TECHNIQUES FOR DISCRIMINATION BETWEEN AGRICULTURE AND SIMILAR LAND COVER TYPES WITH FUZZY LOGIC AND SPECTRAL POLYGON CHARACTERISTICS

David Hansen, GIS Specialist / Soil Scientist GISP
Chris Curlis, Remote Sensing Specialist
Barbara Simpson, GIS Specialist
U.S. Bureau of Reclamation
Mid-Pacific Region
2800 Cottage Way
Sacramento, CA 95825-1898
dhansen@mp.usbr.gov
ccurlis@mp.usbr.gov
bsimpson@mp.usbr.gov

ABSTRACT

This is a review of techniques for separating land cover classes for a change detection project in the Central Valley, California. The Mid Pacific Region of the U.S. Bureau of Reclamation is in the midst of the second cycle in monitoring change over major portions of the Central Valley. In agricultural areas, a considerable amount of spectrally identified change is due to cultural practices. This has led to confusion in separating change due to agricultural practice from change between categories of bare ground, grassland, forbs, and shrub land. Earlier techniques relied on the use of a broadly defined agricultural category to mask out spectral change in agricultural areas. In addition, change from what was active agriculture to another category can be missed. Image differencing of Landsat 7 TM scenes for 2000 and 2005 is being used to identify change. A mask is being developed to isolate and mask out spectral change that is due to agricultural practices. This mask will be developed from MODIS NVDI data for 2005 to identify active vegetation associated with agriculture. This represents the same time period as the Landsat TM scenes. Techniques examined include the use of fuzzy logic for setting thresholds in identifying active vegetation and the uncertainty associated with classes identified in the agriculture mask using Definiens eCognition software and ESRI ArcMap.

INTRODUCTION

The U.S. Bureau of Reclamation (Reclamation) is in the second cycle of a change detection program in the Central Valley, California. This is in support of the Central Valley Project operated in conjunction with the State Water Project of the California Department of Water Resources (DWR). These projects serve water for agricultural and municipal use throughout the valley, San Francisco Bay area, and southern California. This cycle of the change detection program is monitoring change in vegetation and land use between 2000 and 2005. The first cycle covered change between 1993 and 2000 and has been reported on in several publications (Curlis et al. 2003.) The current phase is utilizing several techniques and remote sensing data not available for the earlier cycle. Both change detection cycles rely on identifying spectral change between two dates of satellite imagery.

The valley floor is in an arid to semiarid Mediterranean climate regime with mild winters and an extended growing season. The valley is oriented predominately north to south between about 35 and 40 degree north latitude. Elevation ranges from several meters below sea level to 500 meters with elevations for most of the valley floor below 150 meters. The area contains over 3 million hectares (7 million acres) of developed agricultural land with a wide variety of crops. Field, vegetable, orchard and vineyard crops predominate with most crops under irrigation. Cultural practices and cropping patterns are dynamically driven by market conditions. Due to the diversity of crop types, the variety of planting and harvesting schedules, discrimination of agriculture from other land cover or land use classes with remote sensing is challenging.

A major issue in evaluating spectral change is isolating and masking change due to agricultural practices. In the earlier cycle, a level of uncertainty or fuzzy value was assigned to areas identified as agriculture based on the source of the agricultural assignment. This was useful in screening out change between 1993 and 2000 that was solely due

to agriculture. For the 2005 cycle, fuzzy values are being used based on thresholds of MODIS NDVI values or changes in those values between months in the classification of active vegetation associated with agriculture.

MASKING ACTIVE AGRICULTURAL AREAS FOR CHANGE DATA

Development of an effective mask to eliminate spectral change due to agricultural practices has been an issue for the change cycle of 1993 to 2000 and the 2000 to 2005 cycle. The goal for the 2005 cycle is to develop a mask that recognizes active agricultural areas for the same time period as the 2005 satellite scene used in change analysis. For change between 1993 and 2000, the agricultural mask was developed from available thematic data. The thematic data was developed independently based on analysis of 1997 high altitude imagery and other thematic data. It was not developed for the purpose of masking out active agricultural areas for spectral change analysis. For many areas on the valley floor, the mask for the 2000 change analysis was not very effective in isolating change due to agriculture.

Change between 2000 and 2005 is being developed from Landsat 7 ETM+ SLC off – SLC off Gap filled data. This product fill gaps in the Landsat 7 TM data with earlier or later Landsat passes. Imagery differencing techniques are then used to produce a change product of spectral differences between 2000 and 2005 scenes. The current method is to conduct a principle component analysis (PCA) on both dates for each scene covering the valley. The results of the analysis then undergo image differencing for each of the first three components (ie. 2005 PC 1 – 2000 PC 1). This differencing is conducted in Definiens eCognition software based on the segmented image created from a multi date composite PC image. The PC values are means within each spectral polygon. Thresholds are set individually to create three spectral change images. Each component identifies somewhat different types of spectral change. The three images are combined to identify the full range of spectral change. Much of this spectral change is due to cultural practices in agriculture on the valley floor. Figure 1 shows the overall change study area within California and the main area of interest for the agricultural mask (cross hatched area) on the Central Valley floor.

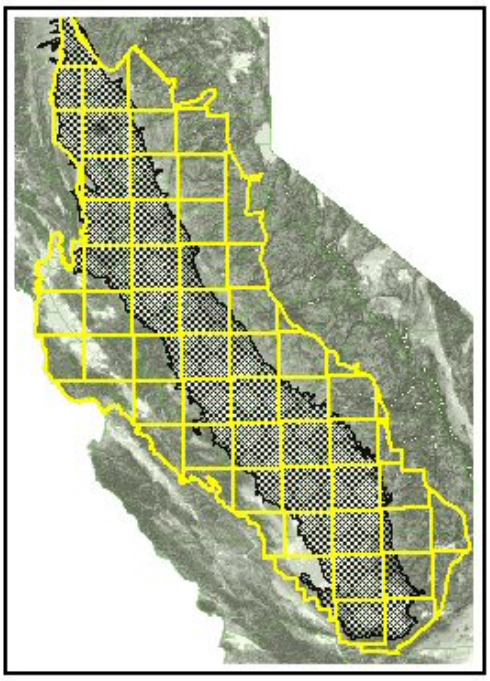

Figure 1. A portion of California with the Central Valley floor (cross hatched) and irregular rectangles defining processing areas used for the change detection project.

To efficiently and effectively manage the processing and review of data for this large area, some 58 processing areas were used in the initial cycle. These processing areas will be used in evaluating change for 2005. These processing areas are shown as irregular rectangles in Figure 1. The agricultural mask will then be used to eliminate change that is due to agriculture. Change remaining after application of the agricultural mask will then be evaluated.

Ideally, the mask will be developed from data that represents the same time period as the 2005 Landsat 7 scenes. This will assist in identifying fields active at about the same time in both the data used to develop the agricultural mask as well as the data used to develop change. In addition, the mask should not include areas identified as agriculture in 2000 that are not active in 2005. The mask will not include lands being fallowed in 2005. There are also significant areas previously identified as agricultural that are no longer in crop production. This includes conversion to other uses such as urban development and land retirement programs. An effective agricultural mask will mask out spectral change between 2000 and 2005 that is due to active cultural practice but does not mask spectral change due to change from agriculture in 2000 to another land use category. A variety of techniques are being examined with Definiens eCognition and ESRI ArcMap software for analysis of 2005 imagery with thematic data to identify areas of active vegetation in 2005. The 2005 agricultural mask will be based on active vegetation areas that were classed as agricultural in 2000. The mask will include a fuzzy ranking based on uncertainty in identifying active vegetation. This discussion will focus on the application of MODIS 250 meter resolution normalized difference vegetation index (NDVI) products for identifying active vegetation associated with agriculture for the mask.

MODIS DATA AND PROCESSING

MODIS Terra data products of vegetation indices for a 10 degree tile covering most of California were downloaded from a NASA site (NASA, 2005) for monthly intervals from March through November 2005. Data is available for two week intervals, but for initial screening only data at monthly intervals were used. Data was then projected using the projection utility provided by NASA and USGS (USGS, 2004) to a common projection used by the change program. The NDVI data is stored as ERDAS IMG files or ESRI grid file format for further analysis in eCognition and ArcMap. This analysis includes both cell by cell analysis between months as well as aggregating and summarizing cell values by a thematic or an image layer.

MODIS NDVI Evaluation and Processing

As identified in the MODIS user guides (Huete, et. al. 1999 and NASA, 2005), MODIS NDVI values are multiplied by 10000. For the area of interest these values range from -3000 to over 9000. This data was trimmed to match the area of interest for the change project. Figure 2 shows the area of interest in California with the July 12, 2005 NDVI product. The nine separate months of NDVI values were then examined in eCognition and ArcMap with other imagery and data layers. The initial review included multi-resolution segmentation in eCognition with the 2005 Landsat 7 TM data to identify thresholds for NDVI values for selected areas of crop and natural vegetation. This review indicated that values below 1000 were of limited interest particularly for the main agricultural areas. NDVI values for all months below 1000 were set to No Data and excluded from further evaluation. Values were not converted to index values or 0.10 to 0.90, but maintained in the 1000 to 9000 range. Figure 2 shows high NDVI values as bright green and low values as red. Besides the bright green of agricultural land in the Central Valley and some interior valleys, other bright green areas are the Sierra Nevada forests and the coastal forests near San Francisco Bay. The red areas are the lower foothills of the Sierra Nevada, and grassland / shrub communities on the valley floors and coastal mountains which were dry in July, 2005. The insert in Figure 2 shows the MODIS data covering one processing area which is on the western edge of the valley floor. The western side of this processing area consists of grass and shrub communities on the lower slopes of the coastal mountain range. The eastern portion of this processing area is predominantly agriculture with a mixed range of MODIS NVDI values.

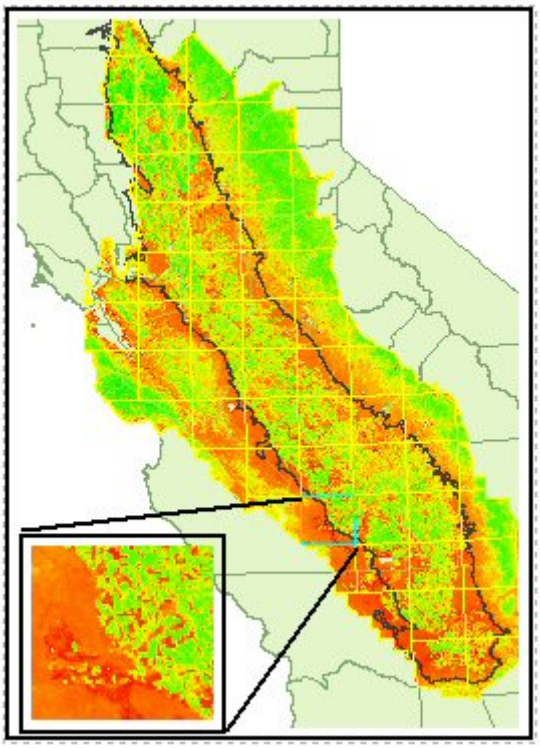

Figure 2. MODIS NDVI 250 meter data for July 12, 2005 with high values in green and low values in red for the program area.

Evaluation on a Cell Basis

Evaluation of NDVI on a cell basis included identifying months of highest values, percent change between high and low values on a monthly basis, the number of months above a particular threshold value, and month of lowest values. This processing could be easily done for the entire area of interest in ESRI grid format. Resulting values were stored as grids. In reviewing NDVI values on a cell basis for agricultural areas, the following characteristics were identified.

- Most agricultural areas show a high percentage change during the growing season
- Some agricultural areas (Particularly rice growing areas) have very high values for NDVI for at least one month between May and September.
- Some agricultural areas have high NDVI values in April dropping rapidly in May and June indicating a spring harvested crop such as winter wheat.
- Many agricultural areas show moderately high values of NDVI from May to September with a lower percentage change from values in March or April.

Based on this evaluation, an initial layer showing areas with active vegetation growth was generated. As part of the evaluation, a fuzzy value was assigned based on the criteria used. Table 1 identifies major elements of the criteria and the fuzzy value or uncertainty assigned for each category.

Table 1. Initial Classification of MODIS Cells into Active Agriculture with Fuzzy Values

Confidence – Fuzzy Value	High NDVI Value	Percent Change	High Month	Low NDVI Value
0.80	> 7000	> 10%	May through September	----------
0.70	> 5000	> 5%	April or May	< 3500
0.60	> 5000	> 20%	Any Month	---------

In applying the criteria in Table 1, classes are evaluated sequentially. Cells meeting the highest confidence criteria are not evaluated again for classes with lower confidence values. Figure 3 shows this initial assignment in an

area of active orchards and vineyards visited in September of 2005 with NAIP 2 meter imagery for 2004 in the background.

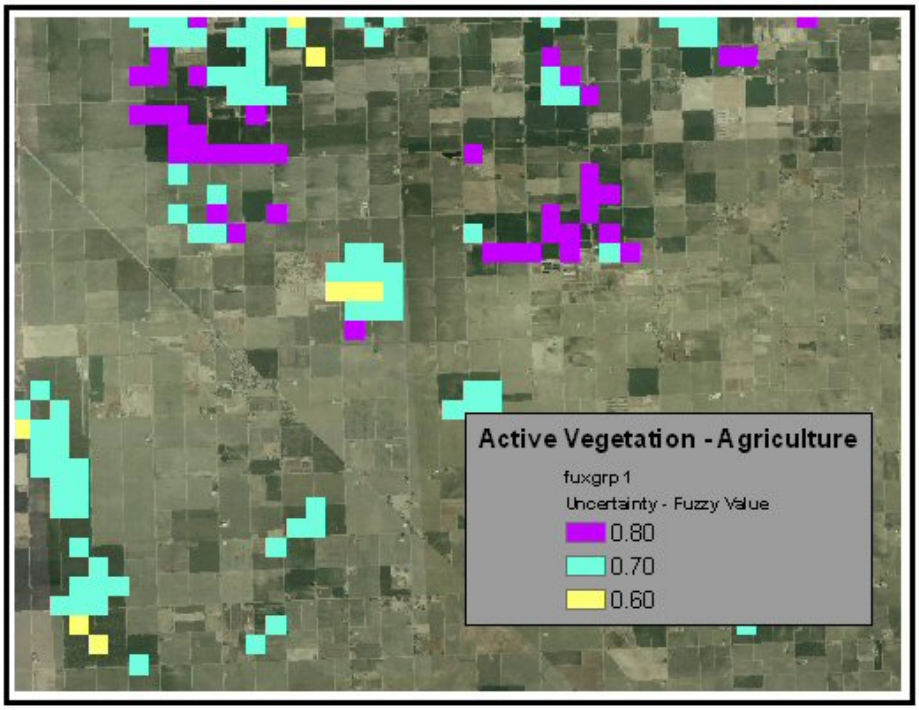

Figure 3. Initial classification of MODIS NDVI values into 3 major fuzzy classes for active agricultural land in a portion of the southern Central Valley displayed over 2004 NAIP imagery.

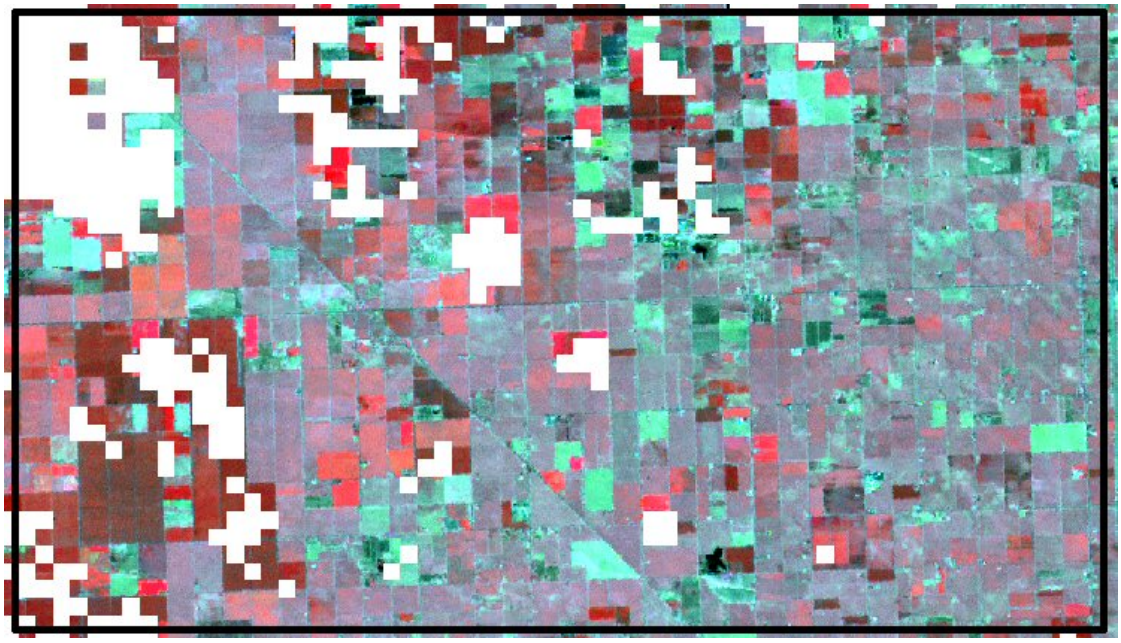

Figure 4. MODIS 250 cells with the same set of major fuzzy classes as Figure 3 shown in white over a Landsat July 2005 scene.

As shown in Figure 3, the initial classification did pick up many agricultural fields that were active in 2005. However, major areas that were known to be active orchards and vineyards during the 2005 time period were not included in this classification. Figure 4 shows the same agricultural area with Landsat 7 TM 30 meter data for July 2005 with near infrared values in red. The initial classification for active vegetation is shown as white. This initial classification is too restrictive in identifying many areas with active vegetation associated with agriculture. Additional classes can be added which do include these areas. However, rankings with lower NDVI values also begin to include more non-agricultural areas including grasslands and wetlands.

Evaluation Based on Aggregation of Cell Values

Figure 4 also raises issues about the registration of MODIS data between thematic data or other imagery as well as between dates of MODIS data. This suggests grouping or aggregation of cells for comparison between dates of interest or for comparison to other data. A technique used for the 2000 cycle was to generate summaries of cell values within polygons segmented from the 2000 TM image. An effective technique used extensively by Reclamation in agricultural areas is to summarize Landsat 30 meter data within field borders. There are several sources of thematic data representing field borders within the Central Valley. There is the Common Land Unit database of the Farm Services Agency, USDA. There are also field border databases developed by the local County Agricultural Commissions for reporting pesticide application and use to the California Department of Pesticide Regulation. Neither source currently provides complete coverage throughout the Central Valley. A surrogate for a complete field border database is a composite of the land use mapping done by DWR. DWR maps crop types on a County basis every few years and most delineations conform to field borders. The composite database of these separate County databases ranges in date from 1993 to 2004. This is an extended period for identifying agricultural field status or current crop type but it does permit effective grouping of major crop categories for generating statistical summaries of MODIS NDVI values.

Evaluation with field borders. The composite database of DWR land use was used to summarize NDVI values for groups of large fields. These summaries were used for initial comparison with selected processing areas that contain a full set of fields. The set of large fields ranges from 1,000 hectares (250 acres) to 10,000 hectares (2,470 acres). Smaller or larger fields were excluded from the valley wide summary. Agricultural field size varies throughout the valley. Average field size is about 500 hectares (120 acres) or much smaller than the fields used for the summary values valley wide. Table 2 and Table 3 show summaries for percentage change of four major crop groups over the entire valley and for one processing area respectively.

Table 2. Summary of percent change in MODIS NDVI values for large fields valley wide.

	Field Crop	Vegetable Crop	Orchard	Vineyard
# of Fields	1183	334	485	396
Total # MODIS cells	34489	9186	14308	13380
Mean % Change – Year	275%	288%	69%	82%
Maximum % Change	621%	565%	408%	386%
Minimum % Change	23%	30%	19%	29%

Table 3. Summary of percent change in MODIS NDVI values for all fields in one processing area.

	Field Crop	Vegetable Crop	Orchard	Vineyard
# of Fields	1372	258	2628	3599
Total # MODIS cells	5284	558	6045	12238
Mean % Change – Year	112%	79%	54%	63%
Maximum % Change	429%	429%	286%	312%
Minimum % Change	6%	10%	5%	9%

Both tables include the number of fields in each crop group and the total number of MODIS cells used for the summary values. Percentage change is based on the month of lowest NDVI values, typically April or October of this year, and the month of highest values. The month of highest values could be any month from May to September. The set of fields covering the entire valley should contain from about 16 MODIS cells to over 200 cells. For individual processing areas, the number of MODIS cells included in each field range from one or a few cells to less than hundred for most areas. Field borders and crop assignments for the full valley range in date from 1993 to 2004. The dates of DWR data for the processing area represented in Table 3 are 2000 or 2003.

The full raster datasets generated from the MODIS NDVI values for the valley can efficiently summarized and evaluated in ArcMap and eCognition by processing area. This permits the inclusion of all fields or representation of fields for analysis in setting thresholds.

Aggregating and summarizing NDVI values by field borders. Summaries generated both at the very large field basis for the Central Valley and for separate processing areas are consistent not only for the four major crop groups identified in the tables but for other major crop groupings. The summaries indicate that additional breaks for active vegetation can be defined within a processing area not identified in the initial mask. The application at the processing area level depends on the use of a field border data base or some other representation of field boundaries. Segmentation of 2005 imagery with field borders and other thematic data in eCognition is used at the processing area level to test thresholds of the summary raster data generated in ArcMap. The summary raster data includes the individual NDVI values for each month and the raster files identifying month of highest value, lowest value, months above a threshold for the full area. Individual raster data files for the separate 58 processing areas do not need to be generated.

The review at the processing area level indicates additional thresholds may be useful for particular crop groups. A study in Kansas (Wardlow, et. Al., 2005) has used the MODIS NDVI data for preliminary classification of crop type. Figure 5 and Figure 6 show the mean, minimum, and maximum NDVI values by month for three major crop types, field crops, vegetables, and orchards for one processing area.

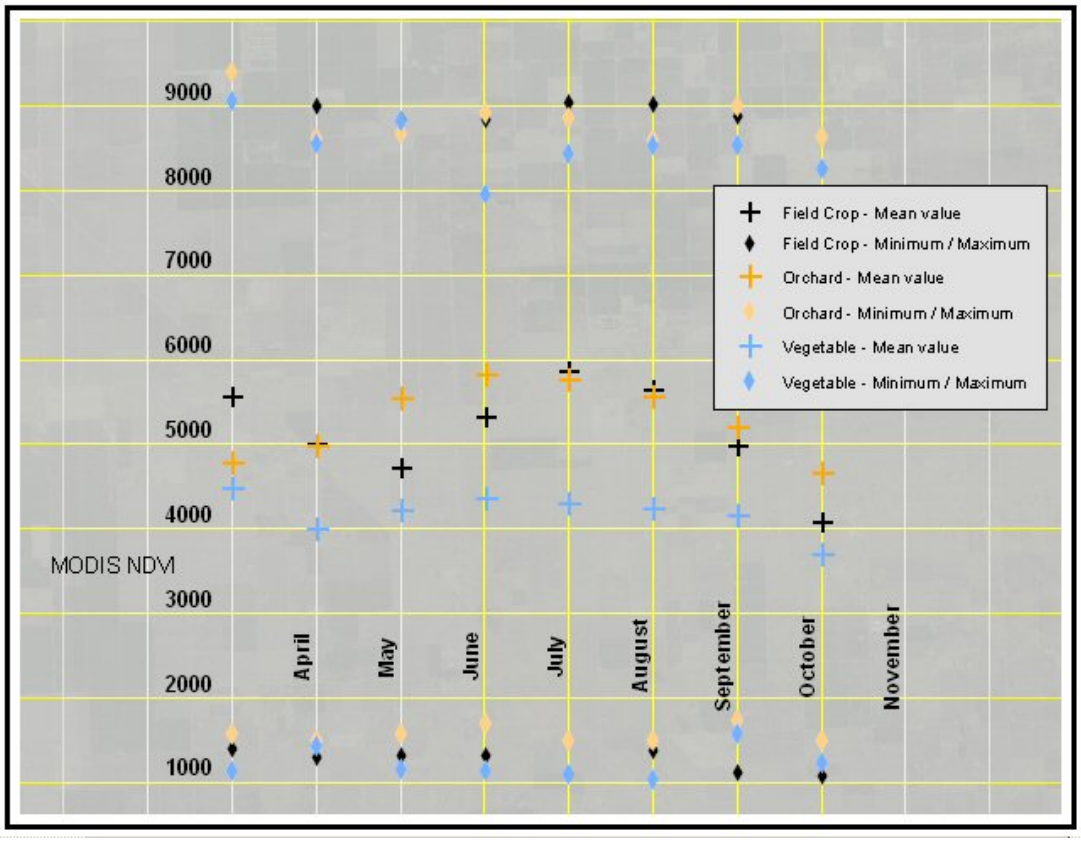

Figure 5. Mean, minimum, and maximum MODIS NDVI values for field, vegetable, and orchards within one processing area.

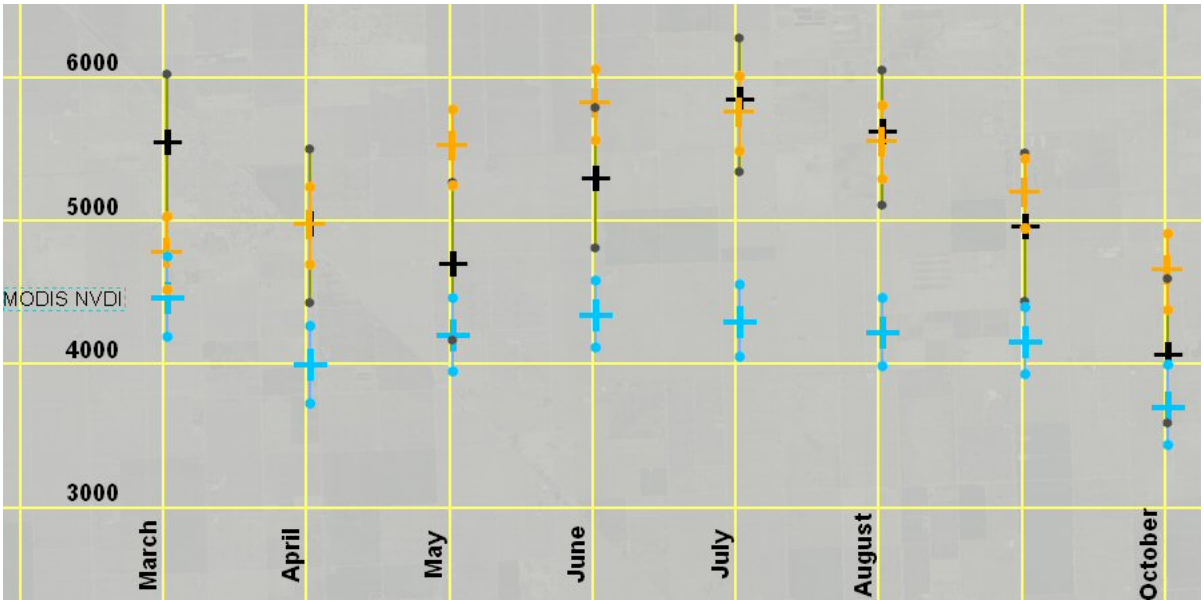

Figure 6. Average mean, minimum, maximum of MODIS NDVI values for three major crop groups in one processing area.

Figure 6 shows the average of these values for the crop types shown in Figure 5. It indicates that some separations may be possible using the MODIS data over the season. These three crop types show some separation in mean NDVI values and the change in these values between months. NDVI values for orchards and field crops are consistently higher than values for vegetable crops. Higher peak values of NDVI for orchards and a more rapid decline in values for field crops at the end of the season may separate field crops from orchards. For this processing area, a threshold value above 4300 should pick up most fields with active vegetation. This approach may be useful for the Central Valley in identifying active major crop groups. Classifying crops with MODIS NDVI values is beyond the intent of developing an agriculture mask. The process is useful for setting thresholds in developing that mask. It will require further investigation at a processing area level. The crop type assignment for these field borders is based on DWR crop mapping in 2000 or 2003. DWR crop mapping for the full valley ranges from 1993 to 2004. It is expected that there will be variation in major crop types between 1993 and 2005. Variation is also expected in field size, climate conditions, cultural practices, and plant conditions throughout the valley. Thresholds initially set within a processing areas will than provide comparisons between other processing areas in identifying the fuzzy value of level of uncertainty associated with the threshold for classifying active vegetation.

SUMMARY

MODIS NDVI values have been useful in the development of an agricultural mask for eliminating change due to cultural practices in the Central Valley of California. The datasets are readily obtainable from several internet sites. The processed data with NDVI are available within a few weeks from the date of capture. The MODIS data cover the same time period as the 2005 Landsat 7 scenes used to develop change. Analysis of the MODIS NDVI values was done on a cell basis between months to identify months with high values, months with low values, and the percentage change between high and low values. These values were then grouped based on a thematic layer representing agricultural field boundaries to produce summary values stored as tables. These values provide a basis for setting a variety of threshold values to identify active agricultural areas. These include threshold values of NDVI for any month, the number of months with NDVI values above or below a particular value, and the percentage change in values between months. These thresholds are being used to improve the agricultural mask and to identify the level of uncertainty in that assignment.

The MODIS data may also assist in identifying some major crop types of interest within the Central Valley. The initial assessment with crop categories identified in the field border database used in this study indicated that major

categories could be identified based on seasonal changes in NDVI values. The field border database used did identify specific crops. However, the data ranges in date from 1993 to 2004. Further exploration of the data will continue for several crop types. This will include the incorporation of climatic information for different parts of the valley for the period of interest.

REFERENCES

Curlis, C., D. Hansen, B. Simpson, D. Adams (2003). Change in the Central Valley, California, 1993-2000. Proceedings from ESRI 2003 International User Conference, San Diego, California.

Huete, A., C. Justice, W. van Leeuwen (1999). MODIS Vegetation Index (MOD 13) Algorithm Theoretical Basis Document, Version 3. University of Arizona, 1200 E. South campus Drive, 429 Shantz Building #38, Tucson, Arizona.

University of Arizona(?). MODIS Product MOD13A3 User Guide. http://tbrs.arizona.edu/project/MODIS/UserGuide_doc.php

NASA (2005). MODIS MOD13Q1v4 Product through the LP DAAC Data Pool. ftp://e0dps01u.ecs.nasa.gov/MOLT/MOD13Q1.004/

USGS (2004). MODIS Reproduction Tool Users's Manual. EROS Data Center. Department of Mathematics and Computer Science, South Dakota School of Mines and Technology.

Wardlow, B.D., S.L. Egbert (2005). State-Level Crop Mapping in the U.S. Central Great Plains Agroecosystem Using MODIS 250 Meter NDVI Data.Proceedings from Pecora 16 "Global Priorities in Land Remote Sensing". October 23-27, 2005. Sioux Falls, South Dakota.