Reprinted from

Ninth International Symposium

Machine Processing of

Remotely Sensed Data

with special emphasis on

Natural Resources Evaluation

June 21-23, 1983

Proceedings

Purdue University
The Laboratory for Applications of Remote Sensing
West Lafayette, Indiana 47907 USA

CATEGORY ANALYSIS OF THE CLASSIFICATION ERROR MATRIX

G.H. ROSENFIELD, K. FITZPATRICK-LINS

U.S. Geological Survey
Reston, Virginia

I. SUMMARY

A classification error matrix typically contains tabulation results of an accuracy evaluation of a thematic classification, such as that of a land use and land cover map. A sample matrix has been made of a land use and land cover map for an accuracy evaluation, where for each sample (or count) the interpretation (classification) is given as the rows (populations), and the verification is given as the columns (responses). The diagonal elements of the matrix represent the counts correct. The remaining elements of the rows represent the errors by commission, and the remaining elements of the columns represent the errors of omission. The proportions of diagonal elements divided by the row sums are considered as the category accuracy relative to errors by commission. This has been the usual designation of the classification accuracy. The proportions of the diagonal elements divided by the column sums represent the category accuracy relative to the errors of omission. Since certain categories are often misinterpreted as certain other categories, the percentages of the points misinterpreted are also given in some cases.[1] This has been the usual extent of the interpretation of the data contained within the classification error matrix. Rosenfield states that the next step in analyzing the classification error matrix would be to use the entire matrix in such studies.[2] A beginning has already been made in this regard.[3]

The classification error matrix is known in statistical terms as a contingency table (and sometimes as a cross-tabulated table) of categorical data. There is a large field of literature developed for the analysis of categorical data and cross-tabulated tables (see Selected References for Categorical Methods). The analysis of categorical data by linear models was started by Grizzle and others, and was expanded by Koch and others.[4,5] Koch explicitly develops the theory for hypothesis testing of the functional results of the methodology. The analysis of categorical data by building hierarchical models for solution by iterative proportional fitting is developed by Bishop and others.[6] Both methods can be used in many applications, and the results will usually be very similar.[7] It is the emphasis on hypothesis testing which differentiates the method of Grizzle and others from that of Bishop and others.[4,6] These contingency table methods allow analysis of the matrix considering all of the data contained, not just the diagonal elements.

The purpose of this study has been to determine how to analyze the entire classification error matrix resulting from the accuracy evaluation of a thematic map, not just the diagonal elements. The expected probabilities of misinterpretation of a category are estimated from the cell frequency divided by the population (row) sum. The misinterpretations among various particular pairs of categories will be analyzed for a number of different land use and land cover maps, and the results inspected for patterns of behavior.

Another question that has arisen in thematic classification deals with the systematic pattern of any one category being misinterpreted in the same manner as another category. This misinterpretation reflects on the ability of the remote sensing system to meet the requirements of the category definition within the classification system. The patterns of misclassification could result from either the remote sensing system or the classification system. A statistical hypothesis test is made on the functional results of the statistical

methodology to test for systematic patterns of misinterpretation. This amounts to testing the coefficients of the response parameters, which in turn represent the category response effects in the statistical model. The null hypothesis becomes: is any one category being interpreted in the same manner as any other category? The alternate hypothesis is that they are not being interpreted the same.

A third question deals with another way of considering the overall accuracy and category comparability problem. This is to consider the problem as one of a measure of agreement between classification and verification, or between classified categories. One measure of agreement that recently began to receive attention in remote sensing applications is that of Cohen's Kappa test for agreement.[8] This statistic has been given attention by Bishop and others.[6]

A number of land use and land cover maps produced under the national land use and land cover mapping program of the U.S. Geological Survey have been analyzed for accuracy.[1,9,10] Classification error matricies resulting from the accuracy tests described in these documents, as well as others obtained from the literature, will be analyzed. The methods of analysis of categorical data and measures of agreement will be used to investigate the information contained within the entire classification error matrices, and the applicability to remotely sensed data will be examined.

Preliminary results indicate that within the error matrix, explanations can usually be made for individual errors. However, the accuracy analysis has been found to be dependent upon having an adequate sample size for the validation of each category. In addition, the sample selection should be based upon some valid methods of allocation: say, proportional allocation supplemented by sampling within two frames.[11] When the error matrix is sparse, systematic errors cannot be explained or deduced from the limited data. Correlations found are the result of having similarly few misclassifications, not the result of having a large number of similar misclassifications.

Statistical testing of the entire matrix of classification errors for comparison between land cover types might reveal systematic errors introduced by misregistration, Sun angle, or similar spectral signatures. Once identified, the causes of these errors could then be examined more carefully.

II. REFERENCES

1. Fitzpatrick-Lins, K., 1978, Accuracy and consistency comparison of land use and land cover maps made from high-altitude photographs and Landsat multi-spectral imagery: Journal of Research, U.S. Geological Survey, v. 6, no. 1, p. 23-40.

2. Rosenfield, G.H., 1981, Analysis of variance of thematic mapping experiment data: Photogrammetric Engineering and Remote Sensing, v. 47, no. 12, p. 1685-1692.

3. Congalton, R.G., 1980, Statistical techniques for analysis of Landsat classification accuracy: Paper presented at meeting of the American Society of Photogrammetry, St. Louis, Missouri, March 11.

4. Grizzle, J.E., Starmer, C.F., and Koch, G.G., 1969, Analysis of categorical data by linear models: Biometrics, v. 25, p. 489-504.

5. Koch, G.G., Landis, J.R., Freeman, J.L., Freeman, D.H., Jr., and Lehnen, R.G., 1977, A general methodology for the analysis of experiments with repeated measurement of categorical data: Biometrics, v. 33, p. 133-158.

6. Bishop, Y.M.M., Feinberg, S.E., and Holland, P.W., 1975, Discrete multi-variate analysis--theory and practice: The MIT Press, Cambridge, Massachusetts.

7. Helwig, J.T., and Council, K.A., eds., 1979, The SAS user's guide: Statistical Analysis System Institute, Inc., Raleigh, N.C.

8. Cohen, J., 1960, A coefficient of agreement for nominal scales: Educational and Psychological Measurement, v. 20, no. 1, p. 37-46.

9. 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. Geological Survey Circular 829.

10. Fitzpatrick-Lins, K., 1981, Comparison of sampling procedures and data analysis for a land use and land cover map: Photogrammetric Engineering and Remote Sensing, v. 47, no. 3, p. 343-351.

11. Rosenfield, G.H., Fitzpatrick-Lins, K., and Ling, H.S., 1982, Sampling for thematic map accuracy testing: Photogrammetric Engineering and Remote Sensing, v. 48, no. 1, p. 131-137.

III. SELECTED REFERENCES FOR CATEGORICAL METHODS

Bishop, Y.M.M., Feinberg, S.E., and Holland, P.W., 1975, Discrete multi-variate analysis--theory and practice: The MIT Press, Cambridge, Massachusetts.

Feinberg, S.E., 1980, The Analysis of Cross-Classified Categorical Data, 2nd edition: The MIT Press, Cambridge, Mass.

Forthofer, R.N. and Lehmen, R.G., 1981, Public Program Analysis - A New Categorical Data Approach: Lifetime Learning Publications, Belmont, Ca.

Haberman, Shelby J., 1974, The Analysis of Frequency Data: The Univeristy of Chicago Press, Chicago, Ill.

Plackett, R.L., 1974, The Analysis of Categorical Data. Griffin, London.

Upton, G.J.G., 1978, The Analysis of Cross Tabulated Data: John Wiley and Sons, New York.

IV. BIOGRAPHICAL DATA

George H. Rosenfield. Mr. Rosenfield is currently employed as a Physical Scientist in the Office of Geographic and Cartographic Research of the U.S. Geological Survey where he in engaged in the application of inferential statistics and hypothesis testing to the solution of problems in remote sensing. He has been employed by the U.S Geological Survey since December 1968, in a number of management, research, and technology support positions for the application of mathematics and statistics to the solution of management and technical problems. Prior to this he worked 3 years with Autometric as a Principal Scientist (Photogrammetry); and almost 9 years with RCA as Photogrammetrist for the Atlantic Missile Range. He received a Master of Science Degree in Photogrammetry from New York State College of Forestry at Syracuse University in 1954; and a BS degree in Forestry from University of Massachusetts in 1949. He has received a number of awards and certificates from the American Society of Photogrammetry and from the Society of Photo-Optical Instrumentation Engineers. He has published more than 33 technical papers in professional journals and authored several chapters in the Manual of Photogrammetry, 3rd Edition.

Katherine Fitzpatrick-Lins. Ms. Fitzpatrick-Lins has been a geographer with the U.S. Geological Survey for the past 10 years. Presently she is working in the Earth Resources Observation System Office applying Landsat image processing to Natural Resource Management. Her involvement in the past was in Multi-disciplinary Resource Management projects and workshops, and in research into the accuracy of Land Use and Land Cover Maps.