Three Proposed Data Collection Models for Annual Inventories

Greg Reams, Bill Smith, Bill Bechtold[1]**, Ron McRoberts**[2]**, Frank Spirek**[3]**, and Chuck Liff**[4]

Abstract.—Three competing data collection models for the U.S. Department of Agriculture Forest Service Forest Inventory and Analysis (FIA) program's annual inventories are presented. We show that in the presence of panel creep, the model now in place does not meet requirements of an annual inventory system mandated by the 1998 Farm Bill. Two data-collection models that use subpaneling are defined, and the pros and cons of using those models are discussed.

The only data-collection model ensuring full compliance with the Farm Bill uses subpaneling with both spatial and temporal controls, resulting in the measurement of a single panel per year, nationally. The same field manual, portable data recorder, edit system, processing system, and estimation methods can be used within and among FIA regions. Such use will result in less duplication of effort and provide national consistency. The FIA program can produce, nationally, an annual database and annual estimates, as well as periodic reports based on 5-year-measurement requirements. Additional benefits will include the means to adjust measurement resources quickly and efficiently in order to measure resource availability by State. Additionally, the true sampling precision per fixed time-period is known, and intensification and detensifications are easy.

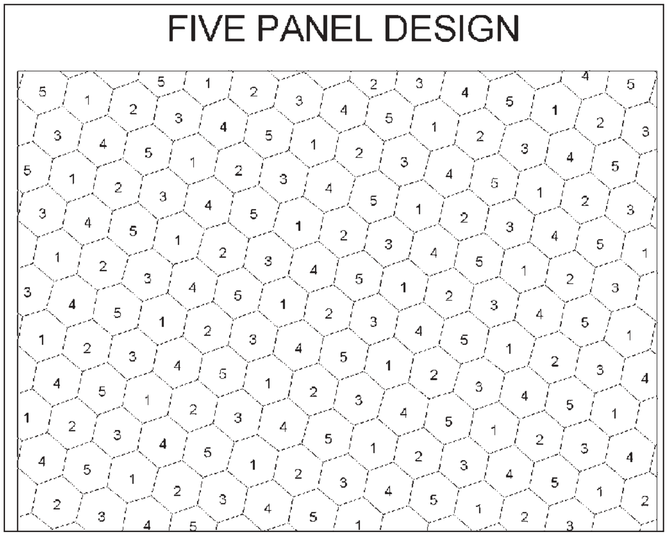
Introduction

The 1998 Farm Bill requires the Forest Inventory and Analysis (FIA) program of the U.S. Department of Agriculture (USDA) Forest Service to measure and process field plots at the rate of 20 percent per year, and to produce reports for each State at 5-year intervals. The legislation was designed to promote annual inventories based on a 5-year remeasurement cycle. Although the 20-percent per year requirement is explicit, the total number of plots by State or region on which the requirement is based was never specified—presumably to avoid micromanagement of the FIA sampling process. Optimistic that historic precision standards (3 percent per million acres of timberland and 5 percent per billion ft^3 of growing stock volume) could be retained while implementing the new requirements, FIA established a systematic national plot network with an overall sampling intensity of 1 plot per 6,000 acres. The legislation required establishment of a database that could be used to produce annual or other estimates and publication of reports based on plots visited during the 5-year measurement periods. Also, advanced technologies such as remote sensing are to be developed and integrated into the program.

The FIA national plot network has been divided into five interpenetrating panels to accommodate the 20 percent per year requirement. Each panel uses overlapping samples (i.e., repeated observations on the same plots). Each panel of plots is characterized by complete and systematic spatial coverage across the population of interest (fig. 1). On completion of all panels, the process is then repeated with the next cycle of panel measurements. Ideally, all of the sample units in a panel are measured in the same way, and all sample units in a panel have the same revisit schedule. Panels can be divided into subpanels to accommodate decreases and increases of sampling intensity. When subpanels are selected systematically, such that each subpanel represents full spatial coverage, they are considered independent samples of the population, and population estimates can be calculated from the completed subpanels of an incomplete panel. Subpanels can be subdivided further into sub-subpanels, and sub-sub-subpanels as needed to accommodate planning and implementation of the survey program.

[1] Project Leader, Research Ecologist, and Research Forester, U.S. Department of Agriculture (USDA) Forest Service, Southern Research Station, Forest Sciences Laboratory, Forest Health Monitoring, 3041 Cornwallis Rd., Research Triangle Park, NC 27709.
[2] Mathematical Statistician, USDA Forest Service, North Central Research Station, 1992 Folwell Ave., St. Paul, MN 55108.
[3] Computer Specialist, USDA Forest Service, Pacific Northwest Research Station, P.O. Box 1148; Corvallis, OR 97339.
[4] Information Management Team Leader, USDA Forest Service, Rocky Mountain Research Station, 507 25th St., Ogden, UT 84401.

Figure 1.—*National five panel design used by FIA. The national base sample has 1 plot per approximately 6,000-acre hex cell. All hexes labeled 1 are panel 1, hexes labeled 2 are panel 2, etc. Each panel comprises 20 percent of the entire FIA national base sample.*



Although the five-panel system was designed to fulfill requirements specified in the 1998 Farm Bill, the FIA program was not adequately funded to measure the target number of plots (1 per 6,000 acres) at the specified rate of 20 percent per year. Even so, this seemed to have no serious negative effects, because several strategies could be used to preserve the integrity of the original design and still satisfy the legislated requirements. This article discusses those strategies and their relationship to the legislation, and it proposes other data-collection models that adequately address the legal stipulations.

Three Data Collection Strategies

The Farm Bill directly or indirectly mandates annual data collection, compilation, and inventory updates. Because FIA always has sought to retain the capacity for design-based estimation, we maintain that the optimal way to satisfy these requirements is by managing data collection efforts to produce temporally consistent panels.

Three data-collection strategies are proposed to mitigate the consequences of inability to measure plots at the rate of one complete panel per year. Under all of the methodologies described below, data collection rules are generally applied at the State level and described as such. It is important, however, to note this is not absolutely necessary. FIA populations of interest, sometimes referred to as "estimation units," usually are defined by political boundaries, i.e., counties or national forests. These estimation units are autonomous and additive, such that State-level estimates of inventory attributes are produced by aggregating data for all the estimation units that comprise a State. Complete and uniform spatial coverage is used to spread the samples evenly over the population to increase the likelihood of unbiased data processing and estimation (Reams *et al*., in press). Because processing proceeds at the estimation-unit level, the sampling rules can be applied at this level. As long as the sampling rules result in complete and systematic spatial coverage for each estimation unit, there is no requirement that the sampling rules be uniform across estimation units within or among States.

Model 1: Creeping Panels

The creeping panel model removes the temporal restriction that each panel must be associated with exactly 1 year. Panels are started and finished based on the availability of funding and personnel. The lack of temporal restrictions allows the time required for panel completion to span several years, or it may proceed in the opposite direction, such that more than one panel is completed in a single year. Whatever the direction, any deviation from the measurement of exactly one complete panel of plots per year results in a situation that has been termed "panel creep." The FIA management team proposed using and has now adapted the operational data-collection system known as the creeping-panel sampling strategy.

There are advantages to using this model. Of all the sampling strategies, creeping panels are the easiest to implement from a data-collection standpoint. Field logistics are simple because no special planning is required. Field crews are rarely required to backtrack over the same territory, except possibly to comply with the requirement that Phase 3 plots be measured during the growing season.

The disadvantages of this model, however, are numerous and inconsistent with the goals of an annual inventory. Because sampling rules vary by State, the lack of temporal control also implies a lack of spatial control such that no systematic annual coverage of a State or county can be assured. At a given time, different States may be measuring different panels, which complicates data retrievals and analyses across multiple States. Lack of spatial and temporal controls create differences in sampling protocols that are difficult to track and manage. Two key inventory attributes, “panel number” and “manual version,” are allowed to vary by year. As field protocols change and new manual versions are released when field protocols change, there is no guarantee of consistency among States, regions, panels, or years regarding when specific manual versions are implemented. Several versions of data recorder software and processing systems must be maintained simultaneously.

Disparity between Phase 2 and Phase 3 sampling schedules may compromise FIA’s ability to combine data. For example, when panel creep is permitted for Phase 2 plots, but not for Phase 3 plots within the same panel, differences in the timing of panel measurements will be necessary. In some cases, field crews may have to use two different manual versions at the same time on a plot. These differences can quickly be exaggerated, as illustrated by the scheduling of panel measurements for the three data-collection models presented in table 1. For example, if data-collection is only funded at 80 percent, each and every Phase 2 panel is confounded with year after the first year, and Phase 2 and Phase 3 schedules are no longer synchronized (table 1).

Meeting significant Farm Bill standards cannot be guaranteed with this model. For example, data cannot be compiled annually, because panels are not scheduled for completion on an annual basis. Systematic spatial coverage across the population of interest, a prerequisite for standard processing, is not achieved using the creeping panel model. Required reports must be based on a variable number of panels if the data included represent a fixed, 5-year time interval. The 5-year reports will not be based on a 5-year interval if the data represent a complete set of panels.

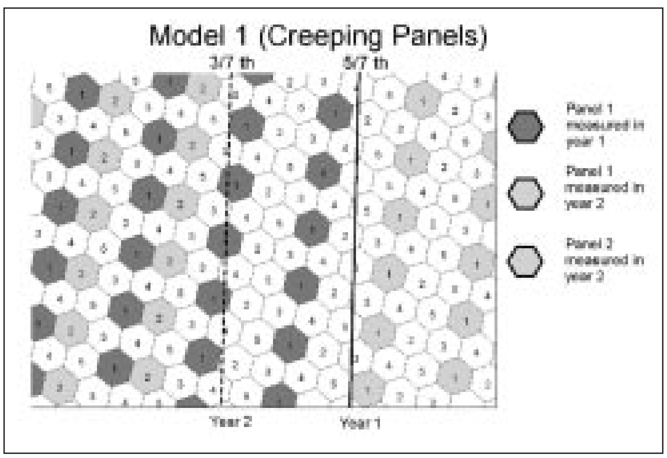
We have additional concerns about this model. If required 5-year reports are not based on a synchronic (5-year) interval, but instead include plots remeasured more than 5 years earlier, then precision estimates can be deceptive because the older data will artificially inflate the sample size. Old data are less reliable, and if substantial change has occurred in the intervening period lead to unknown bias in the inventory estimates. Usually such bias will not be reflected in the standard errors. The longer it takes to complete a panel, the greater the chance of spatially correlated measurement bias. Field crews usually start at one corner of the State, proceed until the field season ends, and then begin data collection where they finished in the next season. Figure 2 illustrates this “clumpy” approach to data collection. If a catastrophic event occurs between the first and second

Table 1.—*Hypothetical data collection schedule for each of the three models. If about 80 percent can be measured, edited, and processed for an annual database, then measure P3 as subpanel 1 and P2 subpanels 2, 3, 4, and 5. Subpanel 1 is of size 1/16th and subpanels 2 through 5 are of size 3/16th. This results in measuring 13/16th or 81.25 percent of the entire full annual panel. For more exact matching of resources and data production, the P2 subpanels could be further subpaneled by size 1/16th .*

Year	P3	P2 Data model		
		Model 1	Model 2	Model 3
1	1.1	begin 1	1.2, 1.3, 1.4, 1.5	1.2, 1.3, 1.4, 1.5
2	2.1	finish 1, begin 2	1.6, 2.2, 2.3, 2.4	2.2, 2.3, 2.4, 2.5
3	3.1	finish 2, begin 3	2.5, 2.6, 3.2, 3.3	3.2, 3.3, 3.4, 3.5
4	4.1	finish 3, begin 4	3.4, 3.5, 2.6, 4.2	4.2, 4.3, 4.4, 4.5
5	5.1	continue 4	4.3, 4.4, 4.5, 4.6	5.2, 5.3, 5.4, 5.5
6	1.1	begin 5	5.2, 5.3, 5.4, 5.5	1.2, 1.3, 1.4, 1.5
7	2.1	finish 5, begin 1	5.6, 1.2, 1.3, 1.4	2.2, 2.3, 2.4, 2.5
8	3.1	finish 1, begin 2	1.5, 1.6, 2.2, 2.3	3.2, 3.3, 3.4, 3.5
9	4.1	finish 2, begin 3	2.4, 2.5, 2.6, 3.2	4.2, 4.3, 4.4, 4.5
10	5.1	finish 3	3.3, 3.4, 3.5, 3.6	5.2, 5.3, 5.4, 5.5
11	1.1	begin 4	4.2, 4.3, 4.4, 4.5	1.2, 1.3, 1.4, 1.5
12	2.1	finish 4, begin 5	4.6, 5.2, 5.3, 5.4	2.2, 2.3, 2.4, 2.5

field seasons, population estimates derived from this panel would not accurately reflect the location or extent of resource damage. Remeasurement intervals may vary widely among panels and cycles. In the creeping panel example provided in table 1, panel 2 is measured during years 2 and 3. It is then remeasured in years 8 and 9. Thus, change will have to be computed using as many as 4 different intervals – years 2–8, 2–9, 3–8, 3–9. Because the data may have been collected in a spatially uneven manner, each interval could be associated with different environmental or cultural effects, yielding inaccurate or intractable trends for the panel as a whole.

Figure 2.—*Collecting data from Panel 1 over 2 years—a spatially clumpy approach.*



Model 2: Spatial Control

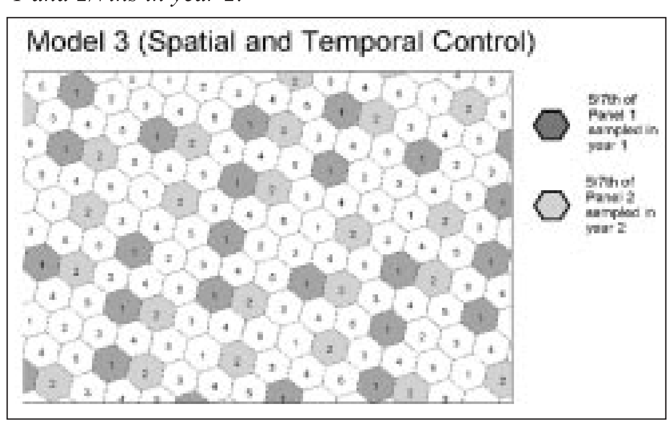
The spatial control model is an adaptation of the creeping panel model and was originally proposed by the FIA Statistics Band members and later advocated by the subteam (D-Team) of the FIA Information Band that is responsible for developing national data processing software. This model relies on subpanels to achieve systematic spatial coverage for the portion of a panel that can be measured in 1 year (Van Deusen 2003). A sufficiently large number of subpanels, each with systematic coverage of the population, are defined a priori. Phase 3 plots may be simply one of these subpanels, which would satisfy the requirement that a Phase 3 subpanel not be allowed to creep. Crews are assigned as many subpanels as can be measured in a field season. Measurement of the rest is postponed until the following year. Figure 3 illustrates a situation where five of seven subpanels are completed in year 1, the remaining two subpanels are done in year 2, and the result is systematic spatial coverage for both years.

This model offers significant improvements over the creeping panel. There is some guarantee that all States will have at least some subpanels measured and completed the same way in any given year, thus guaranteeing systematic coverage and simplifying data retrievals and analyses across multiple States. Differences in sampling protocols can be managed by establishing a rule that manual versions are linked specifically to year, to be implemented only at the beginning of a year.

More significantly, for meeting Farm Bill standards, data can be compiled on an annual basis, and required 5-year reports can be based on a fixed, 5-year interval. Also, precision is more accurately bound to sample size and less influenced by panel creep. The loss of precision caused by inadequate resources is immediately apparent and measurable, so that estimates of precision are not confounded by outdated, unreliable data. Also, annual data compilations reduce the potential for bias caused by catastrophic events. The matching of subpanels to the year in which a catastrophic event occurs eliminates the influence to an entire panel previously measured over a 2-year period.

Nonetheless, Model 2 does have drawbacks. More planning is required, which could complicate field logistics. Without careful prior planning, crews may not complete the prescribed number of subpanels or may have to make an additional pass through the State—if there is time to complete more subpanels. Inability to complete the prescribed number of subpanels might be overcome by revising the number of subpanels in the remaining estimation units to accommodate the shortage of resources. However, this complicates the tracking of which subpanels have been completed, and in which estimation units.

Figure 3.—*Using subpanels to collect 5/7ths of Panel 1 in year 1 and 2/7ths in year 2.*



Further, a less serious form of panel creep is still permitted. Although there is some guarantee that all States will have at least some subpanels completed the same way in a given year, the tracking of which panels and subpanels are measured when and where, is not straight forward. Although manual versions can be linked to years, the version used can still vary within a specific panel. Thus, disparities between Phase 2 and Phase 3 sampling schedules still exist (table 1). Assuming an 80 percent funding of data collection, as with Model 1, each and every panel is confounded with year after the initial year (table 1). Remeasurement intervals still may vary widely among panels, subpanels, and cycles. Remeasurement is not subject to any temporal control, so change estimation is not based on any fixed interval, thus complicating change analyses. Also, FIA may vary the sample size from year to year, which offers an advantage because it allows for annual database production.

Model 3: Spatial and Temporal Control

Model 3 has been proposed by members of the Statistics Band and advocated by members of the D-Team. It incorporates all of the spatial controls available in Model 2, but uses additional subpaneling to establish temporal limits as well, such that panel and year are perfectly coordinated. Model 3 differs from Model 2 in that the unfinished subpanels are simply skipped and will not be measured until the next cycle. Only plots in a single panel are measured in a given year, and the same panel is measured nationally. If a lack of resources makes it impossible to measure an entire panel for a given State, then the number of measured subpanels is adjusted accordingly. For example, the 2/7 subpanel scheduled for year 2 is postponed until the next-scheduled measurement of that panel during the next cycle (fig. 4).

Use of Model 3 reduces sampling intensity to more efficiently use limited resources. For example, the current Federal base intensity for FIA is 1 plot per 6,000 acres. If only 80 percent of that funding were available, then use of Model 3 would temporarily reduce plot intensity to one plot per 7,500 acres. When additional resources become available, the sample intensity can be increased. Intensification might be done with two strategies, or using a combination of those strategies:

1. The plot network for a given State is increased above the base sampling intensity of 1 plot per 6,000 acres.
2. Plots from future panels are temporarily assigned to the current panel.

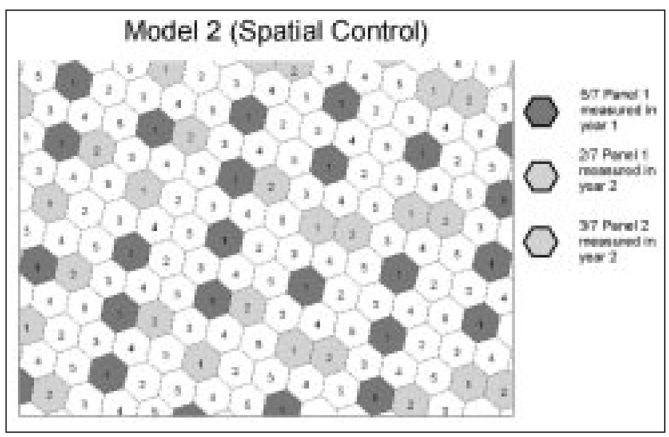
Strategy 2 has the advantage of accelerating change analysis. Suppose that a State had a short-term budget increase to measure plots in two panels for each of the next 3 years. Table 2 shows how panel 4 might be temporarily combined with panel 1, panel 5 with panel 2, and panel 1 with panel 3. This allows FIA to use the maximum number of plots for estimation of current inventory parameters, and makes change analyses possible in year 3. Choosing strategy 1 increases sample size and therefore is favored by those desiring more spatial information and precision, especially if the increased intensity can be maintained.

Model 3 has profound advantages with regard to fulfilling the Farm Bill standards. First, there is some guarantee that all States will have at least some subpanels measured and completed the same way in any given year, thus guaranteeing systematic coverage of the population and simplifying data retrievals and analyses across multiple States. Differences in sampling protocols can be managed by establishing a rule that manual versions are linked specifically to year, which automatically ties them to a specific panel. Disparities between Phase 2 and Phase 3 sampling

Table 2.—*Hypothetical data collection schedule for panel acceleration under Model 3.*

Year	Panel
1	1, 4
2	2, 5
3	3, 1
4	4
5	5

Figure 4.—*Using subpanels to collect $5/7^{ths}$ of Panel 1 in year 1 and $5/7^{ths}$ of Panel 2 in year 2.*



schedules are eliminated. Phase 3 plots are simply one of the subpanels scheduled for measurement in a given year. Most significantly, data can be compiled annually and 5-year reports can be based on a fixed, 5-year interval. Precision is more accurately bound to sample size and less influenced by panel creep. The reduced precision resulting from inadequate resources is immediately apparent and measurable; it is not confounded by outdated, unreliable data. Annual data compilations reduce the potential for bias introduced by the occurrence of catastrophic events. The catastrophic events are restricted to the year(s) in which such events occur, as opposed to influencing an entire panel over multiple years. No panel creep is permitted and no confounding of panel and year occur (table 1). Panels, subpanels, and manual versions are always linked to specific years. Remeasurement intervals are always based on a fixed interval (e.g., 5 or 10 years), if no borrowing of plots from future panels occurs. This model can readily accommodate the borrowing of plots from future panels to accelerate change analysis. The only accounting mechanism needed would be one code to designate temporary panel assignment.

The disadvantages of using Model 3 are fewer than for Model 2, but the two models share the following: More planning is necessary, which could complicate field logistics. In the absence of proper planning, crews may not complete the prescribed number of subpanels or may have to make an additional pass through the State, if there is time to complete additional subpanels. Inability to complete the prescribed number of subpanels could be overcome by revising the number of subpanels in the remaining estimation units to accommodate the shortage of resources. However, this complicates the accounting required to track which subpanels have been completed in which estimation units. Sample size is permitted to vary by year and panel, although this could be considered an advantage.

An FIA Precedent for Subpaneling

Within FIA the precedent for subpaneling, which is required when using Models 2 and 3, already has been established. To date, FIA has used Model 3 to create a Phase 3 subpanel (Subpanel 1) that is always measured annually—without creep. To accommodate the Farm Bill's requirements of annual surveys, using Model 3 we only need to decide on a reasonable Phase 2 subpaneling strategy. To illustrate, consider defining six subpanels per panel. The Phase 3 subpanel is of size 1/16th, and is labeled Subpanel 1. Subpanels 2 through 6 are of size 3/16th each and are labeled as Phase 2, Subpanels 2 through 6. If full funding for all six subpanels is available, FIA measures all six subpanels. If full funding is not available, it measures the Phase 3 Subpanel 1 and as many of the Phase 2 Subpanels 2 through 6 as possible. For example, if data from only about 40 percent of the plots can be measured, edited, and processed for an annual database, then the crew could measure all of Phase 3 (Subpanel 1), and two-fifths of Phase 2 (Subpanels 2 and 3). This results in measuring 7/16th or 43 percent of the entire panel. If data from about 80 percent can be measured, edited, and processed for an annual database, crews could measure all of Phase 3 (Subpanel 1) and four-fifths of Phase 2 (Subpanels 2, 3, 4, and 5). This results in measuring 13/16th or 81.25 percent of the entire annual panel. For more exact matching of resources and data production, the Phase 2 subpanels could be further subpaneled by size 1/16th (table 1). Subpaneling in this manner guarantees production of annual databases, as well as spatially and temporally unbiased, design-based inventory estimates.

Conclusions

The FIA program is now using Model 1 in various regions, clearly in violation of the Farm Bill mandate. Moreover, Model 1 represents the worst possible compromise between annual and periodic inventories. It is not an annual inventory because annual databases and annual design-based estimates are not possible when panel creep occurs. When panel creep occurs, Model 1 is an inefficient periodic inventory where the only advantage gained from requiring crews to backtrack over the same area five or more times during an inventory cycle is the pretense of an annual inventory. Costs for plot production, training, and multiple versions of portable data recorders, as well as editing, processing, and estimation, are excessive.

Model 2 more closely meets requirements of the Farm Bill, although it creates numerous unnecessary challenges for data management and inventory estimation. By using spatial control, Model 2 results in a less serious form of panel creep, annual databases are possible, and systematic coverage is assured, as well as annual and 5-year estimates. Although manual versions are linked to year, they can still vary by panel, and that will lead to a lack of national consistency. Also, disparities between Phase 2 and Phase 3 sampling schedules remain when using this model. Remeasurement intervals can vary widely among panels, subpanels, and cycles.

Model 3 is the only model that ensures full compliance with the Farm Bill. Subpaneling with spatial and temporal controls means one panel per year. The same panel is measured nationally. The same field manual, portable data recorder, edit system, processing system, and estimation methods can be used within and among FIA regions. This results in less duplication of effort and provides national consistency. The FIA program can produce an annual database nationally, annual estimates nationally, and periodic reports based on the required 5-year measurement period.

Additional benefits of Model 3 include the ability it gives FIA to quickly and efficiently adjust available measurement resources by State. Also, the true sampling precision per fixed time period of time is known, and intensification and detensifications are easy.

Recommendations

Data collection Model 3 is the only strategy that meets the requirements of the 1998 Farm Bill. The model provides for the greatest national consistency for sampling, database production, and inventory estimation. It will go a long way in helping FIA do its job.

Literature Cited

Reams, G.A.; Smith, W.D.; Hansen, M.H.; Bechtold, W.A.; Roesch, F.A.; Moisen, G.G. [In press.] The forest inventory and analysis sampling frame. In: Bechtold, W.A.; Patterson, P.L., eds. The enhanced forest inventory and analysis program—national sampling design and estimation procedures. Gen. Tech. Rep. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station.

Van Deusen, P.C. 2003. Issues related to panel creep. In: McRoberts, R.; Reams, G.; Van Deusen, P.; Moser, J., eds. Proceedings of the third annual forest inventory and analysis symposium; 2001 October 17–19; Traverse City, MI. Gen. Tech. Rep. NC-230. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station: 31–35.