Ecological Applications, 29(8), 2019, e02017

Estimating historical forest density from land-survey data: Response

Citation: Baker, W. L., and M. A. Williams. 2019. Estimating historical forest density from land-survey data: Response. Ecological Applications 29(8):e02017. 10.1002/eap.2017

Introduction

Levine et al. (2019; LE hereafter) claim our landscape-scale method for reconstructing historical forests fails, based on updated modern tests in six disjunct small plots, half located in highly altered forests, but here we show our method works even at these inadequate sites. LE is a correction of Levine et al. (2017; LE2017 hereafter). In Baker and Williams (2018; B&W hereafter), we showed that LE2017 did not correctly test our method in these same small plots.

LE admit and correct flaws in LE2017, compelling evidence that LE2017 needs retraction. In LE, new code is used that effectively repudiates all code and findings of LE2017, which had implied in its title, methods, and conclusions that it accurately tested our method (Williams and Baker 2011). However, we showed in B&W that LE2017 did not replicate our method, and was thus an invalid test of it and its application in the Sierra (Baker 2014). LE2017 and its Erratum (LE Appendix S2) deviated so much from our method that they effectively created and tested their own, baseless, Voronoi-based estimator that (1) derived estimates primarily from diameter breast height (DBH; 1.3 m from the ground) tree data, not diameter at stump height (DSH; 30 cm from the ground) data usually recorded in GLO surveys (Williams and Baker 2010), (2) calculated a key measure, mean neighborhood distance (MND) at 21 fixed points, instead of at each corner using its witness trees, (3) analyzed section-corner and quarter-corner data separately, instead of in essential mixed six-corner pools, and (4) averaged values using the arithmetic, not the also-essential harmonic mean. LE reports their new code remedies these differences. However, LE still calculate DSH as about 1.13 × DBH (LE Appendix S2), which still underestimates basal swell (Biging 1984). Our retesting here used 1.23 and led to lower error (Appendix S1: Section S1, Table S1), which LE also demonstrated at YOSE.

Historical tree density, the focus of this exchange, is a key property of forest structure that historically shaped habitat and vulnerability to natural disturbances, yet forest managers setting thinning goals often lack sufficient data on landscape-scale variability in historical tree density. General Land Office (GLO) witness-tree data and our Voronoi-based method (Williams and Baker 2011) can provide historical tree-density reconstructions across very large landscapes (~ 1.9 million ha in B&W). The Voronoi-based method was cross-validated with independent historical sources and in a large modern accuracy test and shown to be quite accurate (B&W).

Here we evaluate LE's updated, but still incorrect, tests. We show our method is highly accurate in much larger modern and historical tests. We show that LE's small modern tests used our equations incorrectly in altered and disjunct stands where these equations are not valid. When we used corrected equations, overestimation disappeared. Our methods thus remain accurate in both key historical tests and these modern tests, new evidence that LE2017 needs full retraction.

Modern Accuracy Testing Much Less Important Than Historical Reconstruction Accuracy

LE appear to infer that their modern testing shows that our method produces historical tree-density estimates (Baker 2014) that are too high for Sierran mixed-conifer forests. LE use three highly altered Sierra stands and three disjunct stands far outside the Baker study area (B&W). Their introduction mentions using GLO data and our method to reconstruct historical forests. Yet, LE is entirely focused in modern forests, does no testing of the ability of our method to reconstruct historical forests, and does not mention or review previous extensive testing of our method in reconstructing historical forests (B&W). LE's testing thus only allows inference about how well our derived equations work in modern forests. However, modern forests frequently have legacies of past land uses (e.g., logging, fire exclusion) that limit their relevance to historical forests. This requires focusing modern testing in less-altered forests, but LE's plots were half in highly altered Sierra forests and half in distant areas not calibrated by Baker (2014). LE's conclusion: "Based on this new analysis of data…MHVD overestimates tree density" lacks the key caveat: "in three highly altered modern stands and three disjunct stands far outside the Baker (2014) study area."

Researchers must also place their work in the context of previous research, but LE claimed that their tests were "robust and powerful," without reviewing past evidence. LE did not even mention or review our much larger modern tests, that had already shown our method had high accuracy (B&W). Williams and Baker (2011) completed large modern tests at 499 corners in three states (Arizona, Colorado, and Oregon, USA) for reconstructions that use equations fitted across all tree species. That testing, unlike LE's, repeated the historical GLO survey method in modern forests across contiguous sets of 21 corners representing 10,878 ha in each state. In the eastern Cascades, Baker (2012) used species-specific equations, and a small modern trial there (15 corners) showed that these equations reduced errors a few percent. In the case of LE's tests, six disjunct 4-ha plots between the Sierra and Baja California cannot provide robust and powerful testing of any method. At most, LE's sample of six corners represents about 1% of 520 total corners (514 in our studies, 6 in LE) where modern testing has occurred. If LE's modern tests correctly found a problem, which we refute, it is confined to a trivial 1% of tests.

Tests of GLO methods in modern forests can be useful, but cross-validation of GLO reconstructions with independent historical records and other reconstructions is really the acid test, as this is the key goal of GLO reconstructions. For 18 total specific historical cross-validations, where an overlay of the GLO reconstruction polygon(s) on the location of an independent historical inventory or other reconstruction was possible, accuracy was high in tests across study areas in Arizona, California, and Oregon (B&W). Mean RMAE (relative mean absolute error, the absolute difference between the GLO estimate and the other historical estimate as a percentage of it) was a low 10.4–11.2% (B&W Appendix S1: Table S4). Eleven of these cross-validations were in the Sierra, where mean RMAE was also a low 10.3%. LE dismiss the need to review this evidence, claiming they have a "robust and powerful design." LE did not review this available evidence that shows that our method works well in these key historical tests. LE mention a new study (Johnston et al. 2018) they think shows our GLO methods are "unreliable," but it is more likely the studies closely agree (Appendix S1: Section S2). LE's study is neither robust nor powerful in testing the ability of our method to reconstruct historical forests, its primary goal.

Here we add new evidence that further refutes LE's implied inference that their modern testing means that our method overestimates historical forest density. As a general historical cross-validation, we pooled and analyzed the distribution of all 30 available independent estimates of historical tree density (B&W Appendix S1: Tables S4, S9). This pooled sample is likely not strictly random, as sources had diverse purposes, but it spans many locations and is likely the best non-GLO sample of historical tree density currently available for Sierran mixed-conifer forests. The distribution (Table 1) shows these forests were generally dense (mean = 257 trees/ha, median = 247 trees/ha), but variable in density, with an SD of 110 trees/ha and a range from 88–684 trees/ha. Corroborating this wide variability are 13 early quotes describing dense forests, 17 describing open forests, and 12 describing scattered trees on rocky sites (Baker 2014, Appendix A: Table A1). Quartiles show one-quarter of estimates were ≥294 trees/ha (Table 1), which indicates very dense forests, although less dense than in many modern forests. In comparison, quartiles of the distribution of 548 GLO estimates (Table 1) of historical tree density in the Sierra (Baker 2014, Table 4) are similar. These two distributions have overlapping 95% confidence intervals for their means (Table 1). If independent estimates were the truth, the 293 trees/ha mean of the GLO distribution has an RMAE of only 14% relative to the 257 trees/ha true mean, and a similar RMAE for medians is only 17%, both indicating high accuracy for the GLO estimates. LE said: "there is no need to refer to historical evidence." However, this new historical validation and our published historical validations (B&W) provide much stronger evidence, than modern testing, of the high accuracy of our method for its key purpose, to reconstruct historical tree density. LE's overestimation finding is thus confined to modern testing and 1% of modern tests.

Table 1. Comparison of attributes of the distribution of estimated historical tree density between independent estimates and GLO reconstructions in Sierran mixed-conifer forests.

Attribute	Independent estimates (trees/ha)†	GLO reconstructions (trees/ha)‡
N	30	548
Mean	257	293
SD	100	477
Minimum	88	47
First quartile	190	139
Median	247	206
Third quartile	294	312
Maximum	684	9147
95% CI for mean	216–298	253–333

† Estimates are from 6 tree-ring reconstructions, 22 early non-timber inventories, and 2 one-chain-wide early timber inventories in or near the Baker (2014) study area (Baker and Williams 2018: Appendix S1: Tables S4, S9).

‡ Estimates are for the 548 GLO reconstructions pooled across the northern and southern Sierra in Baker (2014) with these attributes taken from Table 4, except the 95% CI for the mean, which was estimated from *N*, the mean, and the SD.

LE Modern Test Insufficient in Scale, Pseudo-Replicated, and a Biased Sample

LE's modern test is insufficient in scale. The 4-ha plots in LE are at the same scale as the area covered by one GLO section corner, which is about 1.1 ha based on a typical 300-link (60.3 m) surveyor search radius. LE used 100 samples with a micro-pooling method in each 4-ha plot, but there is only one distinct tree-density estimate. Moreover, the 100 samples were not independent, but pseudo-replicated, as more than two-thirds of individual trees were used more than once (Appendix S1: Section S3, Table S2). Essentially, their 100 samples just refine the error for a single density estimate for one small stand of trees. Even if pseudoreplication had been avoided, our modern study showed that at the one-corner level, mean harmonic Voronoi density (MHVD) had poor accuracy, with a mean RMAE of 62.9% (Williams and Baker 2011). Six-corner pools across a 518-ha landscape are needed to estimate tree density with high accuracy (Williams and Baker 2011). The three Sierran LE plots together would represent only half of one six-corner pool. GLO data from corners are spaced about 800 m; density and other attributes often vary substantially across a landscape, but can be more uniform in a single 4-ha plot. The Williams and Baker (2011) method is calibrated to be a landscape-scale method, not for estimating accurately in small (4 ha) stands. LE's three Sierran plots are also a biased sample of Sierran mixed-conifer forests; all are in the white-fir phase of these forests, which covered just 20% of the Baker (2014) study area. LE is insufficient in scale and mostly irrelevant.

Altered Modern Forests are Inappropriate Test Sites, But Our Method Still Works Well

As explained in the *Introduction*, the Williams and Baker (2011) Voronoi-based method requires collecting calibration data on the relationships of tree DSH, crown radius, MND, and the tree's Voronoi area from a diversity of modern natural forests most likely to be similar to historical forests (Williams and Baker 2011). In the Sierra (Baker 2014), we specifically avoided collecting calibration data in logged or other unnatural forests unlikely to be similar to forests in the historical era. In contrast, LE chose "three Sierra Nevada stands that have been subject to 100+ years of fire exclusion" to test the Williams and Baker (2011) method. There was little to no fire exclusion in historical forests, thus it is fully expected that our GLO equations, calibrated to reconstruct historical forests, might not work well in the highly altered LE test stands.

Relative to historical forests in the Baker (2014) study area (Fig. 1a), the three LE stands (Fig. 1b–d) do appear highly altered, just as LE described them. The three LE stands have very abundant small trees (<20 cm DSH), particularly white firs (*Abies concolor*), which are 80% of trees <20 cm at YOSE, 75% at TEAK, and 61% at PLUM. Small white firs were a much smaller percentage of all white firs in historical forests (Baker 2014: Fig. 3a). White firs, in general, were a small percentage of all trees in historical forests (Baker 2014: Fig. 3a vs. Fig. 3h).

To test whether LE's findings of overestimation also resulted from them incorrectly using equations not calibrated for use in these highly altered modern forests, we

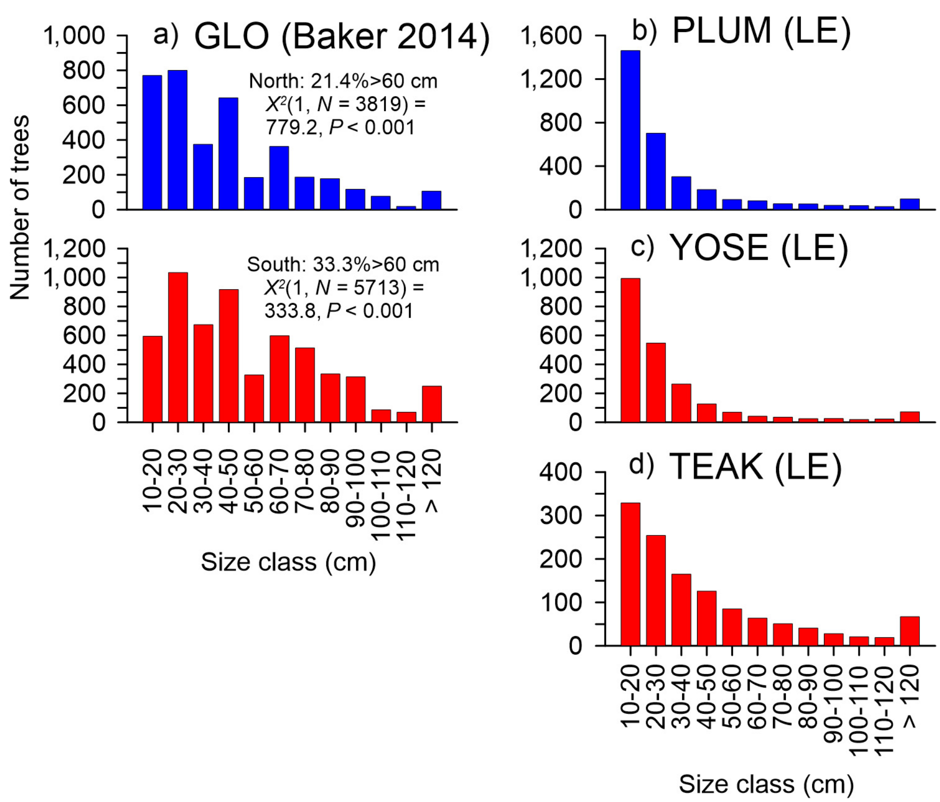

Fig. 1. Diameter distributions (diameter at stump height, DSH) of all trees in the Baker (2014) GLO study area, and at the three Sierra sites used by Levine et al. (2019), (a) reproduced from Fig. 3h in Baker (2014), (b–d) using the data sets from Levine et al. (2019), but with DBH multiplied by 1.23 to approximate DSH, as in Baker and Williams (2018).

refit the Voronoi equations for use in the three stands that overlap with Baker (2014). We ran the LE code to get estimates of crown radius, MND, and actual Voronoi area for trees in 100 runs in PLUM, TEAK, and YOSE. We then collected a stratified random sample of 20 trees from each stand (n = 60 total), using five DSH-class strata (<25 cm, 25–50 cm, 50–75 cm, 75–100 cm, and >100 cm) to ensure coverage of tree sizes. We refit a general Voronoi equation for 59 trees, after removing an outlier ($P < 0.001$, $R^2_{adj} = 0.456$, slope = 0.405, intercept = 1.635), placed these new coefficients in the LE code, and reran it for the three stands. This refitted equation had mean RMAE = 13.0% and RE = 0.92 for the three stands (Table 2). Using this correct refitted equation, LE's reported overestimation disappeared. Overestimation is thus shown to also be an artifact of inappropriate use of the Sierra Voronoi equation (Baker 2014), which was not calibrated for use in highly altered forests like those where LE did their testing. The Williams and Baker (2011) method remains fundamentally sound and accurate, if appropriately fitted equations are used.

Incorrect Equation Also Used in LE Tests in Outside Plots

LE's test also used three plots outside the Baker (2014) study area in physically different settings. Different physical settings mean the relationships of tree DSH, crown radius, MND, and Voronoi area likely also differ. The use of the Sierra Voronoi equation, and possibly the Sierra crown-radius equation, is inappropriate. The only appropriate Voronoi equation, which was fitted across large geographical areas, is the original general equation for mean-based Voronoi area (Williams and Baker 2011; Table 3). There are no available crown-radius equations fitted across large geographical areas, forcing a choice among available crown-radius equations.

To test that the original general Voronoi equation is the correct equation to use for the three stands (BRID, GRAN, META) outside the Baker (2014) study area, we simply placed the equation coefficients for this general Voronoi equation into LE's code. We did the same with two possibilities for the CR equation: Sierran (Baker 2014) and Mogollon (Williams and Baker 2011), since these are the closest areas with GLO reconstructions. When the LE code was rerun with these equation coefficients, the original Voronoi equation and Sierra crown–radius equation reduced RMAE and RE (different but related measures of error) to low levels, with a mean RMAE of only 15.6% and RE of 0.84 (Table 3), showing that these equations work well at all three sites outside the Baker (2014) study area. Even higher accuracy, with RMAE averaging only 9.2% and RE only 1.09, came from using the original Voronoi equation with the crown–radius equation from the closer Mogollon study area (Table 3). Overestimation again was gone, showing that LE's tests failed because inappropriate equations were also used at these three sites.

Poorly Predicted Voronoi Areas Also Due to LE's Use of Incorrect Equations

LE found that estimated Voronoi areas underpredicted true Voronoi areas in all six stands (LE: Fig. 2), but this is an artifact of using incorrect equations. When correct equations are rerun in all six stands, the fit between modeled and actual Voronoi area, shown in LE to be poor and not aligned with a 1:1 line, has good fit and is aligned with a 1:1 line (Fig. 2). The

Table 2. Comparative accuracy of estimated tree density using the Levine et al. (2019) code with the Sierra (Baker 2014) and refitted Voronoi equations at the three Levine et al. (2019) sites inside the Baker (2014) study area.

	Equations		Tree density (trees/ha)		Accuracy	
Site	Crown radius	Voronoi	Plot value†	Median MHVD‡	Relative mean absolute error (%)§	Relative error¶
PLUM	Sierra	Sierra	784	1,416	80.6	1.81
TEAK	Sierra	Sierra	313	568	81.5	1.81
YOSE	Sierra	Sierra	562	1,149	104.4	2.04
Mean					88.8	1.89
PLUM	Sierra	Refitted	784	581	25.9	0.74
TEAK	Sierra	Refitted	313	335	7.0	1.07
YOSE	Sierra	Refitted	562	527	6.2	0.94
Mean					13.0	0.92

† The plot value is from Levine et al. (2019).
‡ The median mean harmonic Voronoi density (MHVD) estimate is the median of the MHVD outputs for 100 runs of the Levine et al. (2019) code.
§ The absolute value of the difference between the median MHVD estimate and the plot value as a percentage of the plot value. Lower values indicate lower error, hence greater accuracy.
¶ The ratio of the median MHVD estimate divided by the plot value, where a perfect match has a ratio of 1.0. Note that the values reported here are lower than in Levine et al. (2019: Appendix S1: Table S1) because we used DSH from multiplying DBH × 1.23.

TABLE 3. Comparative accuracy of estimated tree density at the three Levine et al. (2019) sites outside the Baker (2014) study area.

Site	Equations: Crown radius†	Equations: Voronoi‡	Tree density (trees/ha): Plot value§	Tree density (trees/ha): Median MHVD estimate¶	Error: Relative mean absolute error#	Error: Relative error‖
BRID	Sierra	Sierra	236	539	128.4	2.28
GRAN	Sierra	Sierra	159	318	100.0	2.00
META	Sierra	Sierra	254	869	242.1	3.42
Mean					156.8	2.57
BRID	Sierra	Original	236	191	19.1	0.81
GRAN	Sierra	Original	159	143	10.1	0.90
META	Sierra	Original	254	209	17.7	0.82
Mean					15.6	0.84
BRID	Mogollon	Original	236	255	8.1	1.08
GRAN	Mogollon	Original	159	175	10.0	1.10
META	Mogollon	Original	254	278	9.4	1.09
Mean					9.2	1.09

† Sierra values are species-specific, the Mogollon used one equation for all species.
‡ Sierra values are species-specific, the original Voronoi equation is for all species.
§ The plot value is from Levine et al. (2019).
¶ The median MHVD estimate is the median of the MHVD outputs for 100 runs of the Levine et al. (2019) code.
# The absolute value of the difference between the median MHVD estimate and the plot value as a percentage of the plot value. Lower values indicate lower error, hence greater accuracy.
‖ The ratio of the median MHVD estimate divided by the plot value; a perfect match has a ratio of 1.0

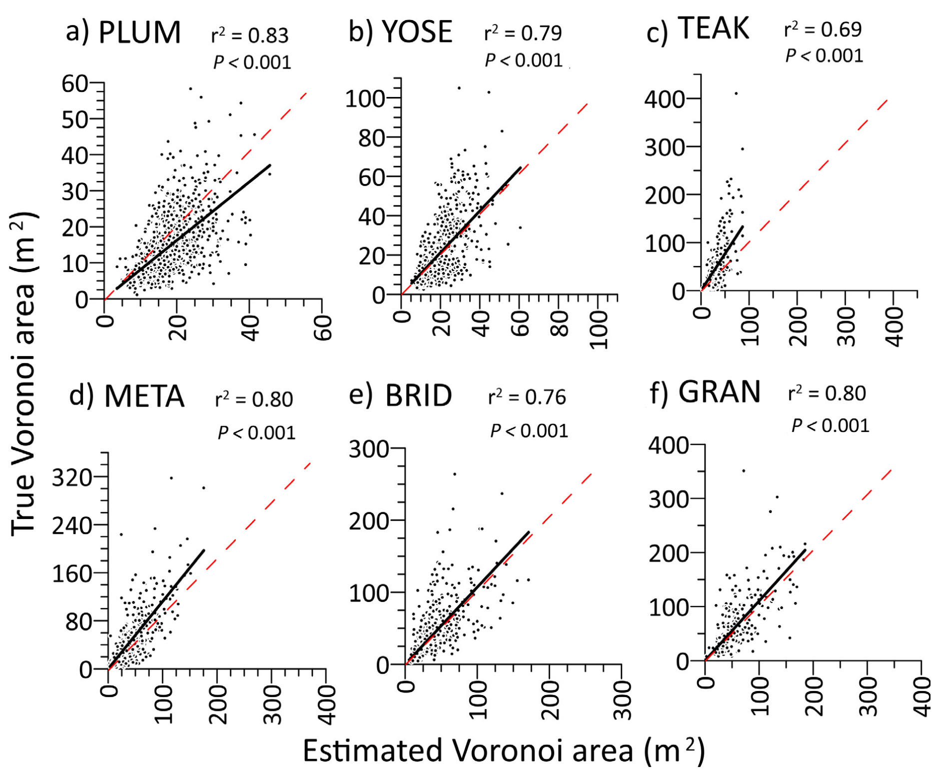

FIG. 2. The relationship between estimated and true Voronoi area for the witness trees in each of the six stands from Levine et al. (2019) using (a–c) the Sierra crown-radius equation and the refitted Voronoi equation for the three stands inside the Baker (2014) study area and (d–f) the Mogollon crown-radius equation and the original Voronoi equation for the three stands outside the Baker (2014) study area. The solid black line is the fitted regression equation and the dashed red line is the 1:1 line. See Levine et al. 2019: Fig. 2 for a comparable figure using the Sierra Voronoi equation incorrectly in all six stands.

Williams and Baker (2011) method, used in Baker (2014), does not overestimate historical tree density, nor does modeled Voronoi area underestimate true Voronoi area even in highly altered modern forests and forests outside the Baker (2014) study area, if correct equations are used.

Conclusion

Here we showed the errors and limitations in the LE critique of the Williams and Baker (2011) method and its application in the western Sierra (Baker 2014). LE and LE2017 rejected our method without testing it correctly, and without acknowledging that a landscape method cannot be adequately tested in small, disjunct plots, or that prior historical and modern tests were strong evidence that a fundamental flaw in the method was unlikely. LE tested equations used in the Sierra (Baker 2014), that were calibrated for use in historical forests, not in highly altered modern forests. GLO methods must be calibrated to particular forests as the allometric relationships of size, tree crown area, and Voronoi area vary with local conditions (Mark and Esler 1970, Williams and Baker 2011). In LE's three highly altered Sierran stands, high accuracy was achieved after we refit the Voronoi equation for these altered stands. Also, high accuracy of our original Voronoi equation in LE's three outside stands, where this original equation is the only appropriate equation, further shows the method is sound. LE's finding of overestimation is an artifact of ignoring past testing and of using an inadequate small, biased sample with incorrect equations. Williams and Baker (2011) and Baker (2014) remain sound.

We agree that testing and validation of methods remains important and could be improved for all methods of reconstructing historical landscapes. Testing landscape methods in small areas is insufficient, but could possibly help improve understanding of the sensitivity of some parts of methods to variations in parameters. For example, the evidence in this exchange demonstrates that crown-radius and Voronoi equations, as well as DSH data, are important to the method.

Overall, our Voronoi-based method remains well validated across western dry forests (Williams and Baker 2011) and in the Sierra study area (Baker 2014) in its ability to reconstruct historical tree density with high accuracy. GLO-based tree-density estimates in Baker (2014) are shown here to be supported by low error ($\sim$14%) in cross-validation with 30 independent estimates of historical tree density, 42 early quotes, and by previous specific and general cross-validations (B&W). There is no other similarly validated method that has provided accurate and abundant tree-density reconstructions for historical dry forests across very large landscapes.

We suspect this exchange with LE is part of an ongoing debate about evidence regarding tree density in historical dry forests of the West: were these forests mostly open or mostly dense, or was there great variability? This debate bears on whether widespread forest thinning is warranted and how much to do, whether modern fires are unnatural, and related issues, potentially affecting land management across large land areas. Natural variability across large historical landscapes likely was central to perpetuating biological diversity and ecosystem functioning (DellaSala and Hanson 2015), thus it is important to have accurate reconstructions across historical landscapes.

Validated GLO reconstructions across large historical landscapes can help guide modern landscape management. Modern forests, like those at YOSE, PLUM, and TEAK, may now be denser than historical forests, but historical Sierran mixed-conifer forests were generally dense and spanned a large range in densities from very open to very dense, as shown previously by reconstructions (Baker 2014), cross-validation of GLO data (B&W), and new analysis here. This body of jointly cross-validated evidence together provides the best available data on natural variability in historical tree density in Sierran mixed-conifer forests. This body of evidence is sufficient to guide modern restoration and management programs, including tree-thinning goals.

Literature Cited

Baker, W. L. 2012. Implications of spatially extensive historical data from surveys for restoring dry forests of Oregon's eastern Cascades. Ecosphere 3:Article 23.

Baker, W. L. 2014. Historical forest structure and fire in Sierran mixed-conifer forests reconstructed from General Land Office survey data. Ecosphere 5:Article 79.

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.

Biging, G. S. 1984. Taper equations for second-growth mixed conifers of northern California. Forest Science 30:1103–1117.

DellaSala, D. A., and C. T. Hanson, editors. 2015. The ecological importance of mixed-severity fires, nature's phoenix. Elsevier, Amsterdam, The Netherlands.

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:Article e02400.

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. Cogbill, B. Collins, A. Larson, J. Lutz, M. North, C. Restaino, H. Safford, S. Stephens and J. Battles. 2019. Estimating historical forest density from land-survey data. Ecological Applications, *in press*.

Mark, A. F., and A. E. Esler. 1970. An assessment of the point-centred quarter method of plotless sampling in some New Zealand forests. Proceedings of the New Zealand Ecological Society 17:106–110.

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 8:63–88.

William L. Baker
Mark A. Williams
Department of Geography/Program in Ecology, University of Wyoming, Laramie, Wyoming 82071 USA
E-mail: bakerwl@uwyo.edu

Manuscript received 23 July 2019; revised 30 August 2019; accepted 16 September 2019. Corresponding Editor: Andrew O. Finley.

Supporting Information

Additional supporting information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/eap.2017/full