

Use of Remote Sensing to Screen Earthen Levees

James V. Aanstoos, Khaled Hasan, Charles G. O'Hara, Saurabh Prasad, Lalitha Dabir, Majid Mahrooghy, Rodrigo Nobrega, Matthew Lee, Bijay Shrestha

Geosystems Research Institute
Mississippi State University
Mississippi State, MS 39762 USA
j.aanstoos@ieee.org

Abstract—Multi-polarized L-band Synthetic Aperture Radar is investigated for its potential to screen earthen levees for weak points. Various feature detection and classification algorithms are tested for this application, including both radiometric and textural methods such as grey-level co-occurrence matrix and wavelet features.

Keywords—synthetic aperture radar, levee screening, earthen levees

I. INTRODUCTION

Earthen levees protect large areas of populated and cultivated land in the US from flooding. The potential loss of life and property associated with the catastrophic failure of levees can be extremely large. Over the entire US, there are more than 150,000 kilometers of levee structures of varying designs and conditions.

Currently, there are limited processes in place to prioritize the monitoring of large numbers of dam and levee structures. Levee managers and federal agencies need to assess levee health rapidly with robust techniques that identify, classify and prioritize levee vulnerabilities with lower costs than traditional soil-boring programs, which can cost many millions of dollars and provide information about the subsurface only in the immediate vicinity of a small-diameter borehole.

This paper gives an overview and presents early results of an ongoing project studying the use of synthetic aperture radar (SAR) in conjunction with other remote imagery as an aid to the levee screening process. Sources of data include the NASA UAVSAR (Uninhabited Aerial Vehicle SAR), a fully polarimetric L-band SAR which is specifically designed to acquire airborne repeat track SAR data for differential interferometric measurements. The instrument is capable of sub-meter ground sample distance.

The L-band SAR measurements can penetrate dry soil to as much as one meter depth. Thus they may be valuable in detecting changes in levees that will be key inputs to a levee vulnerability classification system.

Recently launched space-based SAR platforms have spatial resolutions approaching that of the airborne SAR and also have multi-polar options but use different carrier frequencies. This study is also utilizing data from the German TerraSAR-X satellite, which uses X-band.

Our test study area is a stretch of 230 km of levees along the lower Mississippi River along the western boundary of the

state of Mississippi. Two types of problems that occur along these levees which can be precursors to complete failure during a high water event are slough slides and sand boils.

Sand boils are springs that form on the land side of a levee containing a river at extremely high flood stage. Hydrostatic pressure generated by the column of river water exerts a downward force that is too great for the wall material of the river channel to contain, and thus water is forced through the wall material of the channel. [1]

Slough slides are slope failures along a levee, which leave areas of the levee vulnerable to seepage and failure during high water events. An illustration of slides and sand boils is shown in Fig. 1 [2]. The roughness and related textural characteristics of the soil in a slide affect the amount and pattern of radar backscatter. The type of vegetation that grows in a slide area differs from the surrounding levee vegetation, which can also be utilized in detecting slides [3].

Early detection of the occurrence of these events can assist levee managers in prioritizing their inspection and repair efforts. A remote sensing based solution for their rapid detection would be more efficient and cost effective than frequent on site visits. Furthermore, it may be possible to detect less obvious precursors to the slides and boils themselves by sensing characteristics of the surface soils and vegetation. A working hypothesis of this study is that such characteristics will be manifested in the backscatter of polarimetric radar due to its response to spatially variant soil moisture. In particular, L band radar is known to penetrate dry soils up to one meter in depth, and has been used to map surface soil moisture. [4]

II. DATA

A. UAVSAR Data

The primary remotely sensed data used in this study is from NASA Jet Propulsion Laboratory's UAVSAR (Uninhabited Aerial Vehicle Synthetic Aperture Radar) instrument, a polarimetric L-band synthetic aperture radar flown on a Gulfstream-3 research aircraft. The salient characteristics of this instrument are shown in Table I.

The UAVSAR is normally flown at an altitude of 12.5 km and takes an image swath 20 km wide. Our study area was designed to be collected in two straight-line flight segments structured to capture most of the river levees on both sides of the river. The study area is shown in Fig. 2, with a color composite representation of the radar data overlaid on a base

map of the vicinity. Two data collections were made: the first on June 16, 2009 and the second on January 25, 2010. The flights were flown in a “racetrack” pattern looking toward the river from opposite directions, in order to achieve a range of local incidence angles along the levees.

TABLE I. UAVSAR SPECIFICATIONS [5]

Parameter	Value
Frequency	L-band
Bandwidth	80 MHz
Range Resolution	1.8 m
Polarization	Full quad polarization
Quantization	12 bits
Antenna size	0.5 m range/1.5 azimuth
Power	> 2.0 kW

Although the raw ground sample distance is 1.6 by 0.6 meters, most of our efforts use the multi-look 5 by 7 meter data to minimize speckle effects.

B. Satellite Data

In addition to UAVSAR, data from the German TerraSAR-X satellite was acquired over portions of the same study area. This data also has spatial resolution on the order of one meter, but is only available at X-band. As such it will not penetrate soil of vegetation nearly as much as the L-band. However the ease repeat intervals with which it can be acquired, versus airborne instruments, make it worth testing for this application. Our data collection for this source has only recently been completed and the results for it are not reported here.

C. Ancillary Data

Ancillary data can be used to assist the levee classification process in addition to the remote sensing data. This includes knowledge of the soil characteristics in the vicinity of the levees, the underlying geology of the area, and history of past problems and inspections of the levees.

III. CLASSIFICATION METHODOLOGY

As the radar data was collected, a number of tools were used to assess the statistical distribution of the polarimetric backscatter from the levee features. Features believed to provide good potential for discrimination of the targets of interest were designed, computed, and tested. Both per-pixel and window-based (textural) features were examined. The candidate features were tested with our training data to determine separability between classes of interest. Plots of the Bhattacharyya distance between target classes (target refers to an event of interest on the levee, e.g., landslides, compromised sections of the levee etc.) and the other (background) classes found in the study area were examined to assist in feature selection.

Candidate classification algorithms were run with sets of these features to further narrow the choices. Owing to the relatively small number of ground-truth pixels available, a leave-one-out cross validation technique was employed to estimate overall classification and target detection accuracies. In this approach, all samples but one (and hence features

derived from them) are employed to “train” the classifier, and the sequestered sample is employed for “testing”. This is repeated in a round-robin manner until all samples have been tested and classified. Based on the ground-truth for these test samples, a confusion matrix is estimated and overall accuracy, target detection accuracy and false alarm rate is calculated.

Stepwise Linear Discriminant Analysis (S-LDA) is employed for feature reduction and optimization [6]. In this approach, various features derived from the SAR backscatter imagery are concatenated into a vector, and a forward-selection, backward rejection technique is employed to prune away features that are “less” relevant to the classification problem at hand. S-LDA reduces the feature set by selecting a subset of all available features based on a metric that quantifies the class separation provided by each feature. In this study Bhattacharyya Distance (BD) is used as the metric for calculating the class separation. An LDA based feature “optimization” is then employed on this reduced dimensional subset of features identified by the forward selection, backward rejection search. S-LDA is hence extremely valuable in ensuring that the “most” relevant features are provided to the classifier while ensuring that the classifier is not over-burdened by an excessively high dimensional feature space.

Finally, the “classifier” employed in this work to model class-specific information and label test data (pixels) appropriately is the popular maximum-likelihood classifier. This classifier assumes Gaussian probability distribution functions for each class, and uses training data to learn the mean vector and covariance matrix per class. This information is then employed to find the distance of test vectors from each class model, and a class label is assigned to the sample that maximizes the likelihood value of the test sample being in that class.

Features included in this study are described in detail below. These include per-pixel intensity and related features, as well as features extracted within a window around each pixel – known as textural features.

A. Per-pixel Features

The polarimetric radar data contains three independent channels of backscatter coefficients, those for like-polarized (HH, VV) and cross-polarized (HV) combinations of transmitted and received polarizations. For each, we get complex values giving both magnitude and phase information. The magnitudes of these channels can be used as basic per-pixel features with any classifier. For the airborne UAVSAR radar data we have two different views of the same levees from opposite directions. Due to the high slope of the levee surfaces, the local incidence angles will vary greatly with look direction of the radar. In order to minimize the effects of this, we computed a feature which is the maximum backscatter magnitude for each pixel from both views.

The relationship between the complex backscatter coefficients can reveal details on the nature of the scattering mechanism of the targets, such as relative amount of surface, double-bounce, or volume scattering. Both supervised and unsupervised classification techniques are used with the polarimetric decomposition parameters entropy (H), anisotropy

(A) and alpha (α) derived from the coherency matrix calculated for each pixel [7].

Finally, since models have been developed for estimating soil moisture from SAR images, we implemented these to determine if the resulting patterns of moisture estimates on the levees can be used as a classification feature. Oh *et al.* [8] developed an inversion technique to obtain surface roughness and dielectric constant (related to water content) from the co-polarization ratio p and the cross-polarization ratio q . The model works over limited ranges of roughness and moisture. The co-polarization and cross-polarization ratios are defined as:

$$p = \sigma_{hh}^0 / \sigma_{vv}^0 \quad q = \sigma_{hv}^0 / \sigma_{vv}^0$$

where σ_{hh}^0 is the HH polarized backscatter coefficient, σ_{vv}^0 the VV polarized backscatter coefficient, and σ_{hv}^0 the HV polarized backscatter coefficient. The dielectric constant is obtained by the following empirical equations: [9]

$$\left(\frac{2\theta}{\pi}\right)^{1/3\Gamma_0} \cdot \left[1 - \frac{q}{0.23\sqrt{\Gamma_0}}\right] + \sqrt{p} - 1 = 0$$

$$\Gamma_0 = \left[\frac{1 - \sqrt{\epsilon_r}}{1 + \sqrt{\epsilon_r}}\right]^2$$

where θ is the incidence angle in radians, Γ_0 is the Fresnel reflectivity of the surface at nadir and ϵ_r is the real part of dielectric constant (the imaginary part is ignored). Fig. 3 shows the resulting dielectric constant map using the Oh empirical model for the levee area. In this image red indicates high soil moisture (which is directly related to the dielectric constant) and blue is low moisture. The moisture map has been masked by a 200-meter wide buffer around the levee center lines and overlaid on an aerial photo of the vicinity. Circles mark the locations of slough slides on the levee. It can be seen that a pattern of higher moisture occurs where slides are frequent. Further analysis of this pattern will require taking into account local rainfall statistics, which we have not yet done.

B. Textural Features

In addition to the per-pixel backscatter features, a number of texture features are being explored which utilize the values from pixels in a neighborhood around each pixel being analyzed. Features based on the Gray Level Co-occurrence Matrix (GLCM) [10] and on the discrete wavelet transform (DWT) [11], show promising results in early testing.

A simple texture feature was also tested which is simply the variance of the backscatter magnitude in the neighborhood. Window sizes of 3x3 through 9x9 were tested.

The GLCM methods use estimates of the joint probability distribution of pixel pairs at given distances in given directions. Summary statistics are then calculated on the resulting NxN co-occurrence matrices, where N is the number of quantized gray levels—that is, intensity ranges for a given channel of image data (we use N = 256). The summary statistics can include such

measures as mean, variance, correlation, energy, homogeneity, anisotropy and entropy.

The ability of wavelet analysis to decompose an image into different frequency sub-bands makes it suitable for image classification [12]. In some applications, the energy of each sub-band is used as a texture feature. In others, a feature selection analysis chooses a subset of these which prove effective for a given texture-based classification. Other parameters to be determined include the choice of mother wavelet function, and the neighborhood window size.

For the sand boil and slide detection application, we used wavelet features with one decomposition level from each of the radar polarization channels. We tested these features using different sliding window sizes (5, 7, 8, 9) and mother wavelets (Daubechies, Haar, Symlet, and Biorthogonal). Fig. 4 shows the Bhattacharyya distance between sand boil wavelet features and those of other classes (woods, agriculture, building, grass, water, dry wood, and sand) using different mother wavelets and the SAR HH image data. As can be seen, the Bhattacharyya distances for the building and woods classes are very high compared to the sand and grass. Thus, these wavelet features can easily distinguish buildings and woods from the sand boils, while having difficulty discriminating sand and grass from them.

IV. TRAINING DATA

For this application, the availability of “ground truth” data for training of supervised classification processes is a challenge since the targets of interest are portions of levees that show signs of impending failure, and once these are detected they are quickly repaired depending on their severity. Our study area is one in which the levees are managed by the US Army Corps of Engineers and are well-monitored. Therefore there do not exist a significant number of existing problem areas on these levees. However, the Corps does maintain a good cumulative history of past problems and has identified particularly problematic sections of levees in the study area. These are thus used as training samples. The training data we are using falls into the following categories:

- (1) Existing sand boils
- (2) Existing slough slides
- (3) Locations of historically frequent slides or boils
- (4) *In situ* measurements of soil properties

V. PRELIMINARY RESULTS

Preliminary results are promising. At this time we have not yet received the bulk of our *in situ* soil properties data, so these results reflect the historical training data samples of boil and slide locations.

Some results of using the S-LDA feature selection method on our texture features are shown in Fig. 5. This shows the result of classifying a subset of our study area which contains two slough slides using a 2-class problem (slide or normal levee). The S-LDA algorithm selected the two best features for this example, and those were both based on the HH radar data channel: (1) the GLCM entropy measure from a 135 degree

direction, and (2) the variance of the 5x5 pixel window. Note that the process is trained with data from a masked-out portion of the image which only contains the levee itself and not any of the trees or agricultural fields nearby; however in the output image shown the mask was not applied, hence the large areas of apparent false positives. We are only concerned with the classified targets on the levee itself, and these correspond very well with the actual slide locations. There is a region to the south of the slides with is also a false positive; however, we believe it has similar soil characteristics to the slides and therefore may be an area of interest that could be a potential future slide. Investigation of this theory is ongoing using intense *in situ* soil property measurements.

Since our ground truth data is limited, we seek out areas of the levee that might have characteristics of interest but are not currently known to the levee monitoring agencies. In order to guide our search for such areas, we examined the statistics of promising features using unsupervised clustering which is focused on the levee itself by masking it off from surrounding terrain. Since normal “healthy” levee segments are fairly uniform in soil characteristics and vegetation, any clusters of anomalous values might indicate areas of interest. Fig. 6 shows the results of running an unsupervised clustering algorithm using the k-means classifier on our region of interest with a total of 9 features, some texture based and some pixel based. The magnitudes of the backscatter coefficients in each channel HH, VV, and HV were used, along with the mean and variance of each in a 5x5 pixel window. The number of clusters was set to 8. In the figure it can be seen that the medium blue cluster appears to match the dominant levee class and could correspond to our “healthy levee” areas, while the orange and red clusters are associated with known slide locations (marked by black circles, of which there are five in this image). There are other regions of orange not aligned with the known slide locations, some of which match the GLCM “false positives” of Fig. 5. As mentioned, we will be investigating these areas more closely.

As stated, these results are quite preliminary and cannot yet be rigorously assessed or validated. More data is needed, and will be forthcoming with the results of a recently completed intensive *in situ* soil analysis campaign that covered 5 km of levees including both known problem areas and healthy levee segments.

VI. FUTURE WORK

With the pending acquisition of the intensive *in situ* soil measurements, more extensive training and evaluation of the classifiers will be completed. We will also process the TerraSAR-X data and be able to ascertain the effectiveness of this X-band data versus the L-band airborne SAR.

The soil moisture models will be supplemented by vegetation index calculations from optical multispectral imagery, which should improve the resulting estimates.

ACKNOWLEDGMENT

The authors acknowledge the contributions of two partnering organizations in this research: The US Army Corps of Engineers, and Soil and Topography Information, LLC .

REFERENCES

- [1] J. G. Williams, Sand boils: A modern analogue of ancient and volcanos, Arkansas Academy of Science Proceedings, Vol. XXVII1, 1974, pp 80-81.
- [2] J. Dunbar, “USACE’s lower Mississippi valley engineering geology and geomorphology mapping program for levees,” presentation at Vicksburg, MS, April 16, 2009.
- [3] A. K. M. A. Hossain, G. Easson, K. Hasan, “Detection of Levee Slides Using Commercially Available Remotely Sensed Data”, Environmental and Engineering Geoscience, v.12; no. 3; pp. 235 – 246, 2006.
- [4] Jiancheng Shi, J. Wang, A. Hsu, P. O’Neill, E.T. Engman, “Estimation of bare soil moisture and surface roughness parameter using L-band SAR image data,” IEEE Trans. Geoscience Remote Sensing, Vol. 35, No. 5, 1254-1266, 1997
- [5] Kevin Wheeler, Scott Hensley, Yunling Lou, Tim Miller, Jim Hoffman, "An L-band SAR for repeat pass deformation measurements on a UAV platform", 2004 IEEE Radar Conference, Philadelphia, PA, April 2004.
- [6] R. O. Duda, P. E. Hart and D. G. Stork, *Pattern Recognition*, 2nd ed. ed. Hoboken, NJ: Wiley-Interscience, 2000.
- [7] Dabbiru, L., & Aanstoos, J.V. (2010). Classification of Levees using Polarimetric Synthetic Aperture Radar (SAR) Imagery. Proc. 2010 39th IEEE Applied Imagery Pattern Recognition Workshop. Washington, DC: IEEE.
- [8] Yisok Oh; , "Quantitative retrieval of soil moisture content and surface roughness from multipolarized radar observations of bare soil surfaces," Geoscience and Remote Sensing, IEEE Transactions on , vol.42, no.3, pp. 596- 601, March 2004.
- [9] Y. Oh, K. Sarabandi, and F. T. Ulaby, "An empirical model and an inversion technique for radar scattering from bare soil surfaces", IEEE Trans. Geosci. Remote Sensing, vol. 30, pp.370 -382 1992.
- [10] R.M. Haralick, K. Shanmugam and I. Dinstein. "Textural features for image classification," IEEE Trans. Syst. Man Cybern. 3 (1973), pp. 610–621.
- [11] C. S. Burrus, R. A. Gopinath, H. Guo, Introduction to Wavelets and Wavelet Transforms: A Primer, Prentice Hall, August 24, 1997.
- [12] Huang, K., Aviyente, S., "Wavelet Feature Selection for Image Classification," Image Processing, IEEE Transactions on, vol.17, no.9, pp. 1709-1720, Sept. 2008.

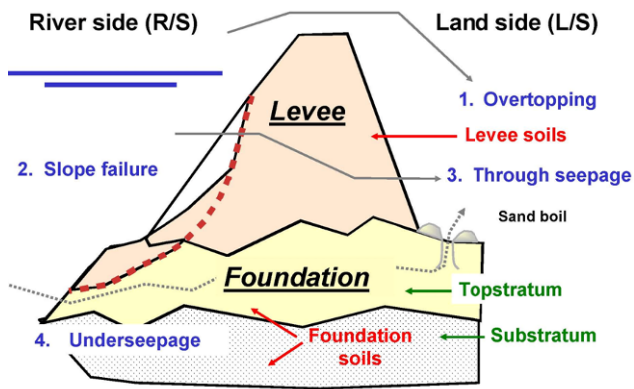


Figure 1. Illustration of levee failure mechanisms, including slough slides and sand boils

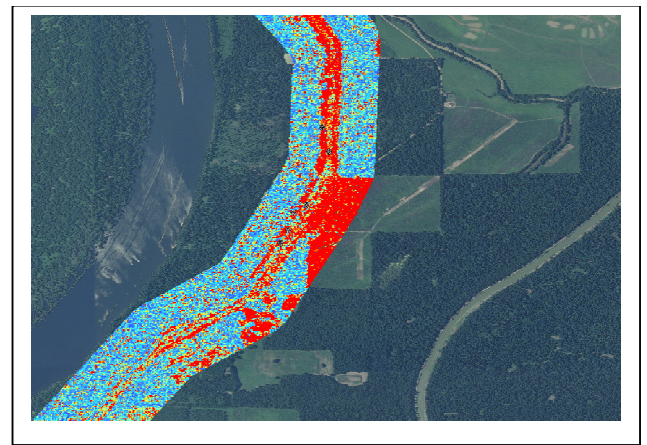


Figure 3. Dielectric constant (related to soil moisture) estimated by the Oh model from the June 2009 radar data.

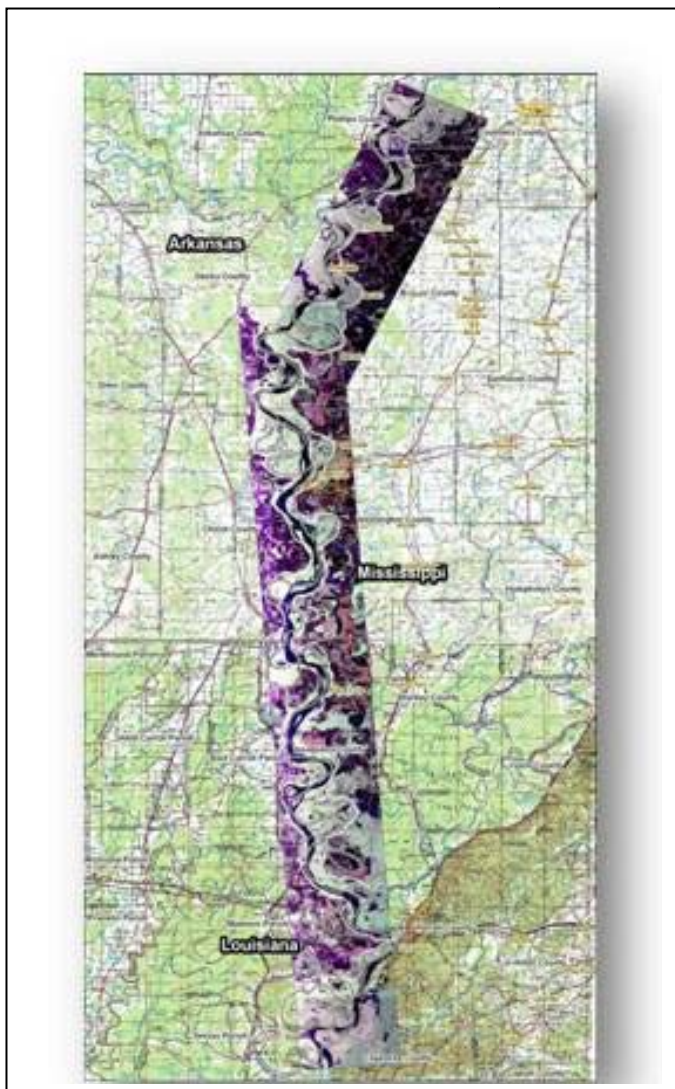


Figure 2. Study area with radar color composite image overlaid on base map. Radar image swath is 20 km wide and total length is 238 km of the lower Mississippi River valley bordering the states of Arkansas, Mississippi, and Louisiana.

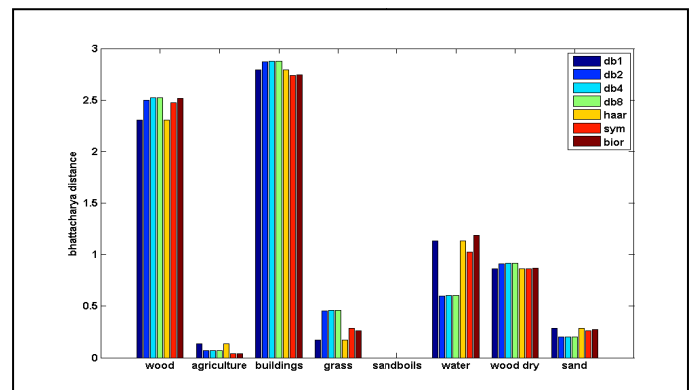


Figure 4. Bhattacharyya distance between sand boil wavelet features and other classes using different mother wavelets with the SAR HH channel.

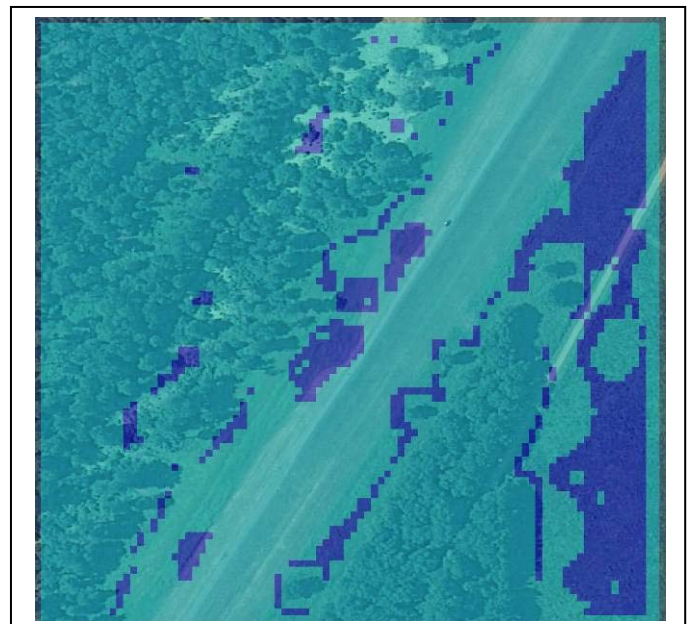


Figure 5. Levee slide classification using best 2 GLCM features selected by S-LDA algorithm.

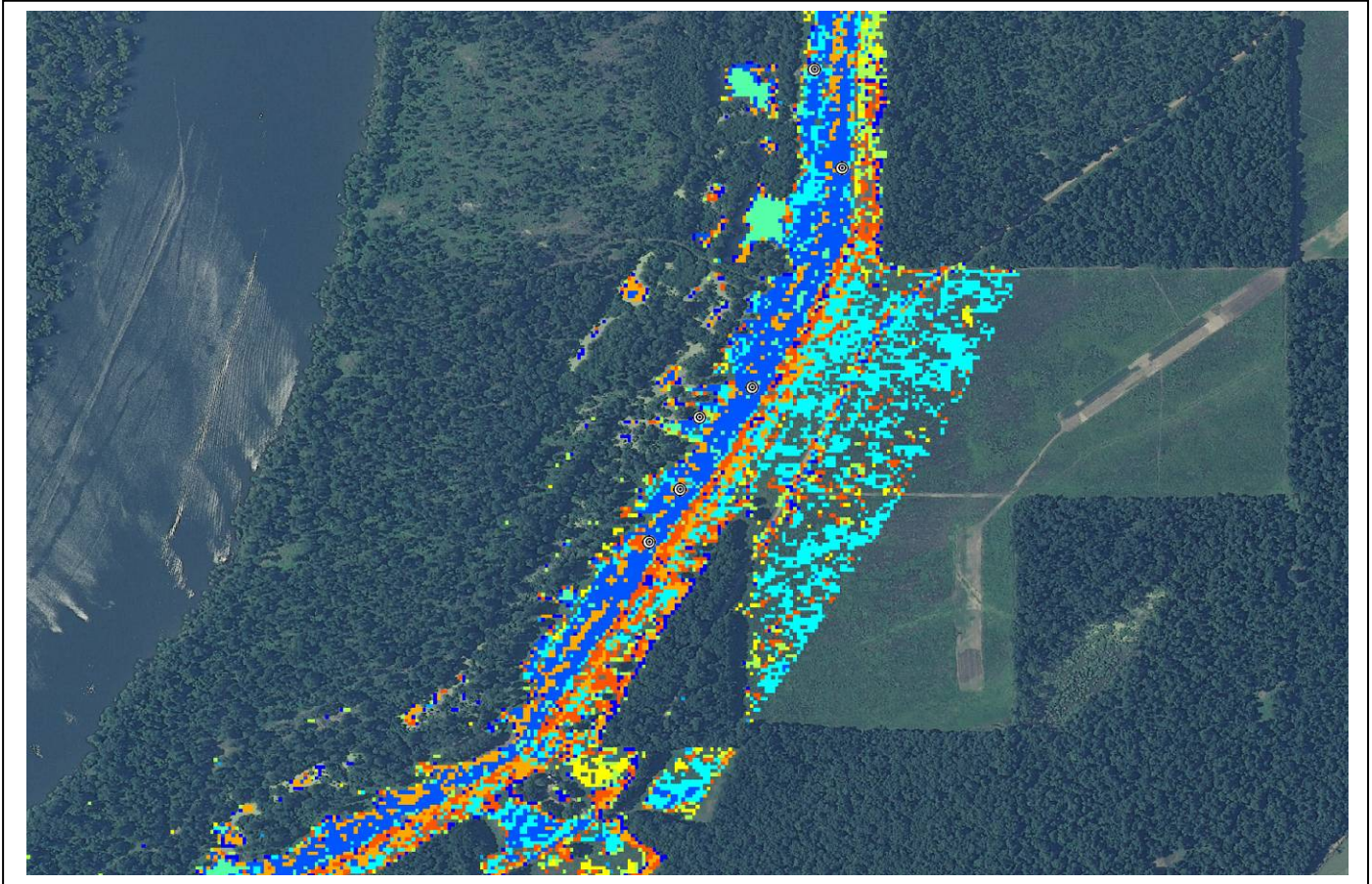


Figure 6. Cluster map using an 8-cluster k -means unsupervised clustering approach, using 9 features: HH, VV, and HV backscatter magnitudes, plus mean and variance of each in a 5x5 window. Statistics were drawn from the buffered levee area with trees masked out. Black circles (5 total) indicate locations of known slough slides.