MICHAEL E. GINEVAN*
Space Technology Center
University of Kansas
Lawrence, KS 66045

Testing Land-Use Map Accuracy: Another Look

Acceptance sampling, together with the binomial probability density function, provides a basis from which a sound statistical methodology for map accuracy validation may be developed.

ABSTRACT: *Specification and statistical validation of classification accuracy of image-derived land-use maps is a persistent problem in remote sensing studies. The present discussion points out shortcomings in statistical procedures proposed in earlier papers, and suggests that use of concepts from the branch of statistics known as acceptance sampling, together with the binomial probability density function, provides a basis from which a sound statistical methodology for map accuracy validation may be developed.*

INTRODUCTION

THE PRODUCTION of land-use maps has become a relatively simple matter with the recent wide availability of both high altitude photography and satellite imagery. Unfortunately, statistical specification of the classification accuracy of such maps remains a problem.

Two recent papers (Genderen and Lock, 1977; Hord and Brooner, 1976) point out the many advantages of image-derived land-use maps, and suggest procedures which might be used to statistically establish classification accuracy. However, the methodology suggested in both papers is questionable from either an operational or a statistical standpoint.

The sampling problem as defined here is the determination of the optimal number, N, of ground truth samples and an allowable number, X, of misclassifications of these samples. Once these have been determined, the results of image interpretation are checked against the N ground truth samples and the map is accepted as accurate if X or fewer of the ground truth samples are misclassified. It is further assumed that misclassification of a given area can be unambiguously determined (i.e., it is either right or wrong). This definition of the problem is essentially the same as that of both van Genderen and Lock (1977) and Hord and Brooner (1976). As suggested by Genderen and Lock (1977) ground truth samples should, if possible, be collected using stratified random sampling.

Any sampling scheme/decision rule adopted for this problem should satisfy three criteria:

- It should have a low probability of accepting a map of low accuracy;
- It should have a high probability of accepting a map of high accuracy; and
- It should require a minimum number, N, of ground truth samples.

With the first and third criteria in mind Genderen and Lock (1977) suggest that accepting only those maps which show no er-

* Now with the Division of Biological and Medical Research, Argonne National Laboratory, Argonne, Illinois 60439.

PHOTOGRAMMETRIC ENGINEERING AND REMOTE SENSING,
Vol. 45, No. 10, October 1979, pp. 1371-1377.

rors in a sample of 20 will result in a probability of less than 0.05 ($P < 0.05$) of accepting a map whose parametric accuracy proportion (Q) is 0.85 or less ($Q \leq 0.85$) and that accepting no errors in a sample of 30 will result in $P < 0.05$ of accepting maps in which $Q \leq 0.90$. The authors further point out that these are nearly the minimum number, N, of ground truth samples for which one can obtain $P < 0.05$ when given the respective Q values.

While the authors' arguments are correct, their procedure ignores the second criterion. The probability of rejecting a map of accuracy Q may be determined by

$$P = 1 - Q^N. \quad (1)$$

Thus, if one accepts only those maps which show zero errors in a sample of 30, $P = 0.785$ for $Q = 0.95$ and $P = 0.260$ for $Q = 0.99$. That is, the probability of rejecting a map which is 95 percent accurate is roughly 8 in 10 and the probability of rejecting a map which is 99 percent accurate is approximately 1 in 4. This feature could well result in large amounts of time and money spent rechecking maps which were acceptably accurate.

Hord and Brooner (1976) take a somewhat different approach in suggesting that one obtain a 95 percent confidence interval about Q based on the sample size N and $\hat{Q}$, the observed proportion correct. The observed proportion correct is calculated as

$$\hat{Q} = \frac{N - X}{N}. \quad (2)$$

The 95 percent interval is expressed as

$$P\,(Q_1 > Q > Q_2) = 0.95$$

where Q_1 and Q_2 are the upper and lower proportion values, calculated from $\hat{Q}$ and N, between which the parametric proportion Q lies with probability 0.95. Once N has been determined one is instructed to select a $\hat{Q}$ (and, thus, an X value) value such that the lower limit of the interval, Q_2, is greater than some objectionable value, say 0.80.

This procedure also satisfies the first criterion in that the probability of accepting a map of low accuracy is small. However, it does not address the second criterion, and, since it assumes N to be previously determined, provides no guidance with regard to the third criterion. A further difficulty associated with this methodology is that it recommends the normal distribution as an approximation to the binomial. When, as is the case here, the parametric proportion Q is of the order 0.90, this procedure gives a relatively poor approximation to the true confidence interval unless N is large (600) (Cochran, 1977). In this regard, a table of exact confidence intervals may be found in Rohlf and Sokal (1969).

TABLE 1. VALUES OF α AND "EXACT β" FOR $N = 30 - 50$; $Q_2 = 0.85$; $Q_1 = 0.90, 0.95, 0.99$; $\beta = 0.05$. NOTE THAT FOR EACH CRITICAL VALUE X ($X = 1, 2, 3$) THE SMALLEST ASSOCIATED VALUE OF N YIELDS THE SMALLEST α PROBABILITIES AND THE "EXACT β" VALUE NEAREST THE NOMINAL β (0.05).

				α Values	
N	X	Exact β	$Q_1 = 0.90$	0.95	0.99
30	1	0.0480	0.8163	0.4465	0.0361
31	1	0.0420	0.8304	0.4634	0.0384
32	1	0.0366	0.8436	0.4800	0.0407
33	1	0.0320	0.8558	0.4964	0.0430
34	1	0.0279	0.8671	0.5123	0.0454
35	1	0.0243	0.8776	0.5280	0.0479
36	1	0.0212	0.8874	0.5433	0.0503
37	1	0.0184	0.8964	0.5582	0.0529
38	1	0.0160	0.9047	0.5728	0.0555
39	1	0.0139	0.9124	0.5871	0.0581
40	2	0.0486	0.7772	0.3233	0.0075
41	2	0.0431	0.7914	0.3371	0.0080
42	2	0.0382	0.8049	0.3510	0.0086
43	2	0.0339	0.8176	0.3648	0.0092
44	2	0.0300	0.8296	0.3786	0.0098
45	2	0.0265	0.8410	0.3923	0.0104
46	2	0.0234	0.8516	0.4060	0.0110
47	2	0.0207	0.8617	0.4195	0.0117
48	2	0.0183	0.8711	0.4330	0.0124
49	2	0.0161	0.8800	0.4463	0.0131
50	3	0.0460	0.7497	0.2396	0.0016

TABLE 2. OPTIMAL SAMPLE SIZES (N), THEIR ASSOCIATED CRITICAL VALUES (X), AND VALUES OF α FOR $Q_1 = 0.90, 0.95, 0.99$, WHEN $Q_2 = 0.85$, $N = 1{,}400$. PART 1 OF THE TABLE IS FOR $\beta = 0.05$; PART 2 IS FOR $\beta = 0.01$. TO USE THIS TABLE TO DETERMINE AN OPTIMAL SAMPLE SIZE AND CRITICAL VALUE (GIVEN $Q_2 = 0.85$; IF $Q_2 = 0.90$, USE TABLE 3.) FIRST DETERMINE THE DESIRED β VALUE. IF $\beta = 0.05$ USE PART 1; IF $\beta = 0.01$ USE PART 2. THEN DETERMINE THE Q_1 AND α VALUES OF INTEREST. READ DOWN THE COLUMN HEADED BY THE APPROPRIATE Q_1 VALUE UNTIL AN α SMALLER THAN THAT DESIRED IS ENCOUNTERED. FOLLOW THIS ROW TO THE RIGHT TO FIND THE DESIRED N AND X VALUES.

Part 1: $\beta = 0.05$

N	X	α Values $Q = 0.90$	0.95	0.99
19	0	0.8649	0.6226	0.1738
30	1	0.8163	0.4465	0.0361
40	2	0.7772	0.3233	0.0075
50	3	0.7497	0.2396	0.0016
59	4	0.7152	0.1719	0.0003
68	5	0.6859	0.1242	0.0001
76	6	0.6467	0.0856	0.0000
85	7	0.6247	0.0624	0.0000
93	8	0.5919	0.0432	0.0000
102	9	0.5746	0.0318	0.0000
110	10	0.5464	0.0221	0.0000
118	11	0.5203	0.0153	0.0000
126	12	0.4959	0.0107	0.0000
134	13	0.4731	0.0074	0.0000
142	14	0.4518	0.0052	0.0000
150	15	0.4318	0.0036	0.0000
158	16	0.4130	0.0025	0.0000
166	17	0.3954	0.0018	0.0000
174	18	0.3787	0.0012	0.0000
182	19	0.3630	0.0009	0.0000
190	20	0.3481	0.0006	0.0000
197	21	0.3252	0.0004	0.0000
205	22	0.3122	0.0003	0.0000
213	23	0.2998	0.0002	0.0000
220	24	0.2802	0.0001	0.0000
228	25	0.2693	0.0001	0.0000
236	26	0.2589	0.0001	0.0000
243	27	0.2421	0.0000	0.0000
251	28	0.2329	0.0000	0.0000
259	29	0.2242	0.0000	0.0000
266	30	0.2097	0.0000	0.0000
274	31	0.2020	0.0000	0.0000
281	32	0.1890	0.0000	0.0000
289	33	0.1821	0.0000	0.0000
296	34	0.1704	0.0000	0.0000
304	35	0.1643	0.0000	0.0000
311	36	0.1537	0.0000	0.0000
319	37	0.1483	0.0000	0.0000
326	38	0.1388	0.0000	0.0000
334	39	0.1339	0.0000	0.0000
341	40	0.1253	0.0000	0.0000
349	41	0.1210	0.0000	0.0000
356	42	0.1133	0.0000	0.0000
364	43	0.1094	0.0000	0.0000
371	44	0.1024	0.0000	0.0000
379	45	0.0989	0.0000	0.0000
386	46	0.0926	0.0000	0.0000
393	47	0.0867	0.0000	0.0000

DISCUSSION

A procedure for determining a sample size, N, and an allowable number of errors, X, which satisfies the three criteria may be developed from a branch of statistics known as acceptance sampling (Guttman *et al.*, 1971; Vaughn, 1974) which is concerned with statistical procedures for determining whether large lots of manufactured articles

TABLE 2.—*Continued*

		Part 2: $\beta = 0.01$		
			α Values	
N	X	$Q = 0.90$	0.95	0.99
29	0	0.9529	0.7741	0.2528
42	1	0.9322	0.6276	0.0662
53	2	0.9102	0.4982	0.0162
64	3	0.8937	0.3986	0.0039
74	4	0.8738	0.3112	0.0009
84	5	0.8565	0.2434	0.0002
93	6	0.8333	0.1839	0.0000
103	7	0.8197	0.1442	0.0000
112	8	0.7994	0.1088	0.0000
121	9	0.7801	0.0821	0.0000
130	10	0.7619	0.0619	0.0000
138	11	0.7358	0.0445	0.0000
147	12	0.7194	0.0335	0.0000
156	13	0.7038	0.0252	0.0000
164	14	0.6800	0.0181	0.0000
173	15	0.6659	0.0136	0.0000
181	16	0.6435	0.0097	0.0000
190	17	0.6308	0.0073	0.0000
198	18	0.6098	0.0052	0.0000
214	20	0.5697	0.0027	0.0000
223	21	0.5594	0.0020	0.0000
231	22	0.5408	0.0014	0.0000
239	23	0.5229	0.0010	0.0000
247	24	0.5056	0.0007	0.0000
255	25	0.4888	0.0005	0.0000
263	26	0.4727	0.0004	0.0000
271	27	0.4571	0.0003	0.0000
279	28	0.4420	0.0002	0.0000
287	29	0.4274	0.0001	0.0000
295	30	0.4134	0.0001	0.0000
303	31	0.3998	0.0001	0.0000
311	32	0.3867	0.0000	0.0000
319	33	0.3740	0.0000	0.0000
327	34	0.3618	0.0000	0.0000
335	35	0.3501	0.0000	0.0000
343	36	0.3387	0.0000	0.0000
350	37	0.3212	0.0000	0.0000
358	38	0.3108	0.0000	0.0000
366	39	0.3007	0.0000	0.0000
374	40	0.2909	0.0000	0.0000
382	41	0.2815	0.0000	0.0000
389	42	0.2668	0.0000	0.0000
397	43	0.2582	0.0000	0.0000

are of acceptable quality. The case where such articles are either defective or non-defective provides a direct analogy to the map accuracy problem because we can consider any land-use map to be made up of a large number of potential ground truth samples (articles) which are either correctly classified (non-defective) or misclassified (defective).

In the map accuracy problem, where the number of articles in the lot is virtually infinite, a sampling plan may be based on the binomial probability density function (p.d.f.). The binomial p.d.f. is given by

$$f(Y;N,Q) = \frac{N!}{(N-Y)!\,Y!} Q^{N-Y} (1-Q)^{Y}. \quad (3)$$

This function describes the probability of getting exactly Y misclassifications in a sample of N drawn from a population with a parametric accuracy proportion Q.

One first determines a low accuracy proportion, Q_2, which one wishes to reject with probability $(1 - \beta)$. The quantity, β, which is equal to the probability of accepting an inaccurate map is known as "consumer's risk" in acceptance sampling.

Once Q_2 and β have been determined one

TABLE 3. OPTIMAL SAMPLE SIZES (N), THEIR ASSOCIATED CRITICAL VALUES (X), AND VALUES OF α FOR $Q_1 = 0.95$, 0.97, 0.99, WHEN $Q_2 = 0.90$, $N = 1{,}400$. PART 1 OF THE TABLE IS FOR $\beta = 0.05$; PART 2 IS FOR $\beta = 0.01$. TO USE THIS TABLE TO DETERMINE AN OPTIMAL SAMPLE SIZE AND CRITICAL VALUE (GIVEN $Q_2 = 0.90$; IF $Q_2 = 0.85$, USE TABLE 2.) FIRST DETERMINE THE DESIRED β VALUE. IF $\beta = 0.05$ USE PART 1; IF $\beta = 0.01$ USE PART 2. THEN DETERMINE THE Q_1 AND α VALUES OF INTEREST. READ DOWN THE COLUMN HEADED BY THE APPROPRIATE Q_1 VALUE UNTIL AN α SMALLER THAN THAT DESIRED IS ENCOUNTERED. FOLLOW THIS ROW TO THE RIGHT TO FIND THE DESIRED N AND X VALUES.

Part 1: $\beta = 0.05$				
			α Values	
N	X	$Q_1 = 0.95$	0.97	0.99
29	0	0.7741	0.5866	0.2528
46	1	0.6768	0.4032	0.0775
61	2	0.5939	0.2767	0.0234
76	3	0.5307	0.1943	0.0072
89	4	0.4606	0.1297	0.0021
103	5	0.4110	0.0901	0.0006
116	6	0.3607	0.0608	0.0002
129	7	0.3178	0.0412	0.0001
142	8	0.2809	0.0280	0.0000
154	9	0.2429	0.0183	0.0000
167	10	0.2157	0.0125	0.0000
179	11	0.1871	0.0082	0.0000
191	12	0.1624	0.0054	0.0000
203	13	0.1411	0.0035	0.0000
215	14	0.1227	0.0023	0.0000
227	15	0.1068	0.0015	0.0000
239	16	0.0931	0.0010	0.0000
251	17	0.0811	0.0007	0.0000
263	18	0.0708	0.0004	0.0000
275	19	0.0618	0.0003	0.0000
286	20	0.0524	0.0002	0.0000
298	21	0.0458	0.0001	0.0000
310	22	0.0400	0.0001	0.0000
321	23	0.0339	0.0000	0.0000
333	24	0.0297	0.0000	0.0000
345	25	0.0260	0.0000	0.0000
356	26	0.0221	0.0000	0.0000
368	27	0.0193	0.0000	0.0000
379	28	0.0164	0.0000	0.0000
391	29	0.0144	0.0000	0.0000

picks a value of N and, using Equation 3, finds the largest value, X, such that

$$\sum_{Y=0}^{X} f(Y;N,Q_2) \leq \beta \qquad (4)$$

The value X as defined by Equation 4 is our critical value, because we will reject our map as being inaccurate if it shows more than X misclassifications in our N ground truth samples. Having determined X, one may now use Equation 3 to calculate the probability, α, of rejecting a map of some high accuracy proportion, Q_1. The quantity α, known as "producer's risk" in acceptance sampling, is given by

$$\alpha = \sum_{Y=X+1}^{N} f(Y;N,Q_1). \qquad (5)$$

It may be noted that, because the binomial p.d.f. is discrete, several values of N have the same associated critical value X. Further, for any fixed critical value X, low accuracy Q_2, and consumer's risk β, the smallest associated N value always yields the lowest producers risk, α. Table 1 shows values of N from 30 to 50 and their X values, for $\beta = 0.05$, $Q_2 = 0.85$, together with the α values associated with $Q_1 = 0.90$, 0.95, 0.99. Considering the column labeled "Exact β", which is equal to the left side of Equation 4, we see that, by increasing N for fixed X, we are in effect reducing our true consumers risk (Exact β) far below its nominal value, at the expense of inflating the producers risk (α). Thus, for fixed β and X the minimum associated N value may be considered optimal since it also minimizes producers risk.

Table 2 presents the optimal values of N which occur between $N = 1$ and $N = 400$ for

TABLE 3.—*Continued*

Part 2: $\beta = 0.01$				
			α Values	
N	X	$Q_1 = 0.95$	0.97	0.99
44	0	0.8953	0.7832	0.3574
64	1	0.8361	0.5759	0.1346
81	2	0.7766	0.4398	0.0480
97	3	0.7203	0.3323	0.0166
113	4	0.6717	0.2522	0.0058
127	5	0.6140	0.1832	0.0018
142	6	0.5692	0.1364	0.0006
156	7	0.5217	0.0989	0.0002
170	8	0.4787	0.0718	0.0001
183	9	0.4328	0.0505	0.0000
197	10	0.3977	0.0366	0.0000
210	11	0.3597	0.0257	0.0000
223	12	0.3254	0.0180	0.0000
236	13	0.2943	0.0127	0.0000
249	14	0.2662	0.0089	0.0000
262	15	0.2409	0.0062	0.0000
275	16	0.2180	0.0043	0.0000
287	17	0.1934	0.0029	0.0000
300	18	0.1750	0.0021	0.0000
312	19	0.1552	0.0014	0.0000
325	20	0.1405	0.0010	0.0000
337	21	0.1245	0.0006	0.0000
350	22	0.1128	0.0005	0.0000
362	23	0.0999	0.0003	0.0000
374	24	0.0885	0.0002	0.0000
386	25	0.0783	0.0001	0.0000
398	26	0.0693	0.0001	0.0000

$Q_2 = 0.85$, $\beta = 0.05, 0.01$, together with their associated critical values (X) and α values for $Q_1 = 0.90, 0.95, 0.99$. Part 1 of the table is for $\beta = 0.05$, part 2 for $\beta = 0.01$. Table 3 presents the same information except $Q_2 = 0.90$ and $Q_1 = 0.95, 0.97, 0.99$. (These tables were produced by a FORTRAN program written by the author. Copies of the program listing are available on request.)

In order to use these tables to determine the number of ground truth samples required to establish map accuracy, one should first have an accuracy proportion which is considered low (Q_2). If $Q_2 = 0.85$ use Table 2; if $Q_2 = 0.90$ use Table 3. One must then determine the desired probability of accepting a map of low accuracy (consumers risk or β). If $\beta = 0.05$, use part 1 of the table; if $\beta = 0.01$, use part 2. One now should pick an accuracy value which is considered high (Q_1) (for Table 2, $Q_1 = 0.90, 0.95, 0.99$; for Table 3, $Q_1 = 0.95, 0.97, 0.99$) and some objectionably large consumers risk, α. Having done this, one reads down the column headed by the appropriate Q_1 value until one encounters an α value smaller than the one selected above. Moving to the left along the row in which this α occurs one finds the N and X values required to produce this α, given our previously determined Q_2, Q_1, and β values. One then collects N ground truth samples, checks the corresponding points on the map against them, and accepts the map as accurate if X or fewer of the samples are misclassified.

For an example, say we consider an inaccurate map to be one in which the proportion correctly classified is 0.85 or less. Therefore $Q_2 = 0.85$ and we use Table 2. Say further than an acceptable consumers risk (β) is 0.05. We are therefore using part 1 of Table 2. Let us also assume that we want to have a probability of rejecting a map which is 95 percent accurate ($Q_1 = .95$) of less than 0.05 ($\alpha = 0.05$). We therefore enter the column headed by 0.95 and read down until we encounter $\alpha = 0.0432$. Reading across, we find that $N = 93$ and $X = 8$. Thus, we take 93 ground truth samples and accept our map as accurate if 8 or fewer of these samples are misclassified.

CONCLUSION

The choice of Q_2, Q_1, β, and α may be expected to vary with the mapping task at hand. In general $Q_2 = 0.85$ or 0.90 seems a reasonable specification for minimal accuracy requirements, and user satisfaction would seem to preclude consumers risk probabilities of greater than 0.05.

Producers risk and the associated problem of determining just what high accuracy is, are more flexible. If one feels that his maps are essentially perfect, and that specifying accuracy is merely a matter of formal verification, taking the smallest possible sample and accepting the map only if no errors are found (as in van Genderen and Lock, 1977) might be attractive in that this approach does minimize the cost of ground truth sampling. However, if one should reject a map, one must presumably correct it, which might be expensive. Thus, larger ground truth sample sizes which provide protection to good but not perfect maps would seem preferable. For routine mapping tasks, which should be very accurate, taking $Q_1 = 0.99$ and $\alpha = 0.10$ would require no more than 81 samples (Table 3, part 2, row 3) and would give considerable protection against rejection of accurate maps. In cases where map accuracy might be lower, one might want greater protection against rejecting acceptable maps and thus might opt for larger ground truth sample sizes. In any of these cases, given established values for α, β, Q_1, and Q_2, the methodology outlined here will allow one to select the minimum ground truth sample size, N, and allowable number of errors, X, which satisfy these restrictions. Tables 2 and 3 can, in most cases, be used to look up the required N and X directly. For combinations of Q_1, Q_2, α, or β not given, Equations 4 and 5, together with the fact that, for fixed X and β, α is minimized by the smallest N satisfying Equation 4, can be used to find the required solution.

ACKNOWLEDGMENTS

I would like to thank Dr. L. R. Holden and Mr. J. Poracsky for their critical comments. This work was supported by grant no. R 80-54-55-01-0 from the Environmental Protection Agency.

REFERENCES

Cochran, W. G., 1977. *Sampling Techniques*. Wiley, New York. 413 pp.

Guttman, I., S. S. Wilks, and J. S. Hunter, 1971. *Introductory Engineering Statistics*. Wiley, New York. 549 pp.

Hord, R. M., and W. Brooner, 1976. Land-Use Map Accuracy Criteria. *Photogrammetric Engineering and Remote Sensing*, 42:671-677.

Rohlf, F. J., and R. R. Sokal, 1969. *Statistical Tables*. W. H. Freeman, San Francisco. 253 pp.

van Genderen, J. L., and B. F. Lock, 1977. Testing Land Use Map Accuracy. *Photogrammetric Engineering and Remote Sensing*, 43:1135-1137.

Vaughn, R. C., 1974. *Acceptance Sampling*. Iowa State U. Press. Ames. 233 pp.

(Received June 5, 1978; revised and accepted June 25, 1979)

Fourth Australian Cartographic Conference

Hobart, Australia
November 5-7, 1980

The Conference, which is held every two years, will bring together cartographers from all over Australia, New Zealand, Papua-New Guinea, and overseas. Its theme is Change, Challenge, and Communication.

For further information please contact

Conference Convenor
G.P.O. Box 44A
Hobart 7001, Tasmania
Australia