AN EVALUATION OF THE NATIONAL WETLAND INVENTORY IN MASSACHUSETTS[1]

DENNIS J. SWARTWOUT, WILLIAM P. MACCONNELL, JOHN T. FINN[2]

Abstract.--Nine Massachusetts communities were selected as study areas to test the accuracy of wetland maps produced by the National Wetland Inventory. The wetlands were divided into three categories (open water, forested wetlands, and open wetlands). The open water category contained the fewest errors (1 error out of 198 samples), while the forested wetlands and open wetlands categories had 8 and 9 errors out of 198 samples respectively.

INTRODUCTION

This research project evaluates the accuracy of the National Wetland Inventory in Massachusetts. Statistical procedures used in this project include techniques of sampling and determination of classification accuracy.

National Wetland Inventory

In 1974 the Fish and Wildlife Service (FWS) directed its Office of Biological Services (OBS) to design and conduct a new national wetland inventory (NWI). A new classification system was evolved for this purpose, Classification of Wetlands and Deepwater Habitats of the United States (Cowardin et al., 1979). This new system replaces the earlier FWS's classification system (Martin et al., 1953) that was used for the FWS 1954 inventory of wetlands, as reported in Circular 39, Wetlands of the United States (Shaw and Fredine, 1956).

The need for a new wetland classification and inventory was due to a better understanding of wetland characteristics and values. Since the 1954 inventory, new Federal and State wetland regulations have led to many regional and statewide wetland inventories using a variety of classification systems, which though appropriate for local needs, could not be used effectively for a national survey.

The new classification system was designed to meet four main objectives:

1. To describe areas that have similar national attributes.
2. To arrange these areas in a system that will aid decisions about resource management.
3. To furnish units for inventory and mapping.
4. To provide uniformity in concepts and terminology throughout the United States.

The system is hierarchical or vertical, compared to the Martin et al. (1953) system which identified 20 wetland types, all of equal rank. Five ecological systems are defined: marine, estuarine, riverine, lacustrine, and palustrine. Subsystems are defined at the second level of the hierarchy. The subsystems--such as limnetic vs. littoral--follow traditional ecological concepts. The third level of the hierarchy is represented by 11 classes that may appear under one or more of the systems. The classes are based on life form of plants for vegetated areas or substrate type for unvegetated areas, and are designed to be recognizable on aerial photographs. The classes are, in some cases, separated into subclasses at the fourth level of the hierarchy. The fifth and lowest level of the hierarchy is the dominance type, as defined by the dominant plant or animal species. The system is open-ended at the dominance level, assuming that users will provide dominance types as wetlands are described in the future. In add-

[1]Paper presented at the In-Place Resource Inventories workshop (University of Maine, Orono, August 9-14, 1981).

[2]Dennis J. Swartwout is a graduate student in the Department of Forestry and Wildlife Management, University of Massachusetts, Amherst, Mass. William P. MacConnell is Associate Professor of Forestry, University of Massachusetts, Amherst, Mass. John T. Finn is Assistant Professor of Wildlife Biology, University of Massachusetts, Amherst, Mass.

ition to the hierarchy of the classification, modifying terms for water regime, water chemistry, soils, and man-made disturbances, may be added at any level in the classification.

The NWI will produce two series of topical wetland maps for the entire country: intermediate-scale (1:100,000) maps and large-scale maps (1:24,000).

The NWI classifies wetlands on high altitude, small-scale (1:60,000-1:130,000), quad-centered aerial photographs. Ideally, NWI uses 1:80,000 color infrared photographs. However, in many parts of the country, this imagery is not available, and NWI utilizes the best available high altitude imagery. Black and white photos at 1:80,000 taken by USGS for orthophoto quads are often used. The photos are examined with seven power stereoscopes and all available collateral information is used in the interpretation work.

Contractors (State agencies, universities, private industry, or branches of the Federal Government) perform the delineation and classification of wetlands on air photos provided by NWI. The minimum mapping unit varies with the scale of imagery used-- 3-5 acres for 1:80,000 imagery and 5-7 acres for 1:130,000 scale imagery. Collateral wetland data sources, such as wetland or soil maps, are used as aids in locating and identifying wetlands.

Objectives of Study

The objectives of this research project are to evaluate the accuracy of the NWI in Massachusetts at three levels:

1. Accuracy of the wetlands located on 1:24,000 scale maps by NWI (are the wetlands shown correctly classified - errors of commission).
2. Accuracy in identifying all wetlands (are some wetlands missing - errors of omission).
3. Accuracy of wetland classification by category down to the class level.

METHODS AND MATERIALS

Study Area Selection

Massachusetts cities or towns were used in this study. To select communities, an attempt was made to isolate a factor common to most of the cities and towns in Massachusetts. That factor is the local conservation commission. These boards of 3 to 7 appointed Commissioners were created in 1957 with passage of the Conservation Commission Act. The commissions received regulatory control over local wetlands in 1972 under the Wetlands Protection Act. By 1978, 334 of the 351 communities of the Commonwealth had established conservation commissions (Dawson and Nickerson, 1978). Each public hearing held by a conservation commission is recorded by the Department of Environmental Quality Engineering (DEQE). The number of hearings each commission has held since 1972 was therefore available to assess the activity of the towns in wetland matters. Each town was placed in one of three categories--high activity, medium activity, or low activity--based on the number of public hearings in the town from 1972 to 1979. Each conservation commission was then contacted to determine if the town had suitable wetland maps available for its use.

The last information used for community selection was whether the town had inland, coastal or both types of wetlands. In testing the NWI product, we wanted to evaluate its accuracy in classifying the easier coastal wetlands as well as the more difficult inland wetlands. Table 1 lists the communities selected for this study.

Table 1.--Massachusetts Communities, conservation commission activity and available wetland maps.

	Wetland Types		Commission Hearing Activity			Maps
Town	Coastal	Inland	High	Med	Low	Local Wetland Maps
Barnstable	*	*	*			*
Amherst		*	*			*
Southampton		*		*		
Westport	*	*		*		
Whately		*			*	
Holyoke		*			*	
Sandwich	*	*			*	*
Wilbraham		*		*		*
Concord		*	*			*

Preparation of Test Maps

As mentioned earlier, NWI wetland maps are produced at scales of 1:24,000 and 1:100,000. The 1:24,000 scale wetland maps were selected for evaluation because Massachusetts is a small, heavily populated state where the towns are relatively small and everyone is familiar with working at that scale.

Individual topographic maps of the town were produced by assembling and trimming U.S. Geological Survey (USGS) 7.5 min. topographic quadrangles for each town. This procedure allowed the wetlands for each town to be shown on a single map.

NWI provided its standard photographic product for use in this study since the maps of Massachusetts are not yet complete. The NWI information on annotated 1:80,000 scale aerial photos was transferred to an acetate overlay to fit the USGS base map using a Bausch and Lomb zoom transfer scope. This procedure corrected scale errors which are inherent in aerial photographs. In this very urban state there was abundant control to insure a high level of cartographic accuracy in the map making.

Area Measurement

Acreage values were determined for each wetland using a Numonics planimeter with three averaged measurements. All wetlands were measured to the nearest one-hundredth of an acre.

Determining Minimum Sample Size for Map Accuracy Analysis

Evaluating the accuracy of the NWI maps requires statistical sampling of the classified wetlands to determine the accuracy of the photo-interpretation. Previous research has shown that the best method for determining thematic map accuracy is by selecting sample points according to a stratified systematic unaligned sampling design (Fitzpatrick-Lins, 1978). This method involves stratifying the map into equal blocks and then systematically selecting points from each block until the sample is complete. Grid block size is dependent on the desired sample size. This method is area-weighted, so the larger types have more sample points and the smaller units may not be sampled at all.

A minimum number of points to validate the accuracy of all wetland categories on the map, to insure that all wetlands were included and that none were missed, was needed. According to Van Genderen and Lock (1977), the smallest sample size for meaningful results is 20 checkpoints, even if the sample is error free. Rosenfield et al. (1980) suggest using a one-tailed statistical test to determine minimum sample size to evaluate the probability (Po) that a given category in a thematic map is interpreted correctly. Using binomial probability theory where x = 1 if an interpreted point is verified as being correct, and x = 0 if the point is incorrect, the minimum sample size for each category with 95% confidence, less a selected percentage of error (E), can be found from the formula:

$$n=(1.645)^2 Po(1-Po)+E+(1.645) \text{ times } \sqrt{(1.645)^2 Po(1-Po)^2+2EPo(1-Po)/2E^2}$$

where n = minimum sample size per category;
Po= $E(\bar{x})$ = expected value of the sample mean;
E = error, usually 10% or less.

Table 2 gives the minimum sample size per wetland category with 10% error and an unexpected (estimated) accuracy (Po) ranging from 50% to 95%.

According to Anderson et al. (1976), the minimum level of interpretation accuracy in the identification of thematic map categories from remote sensor data, should be at least 85%. For this research project, the inventory of wetlands (separation of wetlands from uplands) was tested for an expected accuracy (Po) of 85% with 95% confidence (45 points per category, Table 2), and the classification of wetlands (separation of wetlands into the correct NWI classes) was tested for an expected accuracy (Po) of 95% with 95% confidence (22 points per category, Table 2).

Table 2.--Minimum sample size per wetland category to achieve various expected (estimated) percentages of accuracy (Po).

Po	n
0.50	77
0.70	66
0.80	53
0.85	45
0.90	34
0.95	22

Plot Selection to Identify Errors of Omission and Commission

Once the minimum number of sampling points per category was determined, the wetlands from each map were grouped into three distinct categories to test the accuracy of wetland classification. Category A--open water wetlands lacking trees, shrubs or persistent emergent vegetation, category B--forested wetlands; and, category C--open wetlands consisting of shrubs and persistent emergents. Grouping in this study was necessitated by the nature of the NWI system with its potential for an almost unlimited number of wetland types. In testing the wetland identification accuracy, it was necessary to subdivide the wetland maps into two categories--uplands and wetlands, in order to identify errors of omission (missing wetlands) or commission calling uplands wetlands.

The tests for inventory (separation of wetland from upland) and classification accuracies could now be made. The minimum number of points to test the inventory accuracy was 90 points per map. This is derived by multiplying the number of categories times the minimum number of points per category at the expected accuracy (2 categories x 45 points per category = 90 points, see Table 2.) The minimum number of points to test the classification accuracy is determined the same way (3 categories x 22 points per category = 66 points per map).

Next, a 25 dot per square inch grid was placed over the wetland map. Each dot represented 2.35 acres on the prepared maps which was approximately equal to the NWI minimum mapping unit on 1:80,000 black-and-white aerial photographs of Massachusetts. Random pairs of numbers were then generated in the computer and sample plots were selected using these pairs, until 45 wetland and 45 upland sites had been identified. These points located by the dot grid and computer were used for the inventory test. Additional points were selected until each of the three wetland categories had 22 points each. Some of the points from the inventory test (to be sure that wetlands were separated from uplands) could

also be used to determine the accuracy of the classification. The minimum number of total points used in each test town was 111 (45 + 66). Test points located within upland-wetland transition zones were discarded and new points were selected for testing to avoid ambiguity in the test.

Field Verification

Each of the 111 sample points per town was visited to establish "ground truth" and test the accuracy of the prepared maps in each of the nine study area communities. Ten sites could not be reached because the land was posted and permission for access was not granted. Points that fell in this category were verified using a variety of collateral materials such as large scale photography, wetland restriction maps and local wetland maps. All but two points located on Otis Air Force Base were satisfactorily verified.

Critical Level for Assessing Wetland Category Accuracy

The test to determine whether a wetland category meets the expected accuracy value based on the number of check points in the sample is determined by a one-tailed statistical test. This test uses a cricital level (C), defined as "one less than the minimum number of points which must be correctly interpreted from any given sample in order to accept the hypothesis at a given significance level" (Rosenfield et al., 1980). When the number of correct points for the category is larger than the critical level, the category accuracy equals or exceeds the lower confidence limit.

The critical level is used to test the hypothesis that the accuracy (P) of the interpretation of a specific category is greater than or equal to the expected accuracy (Po) (the hypothesis Ho: $P \geq Po$ = 85 or 95 per cent). The formula for determining the critical level is:

$$C = n\,Po - 1.645\sqrt{n\,Po\,(1 - Po)} - .5$$

where n = sample size
Po = E (x) = expected value of the sample mean
C = critical level.

Tables 3a and 3b give critical levels and confidence limits for different sample sizes.

Table 3a.--Critical levels and confidence limits for sample size selection for an expected accuracy (Po) = 85%.

Sample Size n	Critical C Level C	Number of Successes to accept Ho with 95% confidence	Lower 95% C.L.
22	15	16 or more	0.680
45	33	34 or more	0.756
66	50	51 or more	0.770
90	70	71 or more	0.788

Table 3b.--Critical levels and confidence limits for sample size selection for an expected accuracy (Po) = 95%.

Sample Size n	Critcal Level C	Number of Successes to accept Ho with 95% confidence	Lower 95% C.L.
22	18	19 or more	0.86
45	40	41 or more	0.91
66	59	60 or more	0.91
90	81	82 or more	0.91

For example, for the sample size n = 45 (Table 3a), the hypothesis, (Ho:$P \geq$ Po = 0.85 with 95 per cent confidence) is acceptable when the sample mean ($\bar{x}$) is greater than or equal to the critical level, plus one, divided by the sample size

$$\bar{x} \geq (C + 1)/n = 34/45 = 0.756$$

where:

$$C - (45)(0.85) - 1.645\sqrt{(45)\ (0.85)\ (0.15)} - .5 = 33.$$

In simpler terms, this means that Ho is accepted if 34 or more points out of 45 are judged correct and that Ho is rejected if less than 34 points are found to be correct.

RESULTS AND DISCUSSION

Accuracy Assessment of the Wetland Categories

As described earlier, the wetland maps were divided into three categories for the wetland classification test, and two categories for the wetland inventory test. In determining the accuracy of these categories, a two-tailed statistical test is appropriate. Since this is a binomial distribution, let a correct interpretation = 1, and an incorrect interpretation = 0. The accuracy can be found from the following formula:

$$Pa = Ra/Na$$

where Pa = the mean of category A expressed as decimal, or it is the accuracy value for category A expressed as a percentage.
Ra = the number of correct interpretations of category A, and;
Na = the number of correct interpretations of category A.

Using Barnstable, Category B (Table 4) as an example, 19 points out of 22 were verified as being correct. Therefore, 19/22 = 0.86 = $\bar{x}$, thus 86% is the accuracy of this category. Mainland et al., (1956) give tables to be used with binomial distributions in establishing confidence limits. In the above example, with a mean of 0.86, the confidence limits from Mainland's table would be .6506-.9709 with 95% confidence. This means that

Table 4.--Category accuracies and confidence limits.

Town	Category	Points Sampled	Points Correct	Mean ($\bar{X}$)	Confidence Limits
Amherst	Upland	45	44	0.98	.8822-.9994
	Wetland	45	45	1.00	
	A	22	22	1.00	
	B	22	21	0.95	.7718-.9988
	C	22	21	0.95	.7718-.9988
Barnstable	Upland	45	45	1.00	
	Wetland	45	45	1.00	
	A	22	22	1.00	
	B	22	19	0.86	.6506-.9709
	C	22	21	0.95	.7718-.9988
Concord	Upland	45	45	1.00	
	Wetland	45	45	1.00	
	A	22	22	1.00	
	B	22	21	0.95	.7718-.9988
	C	22	22	1.00	
Holyoke	Upland	45	44	0.98	.8822-.9994
	Wetland	45	45	1.00	
	A	22	22	1.00	
	B	22	22	1.00	
	C	22	22	1.00	
Sandwich	Upland	45	45	1.00	
	Wetland	45	41	0.91	.7877-.9752
	A	22	22	1.00	
	B	22	20	0.91	.7083-.9888
	C	22	20	0.91	.7083-.9888
Southampton	Upland	45	45	1.00	
	Wetland	45	45	1.00	
	A	22	22	1.00	
	B	22	22	1.00	
	C	22	22	1.00	
Westport	Upland	45	45	1.00	
	Wetland	45	45	1.00	
	A	22	22	1.00	
	B	22	21	0.95	.7718-.9988
	C	22	21	0.95	.7718-.9988
Whately	Upland	45	45	1.00	
	Wetland	45	44	0.98	.8822-.9994
	A	22	21	0.95	.7718-.9988
	B	22	22	1.00	
	C	22	20	0.91	.7083-.9888
Wilbraham	Upland	45	44	0.98	.8822-.9994
	Wetland	45	44	0.98	.8822-.9994
	A	22	22	1.00	
	B	22	22	1.00	
	C	22	20	0.91	.7083-.9888

unless a one-in-twenty chance has occurred in sampling, the true population mean is somewhere between 65.06% and 97.07% correct.

Each category tested, in both the inventory and classification tests, for all nine towns, was found to be acceptable. Table 5 shows the total number of points sampled, the number of errors, and the percentage correct for all nine communities in the study area.

Accuracy Assessment of the Wetland Maps

The accuracy assessment was divided into two tests. The inventory test deals with errors of omission. Were all wetlands greater than 2.5 acres

delineated on the imagery and no upland areas falsely delineated? The classification test deals with errors of commission. Were the delineated wetlands labelled with the correct wetland designation? The accuracy value of the maps is the ratio (P) of the number of points correct (R) to the number of points sampled (N) for all categories of the map. This value (P) is compared to the critical level (C) from Table 3a or 3b. If the P value exceeds the lower confidence limit, the map is said to be acceptable at the expected accuracy that it was tested for. This does not, however, mean that those maps having a P less than the expected accuracy are inaccurate, but that we have less confidence that they meet the accuracy standards.

The number of points to be tested per map was determined from the formula for binomial probability. With an expected (estimated) map accuracy of 85 per cent and an acceptable error of 10 per cent, the sample size for each category for the inventory test should be at least 45 points. Twenty-two points are needed for each category in the classification test, because of the higher expected accuracy (95%).

The number of points sampled per map was 90 points for 2 categories in the inventory test and 66 points for the classification test (3 categories with 22 points each). The same points used for the category accuracy tests were used in the accuracy assessment of the wetland maps. The results are shown in Tables 4, 5, and 6.

Table 6.--Tested map accuracies (P)

	Inventory[1]			Classification[2]		
Town	P	Accept	Reject	P	Accept	Reject
Amherst	0.99	*		0.97	*	
Barnstable	1.00	*		0.94	*	
Concord	1.00	*		0.98	*	
Holyoke	0.99	*		1.00	*	
Sandwich	0.96	*		0.94	*	
Southampton	1.00	*		1.00	*	
Westport	1.00	*		0.97	*	
Whately	0.99	*		0.95	*	
Wilbraham	0.98	*		0.97	*	

[1] 90 points per map.
[2] 66 points per map.

Overall Map Accuracy

Once the decision to accept or reject the wetland map has been made, the accuracy of the map can be determined using the following random stratified sampling formula:

$$Pm = \frac{\sum_{i=1}^{n} (\bar{x} * n)}{N}$$

where, Pm = map mean;
$\bar{x}$ = number of possible samples per category; and
N = number of possible samples per map.

The number of possible sample points per map was found by dividing the number of acres per map by the number of acres per sample point (2.35). For example, Amherst has a total area of 17,940 acres. This number divided by 2.35 acres per dot equals 7,443 possible sample points for the map. The same method was used in finding the number of sample points per category.

Using the classification data for Whately (Table 4), the actual map accuracy can be found by:

Category A = 0.95 x 132 = 125.40
Category B = 1.00 x 614 = 614.00
Category C = 0.91 x 71/817 = 64.61
804.01

$$\frac{804.01}{817} = 98.41\% \text{ Classification accuracy}$$

As mentioned earlier, when describing accuracy values, the confidence limits about these values should also be given. Confidence limits for these means are found using standard binomial distribution tables (Mainland et al., 1956) and are shown along with the map accuracy values in Table 7.

SUMMARY AND CONCLUSIONS

The NWI overlays at 1:24,000 scale, interpreted from high altitude, black-and-white 1:80,000 imagery, meet the criteria that they be mapped with an inventory accuracy of 85% and classification accuracy of 95%. Testing these overlays involved a scheme for selecting a minimum number of random sample points per category and verifying the wetlands at these points through ground truthing. As expected, the open water category (A), contained the fewest errors (1 error out of 198 samples), while the forested wetlands (B) and open wetlands (C) categories had 8 and 9 errors out of 198 samples respectively. In this case, because the Cowardin system was tested at the class level, the number of overall errors was greatly reduced and the open water category did not play a significant role in the test accuracies of the maps.

This research project has shown that the NWI was highly successful in nine Massachusetts communities, in distinguishing between uplands and wetlands and classifying these wetlands at the class level. A high level of accuracy was maintained in spite of the fact that the 1:80,000 black-and-white photography was over 5 years old and much of it was taken in the fall when water was less abundant for wetland interpretation than it would be in the spring.

LITERATURE CITED

Anderson, J.R., E.E. Hardy, J.T. Roach, and R.E. Witman. 1976. A land use and land cover classification system for use with remote sensor data. U.S. Geol. Surv. Prof. Pap. 964. 28 pp.

Table 7.--Means and confidence limits for inventory and classification map accuracy.

	Inventory		Classification	
Town	Mean	95% C.L.	Mean	95% C.L.
Amherst	98.18	93.96-99.97	98.85	91.83-99.96
Barnstable	100.00		95.93	91.83-99.96
Concord	100.00		97.76	91.83-99.96
Holyoke	98.21	93.96-99.97	100.00	
Sandwich	98.94	93.96-99.97	93.51	89.47-99.63
Southampton	100.00		100.00	
Westport	100.00		96.08	91.83-99.96
Whately	99.71	93.96-99.97	98.41	91.83-99.96
Wilbraham	98.00	93.96-99.97	99.12	91.83-99.96

Cowardin, L.M., V. Carter, F.C. Golet, and E.T. LaRoe. 1980. Classification of wetlands and deepwater habitats of the United States. U.S. Fish and Wildlife Service, Office of Biological Services, Washington, D.C. 103 pp.

Dawson, A.D., and N.H. Nickerson. 1978. Environmental handbook for Massachusetts conservation commissioners. Mass. Assoc. of Cons. Comm., Tufts Univ., Medford, Mass. 118 pp.

Fitzpatrick-Lins, K. 1980. The accuracy of selected land use and land cover maps at scales of 1:250,000 and 1:100,000. U.S. Geol. Surv. Circular 829. 24 pp.

Mainland, D., L. Herrera, and M.I. Sutcliffe. 1956. Statistical tables for use with binomial samples--contingency tests, confidence limits, and sample size estimates. Dept. of Medical Stats., N.Y.U. Coll. of Medicine. 81 pp.

Martin, A.C., N. Hotchkiss, F.M. Uhler, and W.S. Bourn. 1953. Classification of wetlands of the United States. U.S. Fish and Wildlife Service, Spec. Sci. Rep. 20. 14 pp.

Rosenfield, G.H., K. Fitzpatrick-Lins, and H.S. Ling. 1979. Sampling for thematic map accuracy testing. U.S. Geol. Survey Res. Paper. 26 pp. (unpublished manuscript)

Shaw, S.P. and C.G. Fredine. 1956. Wetlands of the United States. U.S. Fish and Wildlife Service. Circular 39. 67 pp.

Van Genderen, J.L., and B.F. Lock. 1977. Testing land-use map accuracy. Photogrammetric Engineering and Remote Sensing. 43(9):1135-1137.