Landfire Accuracy Estimates for the Great Basin Superzone: Comparison of Original Estimates with Poststratified Estimates Adjusted for the Proportion of Area in each EVT Map Class

Steve Stehman
December 12, 2011
Updated February 8, 2012

Summary
This analysis focuses on using the known EVT class map proportions to adjust the LF accuracy estimates. Specifically, the technique of poststratified estimation is applied. In this technique, the known proportion of area mapped for each EVT class is used to provide estimates of overall and producer's accuracies that weight the sample plots from different post-strata (EVT) differently to take into account the relative proportion of each map EVT class (i.e., the proportion of area mapped as that EVT class). Because the LF sample is not a probability sample, it is possible that some EVT classes are represented disproportionately in the sample relative to their proportional representation in the population (full region). The poststratified analysis adjusts the accuracy estimates to what they would be had the sample represented each EVT class in the same proportion that the class is present in the full region (according to the map). The goal is to assess whether the poststratified analysis significantly alters the accuracy estimates obtained from the unadjusted LF estimates originally reported.

The outcome of this analysis is that the primary differences in the estimates from the original and poststratified methods are that some class-specific producer's accuracy estimates are very different when the poststratified estimates are computed, but usually even these large differences are not entirely unexpected given the large standard errors associated with the estimates because of the small sample size. Overall accuracy (per mapzone in the Great Basin) was generally not dramatically affected by the poststratified estimation approach. From a theoretical standpoint, the poststratified estimates are better than the original unadjusted estimates because poststratified estimates are expected to have smaller standard errors (the actual standard errors were not computed in this analysis for the poststratified estimates). These estimation issues are important to the methodology applied in the Landfire validation exercise, but estimation is a secondary problem relative to the more fundamental concerns related to the small sample size and practical constraints affecting the sampling protocol for obtaining the EVT validation data. The more fundamental issue is how to obtain a better sample for estimating accuracy. A few comments on the sampling issues are provided at the end of this document.

Results
The first step in the analysis of the two estimation methods is to compare the overall accuracy estimates for the Great Basin (GB) Superzone and the 5 map zones that comprise this superzone (Table 1). The unweighted (Unwtd) estimates are the estimates originally reported for LF accuracy. The poststratified (Post) estimates incorporate a weighting procedure using the EVT map information to define strata (each EVT class is

potentially a stratum). The standard error of the poststratified estimator is complex and was not computed. Although the standard errors of the poststratified estimates will be smaller than the standard errors for the unweighted analysis, the reduction in standard error would likely be in the range of 10-25%. For example, a standard error of 0.10 for the unweighted analysis might decrease to 0.075 or 0.09 for the poststratified estimate. The last column in Table 1 (Unwtd-Post) is the difference between the unweighted and post-stratified estimates.

Table 1. Comparing the original LF estimates (Unwtd) and poststratified estimates (Post) of overall accuracy for the GB superzone and mapzones within the GB superzone.

Zone	Unwtd (SE)	Post	Difference
Superzone	42.1 (3.2)	38.8	3.3
Mapzone 12	44.3 (6.8)	41.4	2.9
Mapzone 13	40.0 (8.8)	50.4	-10.4
Mapzone 16	68.6 (7.3)	56.0	12.6
Mapzone 17	35.3 (6.5)	29.7	5.7
Mapzone 18	27.6 (6.1)	25.4	2.2

At the Superzone level and for three of the five map zones, the poststratified estimate did not differ greatly from the original (unweighted) estimate. In Mapzone 13 the poststratified estimate resulted in an increase in accuracy and in Mapzone 16 the reverse occurred with the poststratified estimate resulting in a decrease in accuracy. The general tendency was for the poststratified estimate of accuracy to be lower than the corresponding original estimate (but Mapzone 13 is the exception).

The standard errors of the accuracy estimates (computed for the initial estimates) are generally high at the mapzone level. For example, in Mapzone 13 the two estimates differed by 10.4% but the standard error in Mapzone 13 was 8.8%, so the difference in the estimates of 10.4% is approximately the same magnitude as the "noise" or uncertainty of the estimated accuracy.

In cases where the accuracy estimates differ greatly between the unweighted and poststratified methods, the cause of the difference is linked to the estimation weights. For example, suppose for a hypothetical EVT class K very few of the sample plots are mapped as class K, but this class is actually very common in the mapzone. Further suppose that class K is almost always mapped correctly. In the unweighted analysis, class K will have virtually no effect on overall accuracy because it is rare in the sample. But in the poststratified analysis, the sample units from class K will receive a larger estimation weight to "expand" the class K sample plot weights in the overall accuracy estimate. If class K is typically mapped correctly, the poststratified estimate of overall accuracy will then increase. Conversely, if class K is mapped inaccurately, the poststratified estimate of overall accuracy will likely decrease relative to the unweighted estimate because the low accuracy of class K will be weighted more heavily. Changes in overall and producer's accuracy observed between the original LF and poststratified estimates are primarily a function of the association between the estimation weights of different classes and the accuracy of those classes whose weights change considerably in the two estimates.

The analysis was then extended to examine class-specific accuracy estimates, user's accuracy (which is the complement of commission error rate) and producer's accuracy (the complement of omission error rate). This part of the analysis was conducted only at the superzone level (Table 2), not the mapzone level because sample sizes per class within a mapzone are too small to yield reliable accuracy estimates. The poststratified estimate for user's accuracy is the same as the original estimate. This is because user's accuracy does not combine information from different map EVT classes so the different estimation weights associated with different EVT strata (classes) in the poststratified estimator have no effect on user's accuracy.

Producer's accuracy combines data from multiple EVT classes so the poststratified estimates are different from the initial unweighted estimates. Of the 47 classes in the superzone, 10 had a change of over 10% in producer's accuracy relative to the initial reported estimates, with the largest change being 36%. In 8 of these 10 cases the poststratified estimates were lower than the original unweighted estimates. These large differences in the producer's accuracy estimates were not necessarily the result of small sample sizes. In 8 of the 10 cases where a large difference was noted, the number of sample units with a reference (ground) label of the target class was greater than 10. Standard errors for the unweighted estimated producer's accuracies are large (ranging from 8 to 16.6% for the 9 largest from the 10 cases), so large changes in producer's accuracy are still not necessarily far out of line with the uncertainty of the estimates.

Perhaps the more glaring issue apparent from the Table 2 accuracy results is that the overall sample size (795) is still relatively small to obtain a precise accuracy estimate for the many EVT classes. The standard errors for user's and producer's accuracies are relatively high so a large degree of uncertainty is associated with these estimates. For example, the standard errors for estimated producer's accuracy are often near 10%, so a 95% confidence interval of the form "estimate ±2*SE" would be plus or minus 20% on either side of the point estimate, so these intervals would be very wide.

Broader Overview of Landfire Validation

The approach implemented for Landfire validation based on a holdout sample from the Landfire reference plot database represented a practical, low cost option. The results from such an accuracy assessment provide a coarse assessment of the quality of the map, but the approach is not sufficient for a rigorous, detailed assessment. The two general options for improvement are to continue to refine the analysis and to revamp the sampling design. Efforts to refine the analysis will have somewhat limited success because there is only so much that can be done to overcome problems associated with lack of a probability sampling design and the small number of sample plots in the validation sample (relative to the number of EVT classes being assessed). For example, it would be possible to extend the poststratified analysis illustrated here for the Great

Basin superzone to the remainder of the country. But based on the results from the Great Basin analysis, the poststratified estimates are not convincingly better and it is not apparent that the assessment would be markedly improved. Another analysis option is to pursue a model-based approach to estimation in which a model of classification error becomes a key element of the estimation protocol. This approach is often directly linked to the classification method used to create the map and typically assumes a fully automated classifier (i.e., one free of manual interpretation or human input). There is limited practical application of this approach in accuracy assessment, and it is not clear how applicable it would be for Landfire given that the EVT classification is not a fully automated procedure.

Revisiting the sampling design would entail looking at different options for collecting the reference data. These options would still by necessity be linked to the Landfire reference plot database and constrained by practical considerations. An alternative sampling design would need to take advantage of the best features of the existing reference plot database but perhaps also coordinate better with other programs facing similar challenges of validating maps with legends similar to EVT. Ideally, the sampling design would satisfy the conditions required of a probability sample. For example, the FIA plots in the existing Landfire reference database meet this condition, but plots selected purposefully because of convenience or special interest do not meet the requirements. Given the practical and cost challenges associated with collecting field data on private land, the majority of the reference plots probably need to be located on public land. This suggests the importance of coordinating validation with Federal and State agencies with similar data needs. For example, coordinating data collection via a common sampling design within National Parks is one option. Various GAP projects may offer another option. Sampling design alternatives should be explored in the context of data needs for validation. For example, are there options for validation that do not require an “on the ground” field visit to obtain the reference data? Even if such options are available for a partial or “coarser” assessment (i.e., a simplified legend), these options might be highly beneficial.

Landfire encounters the challenge faced by all other large-area mapping efforts in that validation is an expensive and time consuming proposition if the validation is intended to address the many and varied questions users will have. The current approach provides a practical but relatively crude assessment that likely will leave some users dissatisfied. Whether there are cost-effective practical alternatives remains to be discovered.

Table 2. Class-specific accuracy for GB superzone. User's accuracy estimates are the same for the original and poststratified analysis so only a single column is shown. "StratProd" is the poststratified estimate of producer's accuracy and "Prod" is the original estimate of producer's accuracy. The SE columns are standard errors for the original estimates. The "Diff" column is the difference of the poststratified producer's accuracy estimate minus the original estimate. The columns "Map", "Ref", and "Agree" are the number of assessment units (pixels) that were mapped as the EVT class, had the reference label of the EVT class, and had agreement between the map and reference classification. The three columns provide information related to the number of sample pixels that go into the accuracy estimates for that class. The EVT classes are sorted by the absolute value of the "Diff" column.

Code	LF_EVT Class	Users	SE	StratProd	Prod	SE	Diff	Map	Ref	Agree
2082	Mojave Mid-Elevation Mixed Desert Scrub	37.0	9.3	79.8	43.5	8.0	36.3	27	23	10
2054	Southern Rocky Mountain Ponderosa Pine Woodland	90.0	9.5	44.7	75.0	16.6	-30.3	10	12	9
2123	Columbia Plateau Steppe and Grassland	31.3	11.6	7.4	31.3	3.7	-23.9	16	16	5
2145	Rocky Mountain Subalpine-Montane Mesic Meadow	25.0	21.7	15.0	33.3	15.6	-18.3	4	3	1
2146	Southern Rocky Mountain Montane-Subalpine Grassland	100.0	0.0	8.7	25.0	9.5	-16.3	1	4	1
2016	Colorado Plateau Pinyon-Juniper Woodland	87.5	5.9	63.0	77.8	11.1	-14.8	32	36	28
2081	Inter-Mountain Basins Mixed Salt Desert Scrub	15.9	3.9	30.7	42.4	7.8	-11.7	88	33	14
2011	Rocky Mountain Aspen Forest and Woodland	66.7	7.9	72.1	82.8	10.2	-10.7	36	29	24
2001	Inter-Mountain Basins Sparsely Vegetated Systems	100.0	0.0	20.1	9.5	12.3	10.6	2	21	2
2062	Inter-Mountain Basins Mountain Mahogany Woodland and Shrubland	50.0	11.8	58.8	69.2	15.3	-10.4	18	13	9
2154	Inter-Mountain Basins Montane Riparian Systems	6.7	6.4	10.9	20.0	11.0	-9.1	15	5	1
2019	Great Basin Pinyon-Juniper Woodland	79.3	4.5	70.0	77.4	5.9	-7.4	82	84	65
2217	Quercus gambelii Shrubland Alliance	80.0	17.9	59.5	66.7	22.1	-7.2	5	6	4
2055	Rocky Mtn Subalpine Dry-Mesic Spruce-Fir Forest & Woodland	83.3	15.2	69.1	62.5	16.0	6.6	6	8	5
2079	Great Basin Xeric Mixed Sagebrush Shrubland	15.6	5.4	32.4	25.9	9.6	6.5	45	27	7
2183	Introduced Annual and Biennial Forbland	66.7	11.1	28.9	35.3	7.7	-6.4	18	34	12
2153	Inter-Mountain Basins Greasewood Flat	40.7	9.5	49.6	44.0	11.3	5.6	27	25	11
2004	North American Warm Desert Sparsely Vegetated Systems	33.3	11.1	55.5	50.0	15.2	5.5	18	12	6
2208	Abies concolor Forest Alliance	100.0	0.0	61.4	66.7	29.0	-5.3	2	3	2
2126	Inter-Mountain Basins Montane Sagebrush Steppe	44.4	16.6	39.1	44.4	16.0	-5.3	9	9	4
2109	Sonoran Paloverde-Mixed Cacti Desert Scrub	50.0	35.4	9.2	14.3	9.0	-5.1	2	7	1
2181	Introduced Upland Vegetation - Annual Grassland	21.4	11.0	25.6	21.4	12.7	4.2	14	14	3

2125	Inter-Mountain Basins Big Sagebrush Steppe	31.0	8.6	36.1	33.3	9.9	2.8	29	27	9
2107	Rocky Mountain Gambel Oak-Mixed Montane Shrubland	25.0	21.7	52.6	50.0	35.2	2.6	4	2	1
2080	Inter-Mountain Basins Big Sagebrush Shrubland	38.3	4.4	57.9	59.7	6.0	-1.8	120	77	46
2155	North American Warm Desert Riparian Systems	44.4	16.6	81.8	80.0	16.6	1.8	9	5	4
2220	Artemisia tridentata ssp. vaseyana Shrubland Alliance	55.6	9.6	48.0	46.9	9.3	1.1	27	32	15
2087	Sonora-Mojave Creosotebush-White Bursage Desert Scrub	74.2	7.9	68.0	67.6	8.3	0.4	31	34	23
2061	Inter-Mountain Basins Aspen-Mixed Conifer Forest and Woodland	58.3	14.2	53.9	53.8	14.2	0.1	12	13	7
2006	Rocky Mountain Alpine/Montane Sparsely Vegetated Systems	0.0	0.0	0.0	0.0	0.0	0.0	1	1	0
2012	Rocky Mountain Bigtooth Maple Ravine Woodland	50.0	35.4	100.0	100.0	0.0	0.0	2	1	1
2046	Northern Rocky Mountain Subalpine Woodland and Parkland	0.0	0.0	0.0	0.0	0.0	0.0	2	4	0
2050	Rocky Mountain Lodgepole Pine Forest	75.0	21.7	100.0	100.0	0.0	0.0	4	3	3
2051	Rocky Mtn Montane Dry-Mesic Mixed Conifer Forest & Woodland	0.0	0.0	0.0	0.0	0.0	0.0	3	0	0
2052	So. Rocky Mtn Mesic Montane Mixed Conifer Forest Woodland	0.0	0.0	0.0	0.0	0.0	0.0	5	2	0
2057	Rocky Mtn Subalpine-Montane Limber-Bristlecone Pine Woodland	0.0	0.0	0.0	0.0	0.0	0.0	3	0	0
2064	Colorado Plateau Mixed Low Sagebrush Shrubland	66.7	27.2	100.0	100.0	0.0	0.0	3	2	2
2086	Rocky Mountain Lower Montane-Foothill Shrubland	0.0	0.0	0.0	0.0	0.0	0.0	2	10	0
2091	Sonoran Mid-Elevation Desert Scrub	0.0	0.0	0.0	0.0	0.0	0.0	17	1	0
2103	Great Basin Semi-Desert Chaparral	0.0	0.0	0.0	0.0	0.0	0.0	1	0	0
2124	Columbia Plateau Low Sagebrush Steppe	0.0	0.0	0.0	0.0	0.0	0.0	3	1	0
2127	Inter-Mountain Basins Semi-Desert Shrub-Steppe	0.0	0.0	0.0	0.0	0.0	0.0	2	6	0
2135	Inter-Mountain Basins Semi-Desert Grassland	0.0	0.0	0.0	0.0	0.0	0.0	7	11	0
2139	Northern Rocky Mtn Lower Montane-Foothill-Valley Grassland	0.0	0.0	0.0	0.0	0.0	0.0	1	0	0
2159	Rocky Mountain Montane Riparian Systems	0.0	0.0	0.0	0.0	0.0	0.0	6	4	0
2182	Introduced Upland Vegetation - Perennial Grassland and Forbland	0.0	0.0	0.0	0.0	0.0	0.0	8	3	0
2210	Coleogyne ramosissima Shrubland Alliance	0.0	0.0	0.0	0.0	0.0	0.0	1	0	0