

## SMALL AREA ESTIMATION IN FORESTS AFFECTED BY WILDFIRE IN THE INTERIOR WEST

Gretchen G. Moisen, Research Forester  
Jock A. Blackard, Forester  
Forest Inventory and Analysis  
Rocky Mountain Research Station  
USDA Forest Service  
507 25<sup>th</sup> Street  
Ogden, UT 84401  
[gmoisen@fs.fed.us](mailto:gmoisen@fs.fed.us)  
[jablackard@fs.fed.us](mailto:jablackard@fs.fed.us)

Mark Finco, Remote Sensing Analyst  
Remote Sensing Applications Center  
USDA Forest Service  
2222 West 2300 South  
West Valley City, Utah 84119  
[mfinco@fs.fed.us](mailto:mfinco@fs.fed.us)

### ABSTRACT

Recent emphasis has been placed on estimating amount and characteristics of forests affected by wildfire in the Interior West. Data collected by FIA is intended for estimation over large geographic areas and is too sparse to construct sufficiently precise estimates within burn perimeters. This paper illustrates how recently built MODIS-based maps of forest/nonforest and biomass coupled with field collected plot data can be used to produce estimates of forest biomass within small geographic areas. FIA-collected forest attributes were modeled as functions of 250-m resolution digital ancillary variables using nonparametric tree-based methods, then maps were built over diverse ecological mapping zones. A composite estimation approach was applied to balance the potential bias of synthetic estimates (generated solely from the maps) against the instability of a direct estimator (generated from small numbers of FIA plots). Methods are described and illustrated through applications to fires that occurred in the years 2002 and 2003.

### INTRODUCTION

The Forest Inventory and Analysis Program (FIA) of the USDA Forest Service conducts inventories of status and trends in forested ecosystems nationwide (<http://www.fia.fs.fed.us/>). A network of sample plots have been established across the country at an intensity of approximately one plot per 6,000 acres, and data collection is conducted under an annual rotating panel system. The traditional emphasis of the FIA program has been to construct regional and statewide estimates of forest area, tree volume, growth and mortality. Increasingly, resource questions are directed at *small areas* defined as areas for which there are too few sample plots from which to construct reliable population estimates. In the Interior West, a recurring example of this is wildfire. The question “How much burned?” relates not only to how much area has burned, but how much tree volume, biomass, etc., fall within burn areas. Each year, numerous wildfires break out in the Interior Western States. Often there are an insufficient number of FIA plots within burn areas to produce rapid estimates of resources affected by this prominent disturbance. In addition to fire-related questions, FIA is being asked to answer other questions like: “How many acres were infested by insects and what is the rate of spread?”; “How much old growth is left?”; “Is wildlife habitat declining and for what species?”; “How much of a particular resource is in a Ranger District, BLM District; National Park; Reservation, County, Watershed?”, and the list of questions goes on. Too often, these small area estimation problems are handled by either accepting inadequate precision, or collecting more (and very expensive) data.

Our goals in this study were to 1) develop the capability to produce forest resource estimates for any user-defined small area, and 2) use this methodology to estimate how much tree biomass (live, above ground) occurs within MODIS-based burn areas in the Interior West during the 2002 and 2003 seasons in each of 17 ecological zones. Below we give a brief overview of small area estimation strategies, describe methods we applied to fires occurring in the last two years in the Interior West, discuss results of our analyses, and describe future work in the area of small area estimation for forest inventory applications.

## OVERVIEW OF SMALL AREA ESTIMATION STRATEGIES

There are many ways to construct estimates over small areas. In this section we provide an overview of small estimation strategies including direct estimators, indirect estimators, and explicit small area models.

### Direct estimators

*Direct estimators* use values of the variable of interest only from the sample units in the domain. Direct estimation can be conducted under a design-based (eg. Cochran, 1977; Sarndal et al., 1992; Thompson, 1997) or model-based (Valliant et al., 2001) framework. Model-based estimators are known to perform poorly under model misspecification as sample size increases (Hansen et al., 1983) and FIA relies primarily on direct estimators constructed under a design-based framework for estimating population totals.

### Indirect estimators

*Indirect estimators*, on the other hand, borrow strength by using values of the study from sample units outside the domain, thus increasing the “effective” sample size. An indirect *synthetic* estimator uses a reliable direct estimator for a large area to derive an indirect estimator for a small area assuming the small area has similar characteristics. For an example of a synthetic estimator, think of a National Forest over which there is an intensive grid of sample points. Suppose the area of interest is a watershed on the Forest with few sample points, but one wants to estimate biomass within that small area of interest. One synthetic approach is to simply assign the mean value obtained for the Forest to the watershed. A regression-synthetic estimator might involve building a simple model where biomass is a function of cover type, predicting how much biomass is on each pixel in the watershed, then adding up those pixels for a watershed total. More commonly, we work with pre-generated maps. Adding up pixels on a map of a variables of interest is a very common form of synthetic estimation, though seldom is the affiliated error adequately addressed.

### Small area models

Rao (2003) gives a comprehensive account of small area models. A common framework in many small area applications is the general linear mixed model.

$$\begin{aligned} y &= X\beta + Zv + e \\ v &\sim N(0, G) \\ e &\sim N(0, R) \end{aligned}$$

Here,  $y$  is some response (like forest biomass),  $\beta$  a vector of unknown fixed effects with known model matrix  $X$ , and  $v$  is a vector of unknown random effects with known model matrix  $Z$ . An effect may be considered fixed if the inference space is limited to the observed levels of that effect (like cover type if all are sampled, or a continuous variable like elevation). An effect may be considered random if the inference space is applied to a population of levels, not all of which are observed (like watersheds if not all are sampled).  $R$  and  $G$  are unknown covariance matrices that can be modeled using a number of different covariance structures, like spatial or temporal. This is an *explicit* linking model that accounts for contribution of ancillary variables plus area-specific random effects. Through the general linear mixed model, an empirical best linear unbiased predictor (EBLUP) takes the form:

$$\hat{\mu}_i = \hat{\phi}_i \left[ \bar{y}_i + (\bar{X}_i - \bar{x}_i)^T \hat{\beta} \right] + (1 - \hat{\phi}_i) \left[ \bar{X}_i^T \hat{\beta} \right].$$

The first term in square brackets is a survey regression estimator frequently used in direct estimation. The second term in square brackets is a synthetic regression estimator, and the  $\hat{\phi}_i$  and  $(1 - \hat{\phi}_i)$  are weights that sum to one. So, the EBLUP estimator can be expressed simply as a weighted average of the survey regression estimator and the regression synthetic estimator. The weight  $\hat{\phi}_i$  is calculated at the ratio of model variance to total variance. More weight is assigned to the direct regression estimator as sample size increases:

$$\hat{\phi}_i = \frac{\hat{\sigma}_v^2}{\left(\hat{\sigma}_v^2 + \hat{\sigma}_e^2 / n_i\right)}.$$

There are many ways to produce estimates for small areas. EBLUP estimators balance potential bias of a synthetic estimator against instability of a direct estimator. Small area models allow use of model diagnostics, produce area-specific measures of precision, allow complex data structures, and provide a framework for small area inferences.

## METHODS

Recall that our goals in this study were to 1) develop capability to produce forest resource estimates for any user-defined small area, and 2) use this methodology to estimate how much tree biomass (live, above ground) occurs within MODIS-based burn areas in the Interior West during the 2002 and 2003 seasons in each of 17 ecological zones. The process involved defining small areas of interest (AOIs) using MODIS-based active fire maps and then constructing estimates of forest biomass within these AOIs by creating weighted combination of direct estimates (from FIA-data) and indirect estimates (from moderate resolution maps of biomass) through composite estimators.

### MODIS Active Fire Maps

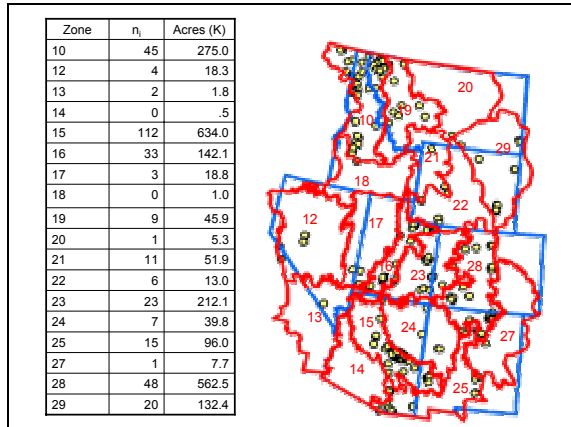
Burn areas in this study were based on MODIS thermal anomaly maps (<http://activefiremaps.fs.fed.us>), produced at the USDA Forest Service Remote Sensing Application Center (RSAC). Thermal anomalies are identified using the MODIS 1-km thermal imagery and the MOD14 algorithm. We recognize that the areas delineated by 1-km imagery are imperfect burn area perimeters; however, they are updated six to eight times daily and provide a rapid assessment of the fire situation nationwide. Work is currently underway to produce national fire perimeters with greater spatial detail and identify levels of burn intensity within these areas. The more detailed fire perimeters will better focus the estimation on resource actually burned.

### FIA data and mapping zones

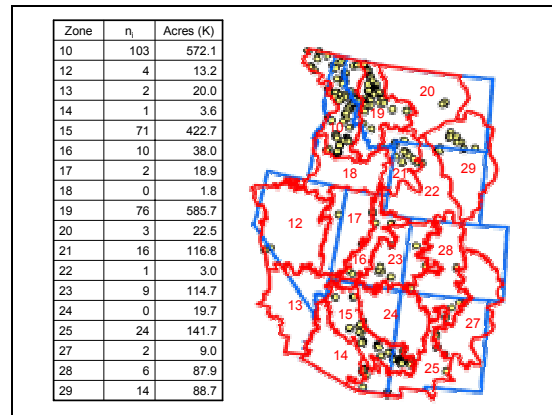
USGS mapping zones based on Homer and Gallant (2001) were used as subpopulations for which individual models and estimates were constructed. The number of sample plots falling within burned areas ranged from as many as 103 down to 0, with the majority falling in the single digits. Figures 1 and 2 illustrate numbers of burned plots by year and mapping zone.

### Moderate resolution maps of forest biomass

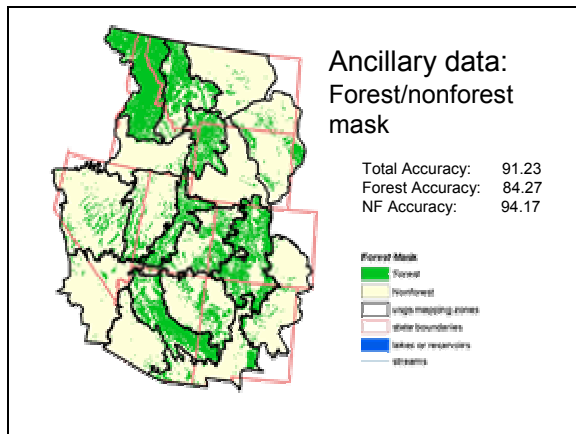
The Nationwide biomass mapping effort of FIA's Remote Sensing Band provided a large set of ancillary data at 250- to 500-m resolution that included MODIS spectral bands, topographic variables, derivatives from the NLCD, climatic variables, with a total of 60+ digital layers. In order to capitalize on these ancillary data layers for estimation, flexible models are needed to relate a field recorded response (like biomass) to the many predictor variables. The relationship is best modeled through committee-based nonparametric models. See Ruefenacht et al. (2004) and Blackard et al. (2004) for more information on these moderate resolution maps. These yield the best predictions but do not fit easily into small area estimation theory. Consequently, we used the predicted forest/nonforest map as a mask (Figure 3), and only predicted biomass (Figure 4) entered into the small area models.



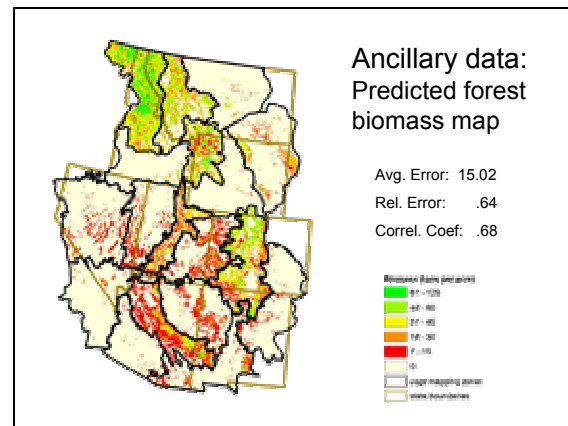
**Figure 1.** Number of FIA sample plots and total area affiliated with MODIS-based burn areas by ecological mapping zone for the 2002 fire season.



**Figure 2.** Number of FIA sample plots and total area affiliated with MODIS-based areas by ecological mapping zone for the 2003 fire season.



**Figure 3.** Moderate resolution map of forest/nonforest.



**Figure 4.** Moderate resolution map of forest biomass (tons per acre).

### Model

A special case of a small area model is a composite estimator:

$$\hat{Y}_i = \phi_i \hat{Y}_{i1} + (1 - \phi_i) \hat{Y}_{i2}.$$

This is also the weighted average of a direct estimator  $\hat{Y}_{i1}$  and a synthetic estimator  $\hat{Y}_{i2}$ . In our fire application, the direct estimator was a survey regression estimator that used FIA plots and the predicted forest biomass map as ancillary information. The indirect estimator used only the biomass map. An approximate optimal weight depends only on the ratio of the mean squared errors (MSE's) of the direct and synthetic estimators, ie:

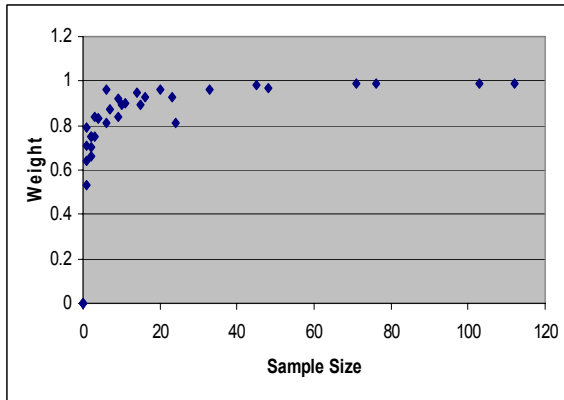
$$\phi_i = MSE(\hat{Y}_{i2}) / (MSE(\hat{Y}_{i1}) + MSE(\hat{Y}_{i2})).$$

So, the weight shifts toward the estimator with the smallest MSE. The MSE of the direct estimator was obtained by simply using the estimated variance of a regression estimator. The MSE of the synthetic estimator was more difficult to construct. There are a number of simple approximations (Rao, 2003), but these can be very unstable and

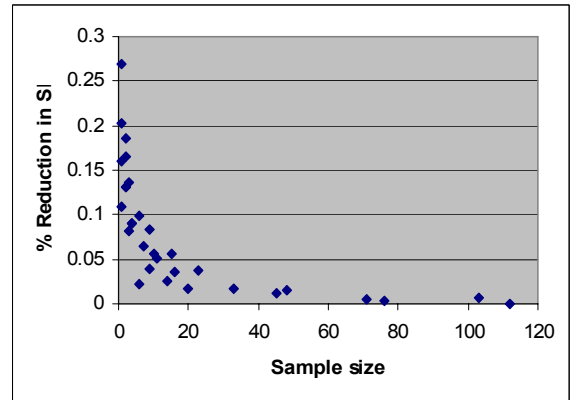
often yield values close to, or even less than, zero. To get around this, empirical MSE's were constructed for each mapping zone using the independent test sets from the biomass mapping effort.

## RESULTS AND DISCUSSION

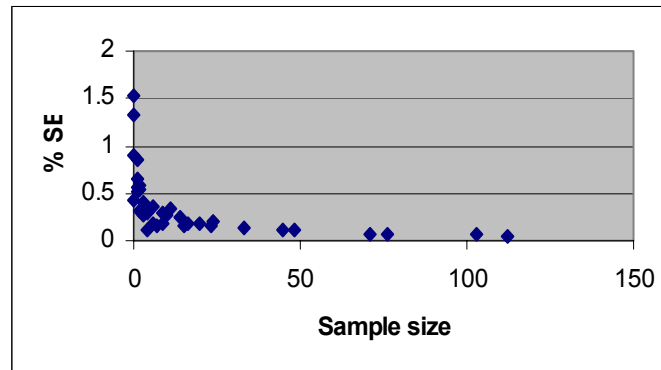
Not surprisingly, the survey regression estimator alone produces more precise direct estimates than simple random sampling. The composite estimator further improves estimates for areas with few sample points by balancing the potential bias of synthetic estimates (generated solely from the maps) against the instability of a direct estimator (generated from small numbers of FIA plots). Figure 5 illustrates how the composite estimator automatically gives more weight to the direct survey regression estimator as sample size increases. For very small sample sizes, almost all the weight is given to the synthetic estimator, but once the sample size approaches 30-40, all the weight goes to the direct estimator relying on FIA plots. This has appeal in that the weighting occurs automatically, and the same algorithm can be applied regardless of the size of the area of interest in a production environment.



**Figure 5.** Weight placed on direct estimator as a function of sample size.



**Figure 6.** Percent reduction in standard error on estimates of forest biomass using a composite estimator vs. a regression estimator plotted against sample size.



**Figure 7.** Percent standard errors in estimates of forest biomass as a function of sample size.

Zone	2002		2003	
	Biomass (K)	% SE	Biomass (K)	% SE
10	13,450.07	10.2	25,235.01	7.5
12	261.60	29.2	471.05	11.7
13	2.15	100+	169.80	58.2
14	2.11	89.8	13.79	51.3
15	20,118.54	4.5	11,109.04	6.8
16	3953.80	12.6	874.86	26.6
17	211.85	40.4	186.54	53.4
18	25.01	100+	39.88	100+
19	1,921.44	16.9	23,461.87	6.3
20	104.86	55.3	611.51	26.7
21	1,246.52	34.2	4,246.44	19.0
22	291.89	17.9	31.90	84.9
23	3,321.37	16.5	2,421.38	42.6
24	814.46	15.7	259.72	29.8
25	1,720.87	14.8	2,364.95	20.5
27	67.07	64.6	126.63	32.4
28	17,830.83	10.4	2,096.73	35.8
29	2,832.07	17.0	1,564.51	24.5

**Table 1.** Estimates of biomass (thousands of tons per acre) within burn areas and percent standard errors by year and mapping zone.

The greatest gains in precision from using the composite estimator over a direct estimator are realized for the smallest sample sizes, as illustrated in Figure 6. Table 1 shows estimates of biomass and percent standard errors by year and mapping zone. Note that despite the fact that the composite estimator yields more reliable estimates than the direct or indirect estimators by themselves, the improvements are not great, and many estimates still have extremely high percent standard errors. This is important to emphasize because when we use only a few sample plots to construct estimates using a direct method, we are aware of the high error. But if we are counting up pixels on an existing map (indirect), we tend to ignore the inherent uncertainty and do not account for it in final estimates. Figure 7 emphasizes this point, as shown by the very high percent standard error for small sample sizes.

#### FUTURE WORK

Our short-term goal is to construct “as-it-burns” resource estimates by accessing limited plot data and pre-built FIA variable maps interactively and link these estimates to the rapid response fire page. Work is underway to produce better estimates of areas actually burned within the fire detection zones.

The use of better ancillary data layers will also improve standard errors in resource estimates. Examples of data sets currently under construction that could be incorporated into the estimation process include: the Nationwide forest type map (Ruefenacht et al., 2004), Nationwide risk map, LANDFIRE-generated maps, and the updated National Land Cover Data Set (NLCD).

Some thought should be given to County/Forest-level estimation. Small area estimates could be tied into global framework for estimation, perhaps through EBLUP. Estimation of resources within counties or National Forests could be constrained such that the sum of the small area estimates equals a total from a direct estimator.

## CONCLUSIONS

Many of the pressing resource questions facing FIA units are directed at small areas. Too often, small area estimation problems are handled by either accepting inadequate precision or collecting more data. There is tremendous need for small area estimation methodologies to be applied in forest inventory applications. In this paper, we illustrated how recently built MODIS-based maps of forest/nonforest and biomass coupled with field collected plot data can be used to produce estimates within small geographic areas. A composite estimation approach was shown to balance the potential bias of synthetic estimates (generated solely from the maps) against the instability of a direct estimator (generated from small numbers of FIA plots). Users that choose to add up pixels on a map to construct small area estimates should be aware that they are constructing a synthetic estimator whose true MSE may be substantially larger than that of an estimation approach that incorporates sample plots as well. Although the composite estimation approach taken here improved the precision on estimates for small areas, many of the small area estimates still were very unreliable. It is important to convey this information to users in the form of valid measures of error.

## REFERENCES

- Blackard, J.A., Moisen, G.G., and Tymcio, R.P. (2004). Modeling investigations into mapping of forest attributes with FIA data for the state of Wyoming. This Proceedings.
- Cochran, W.G. (1977). Sampling Techniques. Wiley, New York.
- Hansen, M.H., Madow, W.G., and Tepping, B.J. (1983). An evaluation of model-dependent and probability sampling inferences in sample surveys. *Journal of the American Statistical Association* 78:776-793.
- Homer, C.G., and Gallant A. (2001). Partitioning the Conterminous United States into Mapping Zones for Landsat TM Land Cover Mapping, USGS Draft White Paper.
- Rao, J.N.K. (2003). Small Area Estimation. John Wiley and Sons, Inc. Hoboken, New Jersey.
- Ruefenacht, B., Moisen, G. G., and J.A. Blackard. (2004). Forest type mapping of the Intermountain West. This Proceedings.
- Särndal, C.E., Swensson, B. and Wretman, J.H. (1992). Model Assisted Survey Sampling. Springer-Verlag, New York.
- Thompson, M.E. (1997). Theory of Sample Surveys. Chapman and Hall, London.
- Valliant, R., Dorfman, A.H., and Royall, R.M. (2001). Finite Population Sampling and Inference: A Prediction Approach. Wiley, New York.