Forestry Advance Access published April 16, 2012

Forestry An International Journal of Forest Research

Institute of Chartered Foresters

Forestry, doi:10.1093/forestry/cps037

Monitoring forest/non-forest land use conversion rates with annual inventory data

FRANCIS A. ROESCH[1]* AND PAUL C. VAN DEUSEN[2]

[1]USDA Forest Service, Southern Research Station, Forest Inventory and Analysis, 200 WT Weaver Boulevard, Asheville, NC 28804-3454, USA
[2]National Council for Air and Stream Improvement, 15 Dunvegan Road, Tewksbury, MA 01876, USA
*Corresponding author. E-mail: froesch@fs.fed.us

Summary

The transitioning of land from forest to other uses is of increasing interest as urban areas expand and the world's population continues to grow. Also of interest, but less recognized, is the transitioning of land from other uses into forest. In this paper, we show how rates of conversion from forest to non-forest and non-forest to forest can be estimated in the US from a continuously improving publicly available annual forest inventory database, under varying definitions of conversion. Two estimation approaches are considered and contrasted. The approaches are a simple ratio estimator and the weighted maximum likelihood estimator. The latter involves a statistical procedure that incorporates the binomial nature of the indicator variables, the transition of mapped plot conditions and an intuitively appealing way to combine data from varying remeasurement periods for a temporally dependent variable.

Introduction

Sustainable forestry practices are meant to ensure that all the benefits of forests are available in perpetuity. These practices on existing forestland are effective as long as that forestland remains forested. The transitioning of forestland to other uses is of vital interest as urban areas expand and the world's population continues to grow. Compensating for this loss, and therefore also of interest, is the transitioning of land from other uses into forest. This interest has led to the development of certification processes (Hansen *et al.*, 2006), often self-imposed by some forest industry entities (Cashore *et al.*, 2004). These certification processes have intensified the necessity to estimate the rate at which forestland is being converted to non-forest and vice-versa. In this paper, we show how rates of conversion from forest to non-forest can be estimated in the US from a continuously improving publicly available annual forest inventory database, developed and maintained by the USDA Forest Service Forest Inventory and Analysis (FIA) Program. Two estimation approaches are considered and contrasted. The approaches are a simple ratio estimator and the weighted maximum likelihood (WML) estimator of Van Deusen and Roesch (2009). The latter incorporates the binomial nature of the indicator variables, the transition of mapped plot conditions and an intuitively appealing way to combine data from varying remeasurement periods for a temporally dependent binary variable. We do not attempt to formally test either of the estimators, as their properties are fairly well known or can be easily deduced from the existing cited literature. Rather we discuss the inferences that can be drawn as a result of similar-seeming applications of the estimators to existing data.

Publicly available remeasured forest inventory plots provide a data source for estimation of conversion rates; however, there are many ways that these data might be used to formulate similar sounding, but quite different estimates of conversion rates. We discuss the reasoning behind some of the different estimands and estimators.

In general, we recognize that each point in a land area observed at two points in time can be classified into one of the four conditions for the problem at hand:

Published by Oxford University Press on behalf of the Institute of Chartered Foresters, 2012.

1. forest to non-forest (*f2nf*),
2. forest to forest (*f2f*),
3. non-forest to forest (*nf2f*), and
4. non-forest to non-forest (*nf2nf*).

In turn, each remeasurement of a forest inventory plot provides observations that can be used to map the plot into at most four sections, one for each of the conditions associated with the points in the plot. These plot observations are related to estimates of annual conversion rates in a temporally dependent fashion. That is, the expected value of each proportion observed on the plot is dependent upon the length of time between observations. This is a trivial complicating factor if the rates of change are stationary through time. Because we are concerned with relatively short windows of observation, we assume that conversion rates are stationary in this paper.

Estimation

Although the term forest conversion means about the same thing to everyone, seemingly slight differences in definition and sampled populations can lead to quite different estimands (the variables to be estimated.) For clarity, we offer Table 1 as an example of a forest/non-forest change matrix over 100 U of land.

Given Table 1, consider the following scenarios:

1. Suppose the above Table 1 represents an area of 100 U of land that is available to be forest from 1 year (time 1) to the next (time 2). That is, cities, industrial sites and the like are not in the population and were not sampled or these areas were filtered out of the sample. Then, the estimands of annual transition probabilities might be formulated as follows:

Annual forest to non-forest – $af2nf_1$ = 5/100 = 0.05.
Annual non-forest to forest – $anf2f_1$ = 10/100 = 0.10.
Annual net loss – net loss_1 = 0.05–0.10 = −0.05 (an increase of 5 per cent).

2. Suppose the sampled population was all land in a defined area and the sample would therefore include plots from locations that could never be forest. In this case, it would make the most sense to define the estimands of conversion on the amount of forestland at time 1 (80 plots).

a*f2nf*2 = 5/80
a*nf2f*2 = 10/80
Net loss_2 = 5/80–10/80 = −5/80

Table 1: An example forest/non-forest table for times 1 and 2

		Time 2		
		Forest	Non-forest	All land
Time 1	Forest	75	5	80
	Non-forest	10	10	20
	All land	85	15	100

Under both scenarios:

$$\text{Net loss}_s = af2nf_s - anf2f_s, \tag{1}$$

where $af2nf_s$ is the annual per cent change from forest to non-forest and $anf2f_s$ is the annual per cent change from non-forest to forest for scenario *s*. The variance of net loss is the sum of the variances of $af2nf_s$ and $anf2f_s$ minus two times the covariance of $af2nf_s$ and $anf2f_s$:

$$V(\text{Net loss}_s) = V(af2nf_s) + V(anf2f_s) - 2C(af2nf_s, anf2f_s).$$

Since the numerators of $af2nf_s$ and $anf2f_s$ arise from distinct subpopulations (time 1 forest and time 1 non-forest) with a constant denominator, they are independent and therefore the covariance is zero. The estimator of the variance of net loss is:

$$\hat{V}(\text{Net loss}_s) = \hat{V}(af2nf_s) + \hat{V}(anf2f_s). \tag{2}$$

Both approaches give a net increase in forest area based on Table 1, but the rate of change would be expressed differently in each case. As long as the assumptions and definitions of the estimands are fully explained, the estimators being used will be understood. The conversion rates obtained from scenario 2 will be consistently higher than those obtained from scenario 1. From a population standpoint, the basis for scenario 1 is easier to identify and more consistent through time than the basis for scenario 2. Scenario 2 might more closely match the expected definition of a rate of change, over a single time interval, because it gives the proportion of previous (i.e. existing) forestland being converted to non-forestland, and the amount of that forestland that has been replaced by formerly non-forestland, when one is interested in rates relative to a particular point in time. The disadvantage to scenario 2 is that the basis (or denominator) changes with each time interval. The basis for scenario 1 must also change as land becomes unavailable for forest use, but it changes more slowly through time than the basis for scenario 2 because many conversions from forest land use do not preclude a return to forest land use. Recognition of the change can be made less often and at convenient times. Here we treat the basis for scenario 1 as fixed over a limited, but non-specified, number of measurement intervals.

Complicating the issue of conversion rates with respect to the scenarios above is the definition of forestland and the determination of land that could be forest. For the former, we rely on the definitions used by FIA and documented in Woudenberg *et al.* (2010) and note that the methods we discuss would not need to be altered by a user who preferred an alternative definition of forestland. For the latter, we note that all land could be classified into some category of current use, say $U = \{A, B, C, D, E, F, G, \ldots\}$, with category F reserved for forest. Assume that for each non-forest category, there currently exists a probability function $f(F \mid U_{-F})$, where the subscript $-F$ represents a particular non-forest category, of a one step ahead transition to forest. For instance, in the US, abandoned farmland is a major source of newly-established forest, so its probability of transition to forest would be higher than all other uses. Also, assume that future social and economic forces can be represented by a function $g(f(F|U_{-F}))$ which serves to act upon and change f. Although it is beyond the scope of this paper, an investigator could estimate each $f(F|U_{-F})$ from existing data and then model $g(f(F|U_{-F}))$ given possible future scenarios.

In this paper, our estimands will correspond to those arising from scenario 1. Limiting the rest of our discussion to scenario 1 allows us to drop the subscript from the expressions for the estimands and the estimators.

The data

We used data from the USDA Forest Service's FIA. FIA has been using a temporally rotating, panelized forest inventory sampling design for slightly more than a decade and the intended monitoring advantages of the design are starting to be realized. For context, we briefly explain the design and refer the interested reader to Reams *et al.* (2005) and Roesch (2008) for a deeper understanding. In this design, the sample plots were located in proximity to a systematic triangular grid consisting of g mutually exclusive interpenetrating panels. The panels were spatially balanced and contained an approximately equal number of sample plots. That is, if the total sample size was n, then each panel consisted of approximately n/g plots. The sequence of panels was measured in order, with one panel measured each year, after which the panel measurement sequence reinitiated. Therefore, if panel 1 was measured in 2001, it would also be measured in 2001+g, 2001+2g and so on. Panel 2 would then be measured in 2002, 2002+g, 2002+2g, etc. The methods described below were applied to the publicly available data arising from this design for 12 states in the southern US.

This database arose from a sample of all land within the boundaries of the US and did not always clearly distinguish between different non-forest uses. Therefore, in our study below, we considered a single non-forest land use category. For the purposes of scenario 1, we assumed that land that 'could be forest' consisted of (1) all land that had been observed to be forest over the time interval of interest in addition to (2) some proportion of the remaining non-forestland. We then investigated a range of reasonable proportions of the non-forestland that could be forestland given the scope of a particular inventory and showed results for the upper end of this range.

The temporally indifferent ratio estimator

Since the initiation of the rotating panel design for FIA, there have been quite a few papers focusing on using trend models for the purpose of improving annual estimates such as Van Deusen (1996, 1999), Roesch *et al.* (2003), Roesch (2007), Johnson *et al.* (2003) and Czaplewski and Thompson (2009). FIA gives many estimates of summary statistics based on the 'temporally–indifferent' assumption given in Patterson and Reams (2005), which ignores trend within an observation period. We used a ratio estimator that would be arrived at under the temporally indifferent assumption, but we point out that conversion rates cannot be temporally indifferent.

For each plot, i, we calculated the area of the plot observed to have transitioned from forest to non-forest ($a_{f2nf,i}$) and from non-forest to forest ($a_{nf2f,i}$) and divided each of these by the remeasurement period for the plot (r_i) to obtain an annual area of each plot transitioning to the other condition. The ratio estimators were formed by taking the sums over the n plots for each of these annualized plot subareas and dividing by the total plot area on land that could be forest (a_F):

$$R_{f2nf} = \frac{\sum_{i=1}^{n} \frac{a_{f2nf,i}}{r_i}}{\sum_{i=1}^{n} a_F},$$

and

$$R_{nf2f} = \frac{\sum_{i=1}^{n} \frac{a_{nf2f,i}}{r_i}}{\sum_{i=1}^{n} a_F}.$$

Note that when considering a rate of change, an assumption of temporal indifference does not equate to an assumption of stationarity. Temporal indifference assumes that an estimator over a number of years has the same expected value as the estimator has for a single year. Stationarity assumes that the expected value for the estimator is the same for each of the years. Therefore, under an assumption of temporal indifference, one could estimate the mean rate over a number of years and assume that it is equal to the annual rate. Under an assumption of stationarity, when a rate is observed over a number of years, one would have to apply a discounted formula in order to establish the annual rate through the observation years.

WML estimator

We compared the ratio estimators above to the WML approach developed in Van Deusen and Roesch (2009). Here we considered the following model for an indicator variable, $I_{j,t}$, for point j at time t and its relationship to an underlying annual conversion probability, P:

$$I_{j,t} \sim P + e_{j,t},$$

where $e_{j,t}$ is an error term that does not follow the normal distribution. The indicator, $I_{j,t}$ changes from 0 to 1 when the status of the condition of interest on the point changes. The model could be applied to any status change, such as a change from forest to non-forest or a change from non-forest to forest. In this case, the $E(I_{j,t} = 1)$ is assumed to be independent of the area of the plot in a particular condition and stationary over the observation period.

Distribution of observed indicators

The complete indicator data for point j on plot i is a sequence of 0s and 1s that is observed at the beginning and end of the plot remeasurement period, r_i. As such, there are only two possible observation outcomes for each point. Either the observation sequence will begin and end with a 0 or it will begin with a 0 and end with a 1. Call the first possibility $h0$ and the other possibility $h1$. Note that $h1$ begins with a 0 and may change to a 1, but it cannot change back to 0. Call the year when it switches from 0 to 1, s_j. It is known that $1 \le s_j \le r_i$, but the actual year when the point status changed is not typically known. The probability of observing $h1$ if s_j is known is $p(h1_j | s_j) = (1-P)^{s_j - 1}$ and the probability of observing sequence $h0$ is $p(h0_j) = (1-P)^{r_i} = Q^{r_i}$.

In practice, the complete indicator data were not observed for a point. However, it was known whether the point sequence had an $h0$ history or an $h1$ history. The unconditional probability of an $h1$ history is simply $1 - p(h0) = 1 - Q^{r_i}$. In the remainder of this paper, we assumed that the actual point transition times were not available.

The unknown value for P can be estimated by finding the value that maximizes the observed data likelihood. As in Van Deusen and Roesch (2009), the plot condition proportion of interest, $a_{x(i)}$, is incorporated as a weight in the likelihood function. This has the effect of allowing plots where $a_{x(i)}$ is large to have the most influence on the estimate of P. The likelihood function, its related Hessian and Jacobian as well as their use in the Newton–Raphson algorithm are given in an Appendix for the interested reader.

Methods

The estimators described above were applied to the publicly available data arising from the FIA design for 12 states listed in Table 2. We used all of the data available on 24 August 2011, for all remeasured (plot intensity 1) plots with time 2 inventory years between 2006 and 2010 and a valid entry in the subplot condition change table, described in Woudenberg *et al*. (2010). We partitioned each plot into at most four sections corresponding to the categories: forest to forest, forest to non-forest, non-forest to forest and non-forest to non-forest.

Table 2: The sum of plot proportions (i.e. the effective sample sizes) for each state in the study at time 1

State	Plot area at time 1
Alabama	2535.4
Arkansas	2948.4
Florida	1039.2
Georgia	4046.0
Kentucky	928.1
Mississippi	928.1
North Carolina	1965.1
Oklahoma	178.8
South Carolina	2084.1
Tennessee	2241.7
Texas	1937.9
Virginia	2946.4

Because FIA conducts an all-land inventory, a large portion of the sampled land will never be forest. For a number of reasons, there is no way to partition, or stratify, the sample into the portion of the sample based on the population of interest in this study (all land that could be forest) and the portion of the sample based on the remaining non-forestland that could never be forest. The most compelling reason is that the true plot locations are not publicly available, and therefore even if a stratification layer were developed to partition all land, the portions of the sample coming from each partition could not be uniquely identified. We considered this by using a contaminated sample approach. We knew that our population of interest was land that could be forest and our sample was based on all land. Therefore, our sample (with respect to our population of interest) was contaminated with observations from land that could not be forest, and we had no sure way of identifying the contaminated observations. Because there were no measures associated with the contaminated elements, i.e. they were simply identified as non-forest, we could delete a portion of plot areas categorized as non-forest to non-forest from the analysis, prior to estimating conversion rates. The decision as to how much land classified as 'non-forest to non-forest' could be forest would be subjective and rather than making an arguable assumption, we investigated a reasonable range of proportions of non-forest land that could never be forest. In each state, for each time 2 inventory year, prior to estimating rates, we deleted 80, 85, 90, and 95 per cent of the plot area classified as non-forest to non-forest, corresponding to the assumptions that 20, 15, 10 and 5 per cent of the land, respectively, that was not observed to be forest, could become forest. This essentially masked out or separated increasing amounts of non-forest land from the sampled population. The rates of land omitted from potential forest reflected the fact that there was some level of maturity in land use in the US. Environmental, social and economic forces have already affected virtually every hectare of land resulting in the current matrix of land uses. We assumed that at least 5 per cent, but no more than 20 per cent of the land that was not observed to be forest could become forest in the near future. The two estimators were then applied to (1) all the data in each state and to (2) the data by final inventory year in each state.

Results

Table 2 gives the overall time 1 forested plot area for each of the 12 states. Figure 1 gives the combined estimates after deleting 80 per cent of the non-forest to non-forest plot area (corresponding to an assumption that 20 per cent of the area classified as non-forest to non-forest could be forest) for each estimator of *af2nf*, *anf2f* and *Net loss*, respectively. Confidence intervals of (+/−) two times the standard error of the WML estimator are shown to illustrate the cases in which the ratio estimator gave estimates which fell outside of the confidence intervals for the WML estimates. For brevity, we do not show the results for the 5, 10 and

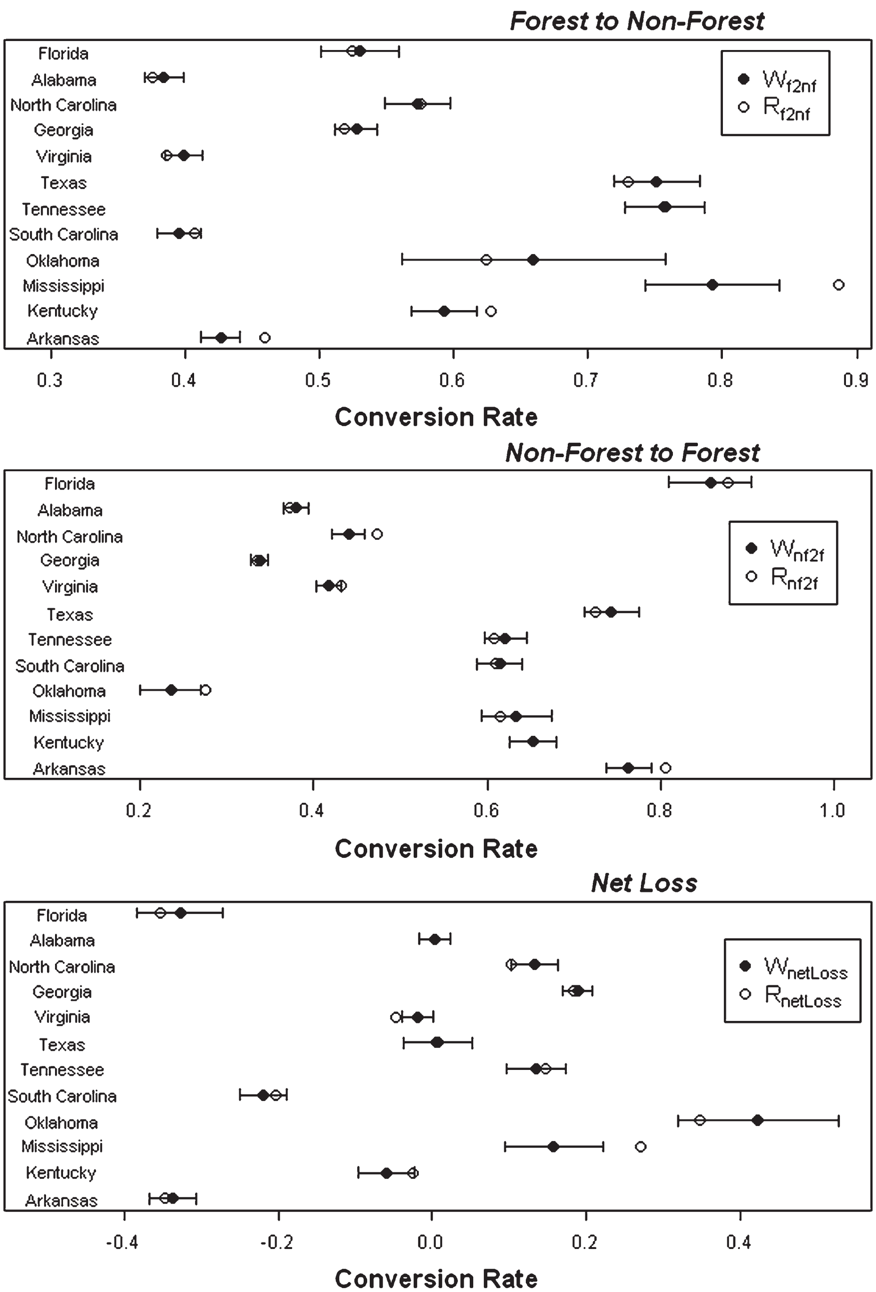

Figure 1. Overall conversion rate estimates from the most recent five panels (if available) of data for each of the 12 states in the study. The rates are based on the assumption that 20% of the observed non-forest to non-forest land could be forest. The forest to non-forest, non-forest to forest and net loss conversion rates appear in the top, center and bottom graphs, respectively.

15 per cent potential forest from non-forest rates since all the graphs have the same shape and differed from the results for the 20 per cent assumption exactly as would be predicted by reducing the denominator in a ratio.

In 8 of those 12 states listed in Table 2, there were enough remeasured panel data for five or more consecutive estimates, allowing for some trend analysis.

Figure 2 gives more detailed per-panel estimates for two of those states (Alabama and Virginia) for the assumption that 20 per cent of the non-forest to non-forest observations over the measurement interval had the potential to be forest. Again, confidence intervals of (+/−) two times the standard error of the WML estimator are shown to illustrate the cases in which the ratio estimator gave results falling outside of the confidence intervals for the WML estimates. Note that consecutive confidence intervals often do not overlap in each of the graphs of Figure 2.

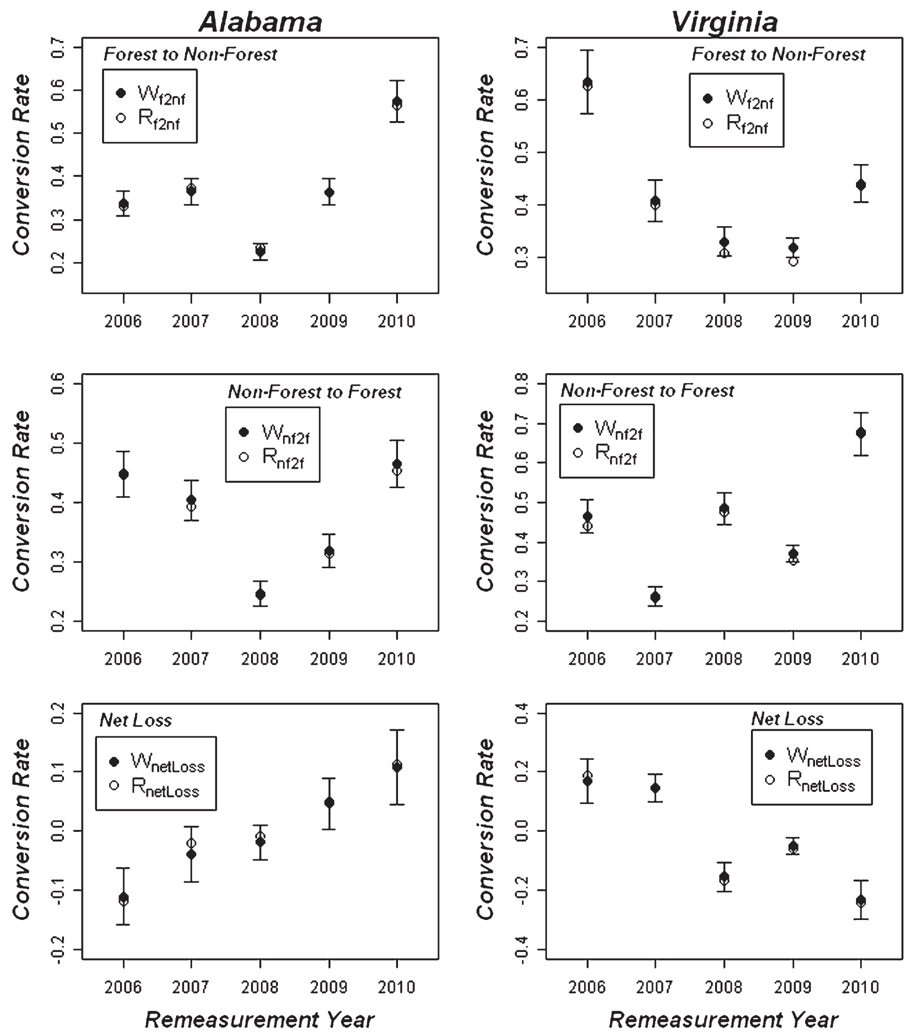

Figure 2. Forest conversion trend estimates. Forest to non-forest (top row), non-forest to forest (center row) and net forest loss (bottom row) conversion rates estimated from each the most recent five panels of data for Alabama (left side) and Virginia (right side). The rates are given based on the assumption that 20% of the observed non-forest to non-forest land could be forest. Confidence intervals of (+/−) 2 (SE) are shown for the WML estimates.

Discussion

Here we used an assumption that conversion rates were stationary (or constant) over small windows of observation. Figure 1 resulted from the assumption that the entire observation period, contributing to five consecutive panel remeasurements (~10 years), constituted a small enough window to support the stationarity assumption. We note that temporal indifference is a stronger assumption in that a result is indifferent to the width of the observation window. A rate variable cannot be both stationary and temporally indifferent because the basis upon which the rate acts changes each year, once the rate has been applied to the previous year's basis. Therefore, if one calculates an average rate over *n* years, under an assumption of temporal indifference, the expected value of the result is not equal to the stationary (or constant) rate that was applied through each of those *n* years. When rates are very low, as in this case, this bias may be hard to detect (and may possibly be unimportant) because the basis changes very little from year to year.

When either of the estimators that we used in this paper is applied to each remeasured panel, a trend in the conversion rate would be easier to detect than when it is applied to pooled multiple panels. For instance, the bottom right graph in Figure 2, for Virginia, shows a decided decrease in net forest loss, observed first between panel remeasurement years 2007 and 2008. The top left graph in Figure 2, for Alabama, indicates a mild increase in forest to non-forest conversions resulting in an increasing trend in net forest loss shown in the bottom left graph of the same figure.

In the bottom graph of Figure 1, the Virginia net loss estimate from the ratio estimator is outside of the corresponding confidence interval for the WML estimate, while in Figure 2, when the ratio estimator is applied within panels, the estimates fall within the confidence intervals for the corresponding WML estimates.

As mentioned earlier, the ratio estimator is predicated on the temporal indifference assumption. As a result, its expected value is different from the expected value of the WML estimator when the assumption is violated, which is the case in this particular application. In cases where these two estimators gave different results, it could have been due to natural variation or it could have been the result of the ratio estimator's expected value being detectably far from the desired estimand. Usually when two estimators appear to yield about the same estimate, the simplest of the two estimators is recommended. In this case, the ratio estimator was the simplest estimator, but we are disinclined to recommend it over the WML estimator because of this potential bias in the estimator. That is, we know that the expected value of the average annual rate estimated by the ratio estimator does not equal the annual rate, whereas the expected value of the WML estimator does equal the annual rate from an assumed stationary process.

Also, note that consecutive confidence intervals for the WML estimators often do not overlap in each of the graphs in Figure 2 for Virginia, which suggests that there are limits to the stationarity assumption. That is, the conversion rates are changing over the observation period. A solution to this problem would be to consider the panel-based WML estimates resulting in Figure 2 to be preliminary estimates and use them as annualized input into a mixed estimator (Van Deusen, 1999), similar to the approach used in Roesch (2007). Caution must be exercised when using any two-stage estimation approach because the preliminary estimates themselves contain errors. However, a two-stage estimation process has an advantage in that it simplifies the handling of the change of support problem (described in Gotway and Young, 2002) because the change of support can be dealt in a simple model at stage 1. Subsequently, at stage 2, one only needs to be concerned with the temporal or spatial trend in the mixed estimator model.

Conclusions

Overall, this study has accentuated the value of the USDA Forest Service's recently implemented annual forest inventory design for the timely evaluation of trends in the nation's forests. We have shown why a discussion and evaluation of conversion rates must be based on well-defined criteria, what some of those criteria might be and how one might use this and similar databases to address those criteria.

Funding

This work was funded by the USDA Forest Service, Southern Research Station, Forest Inventory and Analysis Program.

Acknowledgements

The authors would like to thank three anonymous reviewers and the editor, Dr Euan Mason, for their helpful comments.

Conflict of interest statement

None declared.

References

Cashore, B., Cashore, B.W., Auld, G. and Newsom, D. 2004 *Governing through Markets: Forest Certification and the Emergence of Non-State Authority*. Yale University Press, New Haven, CT.

Czaplewski, R.L. and Thompson, M.T. 2009 Opportunities to improve monitoring of temporal trends with FIA panel data. In *Forest Inventory and Analysis (FIA) Symposium 2008; October 21–23, 2008; Park City, UT. Proc. RMRS-P-56CD*. W.H. McWilliams, G.G. Moisen and R. L. Czaplewski (eds). US Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO. 1 CD.

Gotway, C.A. and Young, L.J. 2002 Combining incompatible spatial data. *JASA*. **97**, 632–648.

Hansen, E., Fletcher, R., Cashore, B. and McDermott, C. 2006 Forest certification in North America. Oregon State University, Extension Service Publication, Corvallis, OR.

Johnson, D.S., Williams, M.S. and Czaplewski, R.L. 2003 Comparison of estimators for rolling samples using forest inventory and analysis data. *For. Sci.* **49**, 50–63.

Patterson, P.L. and Reams, G.A. 2005 Combining panels for forest inventory and analysis estimation. In *Chapter 5 of The Enhanced Forest Inventory and analysis Program-National Sampling Design and Estimation Procedures*. W.A. Bechtold and P.L. Patterson (eds). USDA Forest Service, Asheville, NC. General Technical Report SRS-80.

Reams, G.A., Smith, W.D., Hansen, M.H., Bechtold, W.A., Roesch, F.A. and Moisen, G.G. 2005 The forest inventory and analysis sampling frame. In *Chapter 2 of The Enhanced Forest Inventory and analysis Program-National Sampling Design and Estimation Procedures*. W.A. Bechtold and P.L. Patterson (eds). USDA Forest Service, Asheville, NC. General Technical Report SRS-80.

Roesch, F.A. 2007 Compatible estimators of the components of change for a rotating panel forest inventory design. *For. Sci.* **53**, 50–61.

Roesch, F.A. 2008 An alternative view of continuous forest inventories. *For. Sci.* **54**, 455–464.

Roesch, F.A., Steinman, J.R. and Thompson, M.T. 2003 Annual forest inventory estimates based on the moving average. In *Proceedings of the Third Annual Forest Inventory and Analysis Symposium; 2001 October 17-19; Traverse City, Michigan. Gen. Tech. Rep. NC-230*. R.E. McRoberts, G.A. Reams, P.C. Van Deusen and J.W. Moser (eds). U.S. Department of Agriculture, Forest Service, North Central Research Station, St. Paul, MN, pp. 21–30.

Van Deusen, P.C. 1996 Incorporating predictions into an annual forest inventory. *Can. J. For. Res.* **26**, 1709–1713.

Van Deusen, P.C. 1999 Modeling trends with annual survey data. *Can. J. For. Res.* **29**, 1824–1828.

Van Deusen, P.C. and Roesch, F.A. 2009 Estimating forest conversion rates with annual forest inventory data. *Can. J. For. Res.* **39**, 1993–1996.

Woudenberg, S.W., Conkling, B.L., O'Connell, B.M., LaPoint, E.B., Turner, J.A. and Waddell, K.L. 2010 *The Forest Inventory and*

Analysis Database: Database Description and Users Manual Version 4.0 for Phase 2. Gen. Tech. Rep. RMRS-GTR-245. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO, 336 pp.

Received 18 November 2011

Appendix

Likelihood function

The weighted likelihood function of the observed set of histories is

$$L_W = \prod_i p(h0_{j\varepsilon 0})^{a_{j\varepsilon 0(i)}} \prod_i p(h1_{j\varepsilon 1})^{a_{j\varepsilon 1(i)}},$$

where the first product is over the $h0$ histories and the second product is over the $h1$ histories.

The log likelihood function, in this case:

$$L(P) = \sum_i a_{j\varepsilon 0(i)} r_i \log(Q) + \sum_i a_{j\varepsilon 1(i)} \log(1 - Q^{r_i}).$$

The maximum likelihood estimate is the value where the Jacobian of the log likelihood (L) equals 0. The Jacobian is

$$\frac{\delta L}{\delta P} = g(P) = -\sum_i \frac{a_{j\varepsilon 0(i)} r_i}{Q} + \sum_i \frac{a_{j\varepsilon 1(i)} r_i Q^{r_i - 1}}{1 - Q^{r_i}}.$$

The Hessian of the log likelihood with respect to P is useful for the maximization process and provides an asymptotic variance estimate. The Hessian is

$$\frac{\delta^2 L}{\delta P} = G(P) = -\sum_i \frac{a_{j\varepsilon 0(i)} r_i}{Q^2} - \sum_i \frac{a_{j\varepsilon 1(i)} r_i (r_i - 1) Q^{r_i - 2}}{1 - Q^{r_i}} - \sum_i a_{j\varepsilon 1(i)} \left[\frac{r_i Q^{r_i - 1}}{1 - Q^{r_i}}\right]^2.$$

The estimated variance of $\hat{P}$ is $-1/G(\hat{P})$, which is the negative of the inverse Hessian evaluated at $\hat{P}$.

Newton–Raphson algorithm

The following Newton–Raphson algorithm provides an estimate of P using the Jacobian and Hessian given above:

$$P^{(1)} = P^{(0)} - \lambda \frac{g(\hat{P})}{G(\hat{P})},$$

where λ is a value between 0 and 1 that is used to control the convergence to the maximum likelihood value. In this case, $0 \le P \le 1$ and the maximum likelihood estimate cannot be allowed outside of the known range.