Effect of the Number of Samples Used in a Leave-One-Out Covariance Estimator

Larry Biehl* & David Landgrebe+

School of Electrical and Computer Engineering

Purdue University, West Lafayette, IN

ABSTRACT

Some algorithms, such as Gaussian Maximum Likelihood require the use of the second order statistics, e.g. the covariance matrix, to help characterize the target in addition to the mean. Also, models such as FASSP require the second order statistics of targets to predict the performance of algorithms even though the algorithm may not use the covariance matrix directly. However, many times the number of samples available to make a good estimate of the covariance matrix is small. The Leave-One-Out Covariance (LOOC) estimator can be used to estimate the covariance matrix when the number of samples available is less than the normal minimum required. The normal minimum number of samples needed for a sample class covariance matrix is p+1 samples for p-dimensional data. For the LOOC estimator, in theory, as few as 3 samples are all that are needed. However, what are the affects of using such a low number in practice? This paper presents the results of an experiment that was conducted to measure what the affect may be in one specific instance. Sometimes as few as 0.1p samples produce reasonably satisfactory results; other times 0.4p or more samples are needed.

Keywords: Remote Sensing, Hyperspectral, Leave-One-Out Covariance Estimator

1. INTRODUCTION

Some algorithms, such as the Gaussian Maximum Likelihood classifier, which can be used for the detection and/or identification of targets within a scene, require the use of the second order statistics, e.g. the covariance matrix, to help characterize the target in addition to the mean. Also, models such as Forecasting and Analysis of Spectroradiometric System Performance (FASSP)[1] require the second order statistics of targets to predict the performance of algorithms even though the algorithm may not use the covariance matrix directly. However, many times where the number of spectral features is large and the targets are small relative to the spatial resolution of the imagery, the number of samples available to make a good estimate of the covariance matrix is small.

The Leave-One-Out Covariance (LOOC) estimator can be used to estimate the covariance matrix when the number of samples available is less than the normal minimum required.[2] The normal minimum number of samples needed for a sample class covariance matrix is p+1 samples for p-dimensional data. For the LOOC estimator, in theory, as few as 3 samples are all that are needed.

The LOOC estimator employs a mixing parameter to select an appropriate mixture of the (a) diagonal sample covariance, (b) sample covariance, (c) common covariance, and (d) the diagonal common covariance. The mixture deemed appropriate is the one that achieves the best fit to the training samples in the sense that it maximizes the average likelihood of training samples that were not used in the estimate. The Leave-One-Out technique[2] is to remove a sample, estimate the mean and covariance matrix from the remaining samples, then compute the likelihood of the sample which

* Correspondence: Email: biehl@purdue.edu; WWW: http://dynamo.ecn.purdue.edu/~biehl/; Telephone: 765 494 3529; Fax: 765 494 6951

+ Correspondence: Email: landgreb@ecn.purdue.edu; WWW: http://dynamo.ecn.purdue.edu/~landgreb/; Telephone: 765 494 3486; Fax: 765 494 3358.

was left out, given the mean and covariance matrix estimates. Each sample is removed in turn, and the average log likelihood is computed over all the left out samples. Several different values of the mixing parameter are examined, and the value that maximizes the average likelihood is selected. This covariance estimator is typically non-singular when at least three samples are available regardless of the dimensions of the data, and so it can be used even when the sample covariance or common covariance estimates are singular*.

However, what are the affects of using such a low number in practice? An experiment was conducted to measure what the affect may be in one specific instance.

2. METHODOLOGY/APPROACH

(1) Samples were selected for ten classes in a portion of a June 12, 1992, hyperspectral Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)[4] scene over agricultural crops that had a relatively large number of samples available. See Figure 1. The classes were Farmstead, Corn cleantill EW rows, Corn cleantill NS rows, Corn mintill EW rows, Corn, Soybeans cleantill EW rows, Soybeans drilled, Soybeans mintill EW rows, Soybean NS rows, and Wheat. NS and EW refer to the row direction. ‘Cleantill’ and ‘mintill’ refer to the amount of residue from the previous year’s crop that is on the surface of the ground. Clean tilled ground has very little residue; minimum tilled has a significant amount of residue on the surface. Some of the ten classes such as “Soybeans-Drilled” and

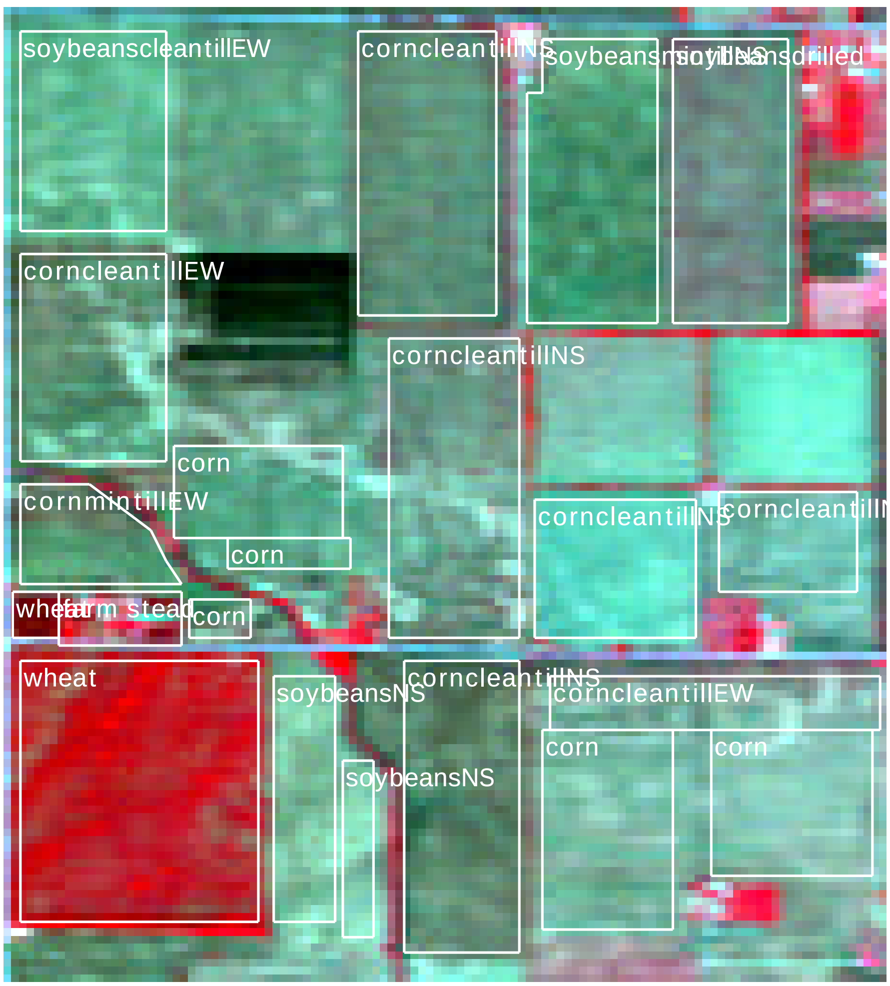

Figure 1. Color infrared image of June 12, 1992 AVIRIS data used for the experiment. The areas used for the selection of training samples for each of the 10 classes are outlined. (original in color)

* This algorithm has been implemented within a remote sensing image processing package called MultiSpec[3]. The application is freely available via http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/.

“Soybeans NS Rows” are very similar spectrally and therefore are difficult to separate. Other classes such as “Wheat” are relatively easy to separate from the other classes with the spectral data available. The AVIRIS scene contained 220 channels of data from 0.4 to 2.4 µm. The spectral classes used in this study were purposely created so that some would be very difficult to separate. Most analyses of this type of data are done to separate only the different species, not different cultural practices within a species.

(2) Random selections of the above samples were used for each of .02n, .03n, .04n, .05n, .1n, .2n, .3n, .4n, .5n, 1n, 2n, and 3n training samples, where n, the number of channels, was 220. For some classes there were only 1n training samples available.

(3) The means and LOO Covariances were computed for each of the sets of training samples selected in step 2.

(4) All samples selected in step 1 were classified using the quadratic maximum likelihood classifier and the statistics generated in step 3. The total number of samples classified was 9,236.

(5) Steps 2, 3 & 4 were run ten times. In other words the experiment was repeated ten times with a different (random) set of training samples for each time.

3. RESULTS

Figure 2 illustrates the results for four of the ten classes with .02p to 1.0p training samples where p is 220 channels (4 to 220 samples). As a reference when all of the samples were used for training the classification accuracy was 99.7% or in other words the error rate was 0.003.

Note some trends that are apparent in Figure 2:

(1) The error rate decreases as the number of training samples increases.

(2) The variation in the error rates of the ten experiments also tends decrease as the number of training samples increase.

(3) There is an inflection point in the error rate curve for spectrally similar classes such as corn and soybeans around 0.3p to 0.4p above which the error rate drops off more slowly. The inflection point for the spectrally different Wheat class is around 0.1p.

4. CONCLUSIONS

The classification of easy to separate classes (Wheat) begins to breakdown, i.e. the error rate and experimental variation increases significantly, when fewer than 0.1p samples were used. The classification of difficult to separate classes (Soybeans NS and Soybeans-Drilled) begins to breakdown when fewer than 0.4p samples were used. These results indicate that for a case with 220 spectral channels, one would want to have at least 20-90 samples of data to estimate the statistics for a class using the LOO Covariance estimator. Twenty samples may be sufficient if the class is spectrally different from all of the other classes in the scene. However, if there are classes that are spectrally similar then one would want to have at least 90 samples.

Note one should not use 20 or 90 as “magic” numbers for 220 channel spectral data. These represent the results for one scene and set of experiments. However, the significant conclusion should be that 3 samples are too few for a robust estimate using the LOO Covariance estimator. Also one needs more training samples when classes within the scene are spectrally similar to each other.

ACKNOWLEDGEMENTS

This research was funded in part by an Army Research Office (MURI) Grant DAAH04-96-1-0444 (1996-1999). This support is gratefully acknowledged. Much of the computer analyses presented in this paper were performed by a student, Eddie Dharma.

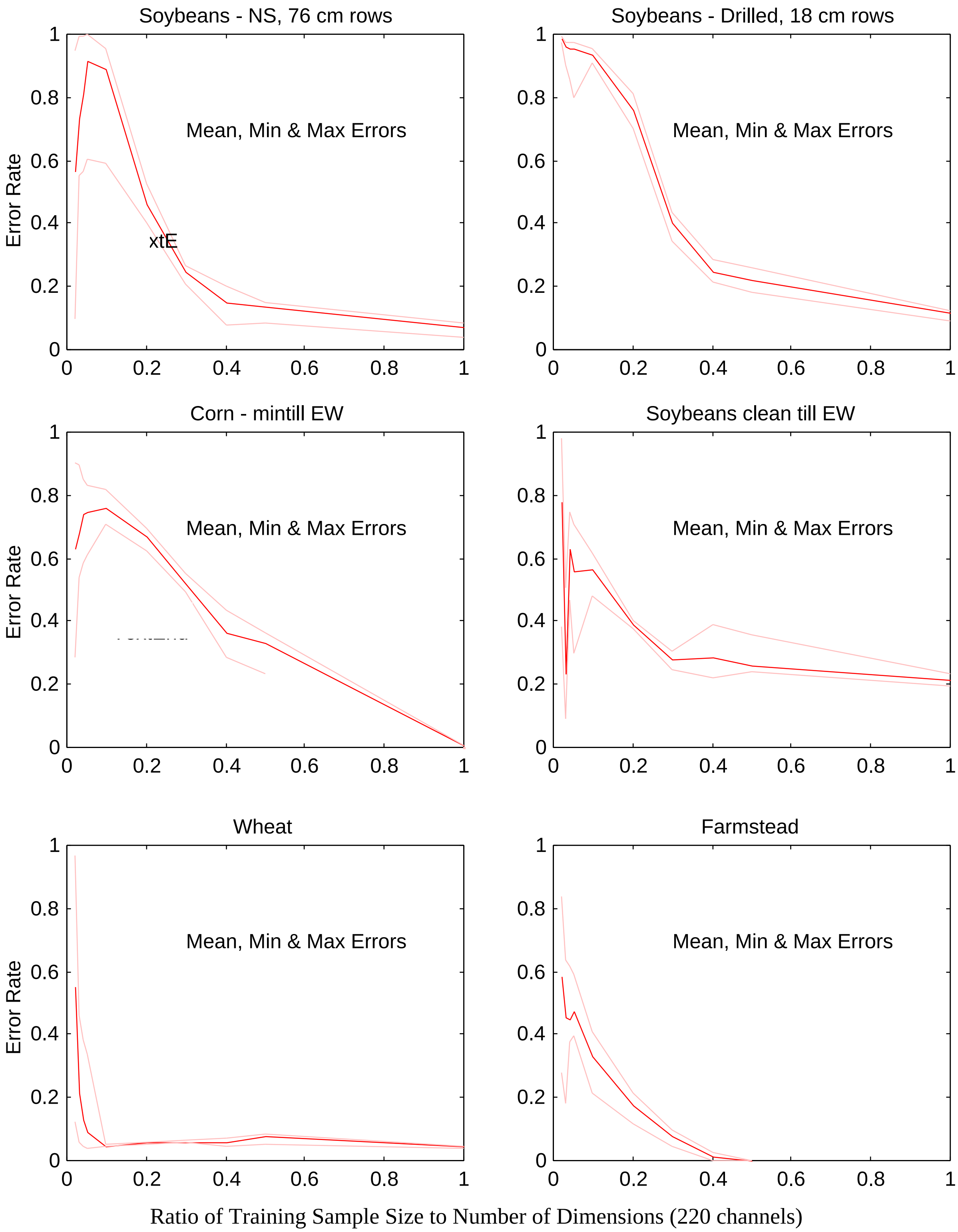

Figure 2. Illustration of the error rate found for four classes using 0.02n, 0.03n, 0.04n, 0.05n, 0.1n, 0.2n, 0.3n, 0.4n, 0.5n, 1.0n sample for training where n was 220 channels. The solid line represents the mean error rate of the ten experiments. The dotted lines represent the minimum and maximum error rate.

REFERENCES

1. J. Kerekes, J. Baum, and K. Farrar, "Analytical Model of Hyperspectral System Performance," Proceedings of Infrared Imaging Systems: Design, Analysis, Modeling and Testing X, SPIE Vol. 3701, pp. 155-166, 7-8 April 1999.

2. J. Hoffbeck and D. Landgrebe, "Covariance Matrix Estimation and Classification with Limited Training Data," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. **18**, no. 7, pp. 763-767, July 1996. (This paper is on the web at: http://dynamo.ecn.purdue.edu/~landgreb/LOOC%20PAMI-web.pdf)

3. L. Biehl, L. and D. Landgrebe, 1996, "MultiSpec - A Tool for Multispectral-Hyperspectral Image Data Analysis", Pecora 13, Sioux Falls, South Dakota, August 1996. (The web site for MultiSpec is: http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/)

4. G. Vane, R. Green, T. Chrien, H. Enmark, E. Hansen, W. Porter, “The Airborne Visible/Infrared Imaging Spectrometer (AVIRIS).” *Remote Sensing of Environment*, Vol 44, pp 127-143, 1993.