A NATIONAL ANALYTICAL QUALITY ASSURANCE PROGRAM: DEVELOPING GUIDELINES AND ANALYTICAL TOOLS FOR THE FOREST INVENTORY AND ANALYSIS PROGRAM

Phyllis C. Adams and Glenn A. Christensen

ABSTRACT

A rigorous quality assurance (QA) process assures that the data and information provided by the Forest Inventory and Analysis (FIA) program meet the highest possible standards of precision, completeness, representativeness, comparability, and accuracy. FIA relies on its analysts to check the final data quality prior to release of a State's data to the national FIA database (FIADB). The analytical portion of the QA process varies considerably from region to region, aided by a variety of tools developed within each region. Recognizing that a QA process consisting of national standards would result in greater consistency and transparency, the national analytical QA task team is developing guidelines for analysts and oversees the development of QA Tools, a desktop charting and graphing application for checking compiled data prior to release. QA Tools includes a rigorous estimation engine based on national standard FIA estimation procedures. Batch processing is utilized to save time and reduce the cost of extensive QA. National consistency results in fewer resources allocated to maintaining multiple processes, thus allowing analysts to spend more time on error resolution.

INTRODUCTION

The U.S. Department of Agriculture Forest Service, Forest Inventory and Analysis (FIA) program is charged with assessing the status and trends in the Nation's forests. The FIA quality assurance (QA) program assures that the inventory produces complete, accurate, and unbiased forest information of known quality. FIA analysts play a critical role in the QA process, with primary responsibility for the quality of data released for publication. This paper outlines the flow of FIA data from the initial field planning stage through the collection, editing, storage, and compilation of data to analytical QA and publication. We highlight the role of QA within each step, with emphasis on the analytical QA stages in which analysts review the data prior to publication. Finally, we introduce QA Tools, a recently developed QA analysis tool.

Quality assurance, the overall system of management activities designed to assure that quality data are generated, is composed of two components: quality control and quality assessment. Quality control includes the operational techniques used to reduce random and systematic errors, such as identifying and adopting standards for producing quality products, training, data collection field checks, developing efficient data flow procedures, data error checking, and assuring consistency through well documented procedures guides. Quality assessment evaluates how well our established standards are met by the program, telling us whether the FIA quality control system is satisfactory. The assessment procedure compares field production data with an independent "blind" measurement to evaluate the relative uncertainty associated with FIA field collected data.

FIA DATA FLOW AND QUALITY ASSURANCE

The FIA QA program includes comprehensive quality control and quality assessment steps from prefield preparation and field data acquisition to processing within the National Information Management System (NIMS) and post compilation analytical QA procedures that prepare the data for posting in public databases. Prior to presenting details about analytical QA, the primary focus of this paper, we describe the overall FIA data flow process. The flow of FIA data and the integral QA components are described through seven stages: 1) Prefield, 2) Data collection, 3) Precompilation data editing, 4) Stratification, 5) Data compilation in NIMS, 6) Analytical QA, and 7) Data delivery (fig. 1). As the FIA program evolved into a national program with the advent of the annual inventory design, procedures became more consistent. For example, a nationally consistent data collection protocol was defined and documented in national FIA field manuals and NIMS was implemented to store and process FIA data. National

Phyllis C. Adams, National FIA Quality Assurance Coordinator and Glenn A. Christensen, Forester, USDA Forest Service, Forest Inventory and Analysis, Portland, OR 97205.

consistency improves operating efficiency and the ability to deliver higher quality data. Furthermore, inconsistent protocols and QA procedures make assessments of the accuracy and repeatability of each regional program difficult to compare. Process documentation that increases consistency across FIA regions is described for each stage and diagrammed in figure 1.

PREFIELD PLANNING

The FIA data flow process begins with the prefield stage. This stage includes activities that occur in advance of field visits to FIA plot locations, such as land-use interpretation based on high-resolution imagery, assembling hard copy materials in support of the field staff, and verifying plot coordinates. Prefield programs have developed independently at each of the four FIA regions, leading to parallel, yet different, procedures, QA processes, tools, and database items. For example, multiple tools have been created to solve similar problems, and overlapping software programs have been developed. FIA land use definitions are not necessarily applied consistently across regions. Regional QA protocols generally include prefield determination of plot visit/nonvisit quality assessment. National prefield protocols and QA procedures are under development and will be documented in a prefield manual, adding consistency and transparency to the process.

DATA COLLECTION

Data are collected on FIA plots using national protocols. Field procedures and protocols are documented in the national FIA Field Manual. QA procedures in this stage have been formalized and are well documented in national and regional FIA QA plans. The QA effort is primarily focused on error control during the field measurement and data collection process. This is accomplished through extensive crew training, documentation of protocols and procedures used in the inventory, and field checks.

PRECOMPILATION DATA EDITING

Raw field collected data undergo three different checks prior to database compilation: personal data recorder (PDR) logic checks, postfield data cleaning edits by field or office staff, and NIMS load error checks. Field data are collected using field data recorders that are programmed with multiple logic checks applied during data entry in the field. In 2007, a national task team was formed to develop the FIA Mobile Integrated Data Acquisition System (MIDAS), a data entry program that is used by all regions to collect both national and regional data. The MIDAS system is being integrated into the overall FIA information management structure. Postfield procedures that include various algorithms are used for checking and editing raw data. These procedures vary from region to region. Finally, prior to populating the NIMS database, additional NIMS load checks are applied to the data by the NIMS system.

STRATIFICATION

Stratification is applied to the inventory data prior to compilation to reduce the variance of estimates associated with forest land area and volume. Field data are grouped into broad strata using the stratification approach outlined in Bechtold and Patterson (2005). The number and type of strata used in the stratification process vary with regional requirements. Strata may include forest/nonforest, public/private ownership, or boundaries such as state, county, and ecological unit. High-resolution imagery and ownership maps are used to classify sampled plots into each stratum. Quality control and assessment of the stratification process is completed regionally, including checking plot locations, matching total area estimates to those provided by the Census Bureau, and double-checking boundary locations on GIS layers.

NIMS DATABASE COMPILATION

NIMS compilation includes calculating a variety of derived variables using estimation equations defined by Bechtold and Patterson (2005). NIMS design and implementation procedures as well as the compilation process are described in various unpublished documents produced by national implementation teams. Data checks of tables are performed informally on the database prior to analyst review. Data checks include an evolving list of data queries to identify errors. These include checks for missing values and checks for consistency between populated values of related variables, such as site index, stand age, and stand size class.

ANALYTICAL QA

After the data are compiled and summarized, FIA resource analysts conduct regional postcompilation checks and audits to ensure that core FIA data are complete and as accurate as possible before publication in FIADB (Miles and others 2001). There is presently no formal review process for all FIA analysts to follow, although each region's QA process generally includes a variety of standard checks and additional analyses that may be needed for unique resource issues. Documentation to define a rigorous nationally consistent process is being drafted. Additionally, tools have been and/or are being developed within each region to help analysts with QA.

DATA DELIVERY

Analyses and reports about status, trends, and location by land use class are generated as annual and 5-year reports and special reports. Data are posted to FIADB, FIA's public Web based database, with its suite of reporting tools. National FIADB tables are populated with regional data and displayed in standardized output tables. Detailed documentation is available in a Database Description and Users Manual (Miles 2008). Additional reporting includes peer-reviewed journal articles, articles in trade journals, extension publications, and other user-oriented outlets.

ANALYTICAL QUALITY ASSURANCE

The primary task of FIA analysts is to interpret and extract meaning from compiled inventory data and to communicate these findings to the public. Each State is assigned to an analyst who checks the State's compiled data prior to release to the public as FIADB tables, as annual and 5-year reports, or as peer-reviewed publications. Analysts must be experts on forestry and ecological issues in their assigned States and regions. They conduct research that adds value to the basic inventory products, and they participate on national FIA task teams.

FIA analysts play a critical role in the QA process; they are responsible for checking compiled data to assure the quality of data released to the public and the accuracy FIADB and FIA's annual and 5-year State reports. This is not a trivial task; in spite of extensive precompilation error checking; some errors are impossible to detect until analysts have attempted to interpret the numbers. After analysts conduct annual QA checks of information, they submit core tables to State contacts for annual review.

This analytical QA process varies considerably from region to region. Typically, FIA analysts use many different approaches, ranging from reviewing summary tables for suspicious values and checking map products carefully to a variety of in depth investigation methods. As the FIA program has expanded and moved to a fully annualized inventory, the need for a more thorough and efficient approach to data QA has increased dramatically. Thus, the national FIA analytical QA task team has proposed a nationally consistent process with standard minimum requirements for checking data prior to release for publication. The team's objective is to increase efficiency with well documented quality control standards and a consistent analytical QA process. Guidelines and checklists for analysts will outline routine QA steps required prior to releasing data to the public.

Analysts' toolboxes contain a variety of tools for data checking, including algorithms developed within each region. For example, FIA analysts may use a statistical application such as SAS (2002) to perform specific data queries and summaries. The same analyst may then use another application such as Microsoft (MS) Access that performs different summaries with the same data. Experienced analysts know what sorts of QA checks to apply to these data, and they generally have extensive knowledge of field collection methods to detect possible errors. But, what happens if the analyst is inexperienced with FIA protocols and the necessary understanding of how to query and summarize these data is not fully understood? Data errors may be missed and summary estimates may be incorrect. Each analyst within the four FIA regions of the United States shares these analytical QA challenges. Similar analytical QA checks are repeated each year on that year's data. In this manner, analyst amasses a collection of queries that can summarize estimates and provide sampling errors from tables within NIMS or FIADB. They must evaluate both measured data collected in the field and compiled estimates (such as volume and biomass estimates). Attempting to streamline this process, the analytical QA team has overseen the development of QA Tools, a charting and graphing application for checking compiled data prior to release. QA Tools is a software application that addresses these issues and moves the FIA program toward a more thoughtful and comprehensive approach to analytical QA.

QA TOOLS: TOWARD A COMPREHENSIVE APPROACH TO ANALYTICAL QA

QA Tools is a desktop application developed as a joint effort by FIA analysts, programmers, and statisticians working toward a single integrated solution that provides a rigorous estimation engine based on national FIA statistical estimation procedures as described in Bechtold and Patterson (2005). The application allows an analyst to methodically, consistently, and efficiently perform QA checks and document the QA checks that were performed. The application can be quickly modified to reflect any changes in data storage, collection, and processing procedures within FIA. Batch mode processing improves efficiency by producing a predefined set of output reports, such as tables of estimates and graphics, during a single session. In setting up a specific QA query, the user can customize any input or output tables and charts. FIA data are brought into the tool directly from NIMS Oracle tables, as FIADB formatted tables, or as custom MS Access tables. User 'wizards' are utilized for quick access and ease of use in defining specific queries and outputs (fig. 2). The current version of the application works with FIADB 4.0 (Woudenberg and others 2010) structure, however it is not currently designed for general public use, but rather accesses prereleased data to perform analytical QA procedures. Any FIA analyst can load standard (or customized) tables and perform the same set of QA checks on data from any State in the country.

ESTIMATION ENGINE

A key strength of QA Tools is the built-in estimation engine that uses estimation procedures as described by Bechtold and Patterson (2005). Prior to QA tools, an analyst who wanted to calculate summary estimates not included as part of standard output tables had to understand the NIMS table structure and estimation procedures well enough to make this calculation. This is not an easy task for new analysts. In addition, a further level of statistical understanding is required to produce error estimates. Each analyst develops their own method to perform these calculations, all with

varying degrees of complexity and ease of customizing for specific inquiries. QA Tools provides the user with easy access to a complex database and a simplified approach for obtaining estimates with sampling errors. Interface 'wizards' guide the user through a choice of either basic estimation or more advanced reporting. Advanced estimation calculations include adding custom tables, user defined variables, and custom sub-groupings for reporting. With the estimation engine in QA Tools an analyst can also more easily evaluate FIA data as a ratio of means (e.g., estimates on a per acre basis).

BATCH PROCESSING

In addition to having the built-in estimation engine, another key strength of QA Tools is the ability for batch processing. The batch process approach allows each analyst to quickly produce tables and charts for error checking and QA documentation. QA Tools includes a 'wizard' to define and simplify a batch process session. Output tables and graphics are defined in report definition files (RDF) that can be saved for subsequent batch processing. These RDF files are loaded into a batch session and processed at one time. Each RDF file is simply a set of Extensible Markup Language (XML) code that tells QA Tools which input data to use, which statistical estimates or summaries are required, any filters applied to the data, format of the output table, and how to build each output graphic requested. The user can select an output table and/or graph for each report (fig. 3), and can designate where all input and output files are stored. Tables are output as MS Excel files and graphics are saved as either Joint Photographic Experts Group (JPEG) or Portable Document Format (PDF) files. Current and future releases of QA Tools include an initial set of RDF files that produce some of the nationally defined FIA tables.

With batch processing in QA Tools, analysts can now easily perform many QA checks by producing output graphics and tables, such as scatter and bar charts, for review. Scatter charts are useful for rapidly evaluating many individual measurement relationships at the tree level, such as diameter and height (fig. 4). Scatter charts also allow the analyst to quickly spot outliers or other unusual patterns in the data. Bar charts are another output graphic type available in QA Tools. Annual estimates of attributes such as forest area, tree volume, and tree biomass and their statistical reliability, can be depicted in bar charts if year by year regional stratification files were produced (fig. 5). A chart 'wizard' within QA Tools is available to help walk the user through the bar chart setup and evaluation process. Bar charts are useful as a QA check for errors or unusual changes over time within summary groups or series of estimates. Each batch process session can be saved as a Batch Definition File (BDF) which is essentially a set of RDF files that are processed at one time. Any BDF can be later loaded into QA Tools and reprocessed with the same dataset, applied to another dataset such as an added panel of field measurements or a dataset from another State.

CONCLUSION

The objective of the analytical QA task team is to promote the collection and dissemination of high quality FIA data through the use of a nationally derived QA process. The task team is developing guidelines to facilitate this process and is promoting the development and use of analytical tools that support this objective. A nationally derived approach to QA promotes a consistent process across all FIA regions leading to the delivery of quality data products with known precision, completeness, representativeness, and comparability. Efficiencies are gained from national consistency because fewer resources go into maintaining separate processes, leaving more resources available to look at the data in greater detail and from many directions.

Consistency facilitates a nationally cohesive program through thorough documentation of FIA processes and methodologies that benefits quality control and leads to greater transparency. Examples include a draft guide to prefield procedures, data collection protocols outlined in a national FIA field manual, a national PDR program (MIDAS) with accompanying documentation and error checking logic, and published FIADB documentation and user guide. In addition, changes and updates in NIMS procedures are applied more consistently when accompanied by documentation. Data delivery procedures and tools, such as standard tables and FIA Web tools have been developed to assist clients in obtaining information.

Consistency is also achieved through the development of QA Tools, a desktop application available to FIA analysts for performing analytical QA of unprocessed data and compiled estimates. The estimation engine within QA Tools provides a quick method to calculate summary estimates and sampling errors. Efficiencies are realized by using QA Tools, especially with the batch processing capability, enabling a large set of predefined analytical QA checks to be completed expeditiously each year.

ACKNOWLEDGMENTS

In addition to the authors, other members of the analytical QA team contributed information for our description of the data flow and analytical QA processes. They include Randy Morin (Northern Research Station), Roger Conner (Southern Research Station), and Mike Thompson (Rocky Mountain Research Station). Olaf Kuegler provided extensive statistical expertise in developing the estimation engine. We especially want to recognize the work of Larry

Potts, the QA Tools programmer who devoted many hours to programming and modifying the application to meet the needs of the analysts.

LITERATURE CITED

Bechtold, W. A.; Patterson, P. L.; eds. 2005. The enhanced forest inventory and analysis program - national sampling design and estimation procedures. Gen. Tech. Rep. SRS¬80. Asheville, NC: U.S. Department of Agriculture Forest Service, Southern Research Station. 85 p.

Miles, P. D.; Brand, G. J.; Alerich, C. L.; Bednar, L. F.; Woudenberg, S. W.; Glover, J. F.; Ezzell, E. N. 2001. The forest inventory and analysis database: database description and users manual version 1.0. Gen. Tech. Rep. NC¬218. St. Paul, MN: U.S. Department of Agriculture Forest Service, North Central Research Station. 130 p.

Miles, P. D. 2008. A simplified forest inventory and analysis database: FIADB-Lite. Gen. Tech. Rep. NRS¬30. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station. 42 p.

SAS Institute. 2002. The SAS system: SAS OnlineDoc®, Version 9.2, HTML format][CD-ROM]. Cary, NC.

U. S. Department of Agriculture Forest Service, Research and Development. 1998. Quality assurance implementation plan. Washington, DC: U.S. Department of Agriculture Forest Service. 24p.

Woudenberg, Sharon W.; Conkling, Barbara L.; O'Connell, Barbara M. [and others]. 2010. The Forest Inventory and Analysis Database Version 4.0: Database Description and Users Manual for Phase 2. Gen. Tech. Rep. RMRS-245. Fort Collins, CO: U.S. Department of Agriculture, Forest Service. Rocky Mountain Research Station. 355 p

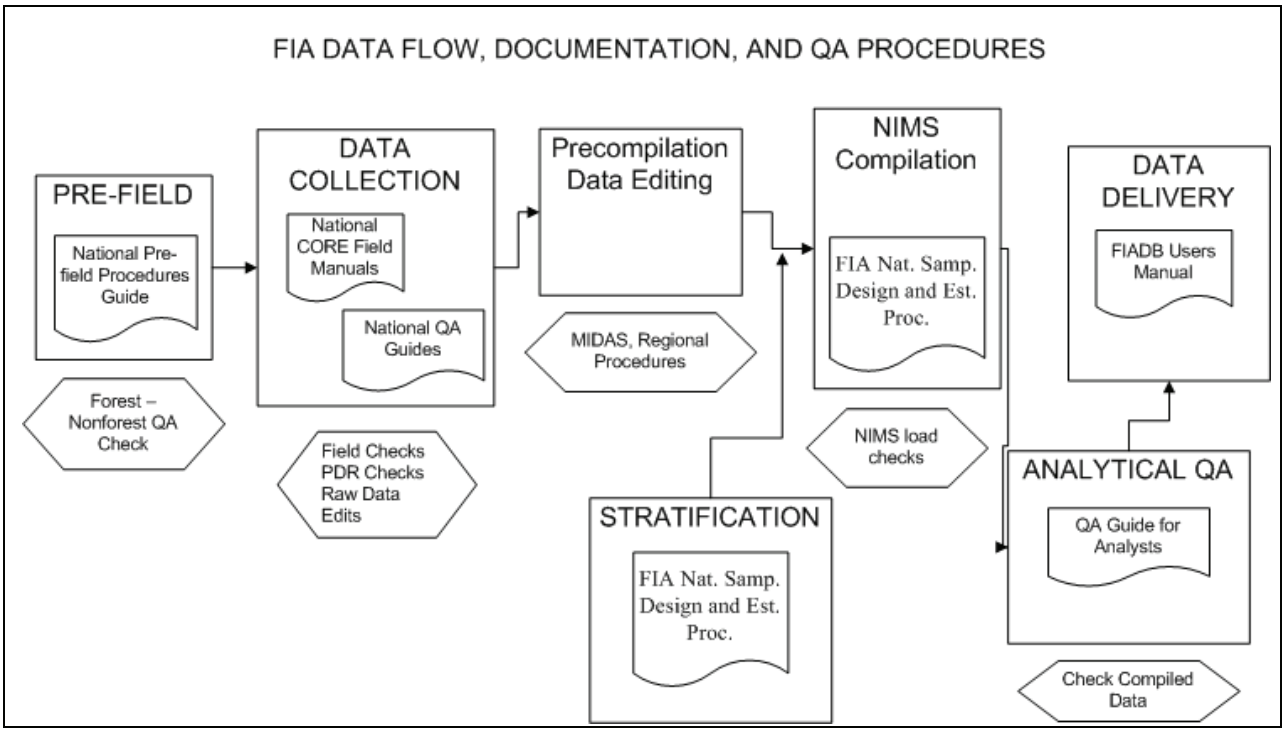

Figure 1—FIA data flow process showing the seven basic stages with examples of QA processes and documentation.

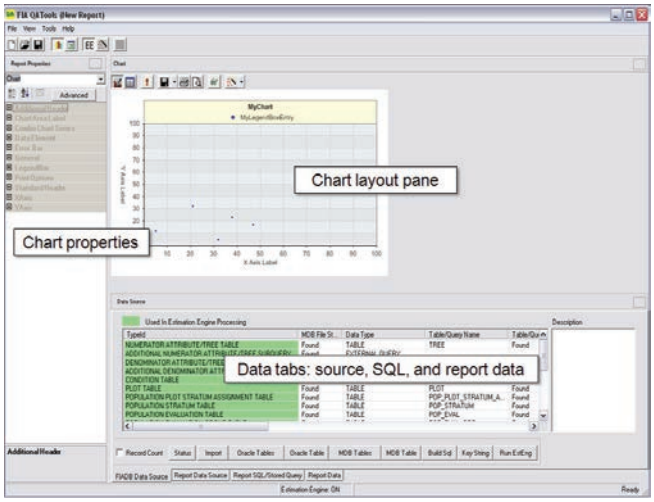

Figure 2—QA Tools “New Report” layout shows the user interface when defining a new analytical QA check.

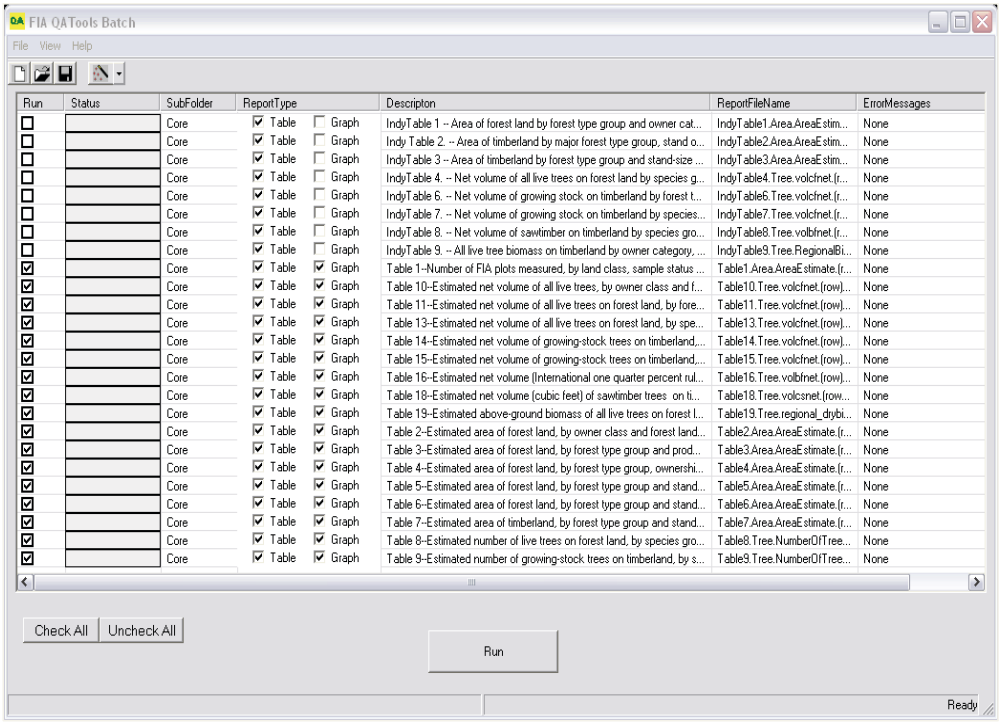

Figure 3—QA Tools “batch process” set-up window shows a group of selected, pre-defined QA checks that will be output as both tables and charts during this session.

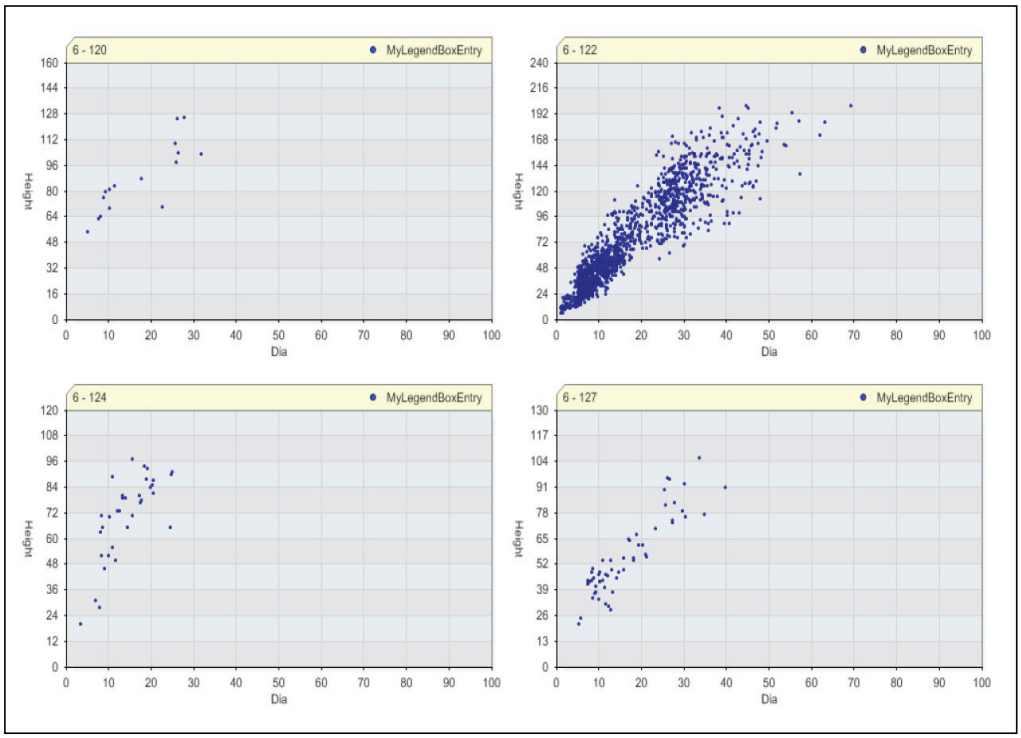

Figure 4—QA Tools scatter charts allow analysts to quickly evaluate many measurements for obvious data errors.

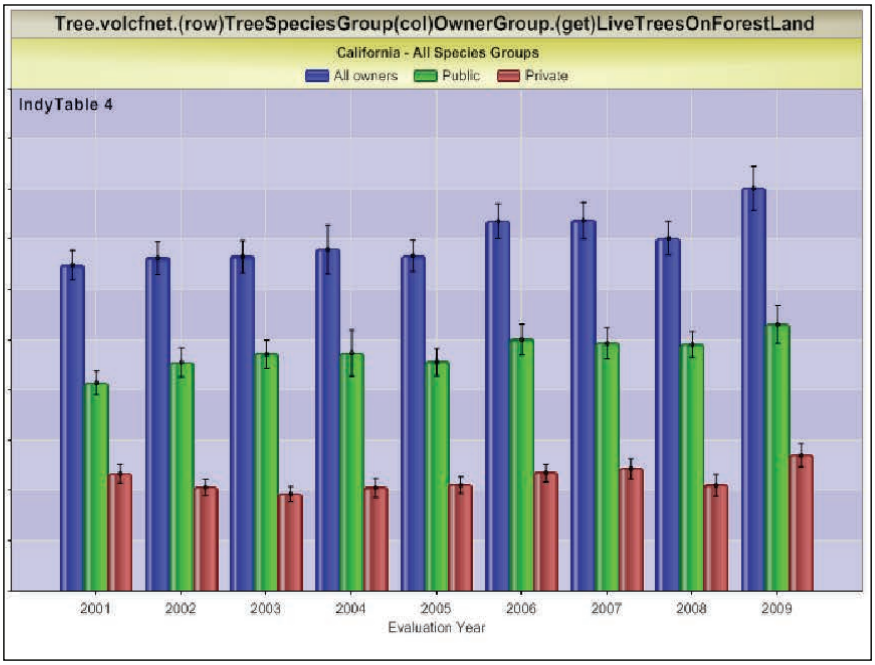

Figure 5— Bar charts from QA Tools enable an analyst to evaluate a current estimate against previously collected data, providing a quick method to spot potential errors.