

---

## An Investigation of Condition Mapping and Plot Proportion Calculation Issues

Demetrios Gatziolis<sup>1</sup>

**Abstract.**—A systematic examination of Forest Inventory and Analysis condition data collected under the annual inventory protocol in the Pacific Northwest region between 2000 and 2004 revealed the presence of errors both in condition topology and plot proportion computations. When plots were compiled to generate population estimates, proportion errors were found to cause underestimation of forested area. We identified reasons for the errors and offered recommendations for error prevention. Modifications were made to the condition proportion module in the compilation system as a result of these findings, and recommendations to reduce error incidence in the field in 2005 have already reduced the error rate by 70 percent.

### Introduction

The annual Forest Inventory and Analysis (FIA) protocol uses the term *condition* for a set of five mutually exclusive land use/land cover classes that describe all landscapes sampled by FIA: accessible forest land, nonforest land, noncensus water, census water, and nonaccessible land. Within accessible forest land, conditions are refined to reflect differences in vegetation density and size classes, or ownership regimes. The spatial extent of conditions is determined by ocular assessments performed in the field following rules from field protocols. Adjacent conditions are delineated by a boundary line with a maximum of two segments. Boundary line vertices are recorded in reference to annular/subplot (partial plot) centers by using azimuth and distance measurements. Where determining the boundary between two conditions is difficult

or impossible due to a gradual transition from one condition to another, the entire annular/subplot component is assigned to the condition present at its center. Field crews use hard copies of plot design templates to sketch the plot conditions before transferring all pertinent information to the data logger. Crew members have the option, usually at the end of the field day or sometimes later, of using a boundary visualization program available on their laptop computers to examine condition boundaries and make any necessary corrections. Conditions recorded in the field are also examined in the office as part of the programmatic quality assurance (QA) effort. Condition mapping errors detected during QA are flagged for follow up and resolution by data collection staff. All plot condition data is loaded into the FIA National Information Management System database (NIMS), where Procedural Language extensions to Structure Query Language (PLSQL) scripts use the condition mapping data to calculate condition proportions. These proportions are used to scale inventory sample plot data and generate State and regional population estimates.

### Problem Identification

Random examination by Pacific Northwest FIA (PNW-FIA) analysts of condition-related data in NIMS for the years 2000 to 2004 revealed topological errors, related either to the locus of condition boundaries or condition labeling. Even for plots with topologically consistent condition data, the condition area proportions calculated in NIMS were often incorrect. Because the condition proportion is the first module in the NIMS compilation and the estimates it produces are widely used by other modules, errors in condition proportion were propagated forward, affecting much of the rest of the compilation.

### Study Objectives

The study had the following objectives:

1. Determine the frequency and type of topological errors in condition mapping.

---

<sup>1</sup> Research Forester, U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, Forest Inventory and Analysis, P.O. Box 3890, Portland, OR 97208. E-mail: dgatziolis@fs.fed.us.

2. Identify circumstances that tend to generate topological errors.
3. Compute the correct condition proportions for all plots.
4. Generate maps of conditions to be used as reference for future plot visits.

To handle the large number of plots to be examined, we generated a collection of linked, custom Arc/Info Workstation Arc Markup Language (AML) scripts. All FIA plots in California, Oregon, and Washington visited during the 2000–04 period were included in the study and their related data processed by the scripts. AML-script output was evaluated and compared to information available in hard copies of plot folders and to digital data contained in the NIMS database. Recommendations for reducing condition error rates were conveyed to field data collection staff as part of field training, and to FIA analysts. All field recommendations were implemented during the 2005 field season. Post-implementation error rates were compared to previous error rates.

### Sources of Condition Errors

Condition mapping and proportion calculation are complex for several reasons. First, dense understory vegetation and/or steep terrain sometimes preclude a clear line of sight between the center and the limiting distance perimeter of the partial plot. Second, field crews must simplify what is sometimes a meandering or indistinct condition boundary to two straight line segments. In either case, condition boundaries recorded are drawn in the field by inference rather than observation, and errors are introduced, most often resulting in intersections of condition boundaries. Intersections also appear in instances in which more than two distinct conditions converge, resulting in condition boundaries that are in close proximity. During computation of condition proportions, the module used originally by NIMS used Euclidean geometry embedded in a small set of heuristic rules to convert geometric boundary information to condition proportions. Condition relationships not considered by the simple rules would produce condition proportion errors that were related to the complexity of the boundary relationship. On the other hand, condition mapping and proportion calculations based on the Geographic Information System (GIS) are analytical (i.e., data oriented),

and, hence, capable of computing proportions correctly, regardless of boundary complexity.

## Results

### Condition Topology

A total of 12,303 plots were processed by the AML scripts. Of those, 11,647 at the annular (macroplot) design and 12,251 at the subplot design had consistent, error-free topology. Nearly 42 percent of all plots contained at least one forested condition. For 17 plots, the condition boundary segment corners were found to be at a distance from partial plot centers larger than or equal to the limiting distance of the partial plot. Table 1 shows the distribution of topology errors types by plot design. Practically all errors (87 out of 98 errors, or 89 percent) occurred in plots with at least one forested condition. Error incidence on the annular design far exceeded the rate at the subplot design, a result that is not surprising given that the much larger annular limiting distance is more likely to encompass a greater number of conditions. Nearly 70 percent of errors involved boundary intersections or erroneous condition labeling. Table 2 shows the error rate by topology error type. The rate computation considered only plots with condition topology that rendered them susceptible to a particular error type. For example, intersections are possible when there are at least two boundaries on a partial plot. The approximately 20 percent of errors classified as

Table 1.—*Distribution of condition topology errors across plot designs. Brackets indicate number of plots with at least one forested condition in each error type class.*

Error type	Design		Total
	Annular plot	Subplot	
Same condition on both sides of boundary or condition label conflict	27 [25]	15 [14]	42
Condition boundary intersection	23 [22]	4 [3]	27
Condition boundary intersects (or is within 0.1 feet from) subplot center	6 [6]	3 [3]	9
Other	14 [14]	6 [6]	20
<b>Errors total</b>	<b>70 [62]</b>	<b>28 [25]</b>	<b>98</b>

other (tables 1 and 2) involved conditions which, if considered independently, violate no field protocol rules. Considered in the context of the whole plot, however, the condition delineations implied the presence of a highly improbable spatial arrangement of conditions. Figure 1 shows the annular design condition map of such a plot where condition 2, classified as accessible forest land, is confined by the remaining three non-forest conditions and cannot reasonably expand outside the plot to reach the 1-acre minimum size required by the protocol rules. Figure 1 also shows that condition 3 appears in annular plot 2 while condition 2 first appears in annular plot 3, in violation of the rule stating that the condition numbering sequence should follow the numbering of partial plots. Checking against the hand-drawn sketches on the plot cards revealed these to be cases of one or more condition labeling errors. Examining the relationship between the number of condition boundaries in a partial plot and the frequency of topology errors (fig. 2) indicates that the probability of an error increases with the number of boundaries in a partial plot. It is unclear whether the probability of an error differs between subplots and annular plots.

Condition topology errors identified via the AML scripts were collated by error type and reviewed by FIA analysts and experienced field crew leaders. The consensus was that the errors were in part due to human mistakes, but also related to ambiguities in the field instructions and poor collection habits (e.g., conversion of measured sloped distances to horizontal by assessing visually, rather than measuring, the slope gradient) carried forward from the periodic inventory protocols that

Table 2.—Error rates per condition topology error type and plot design.

Error type	Error rate (percent)	
	Annular plot	Subplot
Same condition on both sides of boundary or condition label conflict	1.6 [27/1735] <sup>a</sup>	1.4 [15/1086]
Condition boundary intersection	3.5 [23/653]	1.7 [4/236]
Condition boundary intersects (or is within 0.1 ft from) subplot center	0.3 [6/1735]	0.3 [3/1086]
Other	0.8 [14/1735]	0.6 [6/1086]

<sup>a</sup> Fractions represent the number of plots with an error divided by the number of plots where the error was theoretically (based on the number of condition boundaries present) possible.

Figure 1.—Condition mapping of an annular design plot containing logical error in the spatial arrangement of conditions.

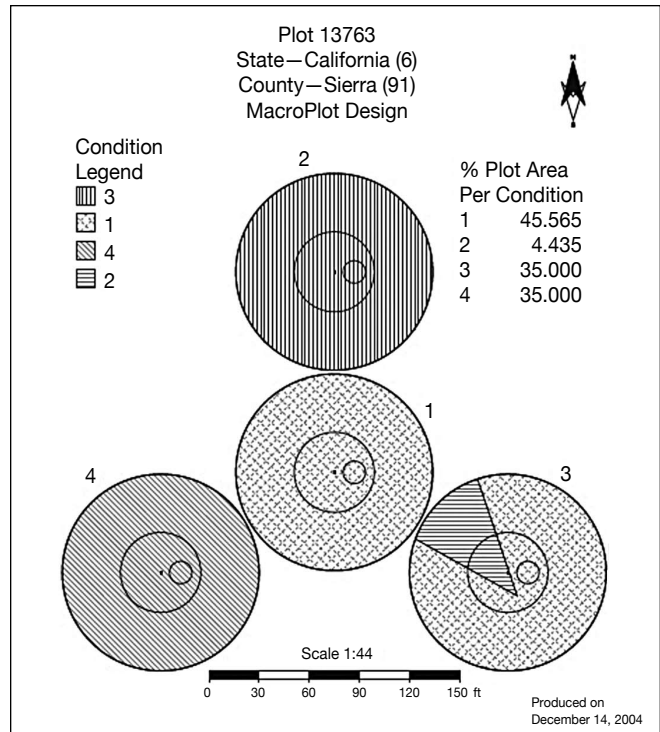
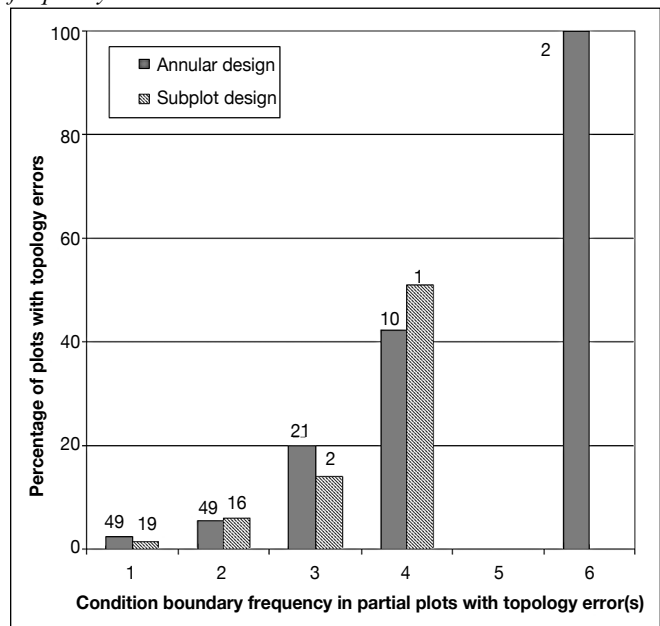


Figure 2.—Percentage of plots with topologically incorrect condition boundary by boundary frequency per partial plot and plot type and design. Numbers above the bars indicate plot frequency.



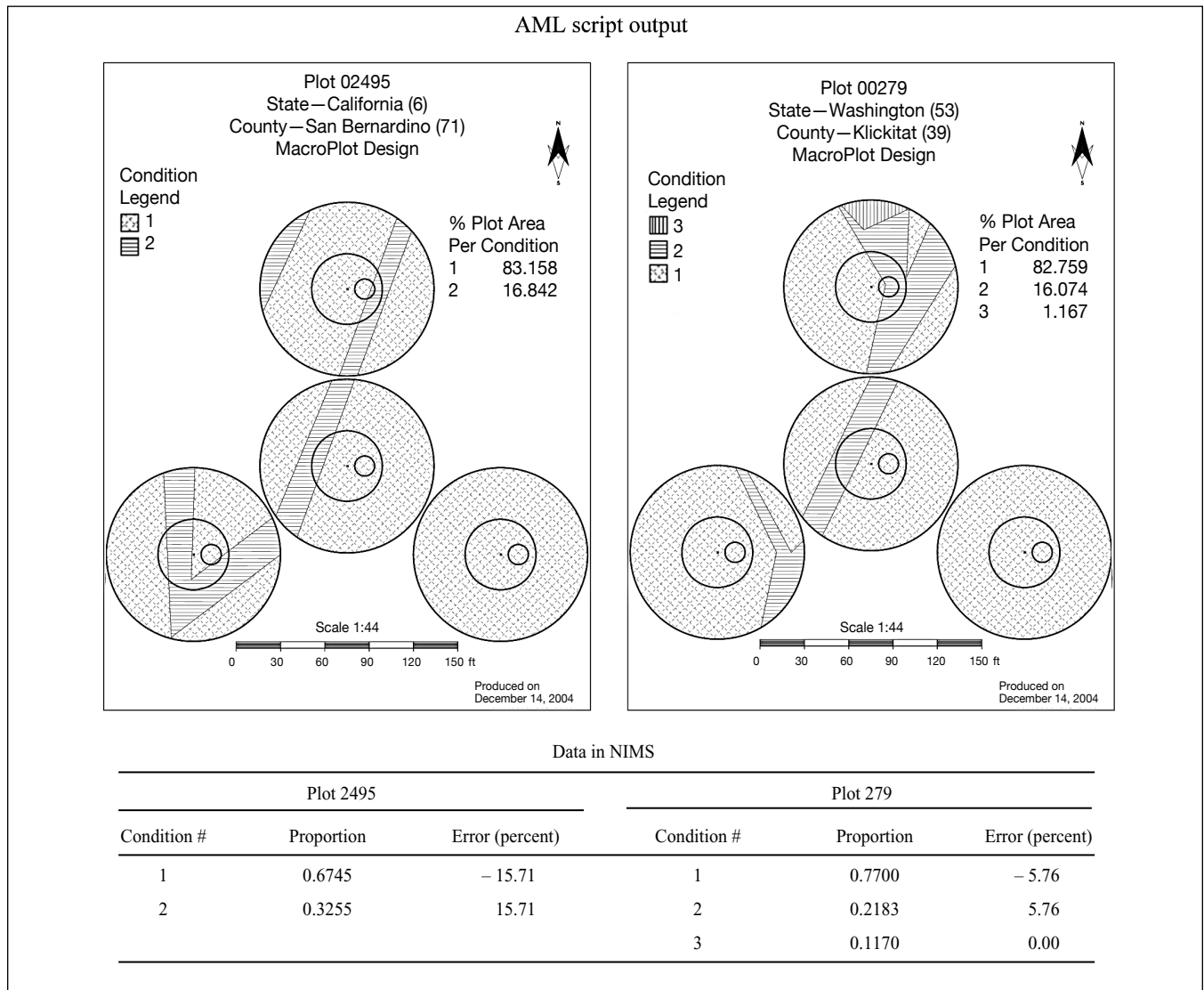
guided field data collection before the start of annual inventory at PNW-FIA. Plot maps of condition topology errors and recommendations on the proper interpretation of protocol rules were conveyed to the field personnel during the annual training sessions in 2005. Preliminary analysis of condition data acquired during the 2005 season suggests that the adoption of the recommendations resulted in a condition topology error rate approximately one-third of the rate in the previous 5 years.

### Proportion Calculation

For plots without topological errors, condition proportions computed via the AML scripts were compared to those

produced by the NIMS compilation. Proportion discrepancies smaller than 0.01 percent (of the total design-specific plot areas) were attributed to rounding error. Discrepancies exceeding 0.01 percent were found at the annular design on 61 plots and at the subplot design on 12 plots. Visual examinations of plot cards from field visits and of the maps produced by the scripts for plots exhibiting a proportion discrepancy reveal that, in all cases, the proportion computed by the AML scripts was correct (fig. 3). Henceforth, proportion discrepancies are referenced as *proportion errors*. Unlike the topology errors, proportion errors appear to be independent of topology complexity (table 3). When examined separately for forested

Figure 3.—Discrepancies between AML- and NIMS PLSQL-computed proportions for two annular design plots.



and nonforested conditions, the distribution of proportion error showed evidence of bias, with forested condition underestimated and nonforested conditions overestimated at both partial plot designs (table 4). The bias was statistically significant. Two-tail t-tests for equal mean proportion error in forested versus nonforested conditions had p-value  $\approx 0$  for the annular plot design and p-value = 0.002 for the subplot design. The test assumed unequal variances.

Documenting condition proportion errors led to rewriting the condition proportion module. The new condition module is capable of detecting the majority of topology errors and computes proportions correctly for all plots with consistent topology. A small number of errors (labeled *other* in tables 1 and 2) pertaining to improbable condition arrangement remain undetected.

Table 3.—*Proportion errors in the NIMS database computed via PLSQL modules by error magnitude and plot design. Quantities in parentheses represent the mean number of condition boundaries present in each plot. At least one forested condition was present in each plot.*

Error magnitude (percent of plot area)	Frequency			
	Annular plot design		Subplot design	
< 1	14	(2.00)	4	(4.25)
1–5	35	(2.35)	4	(1.75)
> 5–10	8	(3.00)	3	(2.50)
> 10	4	(1.75)	1	(3.00)
Total	61	(2.32)	12	(2.88)
Mean absolute error (percent plot area)	3.223		3.522	
Standard deviation	3.183		4.120	

Table 4.—*Distribution moments of proportion calculation errors for forested and nonforested conditions present in the NIMS database calculated by using PLSQL modules.*

Condition	Design	Number of plots	Error moments (percent of plot area)			
			Mean	Standard deviation	Maximum	Minimum
Forested conditions	Annular	61	-2.254	3.887	-15.708	12.378
	Subplot	11	-3.830	4.602	-15.018	0.452
Nonforested conditions	Annular	60	1.796	4.237	-12.378	15.708
	Subplot	10	2.736	4.351	-2.612	15.703

## Discussion

Observing the number of boundary intersections at the subplot design (table 1) could generate the impression that such intersections are rare. When the number of partial plots with at least two condition boundaries is considered (table 2), however, the error rate becomes noticeable, with boundary intersections present at 1.7 percent of such plots at the subplot design and twice as frequent (3.5 percent) at the annular design. This observation could partially be attributed to the reduction of visibility in the field as boundary distance from the partial plot center increases, and partially attributed to the larger area occupied by an annular plot. Errors in condition labeling and instances of boundary intersection with partial plot centers are likely independent from field visibility considerations and plot design sizes. Results indicate that the rate of occurrence for these error types does not differ between designs (table 2). Irrespective of design, an increase in the number of condition boundaries present on a plot increases the probability of condition boundary errors (fig. 2).

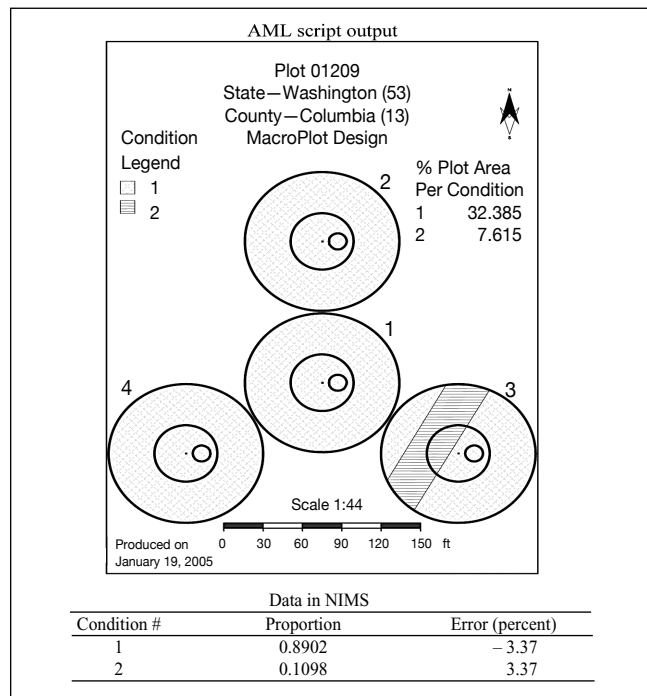
Proportion errors in a plot are complementary; an overestimation in one or more conditions must be balanced by underestimations in others. In absolute value terms, the mean proportion error of 3.2 percent corresponds approximately to one-eighth of the area of a partial plot. Visual examination of condition maps produced by the AML scripts revealed that the majority of plots with a proportion error involve a nonforested, elongated condition splitting a forested condition into two, unequally sized parts (fig. 4). The smaller of the two forested polygons (northwestern part of annular plot 3) in figure 4 is approximately one-eighth of a partial plot. This observation helped trace the cause of discrepancies observed between

condition proportions calculated via the AML scripts and those computed in NIMS (table 4) to an inaccurate PLSQL script structure. Given that only spatial information is used in the computation of proportions and assuming spatially random arrangement of condition types, there is no reason to anticipate condition-type-related bias in the errors detected. Error bias, however, is present. The observed underestimation of forest-condition proportion is likely due to the shape and position of nonforested conditions that are frequently found on partial plots in areas dominated by forested conditions. In the example shown in figure 4, the previous NIMS PLSQL module failed to appropriately account for the smaller of the two forested conditions separated by the nonforested component and, by lumping the small forested condition's area with that of the nonforested condition, generated an area bias toward nonforest.

Despite higher fidelity in condition boundaries mapped and condition proportions computed in 2005, condition errors are still present in NIMS, both in the data and compiled information. Although these errors can potentially be flagged automatically by heuristics embedded in the database compilation procedure, their logical rather than topological nature would always necessitate manual inspection of all pertinent information, an option that is laborious and costly. The best alternative is to identify condition errors while the crew is still in the field. The current data recorder used by FIA could easily employ consistency checks for condition topology and offer visualization of the condition mapping for crews to evaluate before leaving the plot. Rather than relying on abstract perception of geometric relationships between conditions, crew leaders would examine their condition maps on the data recorder while viewing the actual condition breaks on the plot. This process would virtually eliminate all of the condition-related errors identified in this analysis.

Because of differences among FIA regions, for example, in dominant vegetation biomes, land use practices, and ownership regimes, a data compilation procedure that performs well in one region may be problematic in another. Rocky Mountain

Figure 4.—Typical NIMS condition proportion error case involving an annular design plot and resulting from lumping the nonforest area with the smaller of the forested parts.



FIA analysts rarely find two condition boundaries on the same partial plot. The legacy NIMS condition module performed well under such circumstances, but it has proven inadequate for compiling more complex condition arrangements present in the Pacific Northwest. This experience suggests that caution should be exercised every time new data collection and compilation approaches are considered for adoption at the national level to ensure that such changes will not have untoward consequences for regions in which patterns or ranges of data are different.

## Acknowledgments

The author acknowledges Phyllis Adams, Olaf Kuegler, and Alison Nimura, all with PNW-FIA, for their help in this study.