baker (2012b) validated the use of survey section-line data to characterize the modern moderate- to high-severity fire regime in the Colorado Front Range, then analyzed 6904 km of historical section-line records, and found a historical higher-severity fire rotation of 249 years. This estimate is similar to and independent of the WB-method estimate (271 years) from Williams and Baker (2012a) for part of this area, further validating the WB method. Also important, this is independent direct surveyor-recorded evidence of historical moderate- to high-severity fires in historical dry forests. All of this evidence, that does not support their theory, was omitted by H et al.

(B2i). Omission of Odion et al. (2016) that showed FIA data can still reconstruct fire severity

H et al. Table 5 argued that Stevens et al. (2016) had shown that "errors of method and interpretation invalidate inferences about fire severity" from FIA stand-age data. However, H et al. omitted the rebuttal of Stevens et al. by Odion et al. (2016). The Odion et al. (2016) rebuttal of Stevens et al. (2016) found/noted that: (a) with the same definition of high-severity fire, there was 68% agreement between Stevens et al. (2016) and Odion et al. (2014) in terms of classifying historical high-severity fire using FIA stand-age plot-data; (b) 75% of the evidence for historical high-severity fire, which did not pertain to FIA, was not disputed or challenged by Stevens et al. (2016); and (c) while Stevens et al. questioned whether the current occurrence of high-severity fire patches >1000 ha is within the natural range of variation, Stevens et al. (2016) acknowledged that 'High-severity fire was undoubtedly a component of fire regimes in ponderosa pine and drier mixed-conifer forests', including patches >50 ha in area. H et al. omitted all of this evidence, that does not support their theory.

(B2j) Omission of rebuttal and new evidence of historically large high-severity fire patches

H et al. Table 5 argued that Spies et al. (2018) had shown that Odion et al. (2014) documented “only three patches of high-severity fire larger than >1000 ha in OR and WA in the early 1900s.” However, H et al. omitted the rebuttal of Stevens et al. by Odion et al. (2016). Odion et al. (2016) summarized data presented on p. 31 of DellaSala and Hanson (2015), wherein four different sources were discussed regarding historical occurrence of high-severity fire patches >1000 ha in mixed-conifer and ponderosa pine forests of OR and WA. Two of these sources documented individual high-severity fire patches of 14,000 ha and 24,000 ha, while the other two sources documented dozens of occurrences of such patches. Additional data regarding numerous historical high-severity patches of this size in OR and WA, as well as the Sierra Nevada and elsewhere across the western USA, were presented in DellaSala and Hanson (2019), new evidence that was also omitted by H et al. H et al. also omitted that Baker (2014 p. 26) had reported for the Sierra: “...the reconstructions show that contiguous areas of historical high-severity fire commonly exceeded 250 ha and reached as high as 9400 ha.” And, in the Colorado Front Range, H et al. omitted reporting that Williams and Baker (2012b) found that the maximum historical high-severity patch size was 8,331 ha, based on direct surveyor reports along section lines. Thus, H et al. again omitted all this evidence, that does not support their theory.

B3. H et al. Table 6 omitted and mis-interpreted evidence in all four entries in their table, creating a false narrative that high-severity fires have increased in long unburned forests, are preventing adequate recruitment, and are burning higher proportions of forests.

H et al. claimed Odion and Hanson (2006) stood for the proposition that "High-severity fire was rare in recent fires", whereas Odion and Hanson (2006) actually stood for the proposition that long-unburned forests are not experiencing higher fire severity in modern fires. H et al. also cited Safford et al. (2008) as rebutting Odion and Hanson (2006), but failed to mention Safford et al. (2008) was refuted by Odion and Hanson (2008). Odion and Hanson (2008) found Safford et al. had arbitrarily combined two time-since-fire categories, which created a false impression of slightly higher fire severity in long-unburned forests. Odion and Hanson (2008), using the same vegetation severity data, analyzed all time-since-fire categories and found that forests that had not burned in the longest period of time had similar or lower fire severity, not higher severity.

H et al. also cited Spies et al. (2010) as rebutting Hanson et al. (2009) regarding current fire-severity trends, but failed to mention that Spies et al. (2010) was subsequently refuted by Hanson et al. (2010). Hanson et al. (2010) found that a mathematical error, and reliance on an inaccurate anecdotal assertion, had led to an erroneous conclusion that the rate of high-severity fire in old forests of the Pacific Northwest was outpacing the old-forest recruitment rate from growth. Widespread rollbacks of forest protections, and increased logging, were being proposed based on the false data. Spies et al. (2010) did not dispute that the errors had been made, but hypothesized that the initial conclusion might still hold if a much broader high-severity fire definition was used. Hanson et al. (2010) analyzed the Forest Service's own fire-severity field-plot validation data and rates of high-severity fire in old forest from satellite imagery, finding that, even with the broader high-severity fire definition, old forest recruitment still outpaced the rate of high-severity fire in old forest by 7 to 29 times, depending on the subregion, and most mature trees survived fire under this broader definition.

H et al. listed a few studies as rebutting Williams and Baker's (2012a) evidence that severity distributions in some modern wildfires were not different from severity distributions in historical fire patterns they reconstructed. However, H et al. did not mention or cite the many published studies, discussed above, that have refuted these critiques, or the rebuttals and other counter-evidence regarding these few studies. Steel et al. (2015) reported no relationship between time-since-fire and high-severity fire for some forest types. They reported such a relationship for mixed conifer, but the model was based on data for only one narrow time-since-fire category, and the authors excluded from their analysis the most long-unburned forests—those with no recorded history of fire (Steel et al. 2015, Table 4, Figure 4). H et al. omitted evidence in Odion et al. (2010), Miller et al. (2012), and van Wagtendonk et al. (2012), which included the most long-unburned forests, and all time-since-fire categories, and found similar or lower proportions of high-severity fire in the most long-unburned forests. Steel et al. (2015) also reported historical high-severity fire proportions of 4-8% for mixed-conifer forests, based on only a theoretical model, but both Steel et al. (2015) and H et al. omitted mention of numerous studies finding much higher historical proportions of high-severity fire in these forests, based on historical field data, maps, and reports, including Baker (2014), Hanson and Odion (2016a,b), and Baker and Hanson (2017). Steel et al. (2018) reported an increase in high-severity fire proportion since 1984 in some regions, but used a fire-history database that is known to disproportionately omit large, severe fires in the earlier years of the dataset, causing a bias and potential to report false trends (Hanson and Odion 2015). H et al. omitted mention of Hanson and Odion (2015) and Baker (2015a), who used more comprehensive data and found no trends in high-severity fire proportion in the same regions. Guiterman et al. (2015) analyzed a single 38-ha high-severity fire patch, with very limited inferential potential for landscapes. Reilly et al. (2017) reported no increase in high-severity fire proportion in the Pacific Northwest since 1985 but indicated an increase in large high-severity fire patches. H et al., however, omitted DellaSala and Hanson (2019), who found the increase in large high-severity fire patches occurred from the 1980s through 1990s, but there has been no statistically detectable increase over approximately the past two decades.

H et al. cited Safford et al. (2015) as rebutting Hanson and Odion (2014), but neglected to cite or mention that Safford et al. (2015) was refuted by Hanson and Odion (2015). Safford et al. (2015) questioned fire-severity trend analyses reported by Hanson and Odion (2014) for the Sierra Nevada and hypothesized several potential methodological flaws. Hanson and Odion (2015) re-analyzed their initial data, using the new methods proposed by Safford et al. (2015), and found their initial conclusions were robust to re-analysis under Safford et al.'s new methods.

B. Conclusions–abundant multi-proxy evidence of historical moderate- to high-severity fires

Fire-history research has moved beyond old composite-fire-interval (CFI) rate measures, but H et al. cited old debates about CFI, and omitted papers on new methods that use the much sounder fire rotation, and have even corrected old CFI measures to fire rotations (Baker 2017a). These new estimates show frequent low-severity fire was less prevalent than previously thought.

Regarding historical fire severity: (1) research that suggested climate-forcing, not high-severity fires, led to pulses of tree regeneration, did not separate these confounded variables and their conclusions are not valid, (2) research from rare old-growth forests, showing lack of high-severity fires, is not valid evidence that other large parts of landscapes without old trees had severe fires, (3) early timber inventories, reported by H et al. to show low-severity fires dominated, had omitted key documents showing evidence of high-severity fires, (4) H et al. and some of its authors claimed there was no evidence of historically severe fires in dry forests, but omitted abundant published evidence of these fires, including by authors of H et al.

The very large body of evidence omitted by H et al. included hundreds of quotes from early historical documents, many direct observations by land-surveyors and observations by scientists in early forest-reserve reports, detailed mapping in early forest atlases done by the Forest Service, direct newspaper accounts, early oblique photographs, extensive analysis of early aerial photographs, ≥18 tree-ring reconstructions, seven paleo-charcoal reconstructions, eight land-survey reconstructions, and extensive reconstructions using modern forest-inventory and analysis (FIA) age data. Of course, each source has limitations and warrants some critiques, but H et al. omitted nearly all available evidence regarding historically severe fires in dry forests. Omitted evidence clearly shows dry forests historically had infrequent moderate- to high-severity fires.

Moreover, by omitting entire bodies of scientific evidence and rebuttal studies regarding time-since-fire and fire severity trends, H et al. created the false impression that long-unburned forests experience higher fire severity, and that high-severity fire proportion is increasing, when, in fact, the strong weight of scientific evidence indicates that long-unburned forests experience similar or lower fire severity, and high-severity fire proportion is not increasing.

Overall Conclusions–H et al. omitted nearly all evidence that does not support their theory

H et al. framed their review as an independent and objective critique of "dissent in the scientific literature" and "incomplete assessment of the best available science," by providing "a framework for objectively assessing change" (p. 3). This critique-of-dissent approach, however, quickly turned from objectivity and best available science to omission of evidence.

H et al. omitted virtually all evidence, that does not support their theory, in 10 published rebuttals of their papers (Table 6) and in 25 other published papers (Table 7). To elucidate the extent of omission and misrepresentation by H et al. clearly, our review here included: (1) replacement tables (Tables 1, 3-5) that add the evidence omitted by H et al. in their published tables, (2) summary tables that list all omitted rebuttals (Table 6) and omitted published studies with evidence that does not support their theory (Table 7), and (3) extensive text explaining that these omissions left out evidence that does not support H et al.'s theory and conclusions. Together, these show that nearly all of H et al.'s evidence about their theory, including nearly all their table entries, is incorrect and rebutted in publications these authors omitted, and usually did not even cite, much less review. Documented omission by H et al. of highly relevant published evidence, that does not support their theory, shows that H et al.'s conclusions are largely invalid.

This may have occurred before. Earlier we showed (Baker et al. 2018), in a rebuttal that H et al. omitted, that Hagmann et al. (2018) cited 11 papers that purportedly pointed out "errors in methodology or misrepresentation of the work of others" (p. 8), but alleged misrepresentations and errors were never explained. There was no presentation of evidence in nine published studies that specifically rebutted these 11 papers (Baker et al. 2018). These rebuttals were omitted.

Again, it is basic to science, and objectivity in general, that available evidence for and against a hypothesis or theory must be cited and evaluated, including both critiques and corresponding rebuttals of critiques. Methods and evidence must be clear and replicable. The major omissions of evidence by H et al. show that H et al. is not replicable, thus not valid science, and leaves us with a false published review of the state of the science regarding historical dry forests and their historical fires. The second theory, that dry forests had heterogeneous structure and a mixture of fire severities, was not refuted by H et al., and remains supported by the large body of scientific evidence (e.g., Tables 6, 7) that H et al. omitted. Failure of H et al. to reject a false theory (First theory), due to H et al.'s omission of evidence, has significant land-management implications, as thousands of hectares of dry forests may be inappropriately managed each year.

Literature Cited

Baker, W. L. 2006. Fire history in ponderosa pine landscapes of Grand Canyon National Park: is it reliable enough for management and restoration? International Journal of Wildland Fire 15:433-437.

Baker, W. L. 2009. Fire ecology in Rocky Mountain landscapes. Island Press, Washington, D.C.

Baker, W. L. 2012. Implications of spatially extensive historical data from surveys for restoring dry forests of Oregon's eastern Cascades. Ecosphere 3:1–39.

Baker, W. L. 2014. Historical forest structure and fire in Sierran mixed-conifer forests reconstructed from General Land Office survey data. Ecosphere 5:1–70.

Baker, W. L. 2015a. Are high-severity fires burning at much higher rates recently than historically in dry-forest landscapes of the Western USA? PLOS ONE 10:e0136147.

Baker, W. L. 2015b. Historical northern spotted owl habitat and old-growth dry forests maintained by mixed-severity wildfires. Landscape Ecology 30:655–666.

Baker, W. L. 2017a. Restoring and managing low-severity fire in dry-forest landscapes of the western USA. PLoS ONE 12:e0172288.

Baker, W. L. 2017b. The landscapes they are a-changin' - severe 19th-century fires, spatial complexity, and natural recovery in historical landscapes on the Uncompahgre Plateau. Colorado Forest Restoration Institute, Colorado State University, Fort Collins, CO.

Baker, W. L. 2018. Historical fire regimes in ponderosa pine and mixed-conifer landscapes of the San Juan Mountains, Colorado, USA, from multiple sources. Fire 1:23.

Baker, W. L. 2020. Variable forest structure and fire reconstructed across historical ponderosa pine and mixed conifer landscapes of the San Juan Mountains, Colorado. Land 9:article 3.

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 31:1205–1226.

Baker, W. L., and D. Ehle. 2003. Uncertainty in fire history and restoration of ponderosa pine forests in the western United States. Pages 319–333 P. N. Omi, L. A. Joyce (eds) Fire, fuel treatments, and ecological restoration: Conference proceedings. USDA Forest Service Proceedings RMRS-P-29, Rocky Mountain Research Station, Fort Collins, CO

Baker, W. L., and C. T. Hanson. 2017. Improving the use of early timber inventories in reconstructing historical dry forests and fire in the western United States. Ecosphere 8:e01935.

Baker, W. L., C. T. Hanson, and M. A. Williams. 2018. Improving the use of early timber inventories in reconstructing historical dry forests and fire in the western United States: Reply. Ecosphere 9:e023325.

Baker, W. L., T. T. Veblen, and R. L. Sherriff. 2007. Fire, fuels and restoration of ponderosa pine–Douglas fir forests in the Rocky Mountains, USA. Journal of Biogeography 34:251–269.

Baker, W. L., and M. A. Williams. 2015. Bet-hedging dry-forest resilience to climate-change threats in the western USA based on historical forest structure. Frontiers in Ecology and Evolution 2:article 88.

Baker, W. L., and M. A. Williams. 2018. Land surveys show regional variability of historical fire regimes and dry forest structure of the western United States. Ecological Applications 28:284-290.

Baker, W. L. and M. A. Williams. 2019. Estimating historical forest density from land-survey data: Response. Ecological Applications 29:e02017: doi:10.1002/eap.2017.

Battaglia, M. A., B. Gannon, P. M. Brown, P. J. Fornwalt, A. S. Cheng, and L. S. Huckaby.

2018. Changes in forest structure since 1860 in ponderosa pine dominated forests in the Colorado and Wyoming Front Range, USA. Forest Ecology and Management 422:147–160.

Bigio, E. R. 2013. Late Holocene fire and climate history of the western San Juan Mountains, Colorado: results from alluvial stratigraphy and tree-ring methods. PhD dissertation, University of Arizona, Tucson.

Brown, P. M. 2006. Climate effects on fire regimes and tree recruitment in Black Hills ponderosa pine forests. Ecology 87:2500–2510.

Brown, P. M., and R. Wu. 2005. Climate and disturbance forcing of episodic tree recruitment in a southwestern ponderosa pine landscape. Ecology 86:3030–3038.

Brown, P. M., M. R. Kaufmann, and W. D. Shepperd. 1999. Long-term, landscape patterns of past fire events in a montane ponderosa pine forests of central Colorado. Landscape Ecology 14:513-532.

Brown, P. M., C. L. Wienk, and A. J. Symstad. 2008. Fire and forest history at Mount Rushmore. Ecological Applications 18:1984–1999.

Candy, R. H. 1927. Accuracy of methods in estimating timber. Journal of Forestry 25:164-169.

Cogbill, C. V., A. L. Thurman, J. W. Williams, J. Zhu, D. J. Mladenoff, and S. J. Goring. 2018. A retrospective on the accuracy and precision of plotless forest density estimators in ecological studies. Ecosphere 9:e02187.

Collins, B. M., R. G. Everett, and S. L. Stephens. 2011. Impacts of fire exclusion and recent managed fire on forest structure in old growth Sierra Nevada mixed-conifer forests. Ecosphere 2:article 51.

Collins, B. M., J. M. Lydersen, R. G. Everett, D. L. Fry, and S. L. Stephens. 2015. Novel characterization of landscape-level variability in historical vegetation structure. Ecological Applications 25:1167-1174.

Collins, B. M., J. D. Miller, and S. L. Stephens. 2016. To the editor: a response to Hanson and Odion. Natural Areas Journal:234–242.

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.

Colombaroli, D., and D. G. Gavin. 2010. Highly episodic fire and erosion regime over the past 2,000 years in the Siskiyou Mountains, Oregon. Proceeding of the National Academy of Sciences USA 107:18909-18914.

Covington, W. W., and M. M. Moore. 1994. Southwestern ponderosa forest structure: changes since Euro-American settlement. Journal of Forestry 92:39-47.

Delincé, J. 1986. Robust density estimation through distance measurements. Ecology 67:1576-1581.

DellaSala, D. A., and C. T. Hanson. 2015. Ecological and biodiversity benefits of megafires. Pages 23-34 In: DellaSala, D. A., and C. T. Hanson, editors. The ecological importance of mixed-severity fires, nature's phoenix. Elsevier, Amsterdam.

DellaSala, D. A., and C. T. Hanson. 2019. Are wildland fires increasing large patches of complex early seral forest habitat? Diversity 11:157.

Dugan, A. J., and W. L. Baker. 2014. Modern calibration and historical testing of small-area, fire-interval reconstruction methods. International Journal of Wildland Fire 23:58-68.

Dugan, A. J., and W. L. Baker. 2015. Sequentially contingent fires, droughts and pluvials structured a historical dry forest landscape and suggest future contingencies. Journal of Vegetation Science 26:697-710.

Ehle, D. S., and W. L. Baker. 2003. Disturbance and stand dynamics in ponderosa pine forests in

Rocky Mountain National Park, USA. Ecological Monographs 73:543–566.

Farris, C. A., C. H. Baisan, D. A. Falk, M. L. V. 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 22:1021–1033.

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.

Fitch, E. P. 2013. Holocene fire-related alluvial chronology and geomorphic implications in the Jemez Mountains, New Mexico. M.S. Thesis, University of New Mexico, Albuquerque, NM.

Fulé, P. Z., T. A. Heinlein, W. W. Covington, and M. M. Moore. 2003. Assessing fire regimes on Grand Canyon landscapes with fire-scar and fire-record data. International Journal of Wildland Fire 12:129–145.

Fulé, P. Z., T. W. Swetnam, P. M. Brown, D. A. Falk, D. L. Peterson, C. D. Allen, G. H. Aplet, M. A. Battaglia, D. Binkley, C. Farris, R. E. Keane, E. Q. Margolis, H. Grissino-Mayer, C. Miller, C. H. Sieg, C. Skinner, S. L. Stephens, and A. Taylor. 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.

Grissino-Mayer, H. D. 1995. Tree-ring reconstructions of climate and fire history at El Malpais National Monument, New Mexico. PhD dissertation, University of Arizona, Tucson, AZ.

Guiterman, C. H., E. Q. Margolis, and T. W. Swetnam. 2015. Dendroecological methods for reconstructing high-severity fire in pine-oak forests. Tree-Ring Research 71:67–77.

Hagmann, R. K., et al. 2021. Evidence for widespread changes in the structure, composition, and fire regimes of western North American forests. Ecological Applications, in press.

Hagmann, R. K., J. F. Franklin, and K. N. Johnson. 2013. Historical structure and composition of ponderosa pine and mixed-conifer forests in south-central Oregon. Forest Ecology and Management 304:492–504.

Hagmann, R. K., J. F. Franklin, and K. N. Johnson. 2014. Historical conditions in mixed-conifer forests on the eastern slopes of the northern Oregon Cascade Range, USA. Forest Ecology and Management 330:158–170.

Hagmann, R. K., D. L. Johnson, and K. N. Johnson. 2017. Historical and current forest conditions in the range of the Northern Spotted Owl in south central Oregon, USA. Forest Ecology and Management 389:374–385.

Hagmann, R. K., A. G. Merschel, and M. J. Reilly. 2019. Historical patterns of fire severity and forest structure and composition in a landscape structured by frequent large fires: Pumice Plateau ecoregion, Oregon, USA. Landscape Ecology 34:551–568.

Hagmann, R. K., J. T. Stevens, J. M. Lydersen, B. M. Collins, J. J. Battles, P. F. Hessburg, C. R. Levine, A. G. Merschel, S. L. Stephens, A. H. Taylor, J. F. Franklin, D. L. Johnson, and K. N. Johnson. 2018. Improving the use of early timber inventories in reconstructing historical dry forests and fire in the western United States: Comment. Ecosphere 9:e02232.

Hanson, C. T., and D. C. Odion. 2014. Is fire severity increasing in the Sierra Nevada, California, USA? International Journal of Wildland Fire 23:1–8.

Hanson, C. T., and D. C. Odion. 2015. Sierra Nevada fire severity conclusions are robust to further analysis: a reply to Safford et al. International Journal of Wildland Fire 24:294-295.

Hanson, C. T., and D. C. Odion. 2016a. Historical forest conditions within the range of the Pacific fisher and Spotted owl in the central and southern Sierra Nevada, California, USA.

Natural Areas Journal 36:8–19.
Hanson, C. T., and D. C. Odion. 2016b. A response to Collins, Miller, and Stephens. Natural Areas Journal 36:229-233.
Hanson, C. T., D. C. Odion, D. A. Dellasala, and W. L. Baker. 2009. Overestimation of fire risk in the Northern Spotted Owl recovery plan. Conservation Biology 23:1314–1319.
Hanson, C. T., D. C. Odion, D. A. DellaSala, and W. L. Baker. 2010. More-comprehensive recovery actions for Northern Spotted Owls in dry forests: Reply to Spies et al. Conservation Biology 24:334-337.
Hanson, C. T., R. L. Sherriff, R. L. Hutto, D. A. DellaSala, T. T. Veblen, and W. L. Baker. 2015. Setting the stage for mixed- and high-severity fire. Pages 3-22 In: DellaSala, D. A. and C. T. Hanson, editors. The ecological importance of mixed-severity fires, nature's phoenix. Elsevier, Amsterdam.
Hessburg, P. F., R. B. Salter, and K. M. James. 2007. Re-examining fire severity relations in pre-management era mixed-conifer forests: inferences from landscape patterns of forest structure. Landscape Ecology 22:5-24.
Huffman, D. W., T. J. Zegler, and P. Z. Fulé. 2015. Fire history of a mixed conifer forest on the Mogollon Rim, northern Arizona, USA. International Journal of Wildland Fire 24:680–689.
Jenkins, S. E., C H. Sieg, D. E. Anderson, D. S. Kaufman, P. A. Pearthree. 2011. Late Holocene geomorphic record of fire in ponderosa pine and mixed-conifer forests, Kendrick Mountain, northern Arizona, USA. International Journal of Wildland Fire 20:125-141.
Johnston, J. D., C. J. Dunn, M. J. Vernon, J. D. Bailey, B. A. Morrissette, and K. E. Morici. 2018. Restoring historical forest conditions in a diverse inland Pacific Northwest landscape. Ecosphere 9:e02400.
Knight, C. A., C. V. Cogbill, M. D. Potts, J. A. Wanket, and J. J. Battles. 2020. Settlement-era forest structure and composition in the Klamath Mountains: reconstructing a historical baseline. Ecosphere 11:e03250.
Kou, X., and W. L. Baker. 2006a. A landscape model quantifies error in reconstructing fire history from scars. Landscape Ecology 21:735-745.
Kou, X., and W. L. Baker. 2006b. Accurate estimation of mean fire interval for managing fire. International Journal of Wildland Fire 15:489-495.
Landres, P. B., P. Morgan, and F. J. Swanson. 1999. Overview of the use of natural variability concepts in managing ecological systems. Ecological Applications 9:1179-1188.
Levine, C. R., C. V. Cogbill, B. M. Collins, A. J. Larson, J. A. Lutz, M. P. North, C. M. Restaino, H. D. Safford, S. L. Stephens, and J. J. Battles. 2017. Evaluating a new method for reconstructing forest conditions from General Land Office survey records. Ecological Applications 27:1498–1513.
Levine, C. R., C. V. Cogbill, B. M. Collins, A. J. Larson, J. A. Lutz, M. P. North, C. M. Restaino, H. D. Safford, S. L. Stephens, and J. J. Battles. 2019. Estimating historical forest density from land-survey data: A response to Baker and Williams (2018). Ecological Applications 29:e01968.
Long, C. J., M. J. Power, and P. J. Bartlein. 2011. The effects of fire and tephra deposition on forest vegetation in the central Cascades, Oregon. Quaternary Research 75: 151-158.
Merschel, A. G., T. A. Spies, and E. K. Heyerdahl. 2014. Mixed-conifer forests of central Oregon: effects of logging and fire exclusion vary with environment. Ecological Applications 24:1670–1688.
Meunier, J., N. S. Holoubek, and M. Sebasky. 2019. Fire regime characteristics in relation to

physiography at local and landscape scales in Lake States pine forests. Forest Ecology and Management 454:117651.

Miller, J. D., and H. D. Safford. 2017. Corroborating evidence of a pre-Euro-American low- to moderate-severity fire regime in Yellow pine-mixed conifer forests of the Sierra Nevada, California, USA. Fire Ecology 13:58–90.

Miller, J. D., C. N. Skinner, H. D. Safford, E. E. Knapp, and C. M. Ramirez CM. 2012. Trends and causes of severity, size, and number of fires in northwestern California, USA. Ecological Applications 22:184-203.

Morisita, M. 1957. A new method for the estimation of density by spacing method applicable to nonrandomly distributed populations. Physiology and Ecology 7:134-144.

Moritz, M. A., T. J. Moody, L. J. Miles, M. M. Smith, P. De Valpine. 2009. The fire frequency analysis branch of the pyrostatistics tree: sampling decisions and censoring in fire interval data. Environmental and Ecological Statistics 16:271-289.

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 Pinaleño Mountains, Arizona, USA. Forest Ecology and Management 329:264–278.

O'Connor, C. D., D. A. Falk, A. M. Lynch, T. W. Swetnam, and C. P. Wilcox. 2017. Disturbance and productivity interactions mediate stability of forest composition and structure. Ecological Applications 27:900–915.

Odion, D. C., and C. T. Hanson. 2006. Fire severity in conifer forests of the Sierra Nevada, California. Ecosystems 9:1177–1189.

Odion, D. C., and C. T. Hanson. 2008. Fire severity in the Sierra Nevada revisited: conclusions robust to further analysis. Ecosystems 11:12-15.

Odion, D. C., C. T. Hanson, A. Arsenault, W. L. Baker, D. A. DellaSala, R. L. Hutto, W. Klenner, M. A. Moritz, R. L. Sherriff, T. T. Veblen, and M. A. Williams. 2014. Examining historical and current mixed-severity fire regimes in ponderosa pine and mixed-conifer forests of western North America. PLOS One 9:e87852.

Odion, D. C., C. T. Hanson, W. L. Baker, D. A. DellaSala, and M. A. Williams. 2016. Areas of agreement and disagreement regarding ponderosa pine and mixed conifer forest fire regimes: a dialogue with Stevens et al. PLOS One 11:e0154579, doi: 10.1371/journal.pone.0154579.

Odion, D. C., M. A. Moritz, and D. A. DellaSala. 2010. Alternative community states maintained by fire in the Klamath Mountains, USA. Journal of Ecology 98:96-105.

Pierce, J. L., and G. A. Meyer. 2008. Long-term fire history from alluvial fan sediments: the role of drought and climate variability, and implications for management of Rocky Mountain forests. International Journal of Wildland Fire 17:84-95.

Pierce, J. L., G. A. Meyer, and A. J. T. Jull. 2004. Fire-induced erosion and millennial-scale climate change in northern ponderosa pine forests. Nature 432:87-90.

Polakow, D. A., and T. T. Dunne. 1999. Modelling fire-return interval T: stochasticity and censoring in the two-parameter Weibull model. Ecological Modelling 121:79-102.

Reilly, M. J., C. J. Dunn, G. W. Meigs, T. A. Spies, R. E. Kennedy, J. D. Bailey, and K. Briggs. 2017. Contemporary patterns of fire extent and severity in forests of the Pacific Northwest, USA (1985–2010). Ecosphere 8:e01695.

Safford, H. D., J. D. Miller, and B. M. Collins. 2015. Differences in land ownership, fire management objectives and source data matter: a reply to Hanson and Odion (2014). International Journal of Wildland Fire 24:286–293.

Safford, H. D., J. Miller, D. Schmidt, B. Roath, and A. Parsons. 2008. BAER soil burn severity maps do not measure fire effects to vegetation: A comment on Odion and Hanson (2006). Ecosystems 11:1–11.

Scholl, A. E., and A. H. Taylor. 2010. Fire regimes, forest change, and self-organization in an old-growth mixed-conifer forest, Yosemite National Park, USA. Ecological Applications 20:362-380.

Shinneman, D. J., and W. L. Baker. 1997. Nonequilibrium dynamics between catastrophic disturbances and old-growth forests in Ponderosa Pine landscapes of the Black Hills. Conservation Biology 11:1276–1288.

Spies, T. A., P. F. Hessburg, C. N. Skinner, K. J. Puettmann, M. J. Reilly, R. J. Davis, J. A. Kertis, J. W. Long, and D. C. Shaw. 2018. Chapter 3: Old growth, disturbance, forest succession, and management in the area of the Northwest Forest Plan. In: Spies, T.A.; Stine, P.A.; Gravenmier, R.; Long, J.W.; Reilly, M.J., tech. coords. Synthesis of science to inform land management within the Northwest Forest Plan area. PNW-GTR-966. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station:95–243.

Spies, T. A., J. D. Miller, J. B. Buchanan, J. F. Lehmkuhl, J. F. Franklin, S. P. Healey, P. F. Hessburg, H. D. Safford, W. B. Cohen, R. S. H. Kennedy, E. E. Knapp, J. K. Agee, and M. Moeur. 2010. Underestimating risks to the northern spotted owl in fire-prone forests: response to Hanson et al. Conservation Biology 24:330–333.

Steel, Z. L., M. J. Koontz, and H. D. Safford. 2018. The changing landscape of wildfire: burn pattern trends and implications for California's yellow pine and mixed conifer forests. Landscape Ecology 33:1159–1176.

Steel, Z. L., H. D. Safford, and J. H. Viers. 2015. The fire frequency-severity relationship and the legacy of fire suppression in California forests. Ecosphere 6:art8.

Stephens, S. L., D. L. Fry, B. M. Collins, C. N. Skinner, E. Franco-Vizcaíno, and T. J. Freed. 2010. Fire-scar formation in Jeffrey pine – mixed conifer forests in the Sierra San Pedro Mártir, Mexico. Canadian Journal of Forest Research 40:1497–1505.

Stephens, S. L., J. M. Lydersen, B. M. Collins, D. L. Fry, and M. D. Meyer. 2015. Historical and current landscape-scale ponderosa pine and mixed conifer forest structure in the Southern Sierra Nevada. Ecosphere 6:1–63.

Stephens, S. L., J. T. Stevens, B. M. Collins, R. A. York, and J. M. Lydersen. 2018. Historical and modern landscape forest structure in fir (*Abies*)-dominated mixed conifer forests in the northern Sierra Nevada, USA. Fire Ecology 14:1–14.

Stevens, J. T., H. D. Safford, M. P. North, J. S. Fried, A. N. Gray, P. M. Brown, C. R. Dolanc, S. Z. Dobrowski, D. A. Falk, C. A. Farris, J. F. Franklin, P. Z. Fulé, R. K. Hagmann, E. E. Knapp, J. D. Miller, D. F. Smith, T. W. Swetnam, and A. H. Taylor. 2016. Average stand age from forest inventory plots does not describe historical fire regimes in ponderosa pine and mixed-conifer forests of western North America. PLOS ONE 11:e0147688, doi: 10.1371/journal.pone.0147688.

Tepley, A. J., and T. T. Veblen. 2015. Spatiotemporal fire dynamics in mixed-conifer and aspen forests of the San Juan Mountains of southwestern Colorado, USA. Ecological Monographs 85:583-603.

Van Horne, M. L., and P. Z. Fulé. 2006. Comparing methods of reconstructing fire history using fire scars in a southwestern United States ponderosa pine forest. Canadian Journal of Forest Research 36:855–867.

van Wagtendonk, J.W., K.A. van Wagtendonk, and A.E. Thode. 2012. Factors associated with the severity of intersecting fires in Yosemite National Park, California, USA. Fire Ecology 8:

11-32.

Warde, W. And J. W. Petranka. 1981. A correction factor table for missing point-center quarter data. Ecology 62:491-494.

Wasserman, T. N., M. T. Stoddard, and A. E. M. Waltz. 2019. A summary of the natural range of variability for southwestern frequent-fire forests. Ecological Restoration Institute Workin Paper 42, Northern Arizona University, Flagstaff, Arizona.

Williams, M. A., and W. L. Baker. 2010. Bias and error in using survey records for ponderosa pine landscape restoration. Journal of Biogeography 37:707-721.

Williams, M. A., and W. L. Baker. 2011. Testing the accuracy of new methods for reconstructing historical structure of forest landscapes using GLO survey data. Ecological Monographs 81:63–88.

Williams, M. A., and W. L. Baker. 2012a. Spatially extensive reconstructions show variable-severity fire and heterogeneous structure in historical western United States dry forests. Global Ecology and Biogeography 21:1042–1052.

Williams, M. A., and W. L. Baker. 2012b. Comparison of the *higher*-severity fire regime in historical (A.D. 1800s) and modern (A.D. 1984-2009) montane forests across 624,156 ha of the Colorado Front Range. Ecosystems 15:832-847.

Williams, M. A., and W. L. Baker. 2013. Variability of historical forest structure and fire across ponderosa pine landscapes of the Coconino Plateau and south rim of Grand Canyon National Park, Arizona, USA. Landscape Ecology 28:297-310.

Williams, M. A., and W. L. Baker. 2014. High-severity fire corroborated in historical dry forests of the western United States: response to Fulé *et al.* Global Ecology and Biogeography 23:831-835.

Wu, R. 1999. Fire history and forest structure in the mixed conifer forests of southwest Colorado. M.S. Thesis, Colorado State University, Fort Collins, CO.

Yocum Kent, L. L., and P. Z. Fulé. 2015. Do rules of thumb measure up? Characteristics of fire-scarred trees and samples. Tree-Ring Research 71:78-82.

Table 1. Hagmann et al. (2021) Table 3 about historical tree density is replicated on the left, with our omitted rebuttals and other published evidence added on the right and highlighted with a dark border, to show Hagmann et al. omitted essential published evidence and made incorrect conclusions as a result.

Counter-evidence		**Evaluation of counter-evidence**		**Omitted rebuttals and other published evidence also essential to evaluation of counter-evidence**	
Citations	*Counter-premise*	*Citations*	*Implications of evaluation*	*Citations*	*Implication of omitted evidence*
Williams and Baker (2011) Baker and Williams (2018)	Novel methods provide estimates of tree density from point data, *i.e.,* General Land Office (GLO) records of bearing trees	Levine et al. (2017, 2019)	Multiple existing plotless density estimators (PDE) provided less biased estimates than the PDE developed by Williams and Baker (2011) which overestimated known tree densities by 24-667% in contemporary stands	Omitted Rebuttal evidence in Baker and Williams (2018)	Levine et al. (2017) incorrectly coded and applied the WB method, producing spurious results that had no bearing on the WB method.
				Omitted Rebuttal in Baker and Williams (2019), Omitted evidence in Williams and Baker (2011)	Levine et al. (2019) corrected their flawed 2017 code, but then here used incorrect equations. Baker and Williams (2019) used corrected equations with their code at their sites, and showed the WB method worked well. Williams and Baker (2011) had shown that Voronoi-based estimators work better than existing PDEs, and do not overestimate in western dry forests.
		Knight et al. (2020)	Methods supported by PDE sampling theory and multiple accuracy assessments further demonstrate the potential for misrepresentation of historical tree density by biased estimators used at resolutions substantially smaller than the minimum recommended for ~50% accuracy	Omitted evidence in Williams and Baker (2011)	Knight et al. (2020) did not use or test the WB method at all. They used old point-pattern measures that Cogbill et al. (2018) had already shown were inaccurate, require large samples, and underestimate. The WB method was designed to overcome these known limitations, and had already been validated (Williams and Baker 2011) to be able to accurately estimate tree density at the ~518 ha scale in western dry forests.

Williams and Baker (2012a)	Historical forests were denser than previously documented	Johnston et al. (2018)	Existing method for estimating tree density from point data (Morisita 1957, Warde and Petranka 1981) yielded densities less than half as large as estimates using Williams and Baker (2011) methods	Omitted Rebuttal in Baker and Williams (2019) Appendix S1	This study roughly estimated Voronoi-based tree density of 89.6 trees/ha for Johnston et al.'s sites, a modest error of 20% if a Johnston et al. estimate of 112 trees/ha is considered truth. This is within expected accuracy for the WB method (Williams and Baker 2011). Their estimate is not from a random sample and is too small to compare, as they did, with the mean for the whole study area, but is within one s.d. of the reconstructed historical mean (Williams and Baker 2012a), so is congruent with historical variability, as found in the reconstruction.
Williams and Baker (2012a) Baker (2012, 2014, 2015*a, b*)	Historical forests were denser than previously documented	Hagmann et al. (2013 2014, 2017, 2019), Collins et al. (2015), Stephens et al. (2015, 2018), Battaglia et al. (2018), Johnston et al. (2018)	Consistent with the finding that Williams and Baker (2011) methods overestimate tree density (Levine et al. 2017, 2019, Johnston et al. 2018, Knight et al. 2020) early timber inventory records and tree-ring reconstructions for the same study areas documented substantially lower tree densities than those estimated using Williams and Baker (2011) methods	New evidence here	See above for why Levine et al., Johnston et al., and Knight et al. do not show the WB method overestimates tree density. Regarding early timber inventories, see the last line below. Battaglia et al.'s (2018) study area was ~30 times ours, and was based on sampling 97% in logged forests. There has been no validation that the method they used can accurately reconstruct historical tree density in their region, and there especially has been no validation of their method in heavily logged forests where evidence likely has been destroyed by harvesting and associated activities.
Hanson and Odion (2016a)	Managing for dense, old forest and high-severity fire is consistent with historical conditions	Collins et al. (2016)	Fundamental errors compromise assertions about historical conditions including: (1) inappropriate use of coarse-scale habitat maps and (2) inaccurate assumption that areas lacking timber volume in early inventories indicate past high-severity fire	Omitted rebuttal by Hanson and Odion (2016b)	Collins et al. (2016) is not about tree density or forest density and did not belong in this table. However, Hanson and Odion showed that: (1) Collins et al. thought maps were wrong, but missed that areas that were forested by 1992, having recovered from early high-severity fires, had burned again, and (2) Collins et al. had omitted including essential 1911 field survey notes that directly described high-severity fires.

Odion et al. (2014), Baker (2015*a*, *b*) Baker and Hanson (2017)	Spatially extensive early timber inventories and bias in their use and interpretation misrepresent historical conditions	Stephens et al. (2015), Collins et al. (2016), Hagmann et al. (2017, 2018, 2019)	Fundamental errors compromise conclusions, including: (1) use of previously discredited methods (Williams and Baker 2011) to estimate tree density from GLO data as a baseline comparison; (2) incorrect assumptions about the methodological accuracy of early timber inventories; (3) inappropriate comparisons of studies of vastly different spatial scales, forest types, and diameter limits; (4) unsubstantiated assessment of bias in the locations of early timber inventories; and (5) unwarranted assumptions about vegetation patterns as indicators of fire severity	Omitted Rebuttal in Baker et al. (2018)	H et al. omitted our rebuttal (Baker et al. 2018), where we showed that Hagmann et al. (2018) did not contest Baker and Hanson's (2017) key findings: (1) early timber inventories documented to underestimate, are unreliable, and were abandoned by the 1930s, (2) comparisons between timber-inventory estimates and other sources showed it is timber-inventory estimates that underestimate and need correction, (3) one-chain wide inventories, if available data are used, could be fairly accurate, (4) omission of immature conifers and non-conifers may lead to additional underestimation. In response, we revised our estimates of needed correction multipliers to 1.6-2.3. Hagmann et al. (2018) still contended that inventories do not have biased placement, but we presented more evidence.

Table 2. Reported early timber inventory tree-density estimates and corrected estimates with 1.6-2.3 correction multipliers applied, along with estimated total tree-density (conifer + hardwood). Data are from studies that used early timber inventories to estimate historical tree density in dry forests.

Study area	**Source**	**Tree diameters recorded**	**Trees recorded**	**Reported tree density (trees/ha)**	**Corrected tree density (trees/ha)**	**Estimated conifer + hardwood tree density (trees/ha)**
Two-chain-wide timber inventories documented to underestimate tree density by 16-2.3 times (Baker et al. 2018)						
E. Oregon Cascades-N	Hagmann et al. (2014)	15.0 cm+	Main conifers	66	106-152[a]	106-152[a]
E. Oregon Cascades-S	Hagmann et al. (2013)	15.0 cm+	Main conifers	65	104-150[a]	106-152[a]
E. Oregon Cascades-S	Hagmann et al. (2017)	15.0 cm+	Main conifers	68	109-156[a]	109-156[a]
S. California Sierra	Collins et al. (2011)	15.2 cm+	Only conifers	44-52	70-120[a]	90-155[b]
S. California Sierra	Collins et al. (2015)	15.2 cm+	Only conifers	48	77-110[a]	99-142[b]
S. California Sierra	Scholl & Taylor (2010)	15.2 cm+	All trees	99	158-228[a]	158-228[a]
One-chain-wide timber inventory that is not known to underestimate tree density at this time (Baker et al. 2018)						
S. California Sierra	Stephens et al. (2015)	30.5 cm+	Only conifers	55	244[c]	498[d]

[a] Estimate is calculated, as in the text here, as 1.6-2.3 times "Reported tree density."
[b] Estimate is calculated from direct tallies of trees by species in the land-survey records for the southern Sierra, which found that a mean of 22.4% of total trees were oaks, thus conifer + hardwood tree density is estimated as corrected tree density/0.776.
[c] Stephens et al. (2015) was unique in omitting data for conifers < 30.5 cm dbh. Baker and Hanson (2017) redid the Stephens et al. inventory count of trees for their study area and found that for all conifers, tree density had a mean of 196-292 trees/ha for pine/ponderosa and mixed conifer, which are averaged here to be 244 trees/ha.
[d] Estimate is calculated by the recorded percentages of total trees in the land-surveys that were conifers and non-conifers in ponderosa pine (59.5%) and mixed-conifer forests (38.5%) in the area of the Stephens et al. inventory, which averaged together equals a fraction of 0.49. The corrected tree density is thus divided by 0.49 to estimate conifer + hardwood tree density. Note that 49% non-conifer trees is high, but not historically outside the historical range of variability in the southern Sierra overall, where the third quartile of oaks as a percentage of all trees begins at 34.9% (Baker 2014).

Table 3. Hagmann et al. (2021) Table 4 about rates of fire is replicated on the left, with unreviewed published evidence added on the right, highlighted by a dark border, to show Hagmann et al. omitted published evidence and made incorrect conclusions as a result.

Counter-evidence		**Evaluation of counter-evidence**		**Published evidence, essential to evaluation of counter-evidence, that was omitted**	
Citations	*Counter-premise*	*Citations*	*Implications of evaluation*	*Citations*	*Implication of omitted evidence*
Baker and Ehle (2001, 2003) Ehle and Baker (2003) Kou and Baker (2006*a, b*) Baker (2006, 2017a) Dugan and Baker (2014)	Tree-ring reconstructions misrepresent historical fire regimes by overestimating fire frequency and extent because (1) unrecorded fires (e.g., fires that did not scar trees) increase uncertainty of mean fire interval (MFI); (2) interval between pith (origin) and first fire scar should be considered a fire-free interval and included in calculations of MFI; (3) targeted sampling of high scar densities biases MFI; (4) mean point fire interval (mean of intervals between fire scars weighted by the number of fire scars) may more accurately represent historical fire rotation than MFI (mean between all fire scars)	Collins and Stephens (2007)	Unrecorded fires (fire did not scar the tree) may contribute to underestimation, not overestimation, of fire frequency and extent in frequent fire systems. Probability of scarring decreased when intervals between successive fires were short in areas burned by up to four late 20th-century fires. Absence of scar does not indicate absence of fire.	Omitted evidence in Baker (2017a S1 Text)	It is generally agreed that each fire only scars some of the trees. However, with typical scarring fractions, only ~50 trees or ~1 ha need sampling to detect all the fires in a plot. In a sample of 262 reconstruction sites in dry forests of the western USA, 88% sampled ≥1.0 ha. Thus, underestimation from unrecorded fires is likely rare, and absence of a scar in a particular year likely does show lack of fire. A key problem with "composite fire intervals," the primary source of evidence about historical low-severity fire rates, is the compositing process itself. Compositing makes a single list of all fire years in the plot. This assumes all fires burned the whole plot, which is not true, based on 11 studies (Baker and Ehle 2001). Putting small fires in a single list with large fires reduces the "mean composite fire interval" to a small value, leading to large overestimation of rates of fire. Small fires can be filtered, but filtering is arbitrary, and compositing still destroys the long intervals that were found.

		Brown and Wu (2005), Van Horne and Fulé (2006), Brown et al. (2008), Stephens et al. (2010), Yocum Kent and Fulé (2015), Meunier et al. (2019)	Including origin-to-first-scar interval erroneously inflates MFI. Not all trees that survive fire are scarred. As an ambiguous indicator of fire-free interval, it should not be included in calculations of MFI. Additionally, tree establishment may not indicate a stand-replacing disturbance in dry forests where regeneration is strongly associated with climate	Omitted evidence in Kou and Baker (2006a), Polakow and Dunne (1999), Moritz et al. (2009) Omitted evidence in Dugan and Baker (2015)	Fire-history data typically have incomplete intervals at the start and end of a period of record. Real but long fire intervals have more chance, than of appearing at the beginning or end, and getting left out, than do real but short intervals. Thus, censoring starting or ending incomplete intervals biases the record toward estimates that are too short and have reduced variability (Kou and Baker 2006a), as found in two other independent studies (Polakow and Dunne 1999, Moritz et al. 2009). There is no citation in the counter-evidence list that assumed tree establishment indicates stand-replacing disturbance in dry forests. Brown and Wu (2005) incorrectly assumed a fire scar before a pulse of tree establishment does not indicate moderate- to high-severity fire (Dugan and Baker 2015).
		Fulé et al. (2003) Van Horne and Fulé (2006) Farris et al. (2010, 2013) O'Connor et al. (2014)	Complete, systematic (gridded), and random sampling at stand, watershed, and mountain range scale have repeatedly demonstrated fire frequencies similar to those derived from targeted sampling within forest types and scales. In direct comparison studies, no evidence was found that targeted sampling of fire-scarred trees biased MFI estimates. Targeted sampling reconstructed fire parameters comparable to those derived from systematic sampling of both a subset of the trees and all trees in a study area and from independent 20th-century fire atlases	Omitted evidence in Baker (2017a S1 Text)	Evidence cited by H et al. in Farris et al. (2013) and in Van Horne and Fulé (2006) is not correct. Farris et al. (2013) instead found that using a targeted sample led to CFI estimates that were shorter (80-96%, comparing targeted and probabilistic sample size corrected in their Table 3) than that from a statistical sample. Van Horne and Fulé found that a targeted ITFI estimate was only 83% (inverse of 1.2 from p. 865) of ITFI from a random sample. These studies thus show that targeted samples produce CFI/ITFI estimates that are shorter than estimates from random samples.

		Farris et al. (2010) Huffman et al. (2015)	Rather than overestimating fire frequency as suggested in counter-premise papers, MFI may underestimate fire frequency, especially where small fires were abundant	Omitted evidence in Baker (2017a)	MFI as used by H et al. is just composite fire interval (CFI), which has the well established property of producing estimates that are too short relative to fire rotation, the gold standard, as shown in this monograph on this topic, which was omitted by H et al.
		Van Horne and Fulé (2006) Farris et al. (2013)	Composite mean fire intervals (CMFI, e.g., fires recorded on 25% of samples) are relatively stable across changes in sample area or size. See the section on “Underestimated historical fire frequency” for a more detailed summary of CMFI and the highly problematic and inherently biased alternatives proposed in counter-evidence publications	Omitted evidence in Baker (2017a)	CFI estimates do vary with sample size, but they also definitely produce estimates that are too short relative to fire rotation, the gold standard, as shown in this monograph on this topic, which was omitted by H et al.

Table 4. Hagmann et al. (2021) Table 5 about severity of historical fires is replicated on the left, with unreviewed published evidence added on the right, highlighted by a dark border, to show H et al. omitted published evidence and made incorrect conclusions as a result.

Counter-evidence		**Evaluation of counter-evidence**		**Omitted rebuttals and other published evidence also essential to evaluation of counter-evidence**	
Citations	*Counter-premise*	*Citations*	*Implications of evaluation*	*Citations*	*Implication of omitted evidence*
Shinneman and Baker (1997)	Based on early forest inventory age data sets, “nonequilibrium” areas of extensive, high-severity fires in the Black Hills led to landscapes dominated by dense, closed-canopy forests	Brown (2006)	Tree-ring reconstructions of ponderosa pine forest age structures and fire regimes across the Black Hills found synchronous regional tree recruitment largely in response to pluvials and longer intervals between surface fires, especially during the late 1700s/early 1800s which is when early inventory data report similar patterns of recruitment. No evidence of crown fires was found in relation to past fire dates	Omitted evidence in Brown (2006)	Brown (2006) said: “... dense stands were still present at settlement..., and likely contributed to extensive patches of crown fire noted by early explorers and scientists during the late 1800s (Dodge 1965, Graves 1899)” (p. 2509). This is entirely consistent with early reports of severe fires across the Black Hills by Graves and Dodge (Shinneman and Baker 1997)
Baker et al. (2007)	Most ponderosa pine forests in the Rocky Mountains were capable of supporting high-severity crown fires as well as low-severity surface fires	Brown et al. (2008)	Tree-ring reconstruction of ponderosa pine forests in the Black Hills of South Dakota (included in Baker et al. 2007) demonstrated that roughly 3.3% of the study area burned as crown fire between 1529 and 1893; however, tree density in most stands in 1870 could not have supported crown fire	Omitted evidence in Baker et al. (2007)	This 2007 review emphasized the historical fire regime in the Rocky Mountains included variable fire severities with some areas having mostly low severity and old-growth forests. The 517-ha Mount Rushmore area of Brown et al. (2008) is consistent with this review

Williams and Baker (2012a), Baker (2012, 2014)	Fire severity inferred from tree density by size class estimated from GLO bearing trees (Williams and Baker 2011) and surveyors' descriptions suggests low-severity fire dominated only a minority of ponderosa and mixed-conifer forests	Levine et al. (2017, 2019)	Plotless density estimator used by Williams and Baker (2011) overestimated known tree densities due to a scaling factor that does not correct for the number of trees sampled and therefore systematically underestimates the area per tree relationship	Omitted Rebuttals by Baker and Williams (2018, 2019)	Levine et al. (2017) incorrectly coded and applied the WB method, producing spurious results that had no bearing on the WB method. Levine et al. (2019) corrected their flawed 2017 code, but then used incorrect equations. Baker and Williams (2019) used correct equations with their code at their sites, and showed the WB method worked well, and both Levine et al. studies are fatally flawed.
		Fulé et al. (2014), Merschel et al. (2014), O'Connor et al. (2017)	Substantial errors of method and interpretation invalidate inferences about historical fire severity. These include: (1) tree size is an ambiguous indicator of tree age; (2) tree regeneration is an ambiguous indicator of disturbance severity, particularly in dry forests where climate conditions strongly influence regeneration; and (3) lack of direct documentary evidence (*e.g.,* primary observation) of extensive crown fire in historical ponderosa pine forests has been widely noted for nearly 90 yr.	Omitted Rebuttal in Williams and Baker (2014), Omitted evidence in Baker (2015a, 2017a)	Williams and Baker (2014) showed Fulé et al. mistook the WB method, misquoted WB, misused evidence, and created three new false narratives. Merschel et al. (2014) did not contest the WB method, but said there were no reports of late-1800s high-severity fires, even though extensively quoted in Baker (2012, 2014). O'Connor et al. has no bearing on the WB method. Extensive evidence of crown fires in historical ponderosa pine forests is widely published and reviewed in the text here. Baker (2017a) also showed, using tree-ring reconstructions, that low-severity fire was the primary severity across only ~34% of historical dry forests, mostly in the Southwest.

		Stephens et al. (2015), Huffman et al. (2015), Miller and Safford (2017), Hagmann et al. (2019)	Multi-proxy records documented substantially lower levels of high-severity fire in ponderosa and Jeffrey pine and mixed-conifer forests in overlapping study areas	Omitted evidence in Baker and Hanson (2017)	Baker and Hanson (2017) documented that Stephens et al. lower estimate is because they omitted timber-inventory documents that recorded high-severity fires. Huffman et al. does not overlap our study area. Miller and Safford repeated critiques we already refuted (see above). However, GLO reconstructions identify fires before the mining era, and their finding of few trees at low elevations today likely is due to loss of low-elevation forests. Hagmann et al. (2019) estimate of 6% high-severity similar to Baker (2012) 8.9% historically.
Baker (2012), Baker and Hanson (2017)	Estimates of area burned at high severity in Hessburg et al. (2007) validate estimates derived using Williams and Baker (2011) methods *Note: Baker and Hanson 2017 did not belong here, as it has nothing to do with the Hessburg et al. matter*	Hagmann et al. (2018), Spies et al. (2018)	Inappropriate comparisons are not validation. Baker (2012) limited assessment of high-severity fire to tree mortality in dry forests whereas Hessburg et al. (2007) estimated high-severity fire in the dominant cover type whether that be grass or tree for "moist and cold forest" type, with lesser amounts of dry forests	Omitted and incorrect evidence in Hessburg et al. (2007)	This argument is incorrect. Hessburg et al.'s Table 2 shows that specifically in forest cover types (not grass, shrub), their pooled forest percentages in ESR5 were 20.7% low, 55.0% moderate, and 24.3% high, which is even more similar to the Baker (2012) estimates of 18.1% low, 59.9% moderate, and 23.0% high. Hessburg et al. Figure 4 also shows that ponderosa and Douglas-fir cover types had a mean of about 18% low, 59% moderate and 23% high, almost identical to the Baker (2012) estimates.

Odion et al. (2014)	Modern, high-severity crown-fires are within historical range of variation. Inferred fire severity from current tree-age data for unmanaged forests in the U.S. Forest Service Inventory and Analysis (FIA) program. Compared inferences about modern fire severity to estimates of historical forest conditions and fire severity inferred using Williams and Baker (2011) methods	Fulé et al. (2014), Levine et al. (2017, 2019) Knight et al. (2020)	Overestimation of historical tree density and unsupported inferences of fire severity from GLO records weaken conclusions based on Williams and Baker (2011) methods	Omitted Rebuttals in Williams and Baker (2014), Baker and Williams (2018, 2019)	Fulé et al. (2014) mistook the WB method, misquoted publications, misused evidence, and created three new false narratives. Levine et al. incorrectly coded the WB method (2017), then used incorrect equations (2019), and both are fatally flawed. Knight et al. did not use or test the WB method and has no relevance.
		Stevens et al. (2016)	Substantial errors of method and interpretation invalidate inferences about historical fire severity. These include: (1) FIA stand age variable does not reflect the large range of individual tree ages in the FIA plots and (2) recruitment events are not necessarily related to high-severity fire occurrence	Omitted rebuttal in Odion et al. (2016)	With same definition of high-severity fire, there was 68% agreement between these two studies; 3/4 of evidence of historical high-severity fire not from FIA data & not disputed; Stevens et al. agreed "High-severity fire was undoubtedly a component of fire regimes in ponderosa pine and drier mixed-conifer forests"
		Spies et al. (2018)	In contradiction of the counter-premise, Odion et al. documented only three patches of high-severity fire larger than >1000 ha in OR and WA in the early 1900s, which account for 1% of the area of historical low-severity fire regime managed under the Northwest Forest Plan	Omitted rebuttal in Odion et al. (2016); Omitted evidence in Della-sala and Hanson (2015, 2019)	Two sources in omitted 2015 paper, reviewed in the omitted Odion et al. (2016) paper, found high-severity patches ≥ 14,000 ha in OR & WA, two others found many large patches in OR & WA; Numerous other large patches > 1000 ha reported in OR & WA in omitted 2019 paper.

Baker and Hanson (2017)	Stephens et al. (2015) underrepresented the historical extent of high-severity fire in their interpretation of surveyor notes in early timber inventory. *Note: because they omitted key records of high-severity fire that were readily available in the inventory records.*	Hagmann et al. (2018)	Substantial errors of method and interpretation invalidate inferences about the historical extent of high-severity fire. Inferences were based on (1) inappropriate assumptions about the size and abundance of small trees given the ambiguity of data describing small trees in the 1911 inventory, (2) averaging of values derived from different areas and vegetation classifications, and (3) inappropriate assumptions that the presence of chaparral (common on sites with thin soils and high solar radiation) indicates high-severity fire	Omitted rebuttal in Baker et al. (2018)	Hagmann et al. (2018) did not dispute that Stephens et al. (2015) had omitted most trees, and when omitted trees were included, forests were 7-17 times as dense as they reported, and they also did not dispute the abundant data, from numerous historical sources, showing occurrence of substantial high-severity fire patches, small and large, including in chaparral, presented in Baker and Hanson (2017)

Table 5. Hagmann et al. (2021) Table 6 about severity of modern fires is replicated on the left, with unreviewed published evidence added on the right, highlighted by a dark border, to show Hagmann et al. omitted published evidence and made incorrect conclusions as a result.

Counter-evidence		**Evaluation of counter-evidence**		**Omitted rebuttals and other published evidence also essential to evaluation of counter-evidence**	
Citations	*Counter-premise*	*Citations*	*Implications of evaluation*	*Citations*	*Implication of omitted evidence*
Odion and Hanson (2006)	High-severity fire was rare in recent fires in the Sierra Nevada based on analysis of Burned Area Emergency Response (BAER) soil burn severity maps	Safford et al. (2008)	BAER maps greatly underestimate stand-replacing fire area and heterogeneity in burn severity for vegetation. BAER maps are soil burn-severity maps, not vegetation burn-severity maps.	Omitted rebuttal in Odion and Hanson (2008)	Safford et al. arbitrarily combined two time-since-fire categories, creating slightly higher fire severity in long unburned forests. Odion and Hanson, used all categories and found similar or lower fire severity in long unburned forests.
Hanson et al. (2009)	Change in conservation strategies for northern spotted owl (NSO) were unwarranted due to overestimation of high-severity fire in the NSO recovery plan	Spies et al. (2010)	Use of a higher relative delta normalized burn ratio (RdNBR) threshold substantially increased misclassification errors and reduced estimates of high-severity fire extent. Hanson et al. (2009) used an RdNBR threshold of 798 rather than 574 as recommended in the literature (Miller et al. 2009) they cited as the source of the threshold used	Omitted rebuttal in Hanson et al. (2010)	Spies et al. had cited evidence with a math error and incorrect anecdotal evidence to conclude high-severity fire was outpacing old forest recruitment, but did not dispute these, then tried a broader high-severity fire definition. Hanson et al., however, showed this new definition still led to old forest recruitment outpacing high-severity fire by 7-29 times

Williams and Baker (2012a)	Severity distributions in recent fires do not depart from historical	Steel et al. (2015), Guiternam et al. (2015), Reilly et al. (2017), Steel et al. (2018)	Extent and spatial patterns of fire severity in some recent fires have departed from pre-fire exclusion range of variation for some forest types	Omitted evidence in Odion et al. (2010), Hanson and Odion (2015), Della-Sala and Hanson (2019), and many others (see text)	Steel et al. (2015) based historical high-severity proportions on only a theoretical model. Guiterman et al. was from only one 38-ha patch, with little inferential power. Reilly et al. found no trend in high-severity proportion, but more large, high-severity patches, but H et al. omitted DellaSala and Hanson (2019) who found no such increase over the last two decades. Steel et al. (2018) used a database that Hanson and Odion (2015) showed can produce false trends.
Hanson and Odion (2014)	Previous assessments overestimate extent of high-severity fire in modern fires	Safford et al. (2015)	Use of coarse-scale, highly inaccurate and geographically misregistered vegetation map and averaging across unrelated vegetation types and diverse ownerships undermine confidence in Hanson and Odion (2014)	Omitted rebuttal in Hanson and Odion (2015)	Hanson and Odion re-analyzed Safford et al.'s initial data, using new methods that Safford et al. proposed, and found Hanson and Odion's initial conclusions were robust to re-analysis using Safford et al.'s proposed new methods.

Table 6. Ten published rebuttals omitted by H et al., and the sections and tables containing details of the omitted evidence, which refuted the rebutted articles and H et al.'s conclusions.

Omitted rebuttal	**Article rebutted**	**Section(s)/Table(s)**
Baker and Williams (2018)	Levine et al. (2017)	A3a, Tables 1, 4
Baker and Williams (2019)	Levine et al. (2019)	A3a, Tables 1, 4
Baker and Williams (2019)	Johnston et al. (2018)	A3c, Table 1
Baker et al. (2018)	Hagmann et al. (2018)	A3d, B2d, Table 4
Hanson and Odion (2016b)	Collins et al. (2016)	A3e, Table 1
Williams and Baker (2014)	Fulé et al. (2014)	B2c, Table 4
Odion et al. (2016)	Stevens et al. (2016)	B2i, B2j, Table 4
Odion and Hanson (2008)	Safford et al. (2008)	B3, Table 5
Hanson et al. (2010)	Spies et al. (2010)	B3, Table 5
Hanson and Odion (2015)	Safford et al. (2015)	B3, Table 5

Table 7. Twenty-five published original publications, with evidence of historically heterogeneous forest structure and mixed- to high-severity fires, omitted by H et al.

Omitted evidence in these sources	**Evidence omitted by H et al.**
Williams and Baker (2010)	Omitted all evidence showing low bias and error in land-survey records
Williams and Baker (2011)	Omitted all evidence of validations of the WB method
Williams and Baker (2012a)	Omitted all evidence of validations of the WB method and evidence of historically variable tree density and fire severity in dry forests in Arizona, Colorado, and Oregon.
Williams and Baker (2012b)	Omitted all direct evidence of extensive moderate- to high-severity fire in historical dry forests in the Colorado Front Range, evidence validating the WB method of reconstructing historical moderate- to high-severity fires, and evidence of very large high-severity fire patches (up to 8,331 ha).
Baker and Williams (2018)	Omitted all evidence of validations of the WB method and all evidence of historically variable tree density and fire severity documented in multiple historical sources cited in this paper.
Baker et al. (2007)	Omitted all evidence from tree-ring reconstructions, forest-reserve reports, and other early scientific reports that historical dry forests in the Rocky Mountains had tree densities varying from 17-19,760 trees/ha.
Baker (2012)	Omitted quotes from early forest-reserve reports and other early scientific reports that historical dry forests in the eastern Cascades of Oregon had variable tree density and many direct reports of moderate- to high-severity fire.
Baker (2014)	Omitted 47 quotes from early forest-reserve reports and other early scientific reports documenting that Sierran mixed-conifer forests were highly variable in tree density, but typically dense, and omitted numerous early reports of extensive moderate- to high-severity fire in historical Sierran mixed-conifer forests. Omitted 208 quotes from early forest-reserve reports and other early scientific reports that documented historical moderate- to high-severity fires in Sierran mixed-conifer forests. Omitted evidence of high-severity fire patches commonly > 250 ha and up to 9,400 ha in area.
Baker (2017a)	Omitted all evidence in this monograph analyzing why old CFI-based estimates of historical rates of fire are too short, why moderate- to high-severity fires were seldom found using these old methods, and how these old estimates can be corrected to accurately estimate fire history.
Farris et al. 2010, Dugan and Baker 2015.	Omitted any mention of the development of new methods of conducting fire history studies that overcome the limitations of earlier CFI-based fire-history studies that H et al. cite.

Hessburg et al. (2007)	Omitted evidence of severe fires in northwestern dry forests; even though Paul Hessburg is an author of H et al, and also authored this publication, H et al. did not review its evidence. Hessburg et al. studied 303,156 ha in E. OR and E. WA and found "widespread evidence of partial stand and stand-replacing fire" (p. 5) in mixed conifer forests
Baker (2009)	Omitted evidence in six early photographs of the aftermath of severe fires in dry forests in the Rocky Mountains.
Baker (2017b, 2018, 2020)	Omitted evidence that documented that large late-1800s moderate- to high-severity fires occurred in dry forests on the Uncompahgre Plateau and in the San Juan Mountains, Colorado, based on forest-atlases, land-survey records, early photographs, early scientific publications, and other early records, including newspaper reports.
Pierce et al. (2004), Pierce and Meyer (2008), Colombaroli and Gavin (2010), Jenkins et al., (2011, Long et al. (2011), Bigio (2013), Fitch (2013)	Omitted evidence in these paleo-charcoal studies from Arizona, Colorado, Idaho, New Mexico, and Oregon that infrequent moderate- to high-severity fires occurred historically in western USA dry forests.
DellaSala and Hanson (2015, 2019)	Omitted evidence of numerous large historical high-severity fire patches in OR, WA, CA, and other parts of the western USA