Improving the Accuracy of Historic Satellite Image Classification by Combining Low-Resolution Multispectral Data with High-Resolution Panchromatic Data

Daniel J. Getman[1], Jonathan M. Harbor[2], Chris J. Johannsen[3], Bernard A. Engel[4], & Goufan Shao[5]

[1,2] Department of Earth & Atmospheric Sciences, Purdue University; [3]Department of Agronomy, Purdue University, [4] Department of Agricultural & Biological Engineering, Purdue University; [5] Department of Forestry and Natural Resources, Purdue University

ABSTRACT

Many attempts to observe changes in terrestrial systems over time would be significantly enhanced if it were possible to improve the accuracy of classifications of low-resolution historic satellite data. In an effort to examine improving the accuracy of historic satellite image classification by combining satellite and air photo data, two experiments were undertaken in which low-resolution multispectral data and high-resolution panchromatic data were combined and then classified using the ECHO spectral-spatial image classification algorithm and the Maximum Likelihood technique. The multispectral data consisted of 6 multispectral channels (30-meter pixel resolution) from Landsat 7. These data were augmented with panchromatic data (15m pixel resolution) from Landsat 7 in the first experiment, and with a mosaic of digital aerial photography (1m pixel resolution) in the second. The addition of the Landsat 7 panchromatic data provided a significant improvement in the accuracy of classifications made using the ECHO algorithm. Although the inclusion of aerial photography provided an improvement in accuracy, this improvement was only statistically significant at a 40-60% level. These results suggest that once error levels associated with combining aerial photography and multispectral satellite data are reduced, this approach has the potential to significantly enhance the precision and accuracy of classifications made using historic remotely sensed data, as a way to extend the time range of efforts to track temporal changes in terrestrial systems.

KEYWORDS: image classification, multispectral data, panchromatic data, data accuracy, remote sensing, archival data

INTRODUCTION

Multispectral satellite data have been collected over the United States since the successful launch of the first in the Landsat series of satellites in 1972. Although multispectral satellite data are readily available for both historic land cover analysis and temporal change analysis, their use is limited by relatively low spatial resolution. Analysis that incorporates modern satellite technology with resolutions of 5-meters and better with historic satellite data at resolutions of 80 and 30 meters is limited by the lower resolution of the older data. To overcome this limitation for terrestrial observation, there is considerable interest in improving the classification potential of historic satellite data through processes that augment multispectral data channels gathered at relatively low spatial resolutions with additional data gathered at higher spatial resolutions. Very little attention has been given, however, to the use of aerial photography for this purpose. This is most likely due to the difficult task of accurately combining low-resolution satellite data with high-resolution aerial photography.

Traditionally, satellite image classification has been accomplished with algorithms that incorporate spectral pattern recognition techniques to classify each pixel individually, based on the organization of its spectral response in each of the recorded data channels (Lillesand and Kiefer, 2000). Because each pixel is classified independently, information provided by spatial patterns within the data is unavailable for use by these algorithms. In an effort to take full advantage of the spatial information available in remotely sensed data, algorithms that combine spectral and spatial pattern recognition techniques during image classification have also been developed. The Extraction and Classification of Homogenous Objects (Landgrebe, 1980) algorithm, developed by Kettig and Landgrebe (1976), uses scene segmentation techniques to extract spatial data for use along with spectral properties during image classification. In previous experiments with the ECHO algorithm, Landgrebe (1980) has shown it to be an effective tool for separating spectrally similar land cover types within multispectral data sets.

Because the ECHO algorithm is capable of considering spatial patterns during image classification, it is particularly well suited to the task of classifying a data set in which the multispectral data have been augmented with high-resolution data. The high-resolution data provide the algorithm with additional detail with which to extract spatial patterns that can be used along with the multispectral information to improve accuracy in image classification. Although the ECHO algorithm is an excellent candidate for this analysis, all classification algorithms are susceptible to errors introduced during image registration.

In an effort to test this method of improving the classification potential of historic satellite data, two experiments were undertaken in which the ECHO algorithm was used to classify spatially enhanced multispectral data. In the first experiment, multispectral data channels from Landsat 7 with a ground resolution of 30 meters were combined with the 15-meter resolution panchromatic data channel from the same satellite. Because both data sets were captured simultaneously by the same device over the same geographic area, issues associated with geometric registration and temporal differences between the data were avoided. This combination of data allowed for a direct assessment of the ability of the ECHO algorithm to utilize the additional spatial information in the improvement of overall classification accuracy.

The second experiment was conducted using the same multispectral data, the same training samples, and the same test samples, however, the exercise was taken a step further by substituting the Landsat 7 panchromatic data channel with high-resolution (1-meter) aerial photography collected over the same geographic area. As these data were captured using significantly different equipment at a different point in time, it was necessary to ensure that the aerial photography was mosaiced and geometrically registered to match the multispetral satellite imagery. Also, both data sets were resampled to 1, 5, and 15-meter pixel resolutions in an effort to find an appropriate resolution for use in classification.

The assessment of these methods was accomplished through a comparison of thematic maps generated by both the ECHO algorithm and the Maximum Likelihood technique, with and without the high-resolution panchromatic data set in the first experiment and the high-resolution aerial photography in the second.

MATERIALS AND METHODS

OVERVIEW

Using ERDAS Imagine, the 30m resolution Landsat 7 multispectral data channels were resampled to match the higher resolution of the 15m panchromatic data channel used in the first experiment. For the second experiment, the multispectral data channels were resampled along with the 1m resolution aerial photography to generate 15m, 5m, and 1m data sets. Training samples were selected within ArcView GIS using high-resolution aerial photography, and then transferred to the satellite imagery. Using MultiSpec, four classifications were made at each resolution in which the enhanced and un-enhanced data sets were each classified with both the Maximum Likelihood and ECHO algorithms. The accuracy of these classifications was then determined through statistical analysis of randomly selected training samples.

DATA SET DESCRIPTION

In the process of preparing aerial photography and satellite imagery to be combined for use in classification, alterations to the data are usually necessary. The air photos must be mosaiced, adjusted for differences in contrast, resampled to a compatible pixel size, and geometrically registered to match the satellite imagery. The multispectral satellite data must also be resampled to match the more precise pixel size of the higher resolution aerial photography. Although these efforts are necessary in analysis of multi-resolution data, they are manipulations of the data and thus introduce a variety of potential errors into the data set. Geometric misalignment of data sets and temporal differences between data sets can both result in decreased classification accuracy. For this reason, two experiments were undertaken to explore both the theoretical concept and the practical application of the method. In both experiments, the multispectral data used in this research represented a portion of a Landsat 7 Enhanced Thematic Mapper scene (Table 1) captured in 1999. The scene depicts an area just northwest of downtown Des Moines, Iowa USA that covers approximately $5.3 \times 10^{7}\ m^{2}$ (1.3×10^{4} acres). This area was chosen because it included a range of land cover classes, and under visual inspection the differences between the satellite imagery and aerial photography were minimal.

TABLE 1
SUMMARY OF LANDSAT ETM+ SENSOR CHARACTERISTICS

Data Channel	Portion of EM Spectrum	Spectral Definition	Resolution
1	0.45-0.515 µm	Blue-Green	30m
2	0.525-0.605 µm	Green	30m
3	0.63-0.690 µm	Red	30m
4	0.75-0.90 µm	Near IR	30m
5	1.55-1.75 µm	Mid IR	30m
6	2.09-2.35 µm	Mid IR	30m
Pan	0.52-0.90 µm	Green-Near IR	15m

The first experiment was performed using a data set in which the panchromatic data channel from the Landsat 7 scene was combined with the corresponding multispectral data channels. By substituting the higher resolution multispectral and panchromatic data from Landsat 7 for the historic satellite data and aerial photography, respectively, error associated with the rectification of geometric and temporal differences between the data sets was avoided. In using an ideal data set, the results of this experiment are made more applicable to the first objective of this study, which was to determine the effectiveness with which the ECHO algorithm can utilize additional high-resolution spatial information to improve classification accuracy.

In the second experiment, aerial photography was combined with the Landsat 7 multispectral data channels to better simulate an actual combination of historic satellite data with aerial photography. The aerial photography was captured, georeferenced, and combined into photo mosaics through a joint effort of the USDA Natural Resource Conservation Service, Iowa State University, and the Massachusetts Institute of Technology. Mosaics built at the maximum resolution (1m) were generated from the original photography captured in 1993, and were not resampled, feathered, or adjusted for contrast. By using more recent imagery for this experiment, temporal differences between the Landsat 7 data and the aerial photography were reduced, and the results of the first and second experiments can be more easily compared. This comparison is useful in determining if the methods used within this research can be applied to improve the classification accuracy of historic remotely sensed data.

DATA SET PREPARATION

The differences in the ground resolution of the multispectral data and both the high-resolution panchromatic data channel and aerial photography required the data to be resampled to a resolution that would be appropriate for classification. For the data set used in the first experiment, the 30m multispectral channels were resampled to match the 15m resolution of the panchromatic data channel. Due to the significant difference between the 30m resolution of the multispectral data and the 1m resolution of the aerial photography used in the second experiment, both data sets were sampled to 15, 5, and 1-meter grid cells in an effort to determine the optimum resolution for use in classification.

In resampling the multispectral data set to a higher spatial resolution, it was necessary to use a technique that would replicate the original pixels, leaving the radiometric properties of the data unaltered. As it is computationally simple and does not introduce new values into the data (Dikshit and Roy, 1996), the Nearest Neighbor resampling technique was used. As the image is not rotated or distorted in this process, every pixel in the resampled image is assigned its original value. Each 30-meter pixel from the multispectral data set was subdivided into four15-meter pixels, thirty six 5-meter pixels and nine hundred 1-meter pixels, all of which were assigned the spectral properties of the original pixel. By using this process, the radiometric properties of this data set were unaltered. There was no need to resample the panchromatic data channel, and it was simply combined with the resampled multispectral data for use in the first experiment.

The aerial photography was resampled to a lower spatial resolution for integration into the 15m and 5m data sets, which required that the combined response of several adjacent pixels in the original image be accurately estimated and assigned to a single pixel in the resampled image. As it incorporates a greater number of the original pixels when determining the value assigned to a single pixel in the resampled data than either the Nearest Neighbor or Bilinear techniques (Campbell, 1996), the Cubic Convolution technique was used to resample the aerial photography and produce data sets with 5 and 15-meter pixel resolutions from the original 1-meter pixel resolution data.

Although this technique is a reasonable method for estimating the most appropriate value for pixels in the resampled image, it does not represent exactly what would have been captured with a lower resolution photograph. Because of this, it is assumed that a variety of errors, some detrimental to classification accuracy, were introduced during this process.

After resampling, the aerial photography was geometrically registered to the multispectral data, and combined into a single data set at each resolution for use in the second experiment.

GEOMETRIC REGISTRATION

Because the aerial photographs were already mosaiced, the preprocessing required to combine the two resampled data sets for the second experiment was limited to cropping and geometric registration. The aerial photographs were manually registered to the multispectral data to ensure that the original radiometric properties of the multispectral data were preserved (Munechika et al., 1993), however, the radiometric properties of the aerial photographs were distorted during this process. After registration, the two data sets were cropped to identical geographic dimensions and combined into a single data set.

Error associated with image registration can significantly degrade classification accuracy. To reduce this error, registration of multiple data sources should achieve sub-pixel accuracy when the combined data is to be used in classification (Pohl and Van Gendern, 1998). In the process of registering data at 1-meter pixel resolution to data at 30-meter pixel resolution, sub-pixel accuracy is as difficult to measure as it is to achieve. The assessment of the registration procedure used in this analysis was limited to a visual inspection of the combined data set. Although the two images appeared to line up closely under visual inspection, errors in registration between data with 30m and 1m pixel resolutions are difficult to detect. It was assumed that errors resulting from this process were present in the data during classification, and had a negative impact on classification accuracy.

Temporal differences between the two data sets, which were separated by approximately six years, were also present in the combined data set. These differences were found to be negligible during visual inspection, however, errors associated with temporal registration have an effect similar to that of errors associated with geometric registration on classification accuracy.

CLASSIFICATION AND TRAINING SAMPLE SELECTION

Any selection of a land cover classification scheme is driven by the purposes of the study and the characteristics of the study area. In this case the larger goal of the research related to impacts of land cover change on surface water runoff in an urbanizing area, and thus land cover divisions were based on measures of impervious area. Two residential and four non-residential classes were selected by choosing class boundaries based on the percentage of impervious surface for each class, which was estimated from the average lot size for that class ().

Training samples were selected using the combination of high-resolution aerial photography and the Landsat 7 multispectral data channels created for the second experiment. The aerial photography was interpreted within ArcView GIS and individual lots were measured within each neighborhood to determine the average lot size (Table 2). Once the image was interpreted, and the class assignments were made, the boundaries of the training samples were replicated on the combined data set within MultiSpec. Training samples were selected in rectangular groups of pixels, as the neighborhoods were usually rectangular in shape. The total number of samples selected for each land cover class varied with the percentage of the total area occupied by that class (Table 2).

TABLE 2

LAND COVER CLASS DESCRIPTIONS AND TEST AND TRAINING SAMPLE SUMMARY

Land Cover Class		Sample Selection Parameters	Approximate % of Total Image	# of Training Samples	# of Test Samples
(H)	High Density Residential	Residential Neighborhood with Lot Size < 1000 m^2 (> 40% Impervious)	29.2%	70101 (29.9%)	87 (27.1%)
(M)	Medium Density Residential	Residential Neighborhood with Lot Size > 1000 m^2 (< 40% Impervious)	31.8%	68717 (29.3%)	75 (23.3%)
(C)	Commercial	Commercial or Urban Area with > 90% Impervious Surface	16.5%	36984 (15.7%)	50 (15.5%)
(G)	Field	Fields with < 10% Impervious Surface	11.2%	29876 (12.7%)	50 (15.5%)
(F)	Forest	Forest with < 10% Impervious Surface	8.6%	22735 (9.7%)	50 (15.5%)
(W)	Water	Rivers, Lakes, Wetland Areas	2.7%	6381 (2.7%)	9 (2.8%)

THE MAXIMUM LIKELIHOOD ALGORITHM

The Maximum Likelihood algorithm is the most commonly used supervised classification technique in remotely sensed image classification (Landgrebe, 1980; Michelson et al., 2000; Richards, 1999; Tso and Mather, 1999). The technique is a per-pixel classifier, and therefore assigns a class value to each pixel based on its individual spectral response pattern. Because the algorithm considers the spectral properties of each pixel independently, it is incapable of incorporating the relationships that exist between multiple pixels. This limitation precludes the use of any available spatial information by the Maximum Likelihood technique during image classification.

THE ECHO ALGORITHM

The ECHO algorithm was introduced by Kettig and Landgrebe (1976) in an effort to "exploit a particular type of dependence between adjacent states of nature that is characteristic of the data". This dependence is reflected in two aspects of remotely sensed data. First, pixels that are in spatial proximity to each other are unconditionally correlated, and the degree of correlation decreases with the distance between the pixels (Kettig and Landgrebe 1976). That is to say, the probability that two pixels are representative of the same land cover class is directly related the distance between them. The second aspect is a direct result of the first; if two pixels are spatially correlated, and are very similar spectrally, they have an even higher probability of belonging to the same land cover class. These two aspects of remotely sensed data favor the combination of two existing concepts in image classification, scene segmentation and Bayesian statistical theory.

The ECHO algorithm incorporates scene segmentation and Bayesian statistical theory in a two phase classification process. In the first phase, a scene segmentation technique is used to separate the image into statistically homogeneous objects. The second phase involves the classification of those homogeneous groups of pixels using a Maximum Likelihood Sample Classification technique derived from Bayesian statistical theory. This combination of techniques allows the ECHO algorithm to take advantage of spatial trends in the data that are not available to single pixel classifiers.

Previous testing of this algorithm has demonstrated that it is capable of producing a more accurate thematic map, with less class variance, than the traditional Maximum Likelihood technique (Landgrebe, 1980).

REFERENCE SAMPLING ISSUES AND STRATEGY

DEFINING THE SAMPLING ISSUES

A major contributing factor to error in accuracy assessment lies in organization of the sampling strategy. Unfortunately, many published analyses do not report the details of their sampling strategy, and fewer comment on its role in the assessment of map accuracy (Hammond and Verbyla, 1996). The effective interpretation of research results is dependent on understanding the sampling methods used, and so for this analysis these are described in detail.

The objective of the accuracy assessment should be the dominant factor guiding sampling strategy development (Stehman, 1999; Stehman and Czaplewski, 1998). The objective here was to assess the ECHO algorithm's ability to utilize high-resolution panchromatic data, in addition to standard multispectral data, in separating specifically defined land cover classes. By first determining the effectiveness of this method with both an ideal data set, then experimenting with one in which temporal distortion and geometric registration are introduced, the potential of using the approach to enhance historic satellite data with aerial photography can be more accurately estimated.

Of the six land cover classes used in this analysis (Table 2), medium and high-density residential proved to be the most difficult to separate. Due to the spectral similarities of the two classes and the high degree of variability within each class, training the classifier to differentiate between them was challenging. The level of difficulty in separating these two classes is directly related to the resolution of the data, which makes these classes particularly well suited for the purpose of accuracy assessment in this analysis. The high degree of lot size variation within individual neighborhoods complicated the classification process, and several neighborhoods simply did not fit the parameters specified for either land cover class. Even though high-resolution air photos were used for the interpretation, the vagueness of lot boundaries made the average lot size for each neighborhood difficult to confirm. This resulted in several neighborhoods that could not be classified with a high degree of confidence in the air photo interpretation. Keeping the objective in mind, these areas were not included in the target population used in reference data selection because they were not representative of the specifically defined classes that would be used to determine the separation capability of the ECHO algorithm.

These conditions precluded the use of traditional simple random sampling, which is theoretically the ideal method for accuracy assessment (Congalton, 1988). Because of typical logistical issues in the collection of reference data, this situation is not uncommon (Congalton, 1991; Hammond and Verbyla, 1996; Hord and Brooner, 1976; Stehman and Czaplewski, 1998), and several alternative methods are summarized by Congalton (1988). All of these methods, however, are designed to be used in determining the general accuracy of the entire mapped area (Congalton, 1991; Story and Congalton, 1986), rather than to analyze the performance of a specific capability of the classifier. Sampling methods were thus designed to facilitate an accuracy assessment that fulfilled the requirements of the objectives of this study.

DEFINING THE SAMPLING STRATEGY

The first step in the sampling strategy used in this research was the identification of a target population. Although the target population generally represents the entire mapped area, it can also be a subset of that area (Stehman, 1999). The target population in this study was defined by combining polygons that were identified with a high degree of confidence during the initial air photo interpretation. Due to the importance of limiting errors introduced by misinterpretation of reference samples (Congalton, 1991), areas that were not identified with a high degree of confidence offered no contribution to the interpretation of the accuracy assessment and were, therefore, not included in the random sampling. By limiting the target population in this way, the accuracy assessment is limited in that it defines a single attribute of the classification. In this analysis, that attribute is the ability of the ECHO algorithm to differentiate between the specifically defined medium-density and high-density residential land cover classes within the study area.

It is important to note that the entire mapped area was not included in the selection of random samples for reference data. Any measurement of accuracy made using reference data selected in this fashion (that is to say, it does not permit equal probability sampling for the entire mapped area), could produce an optimistic estimate of the overall accuracy of the entire mapped area (Stehman and Czaplewski, 1998). However, the capability of the ECHO algorithm to produce a map with high overall accuracy has already been established (Landgrebe, 1980), and is not the objective of this research.

By generating random sample points within this selective target area, it was possible to produce an error matrix that was statistically valid for use in assessing the algorithm's ability to separate the land cover classes specifically identified within that area. A comparison of classifications made using multispectral data sets with and without the inclusion of an additional high-resolution data set is, therefore, an appropriate estimate of the algorithm's capability to utilize additional spatial information provided by the inclusion of that data.

The interpretation of the randomly generated reference samples was accomplished using high-resolution aerial photography. This method, combined with the enhanced resolution of the satellite imagery, allowed for single pixel samples to be used. This is the preferred sample size when working with satellite data (Janssen and van der Wel, 1994; Franklin et al., 1991), and is advantageous in that optimistic bias associated with the selection of adjacent pixels is avoided (Hammond and Verbyla, 1996).

Due to the differences between the resolutions of the data sets used in the first and second experiments, the center of each randomly selected test pixel used in the first experiment was selected as the test sample in the second experiment. This was possible because the test samples were interpreted using the 15-meter data, and each pixel within both subsequent resampled data sets was divided into an odd number of pixels (nine for the 5-meter data set and twenty five for the 1-meter data set). The use of the same test samples for every classification made throughout both experiments allowed a comparison of accuracy with the same number of single pixel test samples, at the same randomly selected locations, for classifications in both experiments at each resolution.

Determining a statistically valid estimate of classification accuracy from an error matrix requires a minimum number of samples, which is based on both the total number of land cover classes and the size of the mapped area. A minimum of 50 samples from each land cover class represented in the error matrix has been recommended (Congalton, 1991; Hay, 1979) for classifications containing a small (less than 12) number of land cover categories. Congalton (1991) also suggests that the number of samples for each land cover class be adjusted if necessary, taking into consideration its percentage of the total area, the degree of variability within each class, and its importance to the objectives of the classification.

The number of samples selected for each of the six land cover classes used in this study was based on the relative proportion of each class to the total area, ensuring that each of the major classes had at least 50 samples (Table 2). The medium and high-density residential classes received additional samples due to the level of difficulty in distinguishing between them, and their importance to the goals of the accuracy assessment. The water class received far fewer reference samples due to the limited number of pixels over water, and to the ease with which water is classified.

ACCURACY ASSESSMENT METHODS

In this analysis, the accuracy assessment focused on the comparison of classifications made with and without the inclusion of a high-resolution panchromatic data set. An error matrix (Card, 1982) was generated from each classification and used to derive the Overall Accuracy, Producer's Accuracy, User's Accuracy and an estimate of Kappa referred to as KHAT. The details of these methods are readily available (Cohen, 1960; Congalton et al., 1983; Story and Congalton 1986;) and the discussion is, therefore, limited to a brief description of the reasons for their inclusion in this research.

Overall Accuracy is the most commonly used estimate of accuracy in satellite image classification (Congalton et al., 1983), and was included here as a general reference point for comparisons of the accuracy of each classification. Calculations of the Producer's and User's accuracy (Story and Congalton 1986) were included in this assessment to provide references to the accuracy with which specific land cover classes were classified. This is important in distinguishing the contributions from individual land cover classes to the overall accuracy of the classification and in making comparisons of these values from each classification technique.

Kappa, also referred to as KHAT $\left(\hat{K}\right)$ (Cohen, 1960), can be used to demonstrate the difference between the accuracy of a classified image and that of a map produced by selecting class values at random (Congalton et al. 1983). The technique has become a standard component of accuracy assessment in remote sensing (Congalton, 1991; Congalton et al., 1983; Hudson and Ramm, 1987; Rosenfield and Fitzpatrick-Lins, 1986) and was included in this assessment along with a statistical test, described by Cohen (1960), that can be used to determine the significance of the KHAT statistic for any single error matrix. These calculations were used here to define the significance of the differences between classifications made with various algorithms in order to evaluate specific attributes of their performance (Congalton, 1991).

RESULTS AND DISCUSSION

In both experiments, each data set was analyzed by comparing four classifications; two generated using the ECHO algorithm and two generated using the Maximum Likelihood technique. Of the two classifications made using ECHO, the first (ECHO 6) was generated using the six multispectral data channels and the second (ECHO 7) incorporated additional high-resolution panchromatic data. The same was true of the two classifications made with the Maximum Likelihood technique. The first classification made using the Maximum Likelihood technique (MAX 6) was generated using the six multispectral data channels and the second (MAX 7) incorporated additional high-resolution panchromatic data. Training and test samples were constant, and the resampled multispectral data from Landsat 7 were used in all classifications. Improvements in the accuracy achieved by the two algorithms are, therefore, directly attributed to their ability to utilize the additional spatial information made available with the inclusion of high-resolution panchromatic data and their susceptibility to the errors introduced during the preparation procedures associated with each individual data set.

The complete error matrices, along with the user's and producer's accuracies for each class, the overall accuracy, the KHAT statistic, and the variance of KHAT, are summarized in figures 1-3 and 4-12 for the first and second experiments, respectively. As the class values are abbreviated in each of these figures, a legend is provided in table 2. The thematic maps derived from each method are included in figures 13-15. ***(See appendix A for all figures.)***

ACCURACY ASSESSMENT: EXPERIMENT ONE

The overall accuracy of the six band classification increased by 11.5% when the ECHO algorithm was used in place of the Maximum Likelihood technique (figure 3). This number increased to 15.0% when the panchromatic data channel was included in both classifications, which is an improvement of 29.7% of the initial difference in accuracy between the two classifications (figure 3). Similarly, when the panchromatic data was included in the classification, the difference between the values of the KHAT statistic for each technique increased by 30.0% of the initial value, the difference in the variance of KHAT increased by 44.2% of the initial value, and the difference in its significance increased by 80.7% of the initial value (figure 3).

By examining the change in the difference between the accuracy of classifications made with each method when the high-resolution panchromatic data channel is included, the effectiveness with which the each algorithm can utilize the additional data can be compared. The estimates of thematic map accuracy used in this study represent independent aspects of accuracy, and therefore the amount of improvement varies among them. All of these comparisons in the first experiment do represent significant improvement however, and clearly demonstrate the ability of the ECHO algorithm to incorporate additional spatial information in image classification. The inability of the Maximum Likelihood technique to take advantage of additional spatial information is apparent in these results as well.

Both techniques did quite well in their classification of the four non-residential classes. In these four classes, the largest difference between the two classification techniques occurred in the Field class, in which there was a 4.0% difference in both the user's and producer's accuracies between the thematic maps made with the ECHO and Maximum Likelihood techniques. Compared to the differences in the high and medium-density classes, which were just under 20% for both the user's and producer's accuracies, 4% is a very small difference (figures 1 and 2). That improvements in accuracy occur almost completely within the two residential classes is an observation that further demonstrates the capability of the ECHO algorithm to utilize spatial information to differentiate between spectrally similar land cover classes.

In comparing the differences between the ECHO 6 and ECHO 7 classifications specifically, the ability of the ECHO algorithm to utilize the additional information provided by the panchromatic data is explicitly demonstrated. In this experiment, the overall accuracy of the ECHO 7 classification was 5.3% higher than the accuracy achieved by the ECHO 6 classification. Similarly, the value of the KHAT statistic increased by 8.2% of the ECHO 6 value. The variance of the KHAT statistic for ECHO 7 was 32.5% less, and its significance 31.6% greater, than the ECHO 6 value (figure 3).

The significance of the difference between the KHAT statistics for the ECHO 6 and ECHO 7 classifications was determined with the Z statistic (7) and found to be 2.0 standard deviations from the mean. Because the only difference between the two classifications was the addition of the high-resolution panchromatic data, improvements in the classification can only be attributed to the ECHO algorithm's ability to use the additional information available in the panchromatic data. The high value of the Z statistic indicates a significant difference between the two classifications at a 95% confidence interval.

The ECHO 6 and ECHO 7 classifications separated and identified the four non-residential classes equally well. Within these four classes, the most dramatic difference between the two classification techniques occurred in the forest class, in which there was a 0.1% difference in the user's accuracy, and a 2.0% difference in the producer's accuracy. Compared to the differences in the high and medium-density residential classes, which were approximately 10% for both the user's and producer's accuracies, 2.0% is a small difference (Figures 1 and 2). This demonstrates, just as in the comparison between ECHO and the Maximum Likelihood technique, that improvements in accuracy occur almost completely within the two residential classes, a fact that further supports the importance of the capabilities of the ECHO algorithm to utilize the additional spatial data provided in the panchromatic data channel.

ACCURACY ASSESSMENT: EXPERIMENT TWO

Using the 5-meter data set, the overall accuracy of the six band classification increased by 11.4% (from 73.5% to 85.1%) when the ECHO algorithm was substituted for the Maximum Likelihood technique (Figure 9). This difference in the overall accuracy decreased to 10.6% when the aerial photograph was included in both classifications, this decrease in accuracy represents approximately 8.1% of the initial difference in accuracy between the two classifications (figure 9). In other words, the addition of the aerial photography improved accuracy in both classifications, however that improvement was slightly larger between the classifications made using the Maximum Likelihood algorithm. The difference between the values of the KHAT statistic for each technique decreased by 8.0% of the initial value, and the difference in the variance of KHAT decreased by 0.6% of the initial value (figure 9), when the aerial photograph was included in the classification.

The overall accuracy of classifications made with the 1-meter data set demonstrated differences of less than one half of one percent. Classifications made using the 15-meter data set resulted in differences between the improvements in the overall accuracy achieved by the two algorithms that were almost identical to those described for the 5-meter data (figures 6, 9, and 12).

The effectiveness with which each algorithm can overcome the errors introduced during data preparation and utilize the additional spatial information to improve classification accuracy is compared by interpreting the difference between the accuracy of classifications made with each algorithm when the high-resolution aerial photography is included in the classification. The various estimations of accuracy offered here represent independent interpretations of classification accuracy, and therefore it is expected that the range of possible, but hopefully consistent, values vary among them. Of all of the estimates of classification accuracy at each resolution, however, none represents a difference of more than 1.2% between the accuracy achieved by the two algorithms with and without the high-resolution data. In other words, the ECHO algorithm is no more or less capable, in this case, of using the aerial photography to improve classification accuracy than the traditional Maximum Likelihood technique.

In comparing the differences between the ECHO 6 and ECHO 7 classifications specifically, the ability of the ECHO algorithm to utilize the additional information provided by the panchromatic data is explicitly demonstrated. In this experiment, using the 5-meter data set, the overall accuracy of the ECHO 7 classification was 2.2% higher than the accuracy achieved by the ECHO 6 classification. Similarly, the value of the KHAT statistic increased by 2.9% of the ECHO 6 value and the variance of the KHAT statistic for ECHO 7 was 11.1% less than the ECHO 6 value (figure 9).

The accuracy of classifications made with the 15-meter data set demonstrated differences that were almost identical to those described for the 5-meter data. Classifications made using the 1-meter data set resulted in differences between the improvements in the overall accuracy achieved by the two algorithms that amounted to less than one half of one percent (figures 6, 9, and 12).

The Z statistic (7) was used to estimate the significance of the difference between the KHAT statistics for the ECHO 6 and ECHO 7 classifications specifically. Values of 0.6, 0.7 and 0.1 for the 15m, 5m, and 1m classifications, respectively, all represent an improvement in classification accuracy when the high-resolution data are included, however this improvement is not statistically significant. Because the only difference between the two classifications was the addition of the high-resolution aerial photograph, it is apparent that the ECHO algorithm was not capable of utilizing the additional spatial information provided by the aerial photograph to generate a significantly more accurate classification. One possible explanation for this is that the additional information provided by the aerial photography was largely offset by the error associated with image resampling and registration.

Relationships between the accuracy achieved for each land cover class at each resolution were indistinct. There were differences between each class at each resolution with both the user's and producer's accuracies, and trends within these values were weak at best. There were only two clearly discernable trends in the results. First, variations of individual class accuracy between the ECHO 6 and ECHO 7 classifications fluctuated within both the user's and producer's accuracies from an increase of 3.7% without the aerial photograph to an increase of 10.7% with the high-resolution data. For each resolution, the largest individual increase in accuracy was always associated with the addition of the aerial photograph and not with its absence. This indicates that, although there were negative impacts due to errors introduced during data preparation, improvements gained through the inclusion of the high-resolution data source had a more significant impact on classification accuracy.

The second trend demonstrates this more clearly. An average of the differences in the accuracy estimated for each individual class indicated an overall increase when the aerial photograph was included during classification. This overall increase was greatest, over 8% for both user's and producer's accuracy, in the classification made using the 5-meter data set. The 15-meter classifications demonstrated an increase of over 5%, and the 1-meter classifications demonstrated an increase of over 2%, in both measures when the aerial photograph was included in the data set.

At each resolution, and in both the user's and producer's accuracies, the differences between the accuracy of the two residential classes were either very small or the ECHO 7 value was greater than the ECHO 6 value (Figures 4,5,7,8,10,11). These trends in the individual class accuracies, along with the increase in overall accuracy when the high-resolution aerial photography is included, indicate that there is potential for the ECHO algorithm to utilize the additional spatial information available in photography to improve classification accuracy.

CONCLUSIONS

This study has demonstrated that augmenting multispectral satellite data with panchromatic satellite data at a higher resolution can increase the precision and accuracy with which multispectral data can be classified by the ECHO algorithm. Although the use of aerial photography to enhance multispectral data resulted in a small increase in accuracy at each of the three resolutions tested, these increases were significant at confidence levels of only 40 to 60%. Using spatially enhanced data with the ECHO algorithm has great potential for enhancing classification accuracy, but this potential is limited by the accuracy with which the data can be geometrically registered and by the ability to handle or avoid temporal differences in the data.

It was shown in the first experiment that the ECHO algorithm was capable of utilizing the additional information available in the high-resolution panchromatic Landsat 7 data channel to generate a more accurate classification. The lack of significant improvement when additional spatial information was provided in classifications made using the Maximum Likelihood algorithm confirms that it is the design of the ECHO algorithm that allows for the use of the additional spatial information.

It is apparent in the analysis of the classifications performed in this research that improvements in the accuracy of thematic maps generated with the ECHO algorithm were a direct result of its ability to distinguish between the spectrally similar high and medium-density residential land cover classes. Classification accuracy was measured by estimates of the overall accuracy, the producer's and user's accuracies of the two residential classes, and with the KHAT statistic. Each of these estimates demonstrated an improvement in classification accuracy, and the KHAT statistic demonstrated that the inclusion of the high-resolution data allowed for a statistically significant improvement in the overall accuracy of the classification.

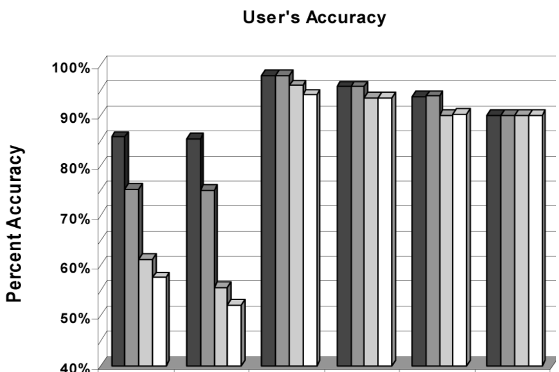
Augmenting the multispectral satellite data with high-resolution aerial photography did not significantly increase the accuracy achieved through classifications made using the ECHO algorithm as measured by the Z statistic. As the only alteration of the data set used by both algorithms was the inclusion of the high-resolution aerial photography, these results suggest that the ECHO algorithm was not capable of utilizing the additional spatial information to generate a significantly more accurate classification. The results of the second experiment do show, however, that the accuracy of classifications made using the spatially enhanced data set were higher than those made using only the multispectral data at each of the three resolutions tested. Although these increases in accuracy were relatively small, they do indicate potential for this method. One possible explanation for this result, which is supported by the results of the first experiment, is that the additional information provided by the aerial photography was largely offset by the error associated with image resampling and registration. Unfortunately, temporal gaps and geometrically imperfect data sets become more common as the age of data increases.

The concept of combining aerial photography with satellite imagery is not well examined in the literature, and more research is needed to explore techniques that can be used to overcome the limitations of the method. Research focused on determining the nature of registration error in data sets with large differences in both temporal and geometric resolution, its potential for degrading classification accuracy, and also in defining scenarios in which the method is appropriate, will be necessary to more fully develop the potential of this approach.

References

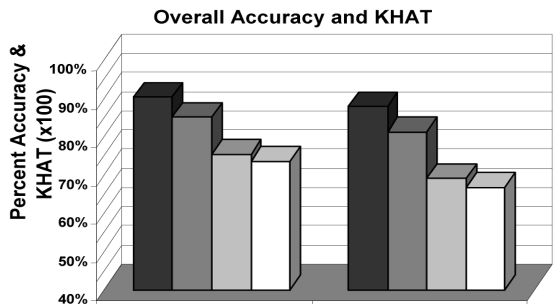
- Campbell, J. (1996). *Introduction to Remote Sensing* (Second Edition). New York: Guilford Press.
- Card, D. H., (1982). "Using Known Map Category Marginal Frequencies to Improve Estimates of Thematic Map Accuracy." *Photogrammetric Engineering and Remote Sensing,* 48(3): 431-439.
- Cohen, J. (1960). "A Coefficient of Agreement for Nominal Scales." *Educational and Psychological Measurement,* 20(1): 37-40.
- Congalton, R. G. (1988). "A Comparison of Sampling Schemes Used in Generating Error Matrices for Accessing the Accuracy of Maps Generated from Remotely Sensed Data." *Photogrammetric Engineering & Remote Sensing,* 54(5): 593-600.
- Congalton, R. G. (1991). "A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data." *Remote Sensing of Environment,* 37(1): 35-46.
- Congalton, R. G. and Green, K., (1999). *Assessing the Accuracy of Remotely Sensed Data: Principals and Practices*. New York: Lewis Publishers.
- Congalton, R. G., Oderwald, R. G., and Mead, R. A., (1983). "Assessing Landsat Classification Accuracy Using Discrete Multivariate Analysis Statistical Techniques." *Photogrammetric Engineering & Remote Sensing,* 49(12): 1671-1678.
- Dikshit, O. and Roy, D. P. (1996). "An Empirical Investigation of Image Resampling Effects Upon the Spectral and Textural Supervised Classification of a High Spatial Resolution Multispectral Image." *Photogrammetric Engineering & Remote Sensing,* 63(9): 1085-1092.
- Franklin, S. E. et al. (1991). "Pixel Sampling of Remotely Sensed Digital Imagery." *Computers & Geosciences,* 17(6): 759-775.
- Hammond, T. O. and Verbyla, D. L., (1996). "Optimistic Bias in Classification Accuracy Assessment." *International Journal of Remote Sensing,* 17(6): 1261-1266.
- Hay, A. M., (1979). "Sampling Designs to Test Land-Use Map Accuracy." *Photogrammetric Engineering & Remote Sensing,* 45(4): 529-533.
- Hord, M. R. and Brooner, W., (1976). "Land-Use Map Accuracy Criteria." *Photogrammetric Engineering & Remote Sensing,* 42(5): 671-677.
- Hudson, W. and Ramm, C., (1987). "Correct Formulation of the Kappa Coefficient of Agreement." *Photogrammetric Engineering and Remote Sensing* 53(4): 421-422.
- Janssen, L. L. F. and van der Wel, F. J. M., (1994). "Accuracy Assessment of Satellite Derived Land-Cover Data: A Review." *Photogrammetric Engineering & Remote Sensing,* 60(4): 419-426.
- Kettig, R. L. and Landgrebe, D. A., (1976). "Classification of Multispectral Image Data by Extraction and Classification of Homogeneous Objects." *IEEE Transactions on Geoscience Electronics*, 14(1): 19-26.
- Landgrebe, D. A., (1980). "The Development of a Spectral-Spatial Classifier for Earth Observational Data." *Pattern Recognition* 12(3): 165-175.
- Lillesand, T. M. and Kiefer, R. W. 2000. *Remote Sensing and Image Interpretation* (Fourth Edition). New York: Wiley.
- Michelson, D. B., Liljeburg, B. M., and Pilesjo, P., (2000). "Comparison of Algorithms for Classifying Swedish Landcover Using Landsat TM and ERS-1 SAR Data." *Remote Sensing of Environment,* 71(1): 1-15.
- Munechika, C. K. et al. (1993). "Resolution Enhancement of Multispectral Image Data to Improve Classification Accuracy." *Photogrammetric Engineering and Remote Sensing,* 59(1): 67-72.
- Pohl, C. and Van Gendern, J. L., (1998). "Multisensor Image Fusion in Remote Sensing: Concepts, Methods, and Applications." *International Journal of Remote Sensing,* 19(5): 823-854.
- Richards, J. A. (1999). Remote Sensing Digital Image Analysis: An Introduction (Third Edition). New York: Springer.
- Rosenfield, G. H. and Fitzpatrick-Lins, K., 1986. "A Coefficient of Agreement as a Measure of Thematic Classification Accuracy." *Photogrammetric Engineering and Remote Sensing* 52(2): 223-227.
- Stehman, S. V., 1999. "Basic Probability Sampling Designs for Thematic Map Accuracy." *International Journal of Remote Sensing* 20(12): 2423-2441.
- Stehman, S. V. and Czaplewski, R. L., 1998. "Design and Analysis for Thematic Map Design: Fundamental Principals." *Remote Sensing of Environment* 64(3): 331-344.
- Story, M. and Congalton, R. G., 1986. "Accuracy Assessment: A User's Perspective." *Photogrammetric Engineering & Remote Sensing* 52(3): 397-399.
- Tso, B. C. K. and Mather, P. M., 1999. "Classification of Multisource Remote Sensing Imagery Using a Generic Algorithm and Markov Random Fields." *IEEE Transactions on Geoscience and Remote Sensing* 37(3): 1255-1260.

Appendix A



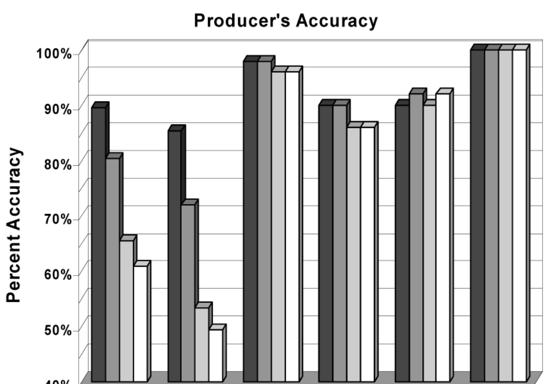
	H-res	M-res	Comm	Field	Forest	Water
ECHO 7	85.7%	85.3%	98.0%	95.7%	93.8%	90.0%
ECHO 6	75.3%	75.0%	98.0%	95.7%	93.9%	90.0%
MAX 7	61.3%	55.6%	96.0%	93.5%	90.0%	90.0%
MAX 6	57.6%	52.1%	94.1%	93.5%	90.2%	90.0%

Figure 1
User's accuracy by individual class for the ECHO 6, ECHO 7, MAX 6 and MAX 7 classifications



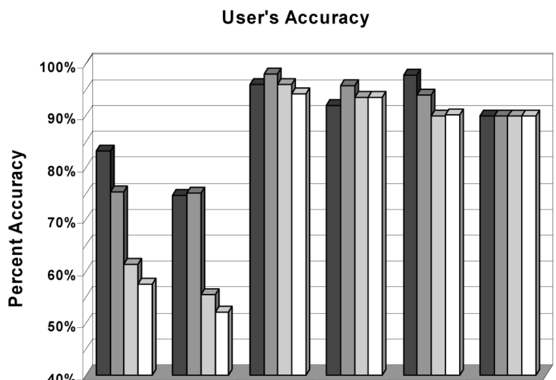
	Overall Accuracy	KHAT (x100)
ECHO 7	90.3%	87.9% Variance: 4.4 E^{-4}
ECHO 6	85.0%	81.2% Variance: 6.5 E^{-4}
MAX 7	75.4%	69.1% Variance: 9.6 E^{-4}
MAX 6	73.5%	66.8% Variance: 1.0 E^{-3}

Figure 3
Overall accuracy and the KHAT statistic for the ECHO 6, ECHO 7, MAX 6 and MAX 7 classifications.



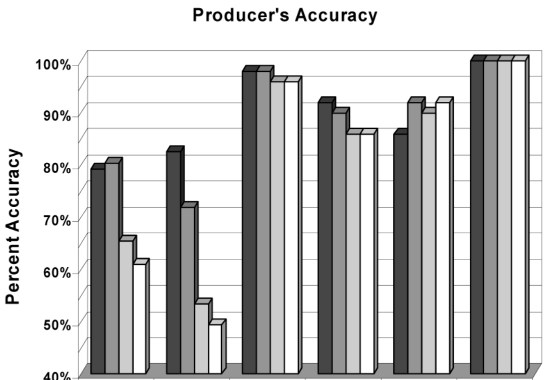
	H-res	M-res	Comm	Field	Forest	Water
ECHO 7	89.7%	85.3%	98.0%	90.0%	90.0%	100.0%
ECHO 6	80.5%	72.0%	98.0%	90.0%	92.0%	100.0%
MAX 7	65.5%	53.3%	96.0%	86.0%	90.0%	100.0%
MAX 6	60.9%	49.3%	96.0%	86.0%	92.0%	100.0%

Figure 2
Producer's accuracy by individual class for the ECHO 6, ECHO 7, MAX 6 and MAX 7 classifications.



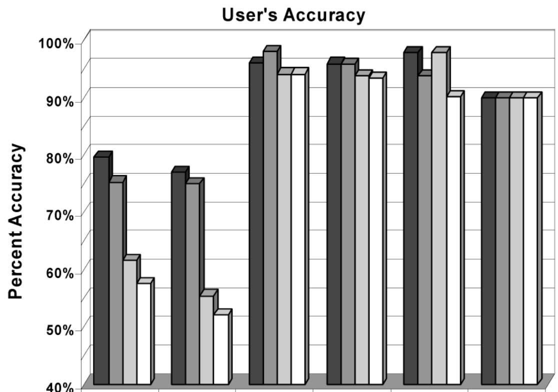
	H	M	C	G	F	W
ECHO 7	83.1%	74.7%	96.1%	92.0%	97.7%	90.0%
ECHO 6	75.3%	75.0%	98.0%	95.7%	93.9%	90.0%
MAX 7	61.3%	55.6%	96.0%	93.5%	90.0%	90.0%
MAX 6	57.6%	52.1%	94.1%	93.5%	90.2%	90.0%

Figure 4
User's accuracy by individual class for the ECHO 6, ECHO 7, MAX 6 and MAX 7 classifications of the 15 meter data set.



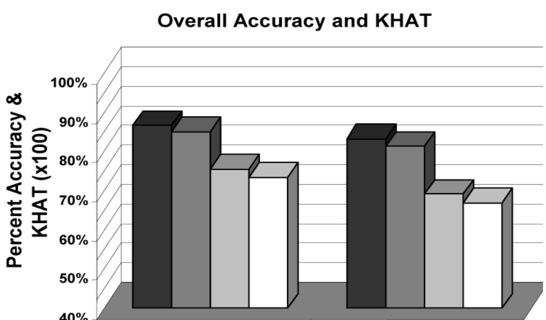
	H	M	C	G	F	W
ECHO 7	79.3%	82.7%	98.0%	92.0%	86.0%	100.0%
ECHO 6	80.5%	72.0%	98.0%	90.0%	92.0%	100.0%
MAX 7	65.5%	53.3%	96.0%	86.0%	90.0%	100.0%
MAX 6	60.9%	49.3%	96.0%	86.0%	92.0%	100.0%

Figure 5
Producer's accuracy by individual class for the ECHO 6, ECHO 7, MAX 6 and MAX 7 classifications of the 15 meter data set.



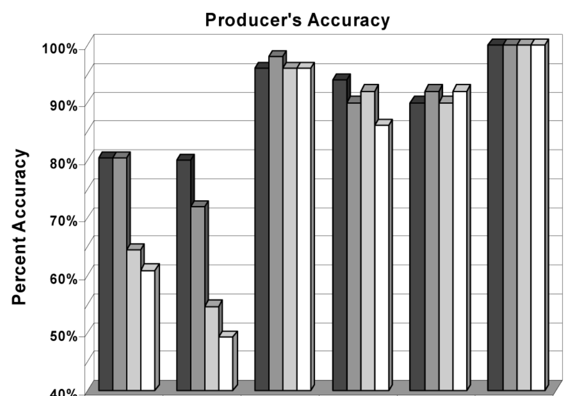
	H-res	M-res	Comm	Field	Forest	Water
ECHO 7	79.5%	76.9%	96.0%	95.9%	97.8%	90.0%
ECHO 6	75.3%	75.0%	98.0%	95.7%	93.9%	90.0%
MAX 7	61.5%	55.4%	94.1%	93.9%	97.8%	90.0%
MAX 6	57.6%	52.1%	94.1%	93.5%	90.2%	90.0%

Figure 7
User's accuracy by individual class for the ECHO 6, ECHO 7, MAX 6 and MAX 7 classifications of the 5 meter data set.



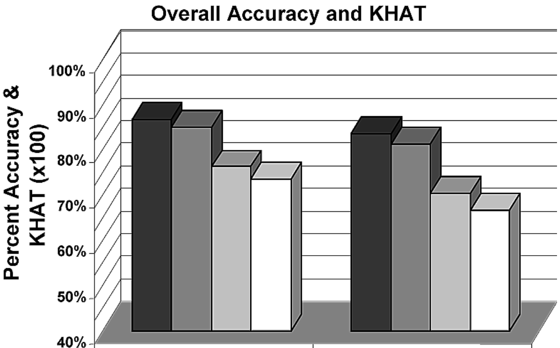
	Overall Accuracy	KHAT (x100)
ECHO 7	86.6%	83.2% Variance: 5.8 E^{-4}
ECHO 6	85.0%	81.2% Variance: 6.5 E^{-4}
MAX 7	75.4%	69.1% Variance: 9.6 E^{-4}
MAX 6	73.5%	66.8% Variance: 1.0 E^{-3}

Figure 6
Overall accuracy and KHAT for the ECHO 6, ECHO 7, MAX 6 and MAX 7 classifications of the 15 meter data set.



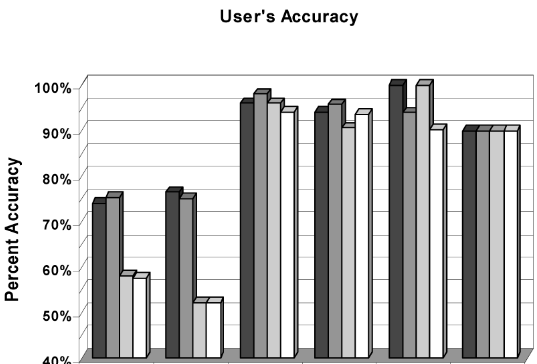
	H-res	M-res	Comm	Field	Forest	Water
ECHO 7	80.5%	80.0%	96.0%	94.0%	90.0%	100.0%
ECHO 6	80.5%	72.0%	98.0%	90.0%	92.0%	100.0%
MAX 7	64.4%	54.7%	96.0%	92.0%	90.0%	100.0%
MAX 6	60.9%	49.3%	96.0%	86.0%	92.0%	100.0%

Figure 8
Producer's accuracy by individual class for the ECHO 6, ECHO 7, MAX 6 and MAX 7 classifications of the 5 meter data set.



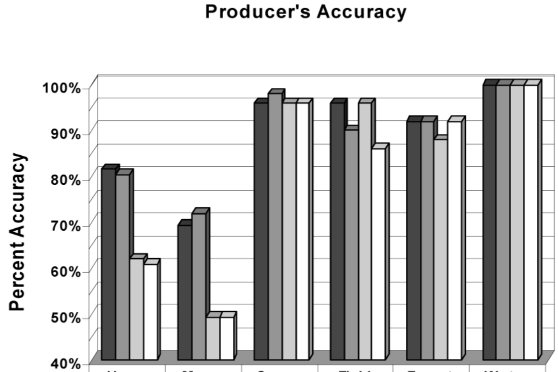
	Overall Accuracy	KHAT (x100)
ECHO 7	86.9%	83.6% Variance: 5.7 E^{-4}
ECHO 6	85.0%	81.2% Variance: 6.5 E^{-4}
MAX 7	76.3%	70.3% Variance: 9.4 E^{-4}
MAX 6	73.5%	66.8% Variance: 1.0 E^{-3}

Figure 9
Overall accuracy and KHAT for the ECHO 6, ECHO 7, MAX 6 and MAX 7 classifications of the 5 meter data set.



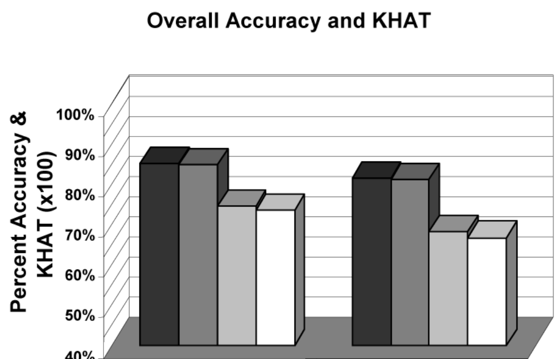
	H-res	M-res	Comm	Field	Forest	Water
ECHO 7	74.0%	76.5%	96.0%	94.1%	100.0%	90.0%
ECHO 6	75.3%	75.0%	98.0%	95.7%	93.9%	90.0%
MAX 7	58.1%	52.1%	96.0%	90.6%	100.0%	90.0%
MAX 6	57.6%	52.1%	94.1%	93.5%	90.2%	90.0%

Figure 10
User's accuracy by individual class for the ECHO 6, ECHO 7, MAX 6 and MAX 7 classifications of the 1 meter data set.



	H-res	M-res	Comm	Field	Forest	Water
ECHO 7	81.6%	69.3%	96.0%	96.0%	92.0%	100.0%
ECHO 6	80.5%	72.0%	98.0%	90.0%	92.0%	100.0%
MAX 7	62.1%	49.3%	96.0%	96.0%	88.0%	100.0%
MAX 6	60.9%	49.3%	96.0%	86.0%	92.0%	100.0%

Figure 11
Producer's accuracy by individual class for the ECHO 6, ECHO 7, MAX 6 and MAX 7 classifications of the 1 meter data set.



	Overall Accuracy	KHAT (x100)
ECHO 7	85.4%	81.6% Variance: 6.4 E^{-4}
ECHO 6	85.0%	81.2% Variance: 6.5 E^{-4}
MAX 7	74.8%	68.4% Variance: 9.8 E^{-4}
MAX 6	73.5%	66.8% Variance: 1.0 E^{-3}

Figure 12
Overall accuracy and KHAT for the ECHO 6, ECHO 7, MAX 6 and MAX 7 classifications of the 1 meter data set.

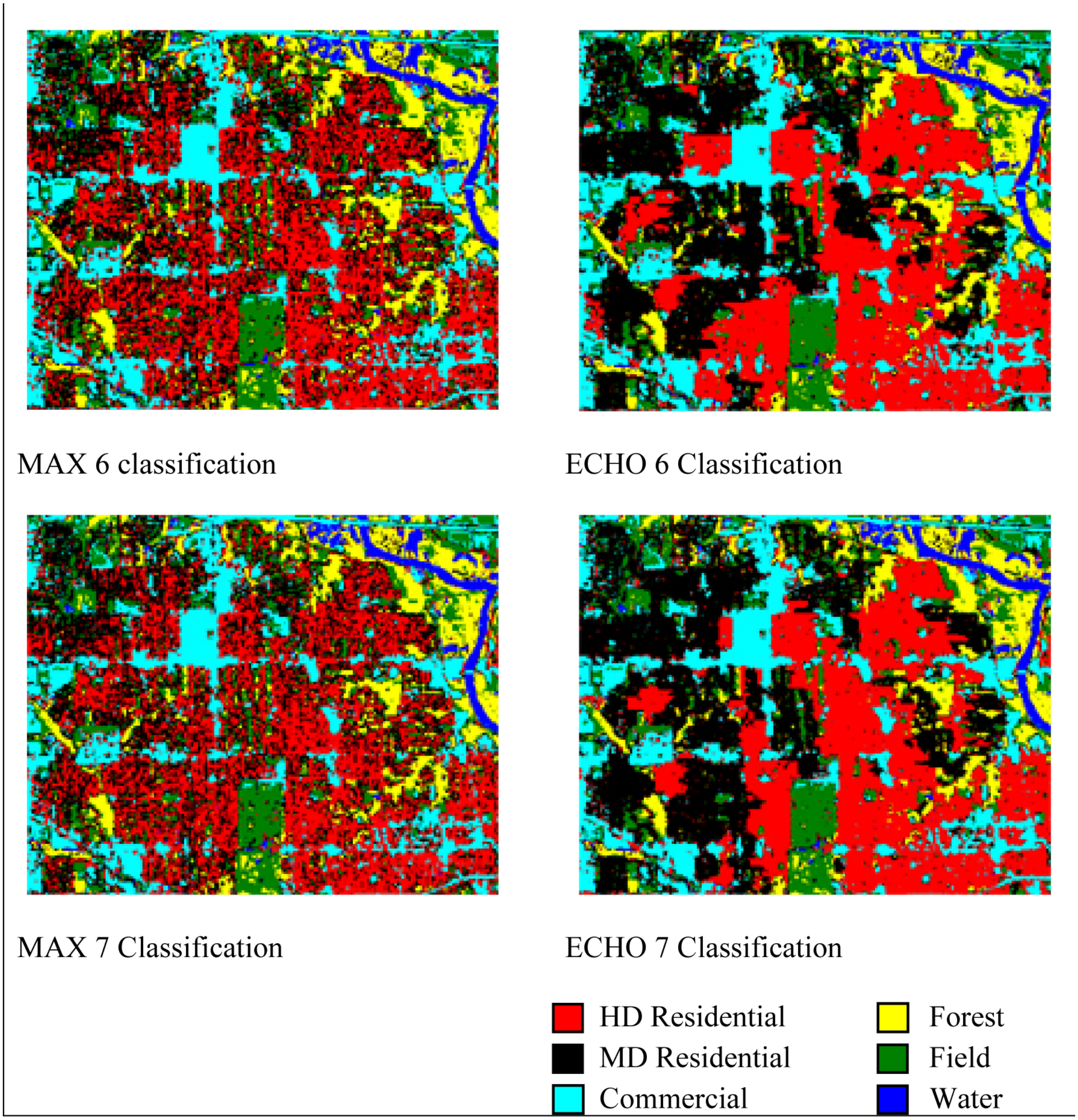


Figure 13
Thematic maps derived from the ECHO 6, ECHO 7, MAX 6 and MAX 7 classifications.

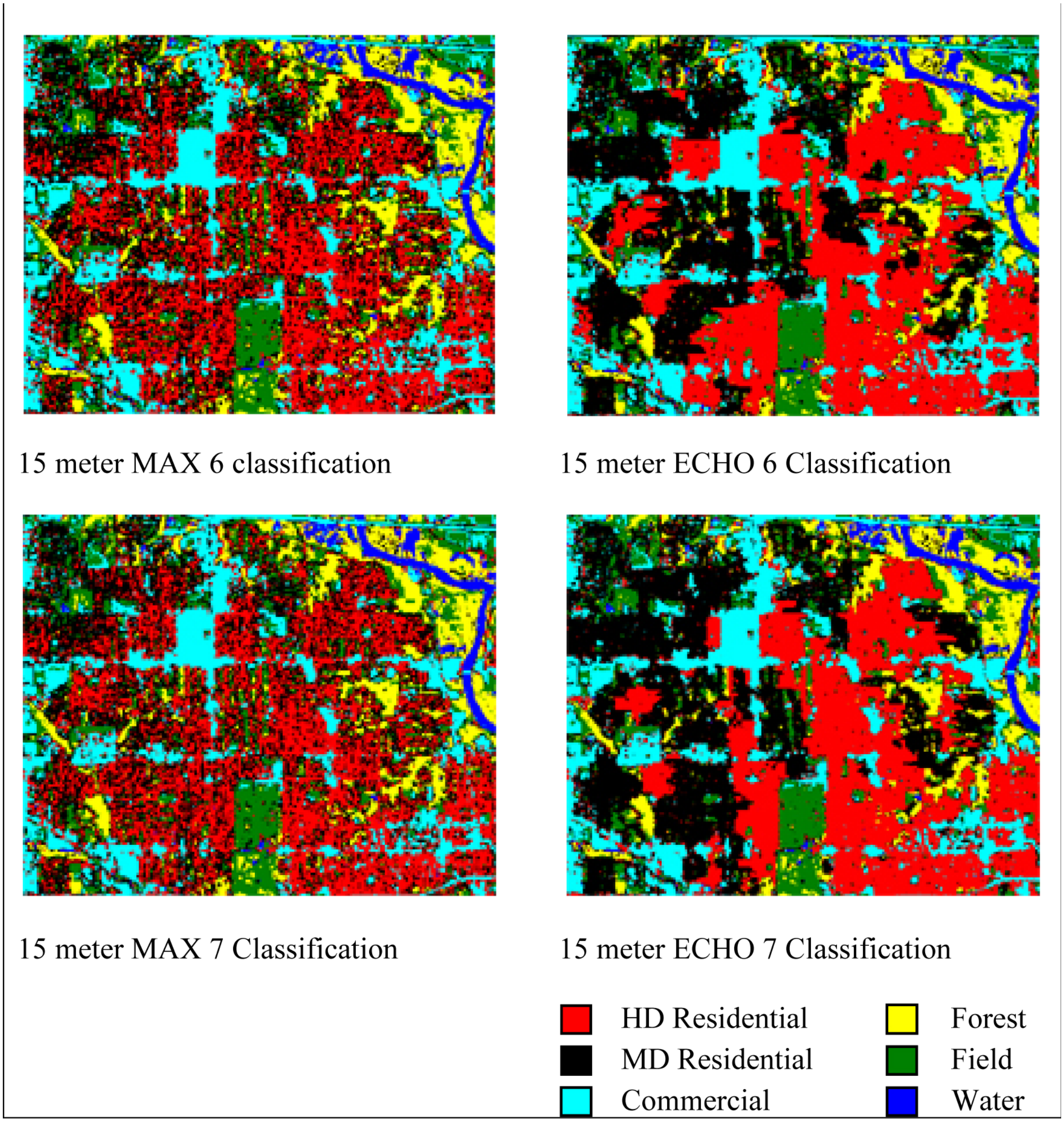


Figure 14
Thematic maps derived from the ECHO 6, ECHO 7, MAX 6 and MAX 7 classifications of the 15 meter data set.

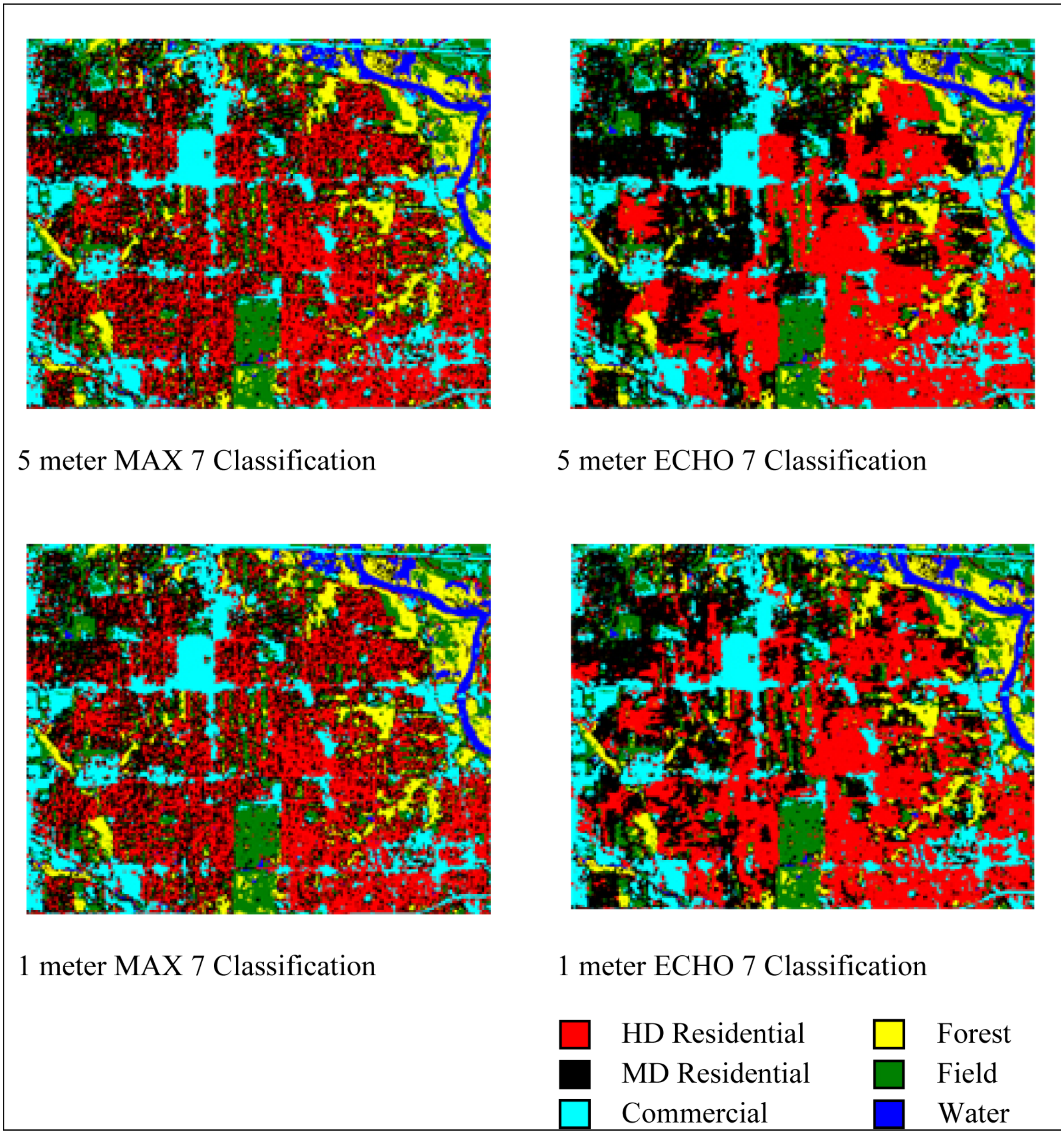


Figure 15
Thematic maps derived from the ECHO 6, ECHO 7, MAX 6 and MAX 7 classifications of the 5 and 1 meter data sets.