COMBINING ACCURACY ASSESSMENT OF LAND-COVER MAPS WITH ENVIRONMENTAL MONITORING PROGRAMS

STEPHEN V. STEHMAN[1], RAYMOND L. CZAPLEWSKI[2], SARAH M. NUSSER[3], LIMIN YANG [4] and ZHILIANG ZHU[4]

[1]*SUNY ESF, 320 Bray Hall, Syracuse, NY 13210 USA (svstehma@mailbox.syr.edu);* [2]*USDA Forest Service, Rocky Mountain Research Station, 240 West Prospect Road, Fort Collins, CO 80526 USA;* [3]*Department of Statistics, Iowa State University, Ames, IA 50011 USA;* [4]*USGS EROS Data Center, Sioux Falls, SD 57198 USA*

Abstract. A scientifically valid accuracy assessment of a large-area, land-cover map is expensive. Environmental monitoring programs offer a potential source of data to partially defray the cost of accuracy assessment while still maintaining the statistical validity. In this article, three general strategies for combining accuracy assessment and environmental monitoring protocols are described. These strategies range from a fully integrated accuracy assessment and environmental monitoring protocol, to one in which the protocols operate nearly independently. For all three strategies, features critical to using monitoring data for accuracy assessment include compatibility of the land-cover classification schemes, precisely co-registered sample data, and spatial and temporal compatibility of the map and reference data. Two monitoring programs, the National Resources Inventory (NRI) and the Forest Inventory and Monitoring (FIM), are used to illustrate important features for implementing a combined protocol.

Keywords: data confidentiality, multiple frame estimation, photo interpretation, probability sampling

1. Introduction

Land cover is an important geographic attribute relevant to many management and research problems in natural resources. Mapping land cover via remotely sensed data is one of the most practical methods for obtaining land-cover information over broad regional scales at relatively frequent time intervals. Users require map accuracy information to determine the adequacy of a land-cover map for their particular objectives. Accuracy is determined by comparing the map land-cover label to the "true" or "reference" land cover at a sample of sites. An accuracy assessment protocol consists of three major components (Stehman and Czaplewski 1998). The response design is the methodology for obtaining the reference land-cover label. The response design may be based on ground visits to the sample locations, or on interpretation of aerial photography or videography. The sampling design is the protocol for selecting the locations at which the reference land-cover label will be determined. Ideally, the sample is a probability sample from the full land-cover map. A probability sample is defined as a sample in which each element (e.g., pixel or polygon) has a known probability of being included in the sample, and all elements of the population have a non-zero probability of being included in the sample. The third

Environmental Monitoring and Assessment **64**: 115–126, 2000.

component of accuracy assessment is the analysis, which includes estimating various accuracy summary measures. Typically an error matrix is constructed and overall, user's and producer's accuracies are reported (Story and Congalton 1986).

Travel, field crew training, and equipment all contribute to the high cost of accuracy assessment. Assessments conducted with a limited budget may suffer from several deficiencies. For example, small sample sizes often result in imprecise estimates for the rarer land-cover classes. Poor quality GPS units may produce larger errors in registering the field sample locations with their corresponding map locations, thus increasing the proportion of apparent classification errors that are, in reality, attributable to registration error. Insufficiently trained field crews or photointerpreters may produce low quality reference data compromising the integrity of the assessment. To save costs, sometimes probability sampling methods are abandoned entirely, creating difficulties for rigorous inference, or restricted to readily accessed locations such as roadsides or public lands not representative of the full map population.

Environmental monitoring programs offer a potential source of high quality reference data to alleviate some of the cost burden of accuracy assessment. We will focus on two such programs, the National Resources Inventory (NRI) conducted by the U.S. Department of Agriculture's Natural Resources Conservation Service (Nusser and Goebel 1997), and the Forest Inventory and Monitoring (FIM) conducted by the U.S. Department of Agriculture's Forest Service (USFS, 1992). Both programs employ probability sampling designs over spatially extensive areas of the United States and adhere to strict data quality and documentation standards.

The ability to integrate environmental monitoring and accuracy assessment protocols would draw from the strengths of both activities and potentially yield a significant cost reduction for accuracy assessment of large-area, land-cover maps. If environmental monitoring programs can contribute suitable quality accuracy assessment data, it is more likely that rigorous accuracy assessment will become a standard component of mapping projects. The purpose of this article is to describe strategies for combining land-cover map accuracy assessment with environmental monitoring sampling efforts. Three general strategies are described and evaluated focusing on the application of the environmental monitoring data to the response design, sampling design, and analysis components of the accuracy assessment protocol. The merits of each strategy depend upon the role of land cover/land use maps to agency objectives. These maps are important inputs to management decisions for public lands and to extension activities that promote improved management of private lands.

2. General Strategies for Combining Monitoring and Accuracy Assessment Protocols

The strategies for combining accuracy assessment and environmental monitoring data represent a gradient of integration from virtually none to almost complete integration. In a non-integrated strategy, no changes are imposed on the monitoring program's sampling and response designs. The monitoring data are used in whatever way possible to contribute to accuracy assessment, but the monitoring agency would not change protocols to accommodate accuracy assessment objectives. This strategy, described in detail in Section 4, is applicable when the land cover mapping project is independent of the activities of the agency conducting the monitoring program.

A more integrated strategy would treat accuracy assessment as an add-on or attachment to monitoring. The primary sampling objectives would remain those of the monitoring program, and these objectives drive sample design planning. The monitoring protocol could be augmented to include an accuracy assessment response design protocol to obtain reference data. Because the monitoring program's sampling design would not necessarily provide sufficient data for the accuracy objectives, additional sampling, either by augmenting the monitoring program's sampling design or by implementing an independent accuracy assessment sampling design, may be required.

A fully integrated accuracy assessment and environmental monitoring strategy could be applied within the three-tier monitoring framework proposed by the Committee on the Environmental and Natural Resources (CENR) (Bricker and Ruggiero 1998, NSTC 1997). The three tiers of the CENR framework are: 1) *remotely-sensed data* (including satellite imagery, aerial photography, and videography) generating geographically extensive land-cover information, 2) an extensive *probability sample* of ground-based measurements, and 3) *intensive study sites* at which experiments are conducted and models developed. In an integrated strategy, both accuracy assessment and environmental monitoring objectives are incorporated when constructing the sampling and response design protocols. An advantage of this strategy is that the accuracy assessment objectives are formulated based on the intended applications of the land-cover map. For example, the accuracy assessment may be designed to evaluate accuracy of landscape indicators of ecological condition or to increase attention on those land-cover classes most critical to the validity of an intensive study site model. This direct link of accuracy assessment objectives to applications is more difficult to forge if the mapping effort is conducted separately from the monitoring program. The integration allows more efficient use of resources. Field crews collecting monitoring data could simultaneously collect reference data for accuracy assessment at sites selected via a probability sampling protocol.

3. Components of an Accuracy Assessment Protocol

3.1 RESPONSE DESIGN

The response design provides the information for assigning a land-cover label or labels to the reference sample location. The land cover/land use classification system is one of the key issues to resolve when combining accuracy assessment and environmental monitoring protocols. The FGDC vegetation subcommittee has proposed a National Vegetation and Information Standard (FGDC 1997) which could serve as a common vegetation classification system for accuracy assessment and monitoring. The National Vegetation and Information Standard is a hierarchical, systematic, and unambiguous classification system for vegetation cover of large geographic extent. The system encompasses a comprehensive list of a large number of vegetation types based on plant physiognomy (life form, structure and phenology) in its upper level classes and the dominant vegetation species for its lower level classes. The advantages of adopting the National Vegetation and Information Standard for a combined accuracy assessment and monitoring strategy is that vegetation data can be organized hierarchically and then aggregated to an appropriate level to meet requirements of accuracy assessment. A limitation of the current National Vegetation and Information Standard is that this system only defines vegetation cover type, and purposely avoids defining land use. Therefore for many land cover/land use mapping projects typical in the remote sensing community, certain types related to land use (e.g., urban areas) are not available in this system.

Another commonly used classification system is the USGS land cover/land use system developed by Anderson et al. (1976). This system is especially designed for land cover/land use mapping from remotely-sensed data, and it includes several land use categories related to urban, agricultural and range lands. A limitation of this system is that it does not differentiate cropland from pasture at either its first or second level. Conceptually, if the goal is to create a comprehensive land cover/land use system for both environmental monitoring as well as accuracy assessment, it is desirable to use a compatible approach based on application of the FGDC system for land cover and the USGS system for land use.

Some experience addressing differences in land-cover classification schemes has been gained from the Oregon demonstration project (Goebel et al. 1998, House et al. 1998) in which the integration of FIM and NRI protocols was evaluated. Differences in definitions between the two programs made it difficult to construct field protocols for determining cover type. Consequently, field crews were instructed to record data in a format that could be used later for classification according to different definitions, including the land-cover classes of both NRI and FIM (House et al. 1998). The use of joint NRI and FIM photo interpretation teams in the Oregon demonstration project was advantageous because of the diversity of experience of the team members, and the resulting common interpretation of the standard protocols (House et al. 1998). A similar cooperative approach will be necessary to implement a protocol allowing monitoring data to be used for accuracy assessment.

In addition to the land-cover class definitions, several other features of the response design need to be resolved. The spatial scale of the monitoring data must be compatible with the spatial scale of the unit chosen for accuracy assessment. For example, if the accuracy assessment unit is a 30 x 30 m pixel and the monitoring data are collected on a 112 m diameter (1 ha) plot, it is unlikely that the data will be appropriate for a site-specific accuracy assessment at the 30 m pixel resolution. In some cases, a non-site-specific accuracy assessment may be possible when the monitoring data represent a larger spatial unit than the desired map unit for accuracy assessment (see Section 4). Selecting a spatial scale compatible for both environmental monitoring and accuracy assessment data depends on several features, including the spatial extent of the land cover map, the minimum mapping unit, and the geometric accuracy of the remotely sensed data. Given the current and emerging technologies and missions acquiring remotely sensed data (e.g., Landsat 7 ETM and the EOS-AM 1 mission), the spatial resolution of optical sensors in general ranges from 30 to 1,000 m. Finer spatial resolution data are likely to be useful only for very localized studies, and the contribution of environmental monitoring data to accuracy assessment in these situations is probably minimal because the necessary spatial detail will be lacking to distinguish the land cover/land use classes at this resolution. Therefore, it can be argued that an assessment unit of approximately 1 acre (about 4 TM pixels or 0.36 ha) is probably a viable size to accommodate accuracy assessment of most large-area land-cover maps.

Precise co-registration of the map and ground coordinates of the reference sample locations is critical to ensure that the comparison between the map land-cover label and the reference label is made for the same spatial location. With the development of improved and lower cost GPS technology, achieving spatial accuracy of less than 5 meters in the horizontal dimension in the field is possible. This degree of geometric accuracy is adequate for meeting large-area, land-cover mapping using medium spatial resolution remotely sensed data (e.g., Landsat) and other coarser resolution data (e.g., AVHRR). The field data collection protocol of the monitoring program needs to be implemented to ensure spatial accuracy to this level.

The temporal compatibility of reference data with mapped data is always an issue for accuracy assessment. It is unlikely that the monitoring data will provide an exact temporal match. The concern is not only with changes in land cover attributable to a difference in years between the mapping date and the monitoring data, but also with differences in land cover within a growing season. Temporal compatibility depends on the land cover/land use type. For example, for crop land and hay/pasture assessment, a within-growing season difference between the reference data and the mapped land cover is probably more critical than a difference between years. Conversely, for assessment of forest versus transitional (clear cut) or in cases of urban expansion, a time difference in years will be more important than a within-season difference. Careful evaluation of the temporal compatibility of data should be made before using the data for accuracy assessment.

3.2 SAMPLING DESIGN

Satisfying the probability sampling criterion is required of the monitoring program if the data are to contribute to accuracy assessment. The population coverage of the monitoring design is another relevant concern. For example, the NRI frame covers all land within the United States, but sample points on Federal land are not evaluated for land cover. FIM coverage focuses on forested land and therefore does not provide complete spatial coverage for accuracy assessment. Because existing monitoring program sampling designs were not constructed with accuracy assessment objectives in mind, some of the rare land-cover classes important to accuracy assessment may not be sampled in adequately large numbers. Consequently, the objective of providing class-specific accuracy estimates may be poorly satisfied because of high variability of these estimates.

Milliken et al. (1998) and Gill et al. (1999) provide a good illustration of using environmental monitoring data for accuracy assessment. Their study assessed the accuracy of several attributes of a vegetation mapping project for the Modoc National Forest in northern California. The allocation of sample size to land-cover types resulting from their use of FIM data is shown in Table I (Gill et al. 1999). Because FIM objectives emphasize forested land and because of the high proportion of the Conifer class in the region, the Conifer class has a much larger sample size than the other classes. The Shrub class is also well represented by the FIM data. However, the other four land-cover classes have sample sizes that are too small for precise estimates of accuracy and the FIM probability sample would need to be supplemented to increase the sample size for these rare classes. Because stratification by mapped land-cover type is the most feasible way to ensure a specified sample size for a rare class, a separate sampling design is probably necessary rather than trying to build upon the existing FIM design for this region.

Table I
Allocation of FIM Sample Data to Land-cover Types in the Modoc National Forest
(from Gill et al. 1999).

Cover Type	Sample size	% of sample	% of mapped population
Conifer	206	66.0	49.4
Herbaceous	6	1.9	1.3
Hardwood	1	0.3	0.3
Mixed tree	0	0.0	1.9
Non-vegetated	1	0.3	3.3
Shrub	98	31.4	43.9
Total sample size	312		

Analogous characteristics may result for an allocation of NRI sample data to land-cover types because NRI focuses on objectives pertaining to agricultural land. Larger sample sizes will occur in the intense agriculture region of the mid-west compared to the smaller area of cultivated land in the western United States. While agricultural lands may be represented by large sample sizes, the NRI data may be sparse for assessing accuracy of other land-cover types. It is possible that by combining data from several monitoring programs, the entire conterminous United States may be well represented. At a minimum, the ability to use FIM data for a few common forest types and NRI data for some agricultural classes may achieve a significant reduction in cost.

The differing objectives of environmental monitoring and accuracy assessment often motivate different sampling designs. Within the monitoring and accuracy projects, multiple objectives will exist. Instead of trying to satisfy every objective, there may exist a "least common denominator" of monitoring and accuracy objectives that can be met given constraints in resources and funding. The goal should be allocation of resources to take advantage of identified common sampling and response design structures to best address high priority objectives. In a fully integrated strategy, a better opportunity exists to construct a sampling design accommodating the dual accuracy assessment and monitoring objectives. A unified design will be much more amenable to incorporating sampling features mutually advantageous to the major objectives of both monitoring and accuracy assessment.

3.3 ANALYSIS

The simplest analyses are available from an equal probability or "self-weighting" design. However, both monitoring and accuracy assessment sampling designs often employ unequal probability sampling, for example via an equally allocated stratified design, and a weighted analysis is necessary to account for the unequal probability structure. Consequently, users must be provided with the appropriate sampling weights for analysis. A similar, though perhaps less critical concern applies to variance estimation. Users interested in computing the standard errors associated with their accuracy estimates will require additional information on the design structures, for example cluster and stratum identifiers, that are relevant to variance estimation. If the sampling design is complex, it may be possible to provide users with simpler, reduced information to allow for variance approximations that may be adequate for most applications. If an independent sample is necessary to augment the environmental monitoring design for accuracy assessment, the analysis will require dual frame estimation procedures (e.g., Hartley 1974, Sarndal et al. 1992). This increases the complexity of preparing the data for individual user accuracy assessments.

Because many monitoring plots occur on private lands, and measurement of those plots depends entirely upon the goodwill of the landowner, monitoring programs must carefully protect private property rights. Therefore, site confidentiality is a major obstacle to creating a combined accuracy assessment and environmental monitoring protocol. If the accuracy assessment data analysis is conducted "in house" by the monitoring agency, the confidentiality problem is diminished. However, because users' interests will span diverse applications, it is advantageous if users have access to the accuracy sample data to conduct their assessments directly. Users interested in accuracy for a subregion of the map will require the spatial locations of the sample to conduct subregional analyses. Other users interested in exploring factors associated with classification error will also need location-specific information.

If the monitoring agency must maintain site confidentiality, no entirely satisfactory option emerges for accommodating both the users and the agency. One approach is for users to submit their analysis specifications to the monitoring agency and have the agency conduct the analysis. But even if the analysis can be conducted without compromising site confidentiality, the analysis burden imposed may be unacceptable to the agency. Administrative mechanisms for permitting these analyses would also be necessary. The possibility exists to reveal sample sites located on public land, thus providing some data for accuracy assessment, though not data representing the full population. This would not impinge on private property rights, but monitoring agencies would be rightly concerned that if public-land sample locations were revealed, land managers would consciously or unconsciously change the management of these sites compromising the sample's integrity for monitoring objectives. The option of non-site-specific accuracy assessment is proposed in the next section as still another approach for dealing with the confidentiality problem.

4. Use of Environmental Monitoring Data in a Non-Integrated Strategy

The approximately 800,000 NRI sample points and 120,500 FIM field plots distributed across the United States represent a vast pool of data potentially useful for accuracy assessment. Even though neither program currently contains provisions for using these data for accuracy assessment, several options are available. The simplest option for using these data is to provide the agency with the land-cover map to be evaluated, and to request the agency to produce the error matrix and accompanying accuracy estimates. This option has the advantage of maintaining site confidentiality but the disadvantage of imposing the analysis burden upon the agency.

Another use of the data would be to conduct a "non-site-specific" accuracy assessment in which comparisons are based on area or proportion of area for each land-cover type aggregated over a spatial region such as a county. For example, FIM data might be used to estimate the total area of forest by county, and these areas would then be compared to the mapped area in forest for each county. The agreement between the mapped areas and those estimated from the FIM data would constitute a measure of non-site-specific accuracy. Merchant et al. (1994)

conducted a non-site-specific accuracy assessment using FIM data and U.S. Department of Commerce Agricultural Census data to evaluate land cover derived from AVHRR data. Their work qualitatively assessed the ability to map forest and agricultural lands at several aggregate spatial units (e.g., county and ecoregion). A motivation for a non-site-specific assessment is that it may circumvent the confidentiality problem of using monitoring data. By aggregating the data, the exact response values measured at a particular sample site will not be known, and the sample location itself does not have to be revealed to conduct the assessment.

Several difficulties arise with non-site-specific accuracy assessment. Map errors may compensate resulting in high agreement in the area of the land-cover class even though the location-specific accuracy may be poor (e.g., a true forest pixel misclassified as agriculture combined with a true agriculture pixel mislabeled as forest produce the same total area of forest and agriculture). Estimates of area in the different land-cover types generated from the monitoring data contribute additional variability not present when the true land-cover areas are known. This increases the variability of the estimated accuracy. The spatial scale to which the reference data can be aggregated depends on the monitoring program's sampling design. The region must be large enough so that the sample provides reasonably precise estimates of land-cover area. Consequently, if the environmental monitoring data are spatially sparse, the aggregate area will need to be large. For example, FIM data are usually too imprecise for reliable county-level estimates, so they would need to be aggregated to larger regions. Even under the best circumstances, a non-site-specific accuracy assessment may provide inadequate information to evaluate the utility of the map for many of its intended uses.

Environmental monitoring data may be used to supplement the accuracy assessment information available for a few land-cover classes. The contribution of FIM data to assessing accuracy of Conifer and Shrub classes was demonstrated by Milliken et al. (1998). Monitoring data may also be useful for assessing rare classes for which reference data are lacking. For example, National Wetlands Inventory (NWI) data might be proposed to assess accuracy of wetland classes. But some difficult problems must be resolved before the NWI data can be put to this use. Are the definitions of the wetland classes compatible between the NWI and the mapping project? Is the NWI data registered adequately with the mapped data? Are the NWI data temporally compatible with the map? For some projects, NWI data are used to "burn in" the wetlands of the land-cover map. Obviously using NWI data for accuracy assessment in these projects is inappropriate.

Environmental monitoring data could be used to evaluate reference data quality. For example, suppose the monitoring program acquires reference data from ground visits, whereas the accuracy assessment response design is based on photointerpreted reference data. Using the environmental monitoring data as "truth", agreement between the photointerpreted land cover and the true land cover at a small sample of locations provides a check on the quality of photo interpretation. The ability to use existing monitoring data for this evaluation is more cost effective than carrying out a separate field test of the accuracy assessment protocol. The monitoring data might also be used with multivariate calibration (Czaplewski and Catts 1992) to improve the accuracy assessment estimates.

Lastly, monitoring data have potential use for assessing accuracy of a classification categorizing a quantitative variable. For example, if the mapped categories are based on forest canopy closure or tree diameter, ground data such as those available from FIM may be used to assess the accuracy of the mapped classes. Quantitative ground data are expensive to obtain, so the ability to use existing environmental monitoring data would be a valuable asset. Gill et al. (1999) employed FIM data for this purpose.

When accuracy assessment is conducted with data from an environmental monitoring program for which accuracy assessment is not a design objective, the monitoring data will obviously not provide an assessment with exactly the desired characteristics. Each of the options described for this non-integrated approach falls short of some accuracy assessment objectives and criteria. The problem of incompatible land cover/land use classes and barriers presented by obligations of maintaining site confidentiality are major obstacles to employing environmental monitoring data for accuracy assessment.

5. Summary

The best scenario for combining accuracy assessment with an environmental monitoring program is when the two can be addressed within a unified framework, and within a single agency or small group of cooperating agencies. The three-tiered monitoring framework (NSTC 1997) offers a structure within which an integrated accuracy assessment and environmental monitoring strategy could be highly effective. In this framework, a common land cover/land use classification scheme could be defined for all three tiers, remote sensing, ground-based probability sampling, and intensive study sites. By including the applications planned for the land-cover data as part of the unified approach, an important built-in mechanism for developing objectives of the accuracy assessment is created. If the applications to which the land-cover map will be subjected are specified, the impact of classification error on these applications is potentially known, and the accuracy assessment objectives may be focused to provide information directly applicable to determining these impacts. Within a unified framework, a single sampling design for both accuracy assessment and monitoring may be feasible. For example, the accuracy assessment objectives may be addressed by increasing the intensity of sampling effort for some land-cover classes while still retaining the basic structure of the overall sampling design. When the accuracy assessment objectives originate in house, the monitoring agency retains control of the sample data and site confidentiality is less of a problem. Although many details would need to be worked out to implement a combined accuracy assessment and environmental monitoring protocol, developing these details would be a productive activity resulting in potential great cost savings for an integrated program such as that proposed in the CENR framework (NSTC 1997).

When the accuracy assessment project originates from outside the agency conducting the environmental monitoring, the confidentiality and land-cover class definition problems of a combined protocol become magnified. Further, it is less probable that the monitoring program sampling design will provide data that will address adequately all objectives of the accuracy assessment. The problem of small sample sizes for estimating rare class accuracy is particularly likely. Still, the advantages of using environmental monitoring data to contribute to accuracy assessment warrant consideration. The probability sampling foundation, high quality data, and cost-savings potentially available from environmental monitoring programs are attractive features. Some agencies have dealt with the barriers of sharing confidential data by establishing data centers with special rights granted to those who become special agency employees and sign a strong confidentiality pledge (e.g., Census Bureau data centers). It may be worth examining this model for sharing environmental monitoring data.

The myriad of land-cover mapping projects and uses of these maps preclude a single approach to a combined accuracy assessment and environmental monitoring protocol. We have described several strategies and identified problems that must be resolved to implement a combined protocol. The solution to these problems requires an innovative, interdisciplinary approach. More research is needed to develop the contributions of monitoring data to scientifically sound, informative, and cost-effective accuracy assessments.

Acknowledgments

We thank Anthony Olsen for suggesting this topic for presentation at the EMAP Symposium on Western U.S. Ecological Systems. Samantha Gill and Jeff Milliken kindly supplied a preliminary version of their manuscript, and Jeff Milliken also provided several helpful suggestions to clarify the presentation. Z. Zhu and L. Yang were supported U.S. Geological Survey contract 1434-CR-98-CN-40274 through Raytheon, an on-site contractor with the U.S. Geological Survey EROS Data Center. S. Stehman was partially supported by a research incentive grant from ESF's Faculty of Forestry.

References

Anderson, J.F., Hardy, E.E., Roach, J.T. and Witmer, R.E.: 1976, *A land use and land cover classification system for use with remote sensor data*, U.S. Geological Survey Professional Paper 964, 28 pp.

Bricker, O.P., and Ruggiero, M.A.: 1998, 'Toward a national program for monitoring environmental resources', *Ecological Applications* **8**, 326–329.

Czaplewski, R.L, and Catts, G.P.: 1992, 'Calibration of remotely sensed proportion or area estimates for misclassification error', *Remote Sensing of Environment* **39**, 29–43.

FGDC: 1997, 'Vegetation classification standard', Subcommittee of vegetation, Federal Geographic Data Committee web site, *http://www.fgdc.gov/fgdc.html*.

Gill, S.J., Milliken, J., Beardsley, D. and Warbington, R.: 1999, *Assessing vegetation map accuracy using Forest Inventory and Analysis (FIA) data*, unpublished manuscript.

Goebel, J.J., Schreuder, H.T., House, C.C., Geissler, P.H., Olsen, A.R. and Williams, W.R.: 1998, *Integrating surveys of terrestrial natural resources: The Oregon Demonstration Project*, Inventory and Monitoring Report No. 2, Fort Collins, CO, USDA Forest Service, Rocky Mountain Research Station, 20 pp.

Hartley, H.O.: 1974, 'Multiple frame methodologies and selected applications', *Sankhya Ser.* **B 36,** 99–118.

House, C.C., Goebel, J.J., Schreuder, H.T., Geissler, P.H., Williams, W.R. and Olsen, A.R.: 1998, 'Prototyping a vision for inter-agency terrestrial inventory and monitoring: A statistical perspective', *Environmental Monitoring and Assessment* **51,** 451–463.

Merchant, J.W., Yang, L. and Yang, W.: 1994, 'Validation of continental-scale land cover data bases developed from AVHRR data', *Proceedings of Pecora XII Symposium,* Sioux Falls, South Dakota, pp. 63–72.

Milliken, J., Beardsley, D. and Gill, S.: 1998, 'Accuracy assessment of a vegetation map of northeastern California using permanent plots and fuzzy sets in natural resources management using remote sensing', *Proceedings of the Seventh Forest Service Remote Sensing Applications Conference, April 1998*, ASPRS Bethesda, MD, pp. 218–229.

NSTC (National Science and Technology Council): 1997, *Integrating the Nation's Environmental Monitoring and Research Networks and Programs: A Proposed Framework,* The Environmental Monitoring Team, Committee on Environment and Natural Resources, National Science and Technology Council, Washington, DC.

Nusser, S.M. and Goebel, J.J.; 1997, 'The National Resources Inventory: a long-term multi-resource monitoring program', *Environmental and Ecological Statistics* **4,** 181–204.

Sarndal, C.E., Swensson, B. and Wretman, J.: 1992, *Model-Assisted Survey Sampling,* New York: Springer-Verlag.

Stehman, S.V. and Czaplewski, R.L.: 1998, 'Design and analysis for thematic map accuracy assessment: Fundamental principles', *Remote Sensing of Environment* **64**, 331–344.

Story, M. and Congalton, R.G.: 1986, 'Accuracy assessment: a user's perspective', *Photogrammetric Engineering and Remote Sensing* **52,** 397–399.

USFS: 1992, *Forest Service Resource Inventories: An Overview,* USGPO 1992-341-350/60861, U.S. Department of Agriculture, Forest Service, Forest Inventory, Economics, and Recreation Research, Washington, DC, 39 pp.