5-2013

Classification of Satellite Time Series-derived Land Surface Phenology Focused on the Northern Fertile Crescent

Brian Embree Bunker
University of Arkansas, Fayetteville

CLASSIFICATION OF SATELLITE TIME SERIES-DERIVED LAND SURFACE PHENOLOGY FOCUSED ON THE NORTHERN FERTILE CRESCENT

CLASSIFICATION OF SATELLITE TIME SERIES-DERIVED LAND SURFACE PHENOLOGY FOCUSED ON THE NORTHERN FERTILE CRESCENT

A thesis submitted in partial fulfillment

of the requirements for the degree of

Master of Arts in Geography

By

Brian Bunker

Brigham Young University

Bachelor of Science in Geography, 2011

May 2013

University of Arkansas

ABSTRACT

Land surface phenology describes events in a seasonal vegetation cycle and can be used in a variety of applications from predicting onset of future drought conditions, to revealing potential limits of historical dry farming, to guiding more accurate dating of archeological sites. Traditional methods of monitoring vegetation phenology use data collected *in situ*. However, vegetation health indices derived from satellite remote sensor data, such as the normalized difference vegetation index (NDVI), have been used as proxy for vegetation phenology due to their repeated acquisition and broad area coverage. Land surface phenology is accessible in the NDVI satellite record when images are processed to be intercomparable over time and temporally ordered to create a time series. This study utilized NDVI time series to classify areas of similar vegetation phenology in the northern Fertile Crescent, an area from the middle Mediterranean coast to southern/south-eastern Turkey to western Iran and northern Iraq. Phenological monitoring of the northern Fertile Crescent is critical due to the area's minimal water resources, susceptibility to drought, and understanding ancient historical reliance on precipitation for subsistence dry farming. Delineation of phenological classes provides areal and temporal synopsis of vegetation productivity time series. Phenological classes were developed from NDVI time series calculated from NOAA's Advanced Very High Resolution Radiometer (AVHRR) imagery with 8 × 8 km spatial resolution over twenty-five years, and by NASA's Moderate Resolution Imaging Spectroradiometer (MODIS) with 250 × 250 m spatial resolution over twelve years. Both AVHRR and MODIS time series were subjected to data reduction techniques in spatial and temporal dimensions. Optimized ISODATA clusters were developed for both of these data reduction techniques in order to compare the effects of spatial versus temporal aggregation. Within the northern Fertile Crescent study area, the spatial reduction

technique showed increased cluster cohesion over the temporal reduction method. The latter technique showed an increase in temporal smoothing over the spatial reduction technique. Each technique has advantages depending on the desired spatial or temporal granularity. Additional work is required to determine optimal scale size for the spatial data reduction technique.

This thesis is approved for recommendation

to the Graduate Council.

Thesis Director:

_____

Dr. Jason Tullis

Thesis Committee:

_____

Dr. Jesse Casana

_____

Dr. Jackson Cothren

THESIS DUPLICATION RELEASE

I hereby authorize the University of Arkansas Libraries to duplicate this thesis when needed for research and/or scholarship.

Agreed _____

Brian Bunker

Refused _____

Brian Bunker

ACKNOWLEDGEMENTS

This research was carried out as part of the research project “Settlement Systems and Environmental Change in the Northern Fertile Crescent” supported by NASA ROSES Space Archeology grant NNX10AM39G.

DEDICATION

This thesis, and the work invested herein, is dedicated to my wife, Eila.

TABLE OF CONTENTS

1. Introduction 1
   1.1 Satellite Imagery Maturity 1
   1.2 Pre-Processed Time Series Imagery Availability 3
   1.3 NDVI Time Series Classification 5
      1.3.1 Flattened-Data 6
      1.3.2 Reduced-Data 6
      1.3.3 Full-Data 7
   1.4 Research Hypothesis 9
2. Background 9
   2.1 Unsupervised Classification 10
      2.1.1 Similarity 11
      2.1.2 User Input 13
      2.1.3 Process 14
      2.1.4 ISODATA 15
      2.1.5 Evaluation 17
   2.2 Image Segmentation 20
      2.2.1 Homogeneity and Heterogeneity 20
      2.2.2 Use Considerations 22
   2.3 NDVI Time Series 25
      2.3.1 Classification 26
      2.3.2 Classification Issues 28
3. Methods 29
   3.1 Study Area 29
   3.2 Remote Sensor Data 31
      3.2.1 AVHRR 31
      3.2.2 MODIS 33
      3.2.3 Landsat and Other High Resolution Imagery 34
   3.3 Data Reduction 34
      3.3.1 Mean Year Method 36
      3.3.2 Segmentation Method 39
   3.3 ISODATA Classification 43
      3.3.1 K Parameter Adjustment and Cluster Validity 44
4. Results and discussion 45
   4.1 Mean Year Reduced-Data 45
      4.1.1 Transformed Divergence 46
   4.2 Segmented Reduced-Data 49
      4.2.1 Transformed Divergence 52
   4.3 Comparison of Mean Year Classes to Segmented Classes 52
      4.3.1 High Resolution Imagery of Selected Classes 59
      4.3.2 Temporal Trajectory of Selected Classes 65
   4.4 Outcome of Work 67
   4.5 Future Directions 68
5. Conclusion 69
6. References 70

7. Appendix 74
7.1 Creation of Mean Year Rasters 74
7.2 Calculation of Segment Means 75
7.2 Creation of Segment Rasters 75

LIST OF FIGURES

Figure 1 Types of reduced-data time series……. 5
Figure 2 Crisp and fuzzy class memberships…….11
Figure 3 Spectral reflectance curve vs. temporal NDVI curve……..15
Figure 4 Segmentation calculation example……..22
Figure 5 Six segmentation scales for time series AVHRR NDVI……24
Figure 6 Northern Fertile Crescent study area……...29
Figure 7 Time series data reduction workflow…….. 35
Figure 8 Temporal data reduction vs. spatial data reduction……36
Figure 9 Mean year time series calculation example……38
Figure 10 Removal of tile segmentation edge artifacts……. 40
Figure 11 Segment reduced time series calculation example…… 43
Figure 12 Transformed divergence for mean year AVHRR and MODIS classifications……. 47
Figure 13 Data reduction vs. average segment size for segment mean method…… 50
Figure 14 Map of mean year AVHRR NDVI classification……...55
Figure 15 Map of mean year MODIS NDVI classification……...56
Figure 16 Map of segment mean AVHRR NDVI classification……... 57
Figure 17 Map of segment mean MODIS NDVI classification…… 58
Figure 18 Point of interest 1 natural color imagery……... 60
Figure 19 Point of interest 1 CORONA image from November 1968…….. 60
Figure 20 Point of interest 1 mean year classification map……... 61
Figure 21 Point of interest 1 segment mean classification map…….61
Figure 22 Points of interest 2 and 3 natural color imagery……62
Figure 23 Points of interest 2 and 3 mean year classification map……62
Figure 24 Points of interest 2 and 3 segment mean classification map……. 63
Figure 25 Points of interest 4 and 5 natural color imagery……63
Figure 26 Points of interest 4 and 5 mean year classification map……64
Figure 27 Points of interest 4 and 5 segment mean classification map……. 64
Figure 28 Points of interest 1 and 2 temporal NDVI graph……...65
Figure 29 Point of interest 3 temporal NDVI graph…….. 66
Figure 30 Point of interest 4 temporal NDVI graph…….. 66
Figure 31 Point of interest 5 temporal NDVI graph…….. 67

LIST OF TABLES

Table 1 Flag values for GIMMS AVHRR NDVI 32
Table 2 Flag values for MODIS NDVI 33
Table 3 Classes intersecting the study area 53
Table 4 Cluster membership of select point of interest locations 59

1. INTRODUCTION

Land surface phenology describes the events connected to seasonal changes in vegetation (Jensen 2007). Researchers have monitored phenological trends and linked them with changes in environmental variables (such as precipitation and temperature) to predict onset of drought in arid climates (Bajgiran et al. 2008; Trigo, Gouveia, and Barriopedro 2010). The Middle East region, including the northern Fertile Crescent, particularly benefits from an increased understanding of vegetation phenology dynamics due to its persistent reliance on scarce water resources, specifically precipitation, for subsistence dry farming. Scholars traditionally believed that the spatial extent of dry farming conformed to a relatively static isohyet between two-hundred fifty and three-hundred millimeters precipitation per year, with minor inter-annual changes driven by precipitation accumulation variation (Beaumont 1996; de Brichambaut and Wallen 1963). This boundary was set using traditional monitoring of vegetation phenology *in situ*. More recently, however, researchers have accessed land surface phenology from remotely sensed images resulting in the ability to monitor large areas at frequent intervals. They have used remote sensing-derived land surface phenology to create zones of similar phenology within a broad area (Gu et al. 2010). This study partially revisits the limit of dry farming proposed by de Brichambeaut and Wallen (1963) by classifying vegetation phenology in the northern Fertile Crescent using remotely sensed imagery.

1.1 SATELLITE IMAGERY MATURITY

Digital satellite remote sensing of the earth has produced records of land conditions for multiple decades, made possible by regular revisit schedules inherent in repeating processional orbits. Multiple programs have launched a series of sensors to maintain uninterrupted time series

image collections spanning many years. Notable examples of such programs are Landsat, jointly sponsored by USGS and NASA, NOAA's Advanced Very High Resolution Radiometer (AVHRR), and NASA's Moderate Resolution Imaging Spectroradiometer (MODIS).

A natural tradeoff between spatial resolution and temporal resolution of global image acquisition exists in satellite remote sensing. High temporal resolution data collection (e.g. daily revisit) requires a large ground area to be covered in each image while high spatial resolution (e.g. ability to identify a specific road) requires focusing on a small ground area. Many new initiatives have lessened the time between revisits while increasing the spatial resolution per pixel. Beginning in 1972, Landsat data collection occurred every 18 days (16 days on later missions) but boasted spatial resolution of 80 × 80 m (reduced to 30 × 30 m on later missions). The AVHRR sensor program, which began in 1978, is able to acquire daily global imagery at a spatial resolution of 1.1 × 1.1 km. MODIS was first placed in orbit in 1999 with the capability of daily image collection at a maximum spatial resolution of 250 × 250 m for data collection in its red and near-infrared bands. As satellite remote sensing embraces maturity, and advances in storage of digital imagery and computational power increase, new imaging programs will likely lessen the tradeoff even more. For a comprehensive review and discussion of long time series satellite imagery programs spanning more than twenty years, readers are directed to Gutman and Masek (2012). Their research includes information about derivation and availability of data products derived from these long time series programs.

The complete archive of successive temporal images acquired by each mission is comprised of multi-decadal records that have garnered significant interest in both academia and in industry. Multi-temporal change detection has achieved widespread adoption to quantify change in land cover using just two images focused on the same spatial region but separated by a

period of time. However, such discretized utilization of rich time series imagery neglects its inherent ability to expose the temporal trajectory of the land changes. In other words, rather than utilize imagery from just two dates, there may be certain advantages in utilizing complete time series when continuous data are available that span multiple decades. A research area where use of such time series information is beneficial is in understanding the processes involved in land cover changes resulting from climate change. Researchers can identify specific changes in year-to-year phenology, examine associated climate and physiographic variables, and model or synthesize processes and interactions that lead to changes on the land surface.

It is important to note that the terms "multi-temporal" and "time series" have slight but significant differences in their definitions. "Multi-temporal analysis" focuses on analysis of spatial changes in a variable among multiple, discretized dates, while "time series analysis" means analysis of temporal trajectory of a variable through a continuum of dates. Change detection between two dates is an example use of multi-temporal imagery. An example use of time series data is classifying continuous vegetation data according to phenological similarity. This research addresses exclusively the use of long (multiple, continuous years) time series imagery for classifying phenological variables.

1.2 PRE-PROCESSED TIME SERIES IMAGERY AVAILABILITY

As a result of the maturity of time series remote sensing, researches have made attempts to produce ready-to-use time series datasets. Traditional preprocessing procedures, among others, require clouds and atmospheric noise data to be removed, bi-directional reflectance correction due to satellite drift, and data calibration because of sensor degradation.

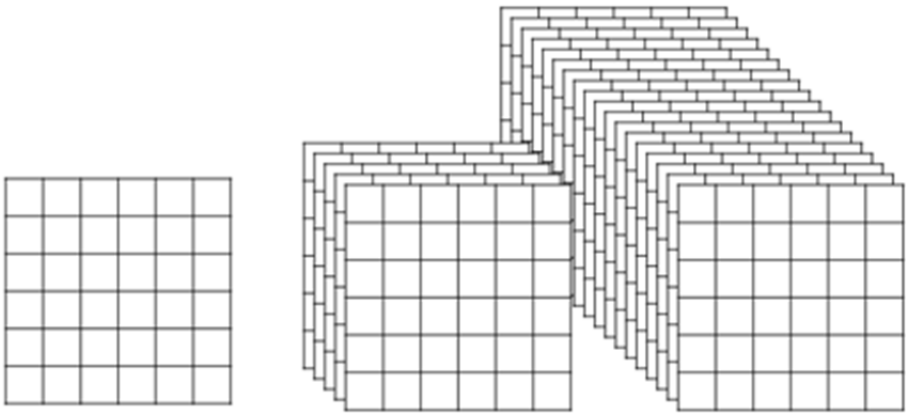
A review of a few examples of research in correcting these issues will provide perspective of their relevance to time series imagery analysis. Holben et al. (1986) presented

Maximum Value Compositing (MVC), which is applied to time series imagery by retaining the maximum reflectance value for each pixel over a compositing period to obtain multi-day composite images. Common compositing periods range from eight days to one month in order to provide enough data to ensure acquisition of maximum reflectance for each pixel in a scene. Later, Vermote and Kaufman (1995) and Los (1998) introduced sensor calibration techniques for AVHRR channels 1 and 2, which normalized collected data that proved valuable contributions for working with time series imagery that requires consistent measurements over time. Vermote et al. (1997) worked to correct degraded AVHRR image data after the Mount Pinatubo eruption spread a large blanket of aerosol particulates across the planet. Left uncorrected, the interference of the aerosols with satellite imagery would have rendered the affected scenes useless in a long time series dataset. In addition, Pinzón et al. (2005) developed a method using empirical mode decomposition to isolate and remove spectral artifacts in time series imagery caused by gradually changing solar zenith angles due to the orbital drift of a satellite platform. These and other techniques have given scientists the ability to create highly consistent time series imagery datasets.

Interest in pre-processed, ready-to-use time series imagery led to the creation of multiple high quality datasets developed using AVHRR or MODIS imagery. The previously mentioned advances removed cumbersome processing steps and point current researchers to explore the depth of information gained from time series analysis of temporal trajectory. This study focuses on investigating the value added to unsupervised classification of the normalized difference vegetation index (NDVI) using the full temporal trajectory time series imagery.

1.3 NDVI TIME SERIES CLASSIFICATION

Scientists have frequently used remotely sensed data to extract thematic information using manual image interpretation or computational statistical pattern recognition algorithms. The process of transforming data into thematic information is known as classification (Jensen 2005).



	Flattened	Reduced	Full
Description	Original time series values used to calculate a representative value or metric per pixel location	Full data time series used to create a set of new rasters of reduced dimensionality in temporal or spatial domain	Minimal data reduction has occurred to produce a regularly spaced time series between start date and end date
Result	Single raster	Raster set with temporal sequence	Raster set with temporal sequence
Examples	Average start or end of season day, day of max or min NDVI, max or min NDVI value, etc.	Mean NDVI per repeating temporal frame (e.g. avg. Jan, Feb, etc. for years included in time series)	Time series NDVI from AVHRR or MODIS

Figure 1 Time series of remotely sensed images have been used as full-, reduced-, and flattened-data time series in digital image classification. Each type of time series targets specific attributes of vegetation phenology in image classification.

Researchers have approached classification of satellite imagery-derived time series NDVI using several methods. The following examples are a sample of the variety of methods used for classifying NDVI time series. The examples were divided into three categories: flattened-, reduced-, and full-data time series. Flattened-data time series are characterized by the

replacement of original data values with single-value metrics derived from the date and NDVI value of important phenological events. One such metric is "length of growing season" calculated as the number of days with an NDVI value greater than a predetermined threshold. Reduced-data time series are designed to reduce the dimensionality of a dataset while still maintaining a temporal sequence. Full-data time series utilize the maximum number of imagery dates to obtain temporal NDVI vectors (array of values) for every pixel location which accurately track the trajectory of data values over time.

1.3.1 FLATTENED-DATA

Gu et al. (2010) demonstrated the use of flattened-data time series to produce a "phenoclass" map of the conterminous United States. Using 16-day composite MODIS scenes for the years 2000 through 2004, they calculated nine phenological metrics: 1) start-of-season time, 2) start-of-season NDVI, 3) end-of-season time, 4) end-of-season NDVI, 5) maximum NDVI, 6) maximum NDVI time, 7) duration of season, 8) amplitude of NDVI, and 9) seasonal time integrated NDVI. Next, they calculated each metric for each year in the dataset and then the median value of each metric for the five years of MODIS data. The result of these processing steps flattened the time series into a set of nine rasters, one for each metric. The authors used the metric rasters as input to image classification to produce forty phenological classes.

1.3.2 REDUCED-DATA

Weiss et al. (2001) aggregated monthly time series AVHRR NDVI data in Saudi Arabia for the years 1982 to 1994 into a one-value-per-pixel-per-year dataset. They calculated one-value-per-pixel as the coefficient of variation (COV), which measured NDVI variation over time (one year period). They then used regression to explore linear trends in COV values over time.

Areas exhibiting negative slopes over the twelve years of data indicated a loss in vegetation while positive slopes showed an increase in vegetation. Next, the scholars classified the COV slopes as declining, stable, or improving per pixel to obtain a desertification status map. The yearly COV values created a reduced dataset but temporal trajectory was a crucial aspect of the study.

Kouchoukos (2001) produced an "agro-ecological" map from fifteen years of time series data in the Fertile Crescent. He calculated monthly averages of AVHRR NDVI data over all fifteen years resulting in twelve NDVI images representing an average year. He then temporally ordered and subjected the monthly averages to unsupervised classification, resulting in a map with seven classes. Kouchoukos reduced the data in this example by averaging corresponding months over time, a method repeated by Al-Bakri & Taylor (2003).

Another example of using reduced-data time series is presented in Benedetti et al. (1994). They performed principle components analysis (PCA) on monthly AVHRR composites of the northern Mediterranean for 12- and 6-month time frames during the years 1986 to 1989. PCA is a common technique to reduce dimensionality of a dataset while retaining the most non-redundant information possible. Next, the researchers conducted unsupervised classification on the first three principle components for both 12- and 6-month datasets. The resulting 6-month PCA class map classified "agro-climatological" regions. Scholars widely employ PCA and other techniques such as Fourier transform, to reduce dimensionality and to remove noise while maintaining the temporal sequence within time series data.

1.3.3 FULL-DATA

Time series remote sensing data are valuable in research like the previously cited studies. Some research questions, however, benefit from the use of a full time series dataset. A full time

series dataset is sometimes referred to as temporal trajectory, a term that denotes the ability to obtain a measurement at any point between a start and end date. Kouchoukos et al. (1998) obtained full time series AVHRR NDVI data for a single hydrologic year, specifically October 1992 to September 1993. After temporally arranging the thirty-six images that comprised the full time series, they classified the data to produce thirteen classes and interpreted the classes according to dominant land cover type.

Nguyen et al. (2012) used Système Pour l'Observation de la Terre (SPOT) time series imagery for 1998 to 2008 to map rice cultivation on the Mekong Delta of southern Vietnam. They ran multiple classifications on the long full time series data and determined the best classification map through the use of a cluster separability index that measured the statistical uniqueness of each class. The resulting map consisted of 77 classes that the scientists examined for rice-cropping activity. Next, they compared the classes with field data and official records, which revealed that their technique was highly successful in identifying rice fields.

Nguyen et al. called the long time series data "hyper-temporal," repeating a term first used by Piwowar and LeDrew (1995) to describe the time series imagery they used to monitor sea ice. The term "hyper-temporal" introduces uncertainty whether the time series is restricted to many images over a long time frame (e.g. 200 images over 5 years) or describes many images over a short time frame (e.g. 200 images over 5 minutes). Therefore, researchers should use the term "long time series" over "hyper-temporal" to avoid ambiguity of temporal duration of time series.

To date, few analyses have made use of full-data long time series. This is partially due to their relatively new availability. Remote sensing missions have entered a phase of maturity, opening opportunities for research in this area. Yet, researchers still face barriers in the analysis

of time series data, which requires specialized computer equipment and programming skills to access the time series data in a synoptic way. Use of full-data time series in studies of temporal change will undoubtedly become more common as reliable time series imagery is extended, computer processing performance increases, and off-the-shelf software capable of analyzing time series becomes available.

1.4 RESEARCH HYPOTHESIS

This study focused on the information content added to classification when long time series are used. Specifically, investigation of the following hypothesis was pursued: there is no spatial difference between unsupervised classifications based on mean year reduced vs. segment mean reduced long time series NDVI. During the investigation of the research hypothesis, the following questions were considered: 1) what are the qualitative differences in the spatial distribution of classes; and 2) what do the varying class boundaries contribute to our understanding of regional and temporal patterns of the classified area? Common ISODATA clustering implementation was used to classify both reduced data time series, and a new use of image segmentation as a data reduction technique was developed.

2. BACKGROUND

At the most elementary level, classification of satellite imagery and associated derived variables is an advanced form of image interpretation and a continuation from the analog era of manual photo interpretation (Congalton 2010). During the analog era, aerial photo analysts manually traced boundaries between objects on a photo to produce a classified thematic map. Such maps were, and are, valuable to professionals needing accurate inventory of land use or land cover over large areas. However, the process of manually creating classified maps was time

and labor intensive, making the mass production of custom classified maps cost and talent prohibitive. The birth of the digital era enabled researchers to transfer the process of classification to computer programs that could create classification maps very quickly with minimal analyst input. The transition from analyst-centric to computer-centric classification, and the associated time reduction for producing a map, coupled with the easy accessibility of digital imagery, has led to a proliferation of classification maps.

2.1 UNSUPERVISED CLASSIFICATION

The starting point for digitally creating a classification map is to group pixels with similar data characteristics into classes. Image classification techniques are commonly divided between supervised or unsupervised algorithms (Jensen 2005). Supervised classification is accomplished using class samples identified from an image via an expert image interpreter. The computer program calculates the similarity of each pixel to each sample class created by the expert. Unsupervised classification clumps the data into classes according to the inherent structure of the data in the measurement space and is usually based on a user-defined *k* parameter that tells the algorithm how many classes the researcher expects to find across the image data. The expert operating the classification software has little to no additional input. The classification software then uses a series of heuristic evaluations of intermediate data class membership to adjust class parameters and reassign data to new classes until the algorithm converges on the best distribution of data given *k* classes.

Unsupervised digital image classification removes one source of human bias from the classification process: false pattern recognition of non-patterns. The human brain can identify visual patterns very easily such as differences in shapes of letters, regions of different colors, differences in faces of separate individuals, etc. However, when faced with a question of subtle

statistical pattern dispersal, the brain tends to pick out patterns that are not statistically significant and creates patterns where randomness exists (Rogerson 2006). Automatic image classification ensures that the label assigned during classification is statistically justified.

Classification using computational methods also provides methods for complex class assignments. Basic classification assigns each pixel in a scene to a specific class label. As a result, each class produced is *crisp* or does not intersect with any other class (also known as *hard*). Advanced classification algorithms can also assign each pixel a probability score for each class it could be assigned to, causing the classes to fade into each other. These soft-boundary classes are known as *fuzzy* classes.

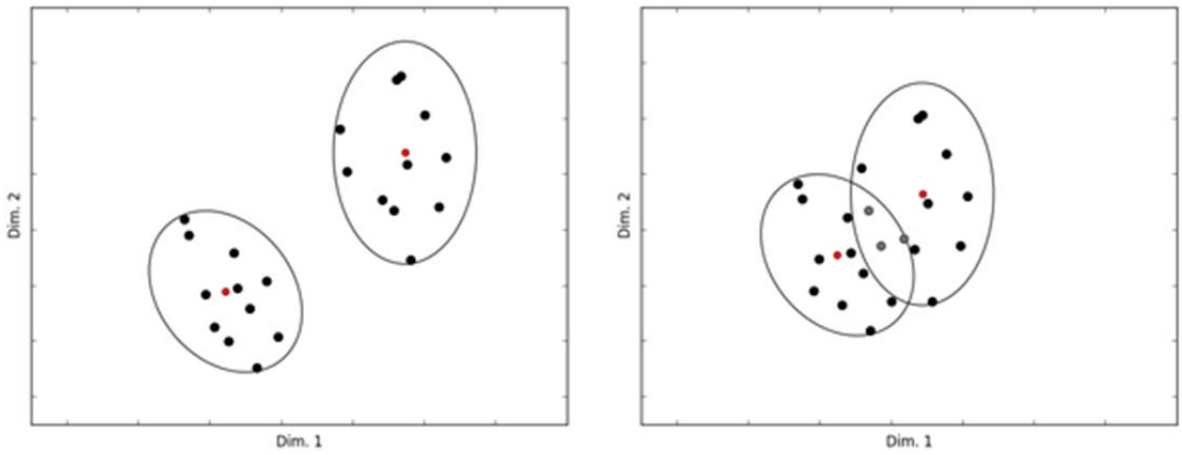


Figure 2 Crisp (left) and fuzzy (right) class distibutions plotted against two input dimensions. The gray data points in the right graph could belong to either class and could have an associated class membership probability value.

2.1.1 SIMILARITY

The goal of unsupervised time series classification is "to determine groups of similar time series" (Liao 2005, 1858). Key to the idea of classification is the concept of similarity. For each implementation of different remote sensing classification algorithms, pixel (and time series vector) calculations determine the level of similarity to a known group or label. The specific

algorithm used to determine similarity is known as a similarity measure. Multiple similarity measures are currently being used to produce class maps. The knowledge of the analyst, structure of the data, and goal of the output class map guide which similarity measure an analyst uses in the process of classification. *D* notation for the following similarity measures represents the distance between two time series vectors, $f(p)$ and $f(q)$.

Minkowski distance is a generalization of Euclidean and Manhattan distances and is the absolute difference between time series values f^p and f^q at each moment t in the vectors (Lhermitte et al. 2011). The distance is calculated using the equation:

$$D_{Mink} = \left(\sum_{t=1}^{n} \left| f_t^p - f_t^q \right|^r\right)^{\frac{1}{r}} \quad [1]$$

When *r*, a unit-less placeholder variable used to generalize the Euclidean and Manhattan distance equations to the Minkowski distance equation, is assigned a value of 1, the equation gives Manhattan distance. Euclidian distance is calculated when *r* is given a value of 2. Many unsupervised classification algorithms rely on vector similarity calculated using Euclidean distances because it is simple to calculate and straightforward to interpret.

More complicated similarity measures calculate the distance between time series after first applying a mathematical transformation such as PCA or Fourier analysis. The PCA distance is measured between a subset of principle components for each time series given by PC^p and PC^q. The distance is calculated by:

$$D_{PCA} = \sqrt{\sum_{s=1}^{m} \left(PC_s^p - PC_s^q\right)^2} \quad [2]$$

PC_s^p and PC_s^q are the s^{th} principle components of the selected subset of *m* principle components. Since PCA is designed to maximize variance in the first principal component (PC) and proceed with reduced variance in subsequent PCs, only the first few PCs may suffice in capturing the necessary information to properly classify each pixel.

Another transformation based similarity measure, ξ, uses relative amplitude and phase of a time series Fourier transform to match the shape of a time series curve to a set of reference curves. The most similarly shaped curve will have the lowest ξ value among the set of references. In direct contrast to a distance-based similarity measure, amplitude differences and phase shifts are ignored and only curve shape is compared. Specific algorithm details can be found in Evans and Geerken (2006).

Scholars have developed other time series similarity measures to address unique issues posed by research problems. A review and comparison of a number of time series similarity measures is presented in Lhermitte et al. (2011). In the current research, Euclidean distance is used to measure similarity of time series vectors.

2.1.2 USER INPUT

Supervised classification is used when *a priori* reference data are available. From a time series remote sensing perspective reference data are pixel vectors of length *n* where *n* is the number of scenes (dates) included in the time series. The *k* parameter, which is the number of vectors in a reference dataset, represents the number of classes in the data. Each vector represents the ideal time series values for a class. When reference data are unavailable, unsupervised classification is used to determine class values.

Unsupervised classification algorithms do not need reference data but usually do require a *k* parameter to be defined by the analyst. The *k* parameter is an estimate of how many distinct classes are present in a dataset of time series imagery. Traditionally, *k* has been derived from a previous classification of the area of interest or interactively adjusted until the analyst determines that significant classes have been identified. Relatively few studies using unsupervised classification of NDVI incorporate the *a posteriori* class similarity statistical analysis described

by Tou and Gonzales (1974) and Swain and Davis (1978) and exemplified by Nguyen et al. (2012). Class similarity will be addressed in greater detail shortly.

2.1.3 PROCESS

The basic classification process traces boundaries between unique areas of a map, grouping together areas that are similar, and applies labels to each group. Similar areas could have simple or detailed labels depending on the variable under scrutinization and the level of classification (e.g. for land cover, "vegetation" vs. "mountain needle leaf evergreen"; or for phenology, "high NDVI" vs. "long high 'green' period with relatively high NDVI during senescent months"). Common classification schemes identify spatial phenomena with a single variable or dimension, using labels such as "grassland," "forest," or "desert."

Classification software programs, on the other hand, can identify patterns in multiple dimensions. Along with spatial patterns, which are defined by proximal pixels of similar value, images can be classified by spectral and temporal patterns. Multi-spectral and hyper-spectral sensors collect image data in either a few (multi-spectral) spectral bands or hundreds (hyper-spectral) of spectral bands. Each pixel represents a vector of data defined by the surface reflectance in each of the sensor's bands. Similar to multi- and hyper-spectral data, each pixel in a time series of image data represent a vector of data (See figure 3).

Unsupervised classification algorithms determine "best" representation of each of k classes and then assign a label to every data point in the input space. The concept of "best" representation of a class has multiple interpretations. The class means representation is most common. Values from the input space are evaluated for similarity against the mean value of data points in each class. Input data are labeled as being members of the class to which they are most similar (similarity defined as smallest difference according to distance measure, D, used).

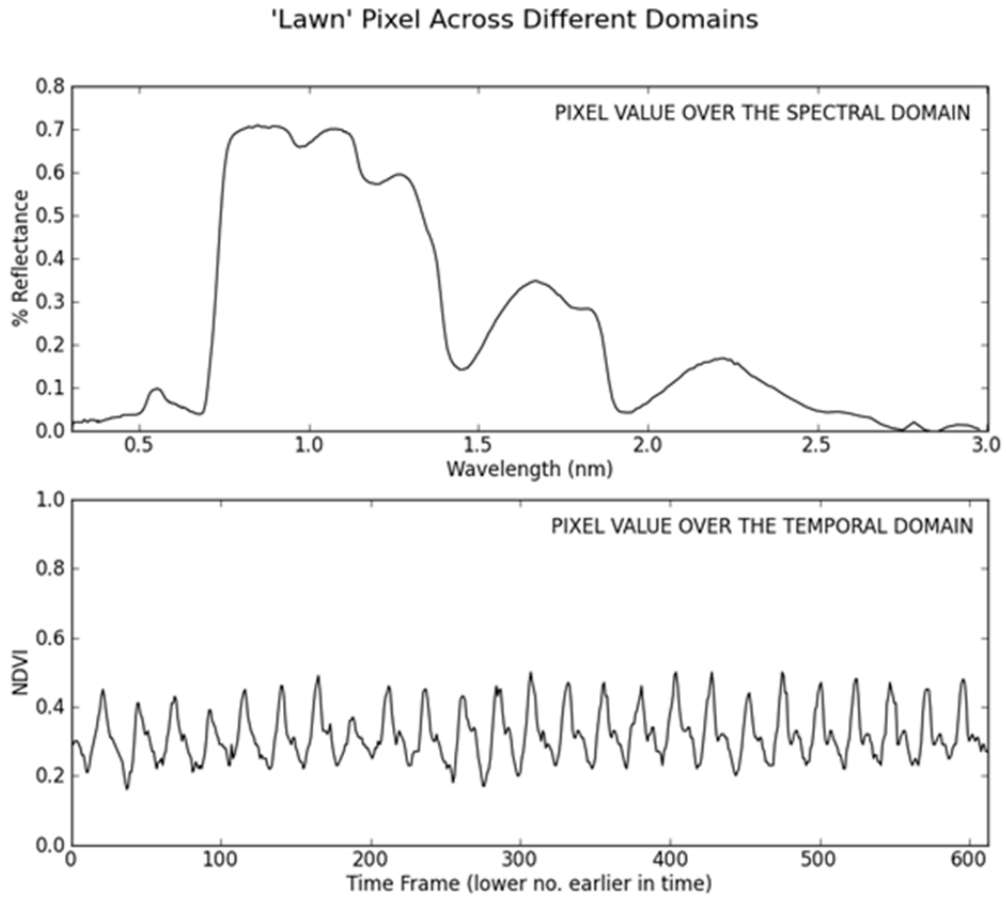


Figure 3 Possible values for a 'lawn' pixel over the spectral (top) and temporal (bottom) domains. Each time frame on the bottom chart represents a half-month time step, taken from time series AVHRR NDVI. Spectral data have unique reflectance values throughout the electromagnetic spectrum allowing classification to group similar spectral curves. Time series data likewise exhibits unique reflectance values throughout a temporal continuum allowing for classification of similar phenological curves. Spectral curve data obtained from the USGS Digital Spectral Library splib06a (Clark et al. 2007).

Another common "best" representation technique is class medoids where, instead of evaluating against the class mean, input values are evaluated against the data vector in the class most similar to the class mean. Many other interpretations of "best" representation exist to satisfy particular research objectives.

In remote sensing applications, each pixel (or pixel vector for time series) is an input data point. During the classification process, each data point is labeled as belonging to a class. The result is a map of nominal data that exhibits the spatial distribution of the computed classes.

2.1.4 ISODATA

The most common classification algorithm adopted in the field of remote sensing is ISODATA developed by Ball and Hall (1965) and commonly referenced in Tou and Gonzales (1974). ISODATA stands for Iterative Self-Organizing Data Analysis Techniques A, suggesting to the user that the process repeats until a threshold or convergence value is met. ISODATA uses a set of heuristic rules to cluster homogenous data based on the internal structure of the data in question.

The ISODATA algorithm is outlined in the following steps (generalized from Tau and Gonzales (1974)):

1. Specify *k* parameter, number of iterations allowed, and other parameters. These vary depending on software packages used.
2. Assign starting class mean vectors.
3. Distribute all data among the classes based on distance to class mean vector. Split or merge classes as needed.
4. Calculate the new class mean vectors from the class membership.
5. Calculate average distance of class members to their class mean vector and the standard deviation for each class membership.
6. Find the class member with the maximum distance to the class mean vector.
7. Split class if calculated values are within a relative threshold.
8. Calculate distance between all class mean vectors.

9. Group classes within a threshold distance.
10. If finished iterating through the data, save class membership to file. Otherwise, repeat process from step 3.

As implemented by Ball and Hall (1965) and as designed in most modern GIS software packages, ISODATA uses Euclidean distance to measure distance (as discussed previously). An illustration of how ISODATA uses Euclidean distance is instructive at this point.

Given a simple one layer raster, the distance from each cell's value to each class mean value is calculated using simple Euclidean distance. For example, if a cell has a value of 30 and the class mean has a value of 45, the Euclidean distance is calculated as $(|30\text{-}45|^2)^{1/2} = |\text{-}15| = 15$. For a three date time series pixel vector with values of [40, 47, 50] and class mean vector of [35, 38, 46] the Euclidean distance is calculated as $(|40\text{-}35|^2 + |47\text{-}38|^2 + |50\text{-}46|^2)^{1/2} = (25 + 81 + 16)^{1/2} = 11.05$. Note that the Euclidean distance for vectors is the same as calculating the norm of the vector produced from differencing the class mean vector (x) and the pixel vector (y), and is mathematically represented as:

$$D_{eucl} = \|x - y\|. \qquad [3]$$

2.1.5 EVALUATION

Methods for evaluating the effectiveness of image classification vary greatly between supervised and unsupervised classification. Available reference data allows supervised classification to be performed and also allows follow-up accuracy assessment to measure how well the classification algorithm assigned pixels to known classes. Time series NDVI do not have corresponding reference data. Therefore, well known accuracy assessment methods such as computing Cohen's kappa coefficient (Cohen 1960) or creating an error matrix (Jensen 2005) are

impossible to use for unsupervised classification of time series NDVI. The process for creating reference data for time series NDVI is undefined.

Cluster validation is used instead of accuracy assessment for unsupervised classification. The goal of cluster validation is to answer the question: did the classification algorithm identify the structure and number of inherent clusters, or classes, in the input data? Methods of answering this question use indices of class cohesion and class separation. Class cohesion describes the compactness of the class members about the class center, while class separation measures the uniqueness of each class by measuring the distance between classes. Researchers have developed many cluster validation indices (CVIs) to manage various scenarios (Arbelaitz et al. 2012).

Since the early days of computational clustering, scholars have demonstrated the need for *a posteriori* evaluation of partitions proposed by classification algorithms. In presenting ISODATA, Ball and Hall suggested that "some evaluation of the significance of a given clustering is possible" based on metrics derived from the clustered data (1965, 51). Tou and Gonzales (1974) provided examples of basic cluster evaluation and pointed to the use of CVIs to guide *k* parameter input. Nguyen et al. (2012) used a CVI called the *Divergence Index* to determine the optimal number of classes, *k*, inherent in MODIS time series NDVI data. Arbelaitz et al. (2013) compared thirty CVIs for performance under varied parameters. Although few indices consistently performed well, the *Silhouette Index* developed by Rousseeuw (1987) is worth noting because it achieved noticeably good results throughout all tests.

Nguyen et al. (2012) used the *Divergence Index* defined by Swain and Davis (1978). The index uses class Signature files to calculate the separability between classes. Signature files

contain a mean vector and covariance matrix for every class in a classification scheme. These values are calculated using the class assignments and the original data.

The Transformed Divergence index (Swain and Davis 1978) is calculated as:

$$TD_{avg} = \sum_{i=1}^{m} \sum_{j=1}^{m} p(\omega_i) p(\omega_j) TD_{ij} \quad [4]$$

where,

$$TD_{ij} = 2000 \left(1 - \exp\left(\frac{-D_{ij}}{8}\right)\right) \text{ (ERDAS 2011)} \quad [5]$$

and

$$D_{ij} = \frac{1}{2} tr\left[\left(\sum_i - \sum_j\right)\left(\sum_j^{-1} - \sum_i^{-1}\right)\right] + \frac{1}{2} tr\left[\left(\sum_i^{-1} + \sum_j^{-1}\right)\left(U_i - U_j\right)\left(U_i - U_j\right)^T\right] \quad [6]$$

Elements of equations 3, 4, and 5 are:

1. TD_{ij} – transformed divergence between classes *i* and *j*
2. $p(\omega_i)$ – *a priori* probability of class membership; equal to m^{-1}
3. m – total number of classes
4. D_{ij} – the divergence between classes *i* and *j*
5. $tr[x]$ – the trace of matrix *x*; calculated as the sum of the elements on the diagonal of *x*
6. $\sum_i$ - the covariance matrix of class *i*
7. U_i – the mean vector of class *i*
8. T – the transpose function

Mean vectors and covariance matrices are obtained from class signature files produced during the classification process.

It is important to note that the minimum Transformed Divergence serves as a complimentary measure to the average Transformed Divergence when evaluating classification schemes. Minimum Transformed Divergence is defined as the minimum Transformed Divergence

measurement between a pair of classes in the set of all class pairs as calculated in equation 4. The minimum Transformed Divergence measures how well each classification scheme separates the most closely related classes. A relatively low value indicates that the closest classes are not well separated, while a relatively high value gives the analyst confidence in the classification scheme. In general, the more classes the data are divided into, the lower the minimum Transformed Divergence. This is due to the fact that as more class means populate in the measurement space, the average distance between classes is reduced. Because of this phenomenon, the analyst must balance the increasing average Transformed Divergence against the decreasing minimum Transformed Divergence to select an optimal classification scheme.

2.2 IMAGE SEGMENTATION

Image segmentation is the process of decomposing an image into homogenous regions called segments. This technology is related to the field of cognitive psychology or the science of understanding how humans interpret the world through sensory perception. As mentioned in the introduction, people are very efficient remote sensors and can accurately and rapidly identify objects. The eyes transmit information to the brain which unconsciously filters noise and applies boundaries and categories to objects. Computer image segmentation mimics this ability by recognizing homogeneity based on proximal data values of similar magnitude and applying boundaries around homogenous areas to create image objects (Baatz and Schäpe 2000).

2.2.1 HOMOGENEITY AND HETEROGENEITY

Image segmentation relies on the ability of the computer to measure contextual information about each pixel in an image and its adjacent pixels. There are two common approaches to image segmentation: top-down and bottom-up. A top-down approach creates an initial segment, which includes all pixels of an image, and subsequently divides the image into

regions and checks for homogeneity. This process is repeated until a threshold homogeneity value is reached. The bottom-up approach starts with pixel-sized objects and merges objects based on homogeneity.

The common implementation of a bottom-up approach is multiresolution segmentation introduced by Baatz & Schäpe (2000). To maximize the homogeneity within segments *and* maximize the heterogeneity between segments, they developed mathematical measures and applied them to image objects. Homogeneity h is defined as:

$$h = \sqrt{\sum_d \left(\frac{f_{1d} - f_{2d}}{\sigma_{fd}}\right)} \qquad [7]$$

f_{1d} is the feature value for segment 1in dimension d, and σ is the standard deviation feature f for all segments in dimension d. Feature f is an arbitrary metric that can be derived from an image object. For instance, f could be the mean of all the pixel values in a segment or it could be the variance of all the pixel values in a segment.

To minimize the heterogeneity within segments, the multiresolution segmentation algorithm performs a merge of two image objects that produces the least change of heterogeneity, h_{diff}, among possible merges. This is measured by the equation:

$$h_{diff} = \sum_d w_d\left(n_1(h_{md} - h_{1d}) + n_2(h_{md} - h_{2d})\right) \qquad [8]$$

h_{1c} is the homogeneity of object 1 in dimension d, h_{md} is the homogeneity of segments 1 and 2 after a virtual merge, n_1 is the size of object 1, and w_d is the weight of dimension d. The virtual merge is a "*what if*" condition: *what* will the homogeneity be *if* segment 1 and 2 are merged? The "*what if*" scenario is repeated to calculate h_{diff} for all pairs of adjacent segments, and the minimum value is chosen to define the segment merge.

Segment form attributes such as shape can also be accounted for during the segmentation process. Adding weight to the form of segments lessens the influence of spectral homogeneity in

the segmentation process. Inclusion of segment form during segmentation is useful to morphologically identify features when there are weak spectral differences between the features or when features may be too influenced by spectral heterogeneity to accurately identify features in noisy imagery.

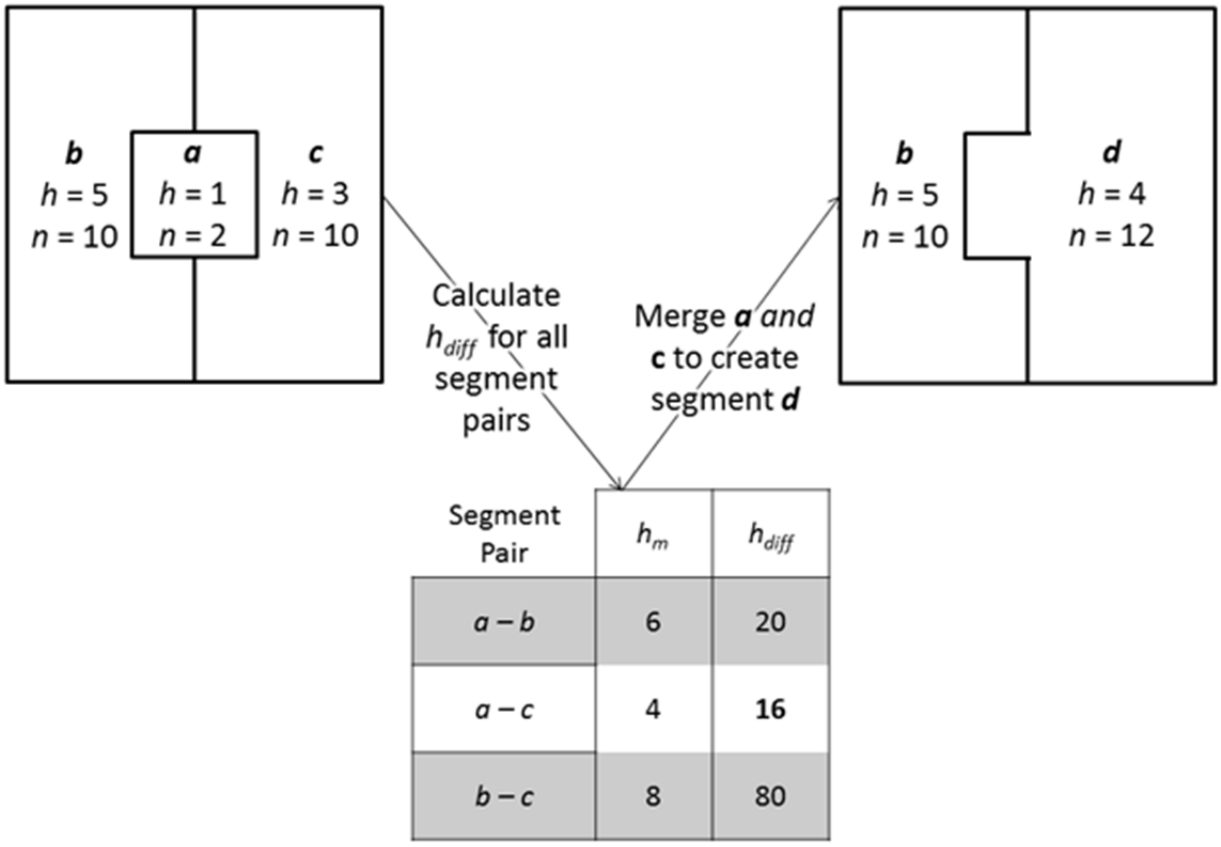


Figure 4 Satellite imagery can be segmented to create image objects bounding neighboring homogeneous pixels. This example illustrates the segment merge decision criteria: minimize potential increase in heterogeneity within a segment.

2.2.2 USE CONSIDERATIONS

Image segmentation can effectively divide an image into image objects that correspond to real world objects visible in the segmented scene (Corcoran and Winstanley 2007). Many

analyses that used image segmentation focused on extracting land use/land cover classes from basic grayscale or multi-spectral imagery where *a priori* class labels were manually assigned by the analyst during the segmentation process. For example, Corcoran and Winstanley (2007) used grayscale imagery of a suburban residential area to compare human-created vs. computer-created image segments. Volunteer test subjects were instructed to draw boundaries around objects in the image, which included streets, vehicles, houses, etc. The optimal computer segmentation strategy produced much smaller segments on average than the human-made segments. This indicates that the computer segments are relatively conservative in their approximation of real-world objects.

The concept of image objects can easily be expanded to many dimensions. Although humans perceive objects in the visible electromagnetic spectrum at a single point of time represented as three components (RGB or IHS), computers can interpret image objects in an unlimited set of components. For example, instead of inputting a single grayscale image, the computer could input a scene with ten co-registered grayscale images collected throughout a year. Segments created from this scene would delineate areas of homogenous temporal spectral value across all layers. Thus, researchers can use the image segmentation process to explore data that is not easily visualized.

Image segmentation software commonly implements functions to increase the fit of segments to real world objects. These functions include sub-segmenting and merging. Sub-segmenting allows researchers to take larger image segments where large areas of homogeneity exist and isolate, into smaller segments, the image segments that contain smaller heterogeneous regions. Merging functions work to select small segments which are homogenous according to

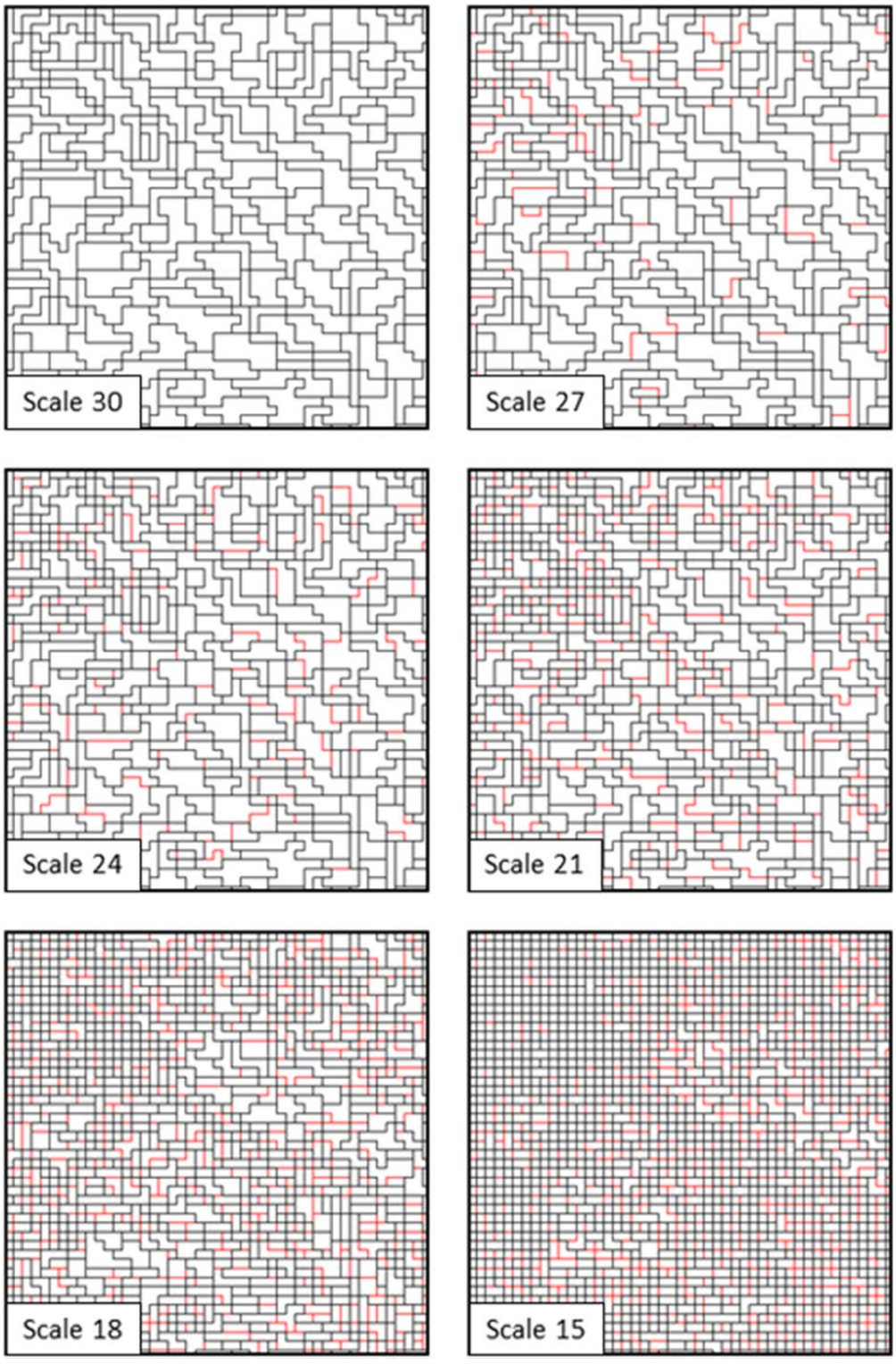


Figure 5 Six segmentation scales for time series AVHRR NDVI. Red lines indicated segment boundaries added from previous (larger) segmentation scheme. Scale 15 (bottom right) is comprised of mostly pixel-level segments which cannot be further subdivided. Image segmentation repeated for multiple scales reveals the optimal segmentation scale.

an expert defined threshold, then merge them into a new larger segment. These processes demonstrate the control the analyst is given over the segmenting process.

Because image segmentation groups homogenous values, it effectively reduces the amount of data needed to represent an image. A similar concept involves the resampling of raster data to a larger cell size. Instead of performing a uniform gridded merging of cells, image segmentation merges cells according to value homogeneity. The mean value of the merged pixels can be used to represent the new, larger area segment. Reduced raster data appear as s segments instead of p pixels, where s is always less than p.

2.3 NDVI TIME SERIES

This research relied on the ability of multi-spectral remote sensing systems to record the physical interaction of insolation with green-leafed vegetation. Reflection spectra of healthy green-leafed vegetation display relatively weak red reflectance (ρ_{red}; due to internal chlorophyll absorption) and strong infrared reflectance (ρ_{NIR}; due to the spongy mesophyll of a plant cell) of incident solar radiation (Jensen 2007). The normalized difference vegetation index (Deering et al. 1975; Tucker 1979) takes advantage of this phenomenon to measure vegetation health as proxy for in situ observation. NDVI is calculated as:

$$NDVI = \frac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + \rho_{red}} \quad [9]$$

Variables ρ_{NIR} and ρ_{red} correspond to AVHRR channels or MODIS bands 2 and 1, respectively.

Valid NDVI values range from -1 to 1, with healthy green vegetation approaching a value of 1 and senescent vegetation closer to a value of 0. Surfaces exhibiting reflected spectra higher in the red than NIR (e.g. water) calculate to negative NDVI, between -1 and 0. In sparsely vegetated arid environments, soils and/or rock that are visible through the vegetation structure

can adversely affect NDVI (Kouchoukos et al. 1998). In relatively small spatial scale time series imagery, researchers found minor negative impacts from background material (Weiss, Marsh, and Pfirman 2001). Schmidt and Karnieli (2000) found that AVHRR NDVI data were generally good indicators of vegetation in arid environments, especially when spatial separation exists between land cover types (annuals, perennials, biogenic crusts/lichens, etc.); the effectiveness decreases as land cover types are mixed (e.g. when low ground vegetation are covered by taller vegetation canopy throughout the course of the year).

2.3.1 CLASSIFICATION

Phenology is indirectly measured through indices of time series remotely sensed data such as NDVI or Enhanced Vegetation Index (EVI). Various researchers have used these indices to classify vegetation into phenological groups. Grouping similar phenological values aids creation of classified maps of agro-climatological regions.

The natural phenological cycle, visible with the naked eye in the seasonal change of vegetation, produces a seasonal oscillation (often referred to as a signal, curve, or profile) in the NDVI calculated from data throughout a year (the lower pane of figure 3 shows the phenological cycle for twenty-five consecutive years). Time series NDVI data spanning the length of a year can serve as a valuable proxy for plant phenology (Geerken 2009). These data are formed by a consistent spatial extent represented in a series of rasterized NDVI layers. Each represents a specific snapshot or aggregation in time. For example, if data are collected daily and composited twice a month, the time series would consist of twenty four (twice a month for twelve months) separate layers where the pixels, in corresponding locations throughout the layers of the time series, reveal the temporal NDVI curve (see figure 2). Inter-annual comparison can occur if time

series data is collected for multiple years and is processed to maintain internal consistency (DeFries, Hansen, and Townshend 1995).

Time series NDVI data are used as input data for classification algorithms to create data partitions. Each input pixel corresponds to pixels in the same spatial location on each of the images in the time series. Therefore, input to the classification algorithm is a vector (array) of data in *n* dimensions where *n* is the number of images in the time series. Each NDVI vector plotted through time reveals the phenological curve at each pixel location. Using a similarity measure, the classification algorithm groups and then assigns class membership labels to pixel vectors of similar structure. The output map indicates to which group each pixel vector was assigned.

Authors have used various names to describe the classes resulting from this process. Benedetti et al. (1994) used the term "agro-climatological" to describe the influence of climatological variables such as precipitation, temperature, and solar radiation on the health of vegetation throughout the time series. Kouchoukos (2001) applied "agro-ecological" to a seven class map. He proceeded to manually interpret the classes based on NDVI curve shape and correspondence to ground phenomena. White et al. (2005) and Hargrove et al. (2009) adopted the term "phenoregion" to describe the classification of NDVI data combined with climate data. Gu et al. (2010) called NDVI time series classes "pheno-classes," which reflects the cyclical nature of the data being classified. Commonly, classification software automatically assigns numerical labels to classes because descriptive labels become cumbersome when describing many small differences between NDVI time series vectors. Classification of NDVI time series vectors is therein limited in its comparison to land cover/land use classification.

2.3.2 CLASSIFICATION ISSUES

Several problems face an effective classification of time series. First, noise in a time series masks the expected value. Noise in a dataset is manifest by erroneous values and may be introduced to a dataset at the sensor or during image pre-processing. During classification, an erroneous value will make a corresponding pixel vector classify differently than an accurate value. Consequently, the partitioning scheme for the entire time series dataset would be affected. Readers are recommended to review Hird and McDermid (2009) to explore issues and proposed algorithms to solve problems arising from noise in time series data.

Second, sampling at irregular intervals and uneven lengths of time series preclude classification due to temporally varying NDVI properties. This concern may comfortably be ignored when working with time series imagery from satellite platforms. All pixels in an image are temporally comparable because each image was collected at the same time. Since each image in a time series must contain a value (possibly "No Data") for each of its pixels, uneven pixel vector lengths are unlikely to occur.

Other concerns are relevant when using statistical methods with assumptions about the structure of the data. Two of these assumptions specific to time series data are temporal serial-correlation (or commonly called temporal autocorrelation) and non-stationarity. These assumptions are not addressed in this work; however readers should note their constraints when performing statistical analysis with time series data.

3. METHODS

3.1 STUDY AREA

The study area is the northern Fertile Crescent, which covers the area extending longitudinally from 28˚ to 50˚ east with a latitudinal span from 28˚ to 42˚ north. The northern Fertile Crescent derives its name from an arc-shaped agriculturally productive zone trending east-west with a central northern apex and southward-bowed ends.

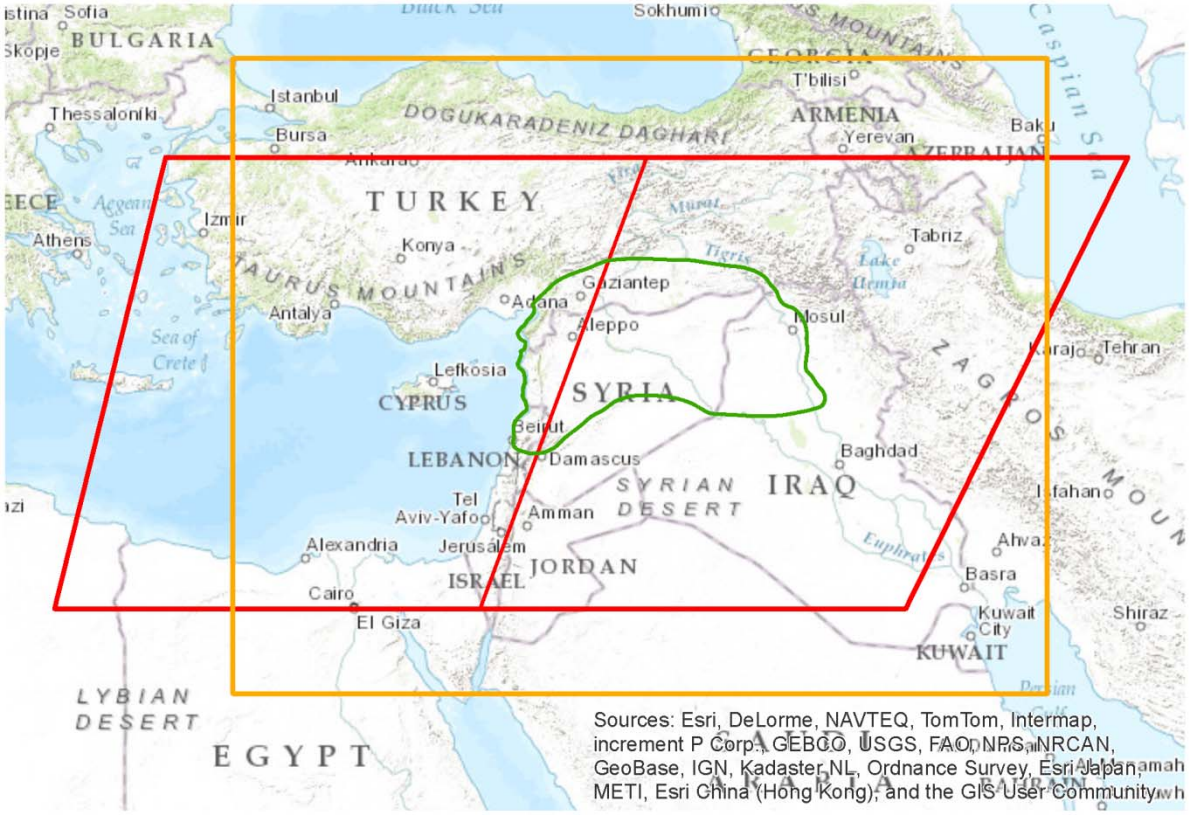


Figure 6 The northern Fertile Crescent study area is outlined in green, the boundary extent of the AVHRR data is in orange, and the boundary extent of the mosaicked MODIS data is in red (columns h20 on left, h21 on right). Esri provided basemap image.

A principle feature of the region is a strong north-south precipitation gradient with high yearly totals in the north and minuscule totals in the south (Hole & Zaitchik 2007). A strong natural vegetation gradient also coincides with the climate gradient. Dense green vegetation is

fed by the high precipitation in the north, while the low precipitation south is dominated by desert and meager scrubland (Beaumont 1996).

Over the timeframe of the data used in this study, widespread changes in terms of agricultural production and water infrastructure occurred in the Fertile Crescent. Large-scale development projects altered traditional dry farming methods by stabilizing the availability of irrigation water. Turkey's Southeastern Anatolia Project (Güneydoğu Anadolu Projesi, GAP) introduced dams and canals designed to increase total agricultural production in southeastern Turkey.

Beaumont (1996) and Hole and Zaitchik (2007) explored the use of remote sensing to monitor changes resulting from the projects. Midway through the study period, Beaumont (1996) analyzed the state of agriculture and water resources. He predicted unsettling consequences for downstream neighbors of the increased irrigated agricultural lands because of the 1) reduced flow of water in existing rivers and 2) increased pollutant runoff from agricultural activities such as fertilizer, pesticides, and other chemicals.

Hole and Zaitchik (2007) followed up Beaumont's (1996) work by reporting observed consequences from the heavy agricultural development. They noted that areas outside of the traditional limits of dry farming were being used extensively for agriculture through the help of irrigation canals and new groundwater wells. In other locations, the arable land area dropped to levels which could not maintain the historic area of farmed land. The reader is encouraged to review Beaumont (1996) and Hole and Zaitchik (2007) for a full explanation of the consequences of the noted development projects. Recent research confirms that water resources were taxed at unsustainable levels during the years following implementation of the development projects (Voss et al. 2013).

3.2 REMOTE SENSOR DATA

3.2.1 AVHRR

This study used the Global Inventory Modeling and Mapping Studies (GIMMS) AVHRR time series NDVI data from the Global Land Cover Facility (GLCF) at the University of Maryland (Tucker, Grant, and Dykstra 2004). GIMMS data were designed to be consistent and comparable throughout 25 years (1981-2006) and between newer earth observation systems such as SPOT-4 Vegetation instrument and Moderate Resolution Imaging Spectroradiometer (MODIS) (Tucker et al. 2005; Brown et al. 2006). Data were obtained from six different AVHRR instruments and identified by the satellite on which they were mounted (NOAA-6, NOAA-7, NOAA-9, NOAA-11, NOAA-14, and NOAA-16). NDVI was calculated per scene with the equation 9, where Channel 1 and Channel 2 are data from the red and infrared portions of the electromagnetic spectrum, respectively. Half-month MVCs (Holben 1986) with a spatial resolution of 8 × 8 km were calculated for all areas of the globe. The researchers applied corrective processing to the dataset in a temporally and spatially targeted manner to create a spatially seamless and temporally continuous product. The reader is referred to Tucker et al. (2005) for a detailed explanation of processing and manipulation techniques used in the creation of the GIMMS NDVI data.

NDVI data were stored in 16 bit integer raster with valid NDVI data range from -1,000 to 1,000. Water was given a value of -10,000 and masked pixels a value of -5,000. A corresponding flag raster was downloaded with the NDVI raster. The true NDVI values (scaled between -1 and 1) were obtained using the equation:

$$NDVI = float\left(\frac{raw}{10{,}000}\right) \qquad [10]$$

The quality flag raster also stored 16 bit integer values that need to be converted to a usable data range of 0 to 6. A summary of the meaning of the flag values is given in table 1. The flag values are designed to be used to exclude data which fails to meet the standards of a given application. For this research, data values associated with flag values of 0 through 5 were used in the analysis. The flag values were obtained from the raw flag data using the equation:

$$Flag = raw - floor\left(\frac{float(raw)}{10}\right) \times 10 \qquad [11]$$

Flag Value	Description
0	Good value
1	Good value, possibly snow
2	NDVI retrieved from spline interpolation
3	NDVI retrieved from spline interpolation, possibly snow
4	NDVI retrieved from average seasonal profile
5	NDVI retrieved from average seasonal profile, possibly snow
6	Missing data

Table 1 Flag values for GIMMS AVHRR MVC NDVI.

Other time series AVHRR NDVI datasets are available and were considered for use in this research. Comparable datasets are: Pathfinder AVHRR Land (PAL) dataset developed by James and Kalluri (1994), Land Long Term Data Record (LTDR) version 3 (v3) developed by Pedelty et al. (2007), and Fourier-Adjustment, Solar Zenith Angle Corrected, Interpolated Reconstructed (FASIR) dataset developed by Los et al. (2000). Beck et al. (2011) compared the four datasets and found that each dataset excels in various applications as listed below. The PAL dataset showed a positive NDVI trend in desert region where NDVI should remain constant. The LTDRv3 dataset contained the lowest total potential atmospheric error and is applied properly when absolute NDVI values are used. The researchers found the FASIR dataset to be similar to the LTDRv3 dataset. Ultimately, they concluded that the GIMMS dataset was more consistent

throughout the time series than the other datasets and recommended GIMMS for use when detecting change or performing trend analysis.

3.2.2 MODIS

In this research, MODIS time series NDVI data were used in conjunction with AVHRR time series NDVI data. MODIS data were obtained for twelve years (2000-2012) from the USGS Land Processes Distributed Active Archive Center (LP DAAC) using NASA's Reverb web interface (http://reverb.echo.nasa.gov). The MODIS time series was processed in a similar manner to the AVHRR data. Individual scenes were examined for cloud and/or atmospheric contamination and poor sun-target-sensor geometry. Pixels without contamination or poor geometry were retained for further processing. MVC was then used to produce a 16 day composite. The full processing procedure is explained by Solano et al. (2010).

The downloadable scenes were multi-layer and conveyed the actual data values for NDVI and EVI along with pixel quality and other information. NDVI was stored as 16 bit integers with a range of -2,000 to 10,000. Actual NDVI was calculated from the stored values using the following equation:

$$NDVI = raw \times 0.0001 \qquad [12]$$

'NoData' values were presented as -3,000. Quality data were stored as 8 bit integers with valid range -1 to 3. A summary of data errors is given in table 2. Data values with corresponding error values of 0 through 2 were used in data analysis for this project.

Value	Summary QA	Description
-1	Fill/No Data	Not Processed
0	Good Data	Use with confidence
1	Marginal data	Useful, but look at other QA information
2	Snow/Ice	Target covered with snow/ice
3	Cloudy	Target not visible, covered with cloud

Table 2 Flag values for MODIS MVC NDVI.

3.2.3 LANDSAT AND OTHER HIGH RESOLUTION IMAGERY

Landsat Thematic Mapper (TM) scenes of the selected region within the study area, specifically from 1984 and 1985, were downloaded from NASA's Reverb web system. Additionally, recent date high resolution aerial and satellite imagery, which is available through Esri's ArcMap® (ESRI 2012; see http://goto.arcgisonline.com/maps/World_Imagery), were accessed to aide manual interpretation of classification results. Classification results influenced the location and timeframe of downloaded scenes. Five locations were selected according to interest relative to class differences. Locations included agricultural areas east of Al Assad Reservoir, areas in the Harran Plain, areas near Ceylanpınar, Turkey, and two adjacent locations between Aleppo, Syria and Al Assad Reservoir.

3.3 DATA REDUCTION

This research explored the differences between optimum classification of reduced-data mean year and reduced-data segment mean time series NDVI. The ISODATA algorithm was used to classify both mean year and segment mean time series using simple Euclidean distance to measure pixel vector similarity.

The AVHRR and MODIS data were processed independently from each other but shared common processing steps. Pre-processing steps were undertaken to arrange and focus the data on the Fertile Crescent. AVHRR data were downloaded as global scenes and clipped to bounds of 28˚ to 50˚ east and 28˚ to 42˚ north. Full-data AVHRR time series was comprised of 612 NDVI rasters corresponding to the 612 compositing periods of the covered twenty-five and a half years.

The MODIS data were downloaded as two adjoining tile scenes per acquisition date. The western scene covered column 20, row 5; the eastern scene covered column 21, row 5. For each

date, the two scenes were mosaicked using the ArcPy® (ESRI 2012) Python package. The MODIS full-data time series was comprised of 294 mosaicked NDVI rasters corresponding to 294 compositing periods covered in the twelve included years.

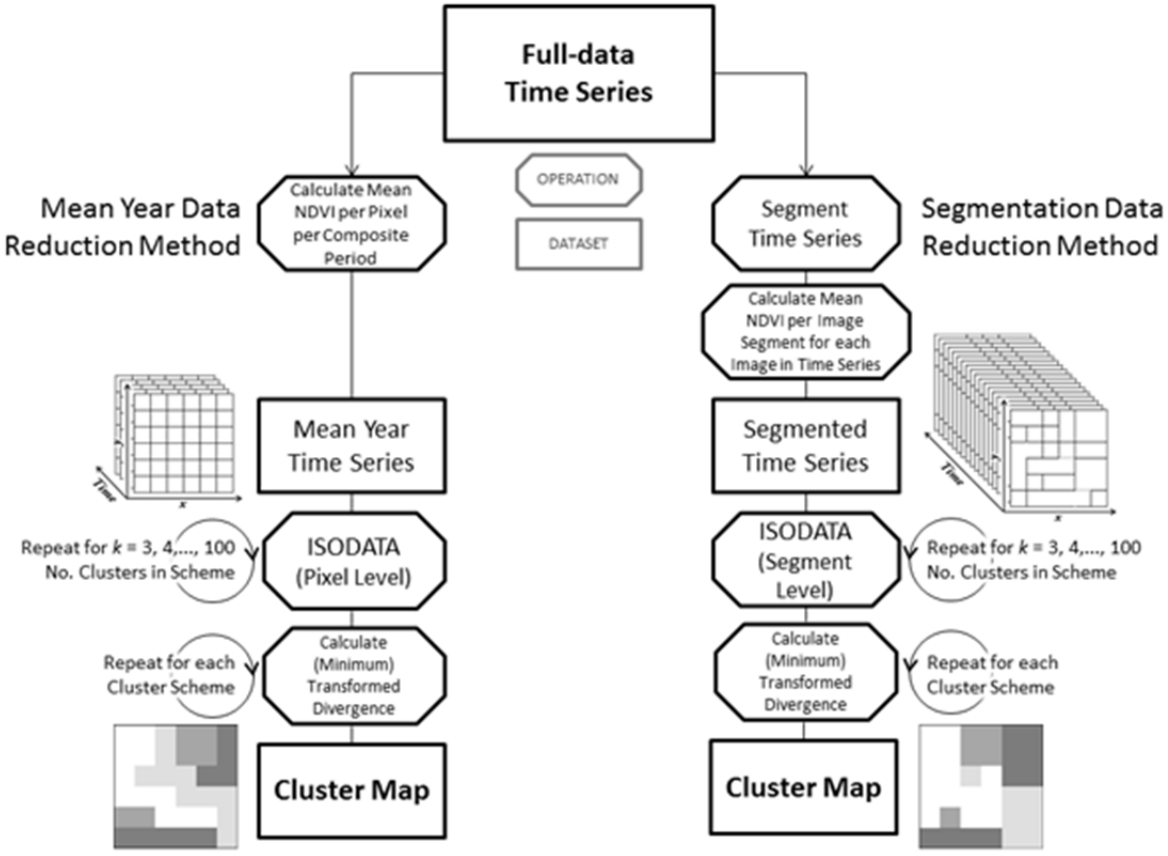


Figure 7 Process workflow comparison for the mean year data reduction method (left branch) and the segment mean data reduction method (right branch). The only difference is the method used to calculate the time series which is used as input to the ISODATA classification function. Both methods were executed for both AVHRR and MODIS data.

Raw data values were converted to actual NDVI using equation 7 for AVHRR. MODIS data were extracted from each associated downloaded HDF file and left in their scaled NDVI integer form to accommodate use of 16 bit depth rasters to reduce the memory cost during

processing. Actual quality flag values for AVHRR were calculated using equation 8. The MODIS quality data did not require any calculation as they were given in actual values.

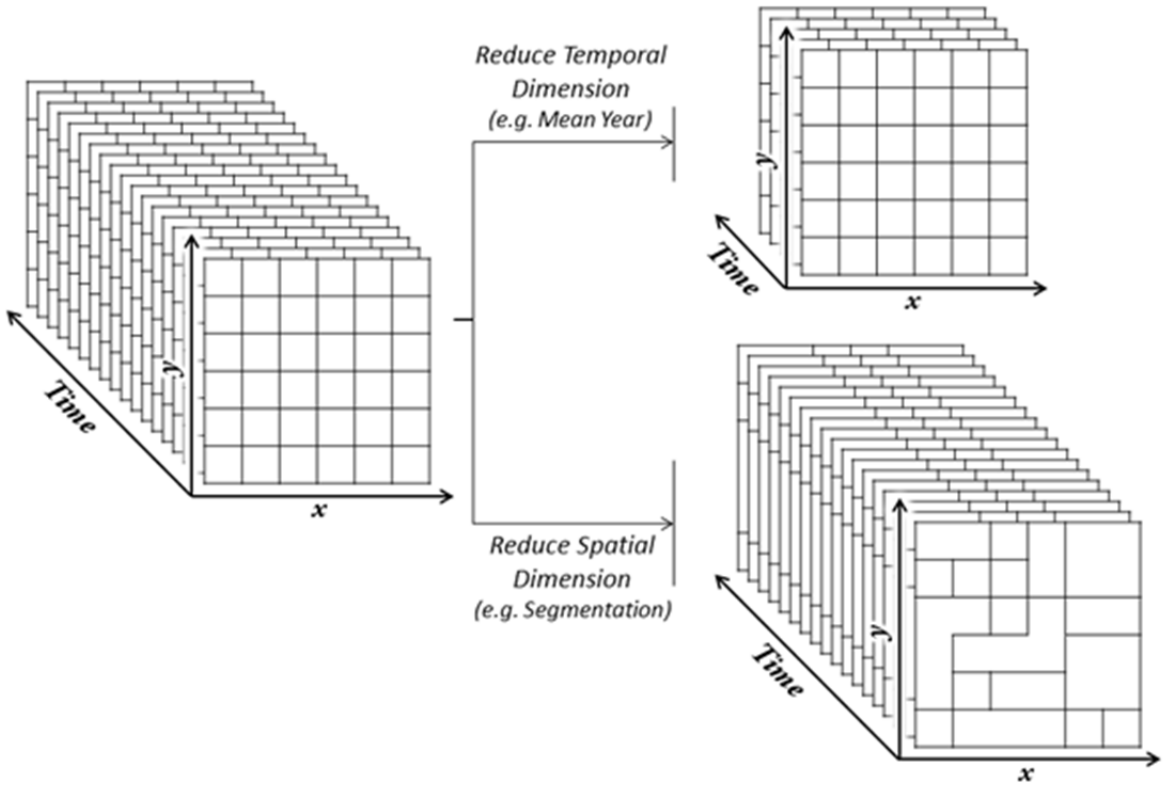


Figure 8 Conceptually, the mean year data reduction method reduces full-data time series through temporal aggregation (top) while the segment mean data reduction method reduces full-data time series using spatial aggregation (bottom).

3.3.1 MEAN YEAR METHOD

Reduced-data mean year time series were created from the AVHRR or MODIS data using the ArcPy® Python package. Pixels in scenes that had undesirable quality flag values were given a value of 'NoData' so as to not influence the calculated mean value. For each time series, corresponding time frames for each year in the time series were averaged to produce a 'mean year' time series. The mean year for the AVHRR time series consisted of 24 average NDVI

rasters corresponding to the 24 bi-monthly composites each year; the mean year consisted of 23 rasters for the MODIS time series since the compositing time of the MODIS data is every 16 days.

3.3.1.1 WRITTEN EXPLANATION

The process used to calculate the mean year time series was as follows:

1. Identify common compositing periods for each year (e.g. Jan 1-15, Jan 16-31… etc.).
2. Group all images in the time series by their compositing period.
3. Add cell-by-cell NDVI value for all images in a compositing period image group which have valid pixel values; valid pixels are pixels which are not associated with an undesirable flag value.
4. Divide cell-by-cell the summed NDVI by the total number of valid pixels that contributed to the summed NDVI to obtain a mean raster for each compositing period.
5. Repeat steps 3 and 4 for each compositing period image group.

The resulting set of mean compositing period rasters were temporally ordered into a mean year time series (24 layers for AVHRR and 23 layers for MODIS).

3.3.1.2 MATHEMATICAL EXPLANATION

Calculation of the mean year time series is represented by the following logic and equation. For all composite periods $c = 1, 2\ldots p$, p being the total number of compositing periods in a year, the mean of composite period c, $\bar{c}$, over y years is given as:

$$\bar{c} = \frac{\sum_{i=1}^{y} Raster_{i_c}}{Contrib.\ Raster_y} \qquad [13]$$

Element $Raster_{ic}$ is the i^{th} year raster in compositing period c. *Contrib. Raster*$_y$ is a raster which contains the number of valid contributions made to the summed raster, per cell. If, for each raster contributing to the sum, no cell is flagged as invalid, the *Contrib. Raster* takes the form of a uniform raster with each cell value equal to the number of years y in the original time series belonging to compositing period c. When pixels are flagged as invalid, the cell values in *Contrib. Raster* vary.

3.3.1.3 GRAPHICAL EXPLANATION

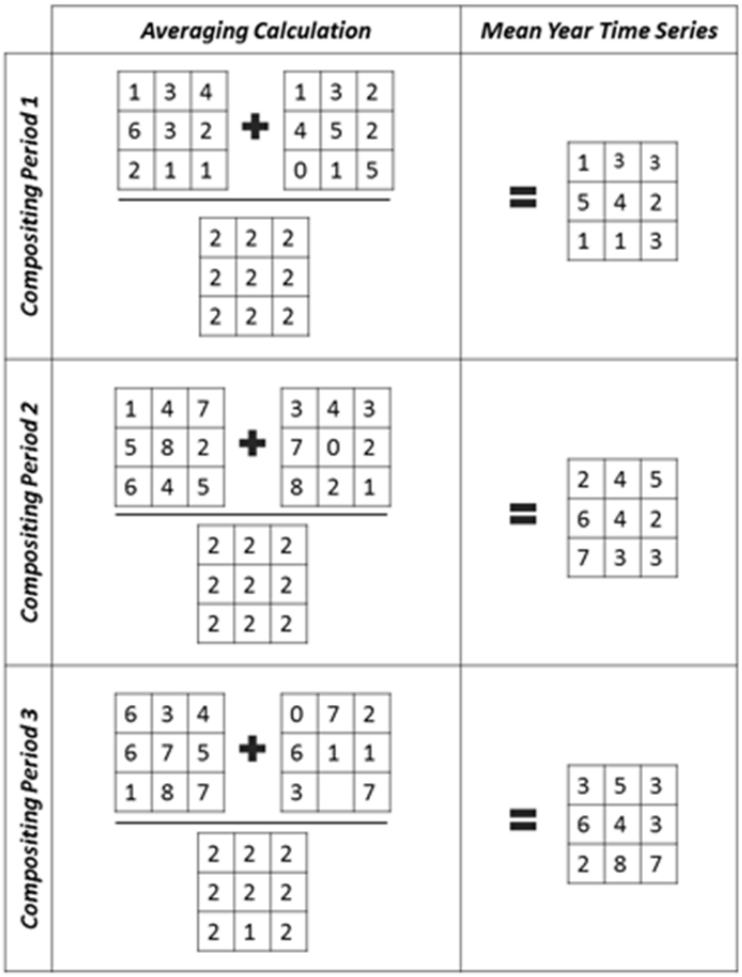


Figure 9 Mean year time series for AVHRR and MODIS were calculated like this example mean year time series calculation for a two year full-data time series with three compositing periods per year. The mean year time series is comprised of the same number of rasters as the number of images acquired in a year.

An example calculation of a mean year time series is given in figure 9. The example original time series is made up of two years with three compositing periods per year, which gives a total of six images in the full-data time series. Each raster is represented by a grid of cells with integer values representing per cell data value. Where data values are missing such as if an undesirable flag value had been associated with a cell value a blank cell appears (top right raster in first column of third row). The associated contributing raster reflects the missing value by only counting valid cells in each raster in the numerator. Each raster in the resulting mean year time series is shown in the right column.

3.3.2 SEGMENTATION METHOD

eCognition® (Trimble 2011) image segmentation software was used to segment both AVHRR and MODIS full-data time series to produce segment reduced-data. Due to the large amount of data existent in the MODIS time series, slightly different processing methods were used for AVHRR and MODIS data. Initially, both AVHRR and MODIS full-data time series were temporally ordered and loaded into eCognition®. Only for the MODIS data, eCognition® split each image in the time series into 600 × 600 pixel tiles to reduce the data input to the segmentation process. Multi-resolution segmentation was executed for the entire AVHRR time series and each MODIS time series tile using image pixels as input. Shape and compactness parameters were set to 0. The MODIS tiles were stitched together using a custom rule set that identified and re-segmented the tile border objects. This was done to remove superimposed linear segments introduced in the tiling phase (see figure 10).

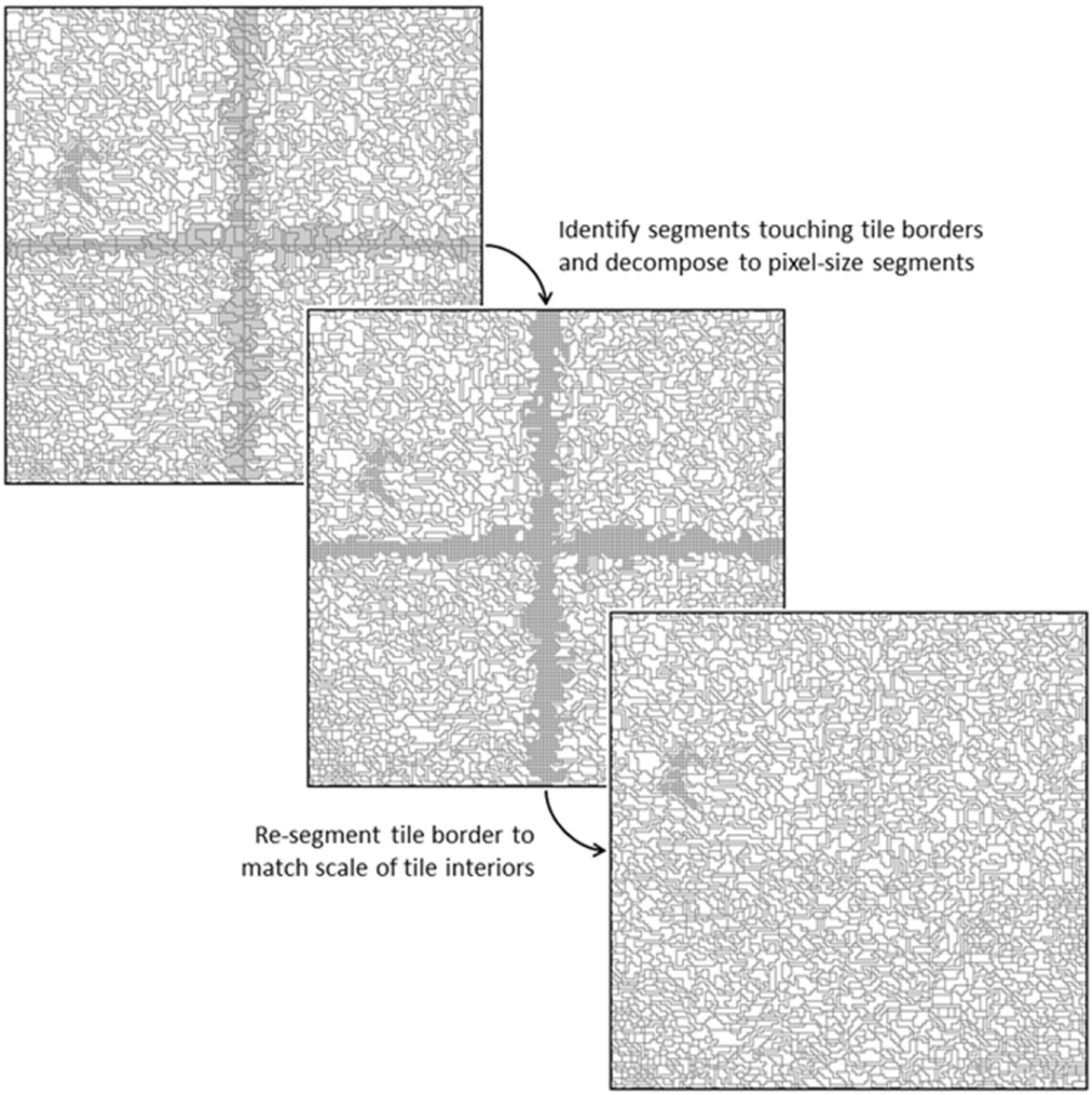


Figure 10 Segmented MODIS tiles were processed to remove linear segment boundaries caused by the tiling proceedure. Final segments exhibited no remnant of the tile borders.

Segmentation was repeated for scale parameter values of 0.05, 0.10, 0.15, 0.20, 0.25, and 0.30 for AVHRR data and 15, 18, 21, 24, 27, 30, 33, 36, 39, and 42 for MODIS data. The vastly different scale parameters for the different data types were used due to the range difference between the actual NDVI values for AVHRR (range -1…1, 32 bit floating point) and scaled

NDVI values for MODIS (range -2000…10000, 16 bit integer). The segments produced at each scale level were compared to the total number of pixels in a single date of imagery to determine the percent of data that was reduced and the average number of pixels per segment that were produced under each proposed segmentation scheme. Manual interpretation and selection of best segment scheme versus amount of data reduction was performed with help of the graphs displayed in figure 13.

3.3.1.1 WRITTEN EXPLANATION

Segments from the best segmenting scheme were exported from eCognition® in Esri shapefile format. Each segment was represented by a polygon that was at least the same size as a pixel. The polygons were used to create a series of spatially un-registered rasters according to the following description. A raster was created for each image in the full-data time series. Each new raster contained a cell for every segment in the shapefile. Each cell contained the mean value of the pixels bounded by a segment and calculated from the corresponding image in the full-data time series. The number of rasters per data set equaled the number of layers in the time series. Each raster was one column by *s* rows where *s* was the number of segments produced in the segmenting step. In terms of procedure, the following steps were taken:

1. Load all images in the time series into eCognition® and segment at a scale of 0.3 for AVHRR or 42 for MODIS with shape and compactness set to 0; export segments.
2. Repeat step 1 for scale parameter values of 0.05, 0.10, 0.15, 0.20, and 0.25 for AVHRR or 15, 18, 21, 24, 27, 30, 33, 36, and 39 for MODIS.
3. Determine best segmentation scheme by a combination of % data reduction and average segment size; subsequent steps apply to the best segmentation scheme only.
4. Calculate mean NDVI value per segment from the first layer in the original time series.

a. Add all pixel values which fall in each segment.
b. Divide by the number of valid pixels contributing to the summed value; valid pixels are defined as pixels not having an associated undesirable flag value.

5. Write the mean NDVI value per segment to a cell in a new raster of dimension 1 x *s*.
6. Save the new raster to file as a member of the segment mean time series.
7. Repeat steps 4 through 6 for all layers in the original time series.

3.3.1.2 MATHEMATICAL EXPLANATION

Mathematically, the mean segment time series is calculated using the following logic and equation. For each segment $i = 1, 2 \ldots s$, where s is the number of segments in the segmentation scheme, the mean of i, $\bar{\imath}$, is calculated using:

$$\bar{\imath} = \frac{\sum_{p=1}^{n} pixel_p}{n} \qquad [14]$$

The variable n is the number of pixels bounded by segment i and $pixel_p$ is the value of pixel p. Furthermore, this procedure is applied to each image in the original time series to obtain the segment reduced time series.

3.3.1.3 GRAPHICAL EXPLANATION

Figure 11 demonstrates the segment averaging calculation. The segment boundaries are represented by the leftmost object in the left column, which is labeled Segments. The segment boundaries are superimposed onto each raster from the full-data time series to identify which cells should be grouped as belonging to each segment. Each segment pixel group is summed then divided by the number of valid cells contributing to the summation. Valid cells are cells that have an NDVI value not masked by the flag data. An example of an invalid cell is given in the rightmost object in the left column, third row. The mean value calculated for the associated

segment excludes the invalid cell from the calculation. If all cells in a segment are invalid, the corresponding cell in the segment mean raster will have a value of 0. Another possible value is 'NoData' represented by the value 32767, which is the maximum positive integer using signed 16 bit storage.

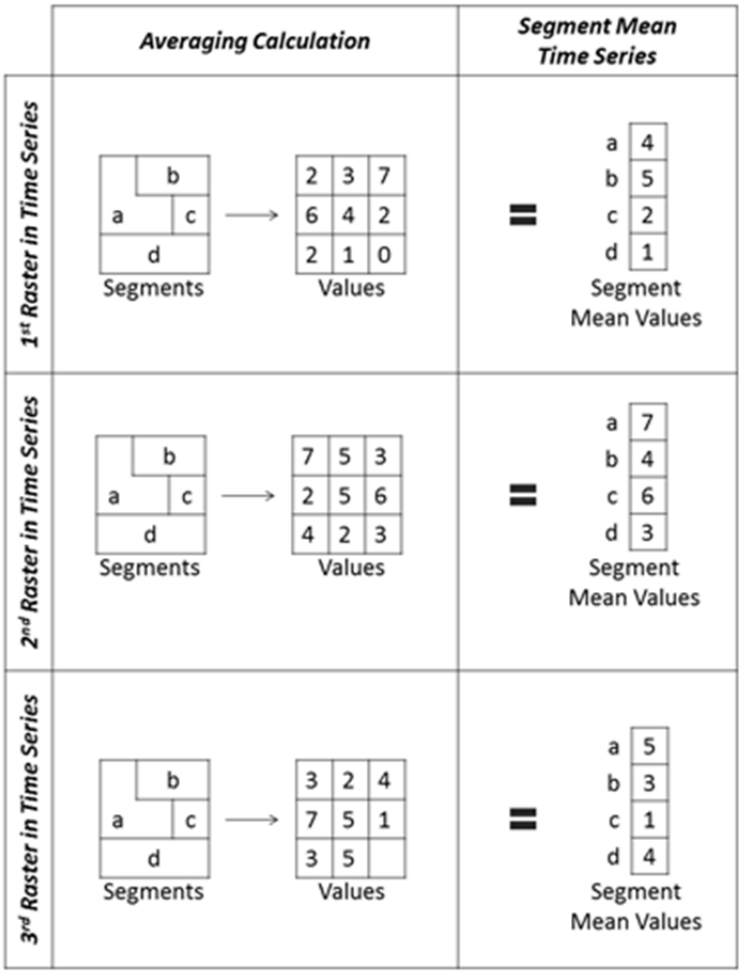


Figure 11 Segment mean reduced-data time series were calculated from AVHRR and MODIS full-data time series according to the scheme presented in this example segment mean time series calculation for a three raster full-data time series. The total number of rasters in each segment mean time series equaled the number of images in the full-data time series.

3.3 ISODATA CLASSIFICATION

Each raster in the mean year reduced-data raster set was arranged in temporal order to obtain the mean year time series. Each pixel location in the mean year time series represented an

NDVI vector spanning a single "mean" year. Layers were arranged so that the beginning of the calendar year was layer 1 and subsequent time frames were assigned ascending layer numbers for a total of 24 ordered images for AVHRR and 23 for MODIS.

Similarly, each raster in the segment mean reduced-data raster set was temporally ordered to produce the segment mean time series. Each pixel location in the segment mean time series represented the temporal trajectory vector of NDVI spanning 612 compositing periods for AVHRR and 294 compositing periods for MODIS. Layers in each data set were temporally arranged so the earliest compositing time and year were assigned to the first layer and subsequent time frames were assigned to ascending layers.

Mean year reduced-data and segment mean reduced-data time series for both AVHRR and MODIS were subjected to unsupervised classification using the ISODATA algorithm as implemented by ERDAS IMAGINE®. ISODATA was used to cluster similar pixel vectors and output a map of the class spatial distribution for each reduced-data time series for each sensor type (total of four maps).

3.3.1 K PARAMETER ADJUSTMENT AND CLUSTER VALIDITY

The optimal classification scheme for each reduced-data time series was identified by iterating through $k = 3, 4 \ldots 100$. Classification was repeated with unchanging parameters except for the k parameter, which suggested the number of classes into which ISODATA should partition the data. Maximum number of classes was limited at 100 to keep the number of classes relatively low to ease interpretation but large enough to extract spatially small classes. Each proposed classification scheme resulting from each different k was scrutinized using cluster validation techniques.

Cluster validity for each ISODATA execution was calculated using the Transformed Divergence index (Swain and Davis 1978) implemented using ERDAS IMAGINE® (ERDAS 2011). The best proposed classification scheme was determined manually using a combination of average Transformed Divergence and the minimum Transformed Divergence (see figure 12).

4. RESULTS AND DISCUSSION

4.1 MEAN YEAR REDUCED-DATA

The structure of mean year reduced-data is inflexible; mean year reduced-data will always consist of a time series with the same number of rasters as compositing periods in the corresponding full-data time series, regardless of years spanned in the full-data time series. This property trades flexibility of amount of data reduction (see section 4.2) for reduced complexity of computation of the reduced-data time series. Thus, the AVHRR time series, which spans 25.5 years, was reduced to 24 rasters, while the MODIS time series, which spans 12 years, was reduced to 23 rasters.

Theoretically, the information inherent in the mean year reduced-data AVHRR time series should reflect its longer span. However, the averaging calculation for the mean year method made retaining discreet contributions of the additional years impossible. Instead, mean year reduced-data were used to obtain the average value for a particular location over time. Another way of conceptualizing this is to think of each cell of a raster as being smoothed over the temporal dimension, which is an effective way of ignoring specific unique values in the full-data time series.

4.1.1 TRANSFORMED DIVERGENCE

Optimal classification was determined from the calculated transformed divergence between classes. Figure 12 reveals that average Transformed Divergence remained relatively high in all classification schemes for both the AVHRR reduced mean time series and MODIS reduced mean time series. The minimum Transformed Divergence became useful in this scenario as selection criteria for better classification schemes. Local maxima of the calculated minimum transformed divergence indicated classification schemes that divided the data relatively well. While the general trend of minimum Transformed Divergence for MODIS and AVHRR were similar, MODIS minimum Transformed Divergence averaged around two hundred points less than AVHRR on classification schemes with greater than 40 clusters. This phenomenon may be influenced by the ISODATA clustering not being able to converge on an ideal classification scheme for a large number of classes in the limited number of iterations allowed (6).

Comparison of minimum Transformed Divergence local maxima showed interesting similarities. Each had an early peak at 7 classes for AVHRR and 9 for MODIS; a peak after a precipitous drop at 21 classes for AVHRR and 20 for MODIS; and a local peak at 73 classes for both AVHRR and MODIS. The coincident peaks may indicate a natural division within the NDVI data. Subsequent analysis focused on these classification schemes.

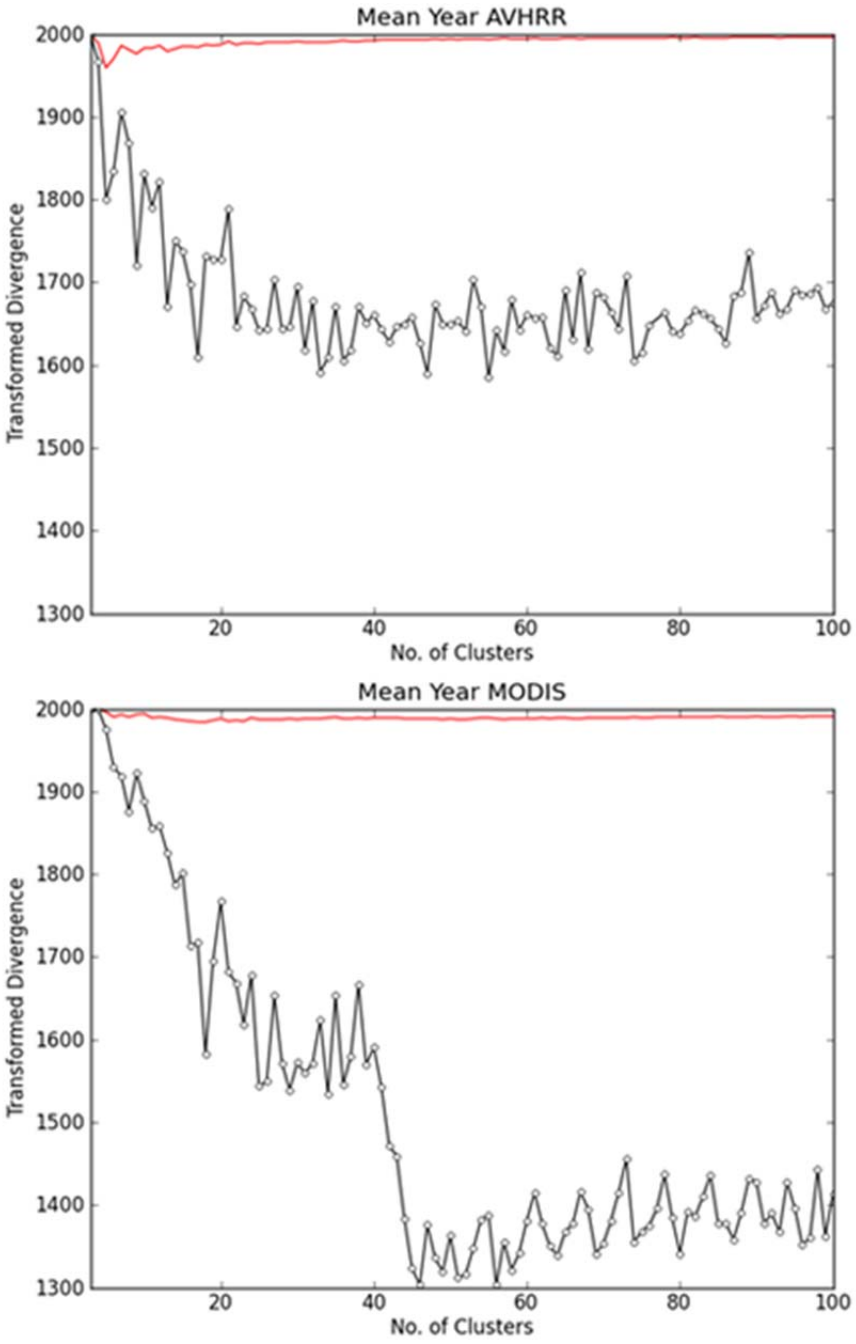


Figure 12 Transformed divergence (red) and minimum transformed divergence (black with circle markers) of possible classification schemes for mean year AVHRR (top) and MODIS (bottom) data were graphed to compare cluster schemes of AVHRR to MODIS and to determine local maximum minimum Transformed Divergence. Common local maximums occur around 7 to 9, 20 to 21, and 73 class (cluster) schemes.

Six of the seven classes from the seven class scheme for AVHRR intersected the Fertile Crescent area of interest. Simple visual inspection of the spatial distribution of the classes revealed coarse boundaries that generally mimicked isolines of the steep vegetation gradient. Of the twenty-one classes in the second scheme, seventeen intersected the area of interest. Seven of the seventeen classes contributed ten or less pixels to the area. The remaining ten classes represented large, compact sections within the area except one, which was fragmented and occurs on the periphery and near rivers. Finally, the seventy-three class scheme divided the area into forty-three classes with many of the included clusters only contributing a few cells to the area. Only eight of the forty-three classes were major contributors to the area, while another eleven to twelve classes were minor contributors, provoking further investigation. The remaining twenty-four classes were on the periphery of the area, comprising only a few cells, or were exclaves from classes with centers distant from the area of interest. The seventy-three class scheme was chosen as the optimal AVHRR classification map for the Northern Fertile Crescent, because the scheme captured the diversity of class structure while the classes can be reduced to ten to twenty-six important classes.

The seven class scheme for the MODIS mean year reduced-data time series was dominated by four classes that hid potentially interesting variation in the vegetation gradient. In the twenty class scheme, more variability was exposed and the resolution enhancement compared to AVHRR became apparent with the ability to pick up intra-segment variation that was diluted by the AVHRR. Class dispersion throughout the study area was dominated by four classes, while only four other classes provided appreciable data for analysis. Remaining classes contributed a low percentage of the area, were focused on water bodies within the area, or did not intersect the area. Lastly, the seventy-three class scheme was examined.

Eleven of the seventy-three classes were major contributors to the area of interest, and another thirteen contributed meaningful spatial segregation and information worth investigating further. Forty-seven other classes intersected the study area but were very minor contributors or extremely scattered throughout the area. Only two classes did not intersect the study area at all. The seventy-three classification scheme was chosen for further development of a MODIS cluster map of the northern Fertile Crescent due to the reasonable expectation of interpreting the twenty-four major and minor contributing classes and the high level of spatial clustering apparent within the area of interest.

4.2 SEGMENTED REDUCED-DATA

Figure 13 shows the percent data reduction and average segment size (in pixels) for the various segmentation schemes investigated. Selection strategy for the best segmentation scheme balanced the need to reduce the amount of data while maintaining appropriate spatial resolution. The application of this strategy varied between individual time series. Average segment size in pixels weighed more heavily than percent data reduction for AVHRR data because of the already coarse spatial resolution. Accordingly, candidate segmentation schemes were limited to schemes with segmentation scales less than 0.30 because schemes 0.30 and above reduced spatial resolution too much for the AVHRR data even though improved data reduction occurs. Visual review of the segments for each remaining candidate segmentation scheme was required to identify differences in segments of the segmentation schemes. Scale 0.15 is selected as best segmentation scheme because it effectively segmented the area of interest while showing improved data reduction over the next scheme, scale 0.10, which had a similar segment size value.

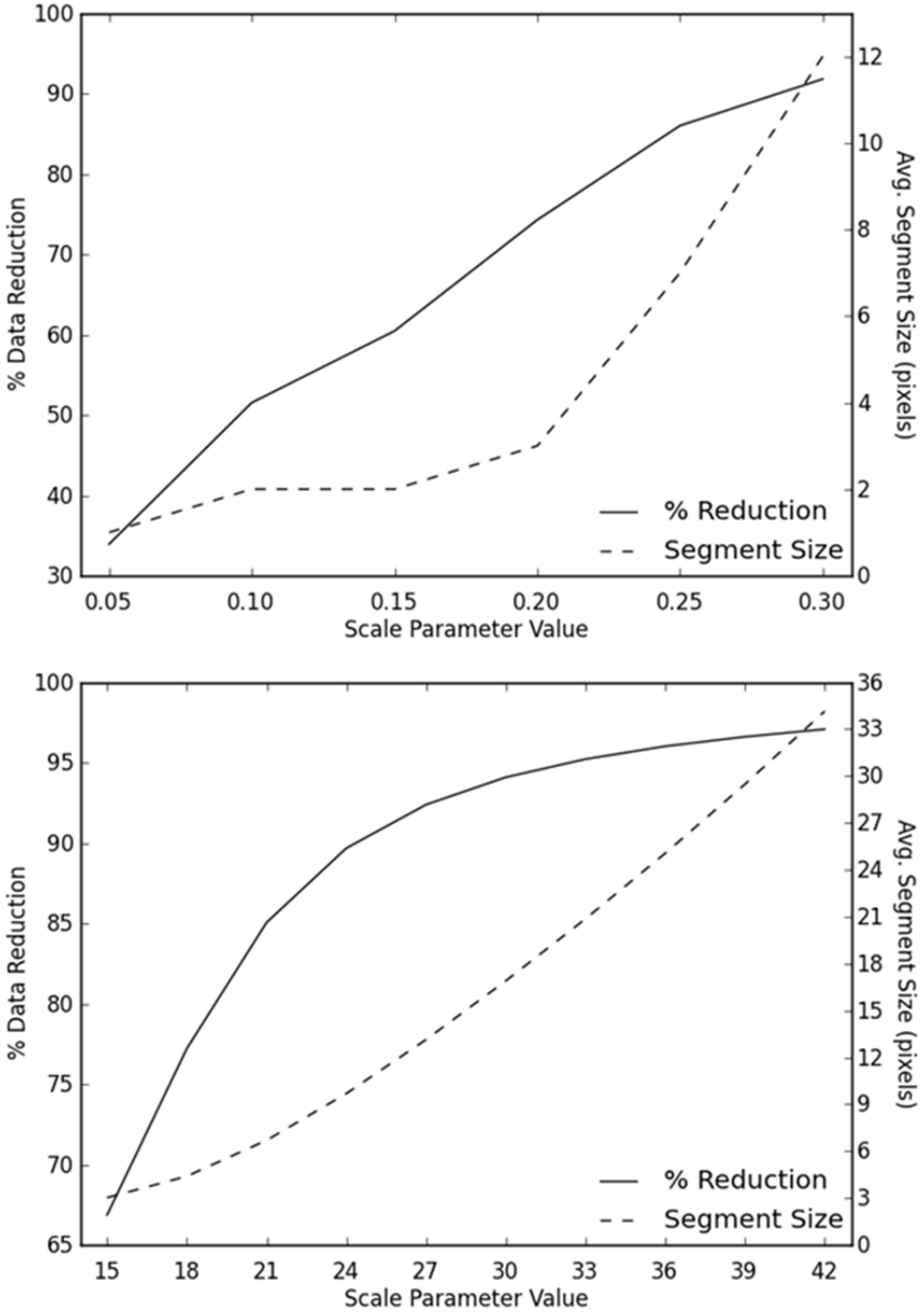


Figure 13 Data reduction and average segment size for various segmentation scales for AVHRR (top) and MODIS (bottom). These graphs were used to guide the selection of optimal segmentation scheme. Maximizing percent data reduction was paramount in the selection criteria for the MODIS time series while maintaining a relatively small segment sizes was decisive for the AVHRR time series.

Selection of best segmentation scale for MODIS data was approached differently than for AVHRR. Percent data reduction was given more weight in the selection process because of the necessity to limit the amount of data processed in the classification procedure. As the data reduction curve approaches the limit of 100%, the value returned by each successive segmentation schemes is reduced as shown by the flattening of the curve. Similarly, as the segmentation scale increases, the value returned by the average segment size is reduced due to the exponential growth of the average segment size. An estimated starting point to eliminate segmentation schemes was at the transition of vertical trend to horizontal trend on the percent data reduction curve, specifically at the 24 scale point. Accordingly, segmentation schemes of scale 24 and below were eliminated from the candidate best segmentation scheme pool.

To further refine the candidate pool of best segmentation schemes, total time series size was considered. To approximate the size of mean year reduced time series, the total data points (pixels) were computed for the mean year time series (1,059,840,000 pixels) and then the total allowed segments in the segmentation scheme were calculated by dividing the total number of pixels in the mean year time series by the number of rasters in the full-data time series. Consequently, the optimal number of segments in a segmentation scheme was at most 3,604,898, which translated to the 92.18 percent data reduction threshold. This constraint eliminated scale 27 from consideration.

Segment size increased dramatically with each successive segmentation scheme while the change in percent data reduction diminished rapidly. Scale 30 was selected as the best segmentation scheme from the pool of remaining candidates due to it being the next smallest scheme after the implemented constraints.

4.2.1 TRANSFORMED DIVERGENCE

Due to the high number of bands used as input to the ISODATA clustering algorithm for both AVHRR and MODIS segment-mean time series, the IMAGINE® transformed divergence separability function did not produce reliable results. The average and minimum transformed divergence for each classification scheme produced a value of 2000, which is the maximum value possible for this metric. Alternatively, the same number of classes from the mean year classification schemes was used for the segment mean classification for both AVHRR and MODIS time series.

4.3 COMPARISON OF MEAN YEAR CLASSES TO SEGMENTED CLASSES

Classes were categorized as major, minor, and scattered to aide interpretation of spatial distribution of each classification scheme. Table 3 shows the major, minor, and scattered class contributions to the study area. Major classes contributed significant portions of the total area. Minor classes were comprised of spatially interesting zones and contribute more than a few pixels to the area. Scattered classes contributed only a few pixels per class to the area. These categories were used to guide the creation of classification maps for each data type and reduction method. As a result, only major and minor classes are shown on the maps in figures 14 through 17.

Figures 14 and 15 depict the spatial distribution of the major and minor classes for AVHRR and MODIS mean year data reduction method, respectfully. Figures 16 and 17 show the distribution of major and minor classes for AVHRR and MODIS segmented mean data reduction method, respectively. A simple comparison of the major and minor class inclusion table revealed differences in the classification schemes. A comparison of the classification maps

confirmed this conclusion. A comparison of AVHRR maps (figures 14 and 16) and MODIS maps (figures 15 and 17) revealed the extent of class distribution differences.

	Major	Minor	Scattered
Mean Year AVHRR	23, 24, 25, 28, 30, 36, 41, 46	22, 26, 29, 37, 39, 40, 42, 45, 52, 60, 65, 66	18, 19, 20, 21, 27, 31, 32, 33, 34, 43, 44, 47, 49, 55, 57, 59, 62, 63, 65, 68, 69, 70, 71, 73
Mean Year MODIS	25, 26, 27, 29, 30, 41, 43, 46, 49, 56, 58	24, 28, 44, 47, 48, 55, 57, 59, 60, 61, 63, 64, 72	2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 31, 32, 33, 34, 35, 37, 38, 39, 40, 42, 45, 50, 51, 52, 53, 54, 62, 65, 66, 67, 68, 69, 70, 71, 73
Segment Mean AVHRR	19, 23, 24, 30, 32, 34	16, 18, 20, 27, 33, 36, 45, 46, 52, 56, 57, 59	13, 14, 15, 17, 21, 22, 25, 26, 28, 29, 35, 37, 38, 40, 41, 43, 48, 50, 53, 55, 60, 61, 63, 64, 65, 68, 70, 72
Segment Mean MODIS	32, 34, 35, 36, 39, 41, 42	37, 38, 40	25, 26, 28, 29, 30, 33

Table 3 Classes intersecting the study area. Major classes contribute significant portions of the total area. Minor classes are comprised of spatially interesting zones and contribute more than a few pixels to the area. Scattered classes contribute only a few pixels per cluster to the area. Analysis focused on the major and minor classes only.

An obvious difference between the AVHRR classification schemes was the number of classes apparent in the interior of the Syrian Desert to the middle center/bottom of the maps. Classification of segmentation means resulted in a more homogenous desert with fewer classes, while the mean year scheme fragmented the desert with more classes. This fragmented pattern was repeated throughout the map comparison. Another example was found near the top border of the study area where classes were layered in concentric crescents. More classes were seen in the mean year classification scheme than the segment mean scheme. Overall, the segment mean scheme had two less major classes but maintained the number of minor classes.

A comparison of the MODIS classification schemes showed similar but more drastic differences. Again the desert was less fractured into multiple classes in the segment mean scheme. The area near the upper border of the study area was also simplified from many classes to only a few classes. Most revealing of the change in classification schemes came from the class inclusion chart. Major classes in the segment mean scheme dropped to seven classes from eleven in the mean year scheme. Minor classes also dropped from thirteen to only three in the segment mean scheme. A driver of this drastic change was the reduced number of valid classes produced during the ISODATA clustering of the segment mean time series. While all other classification schemes produced twenty-five valid classes, the clustering of segment means only produced valid classes starting at class 25 and ending at 73.

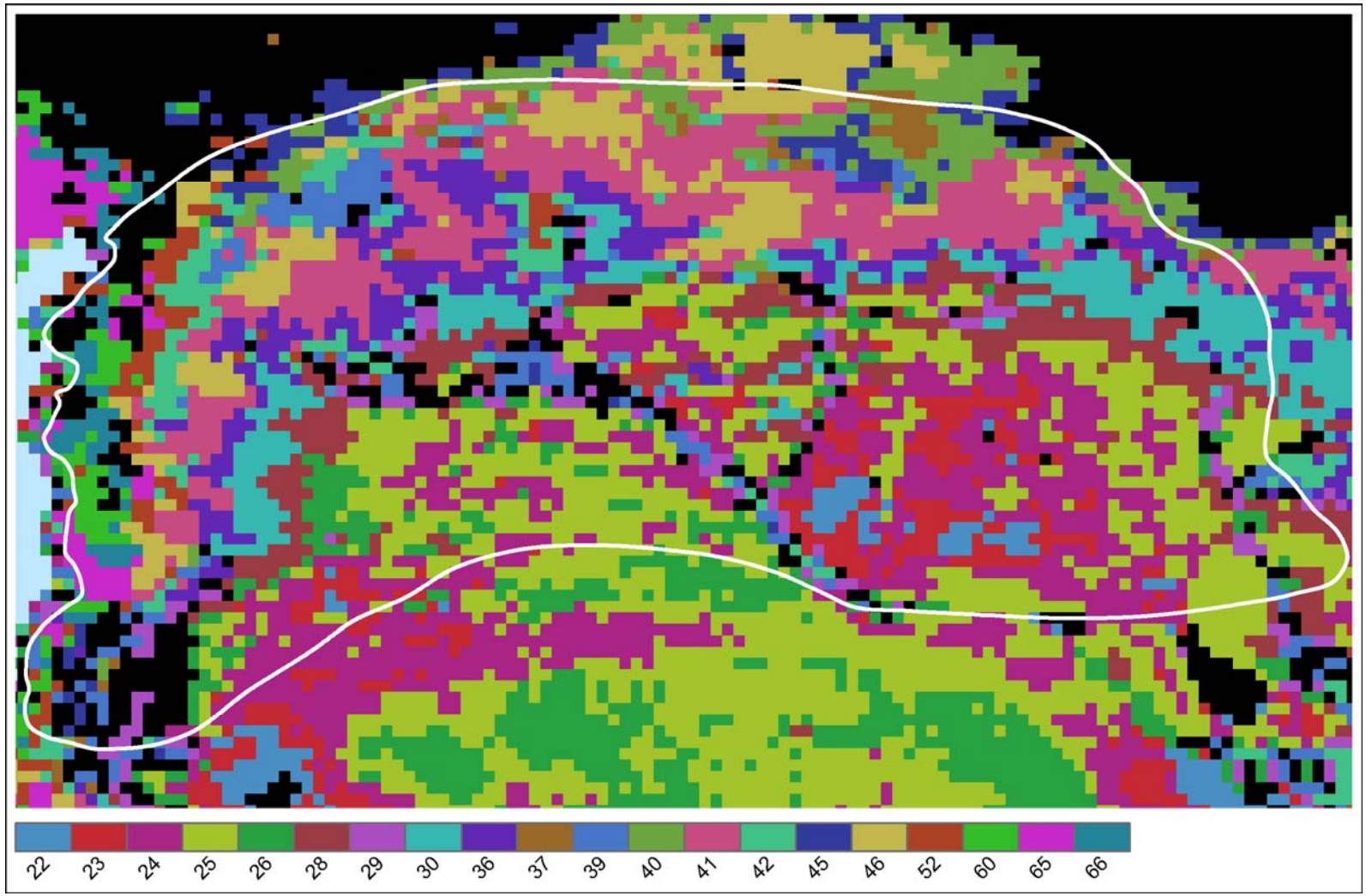


Figure 14 Major and minor clusters from the AVHRR mean year data reduction method. The study area is highly fragmented, but a general pattern of concentric layers following the Fertile Crescent are apparent.

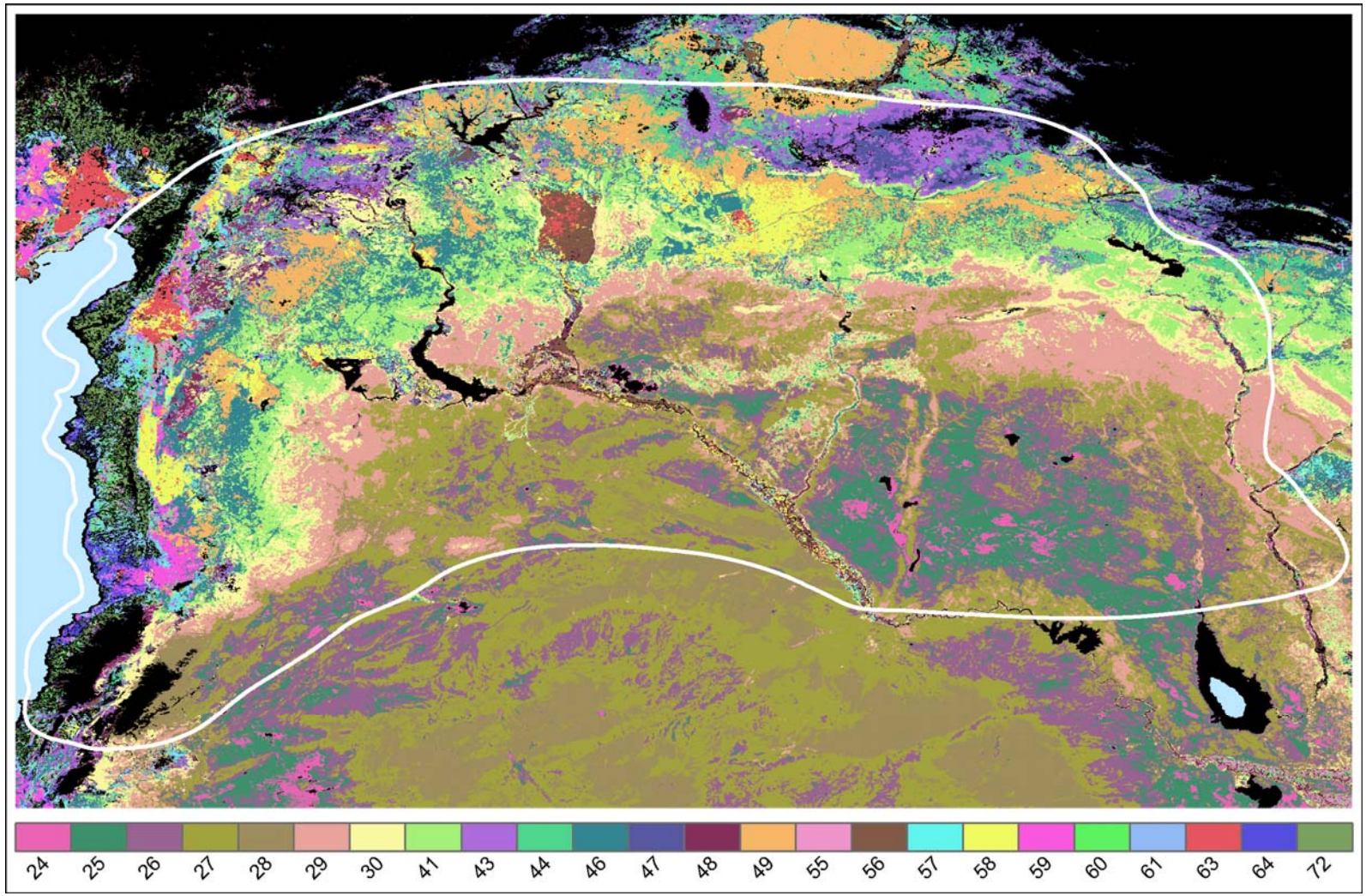


Figure 15 Major and minor clusters from the MODIS mean year data reduction method. Similar fragmentation and patterns as seen in the AVHRR mean year map are shown in greater detail in the MODIS data.

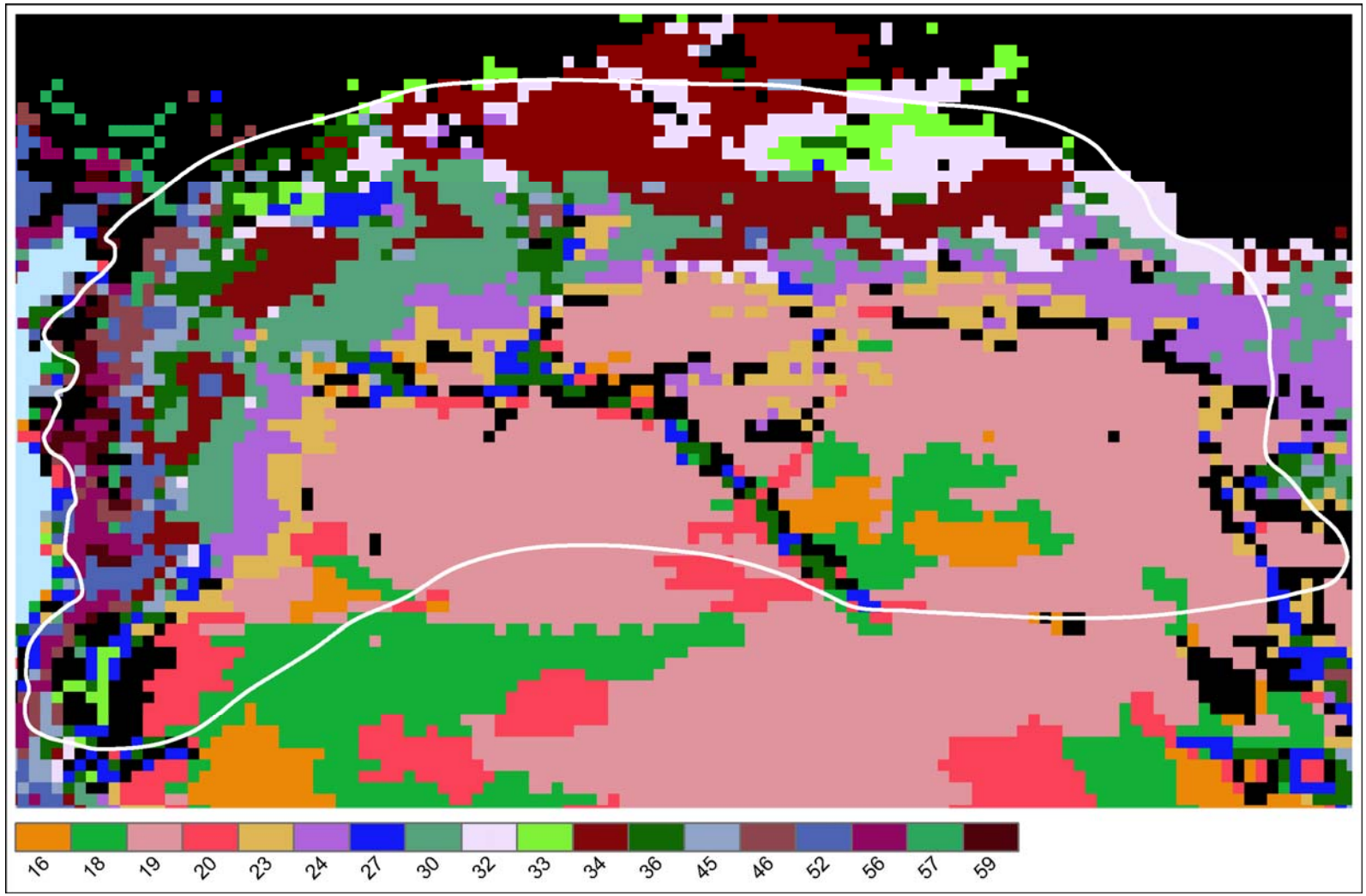


Figure 16 Major and minor clusters from the AVHRR segment mean data reduction method. Less fragmentation is apparent with this method while maintaining the spatial pattern of concentric layers built around the Fertile Crescent.

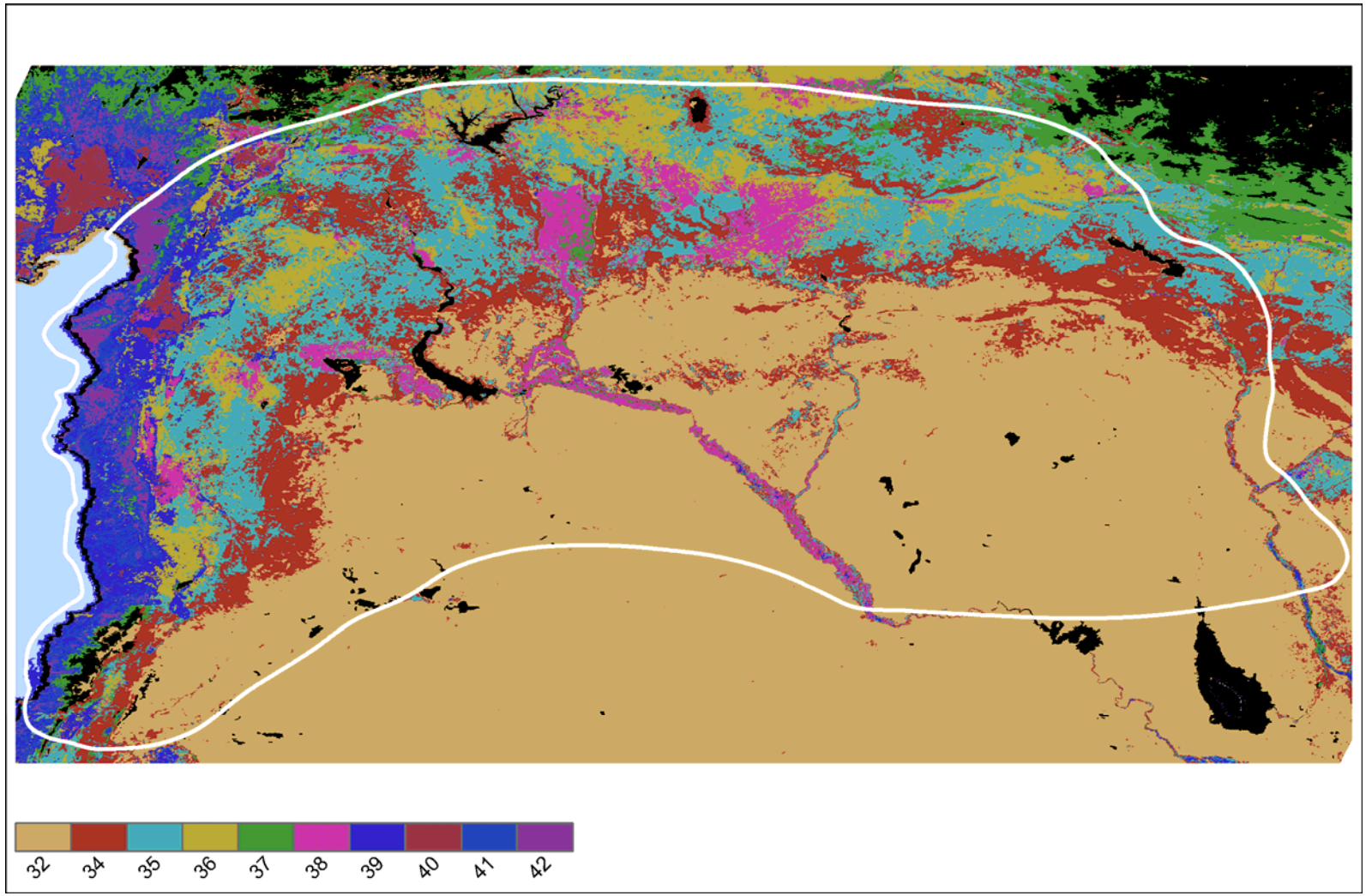


Figure 17 Major and minor clusters from the MODIS segment mean data reduction method. Even more drastic reduction of fragmentation is shown in the MODIS data.

4.3.1 HIGH RESOLUTION IMAGERY OF SELECTED CLASSES

To aide interpretation of the class differences, high resolution historical Landsat Thematic Mapper (TM) imagery was downloaded for select locations and compared to the class maps. Freely available high resolution aerial and satellite imagery were also compared to class maps. Only MODIS-based class maps were used from this point due to the very coarse resolution of the AVHRR class maps. Table 4 shows class membership differences among five locations selected from the study area with a brief description of apparent groundcover based on the imagery. Figures 18 through 27 depict enlarged maps of the selected points of interest. Historical Landsat imagery was used as a reference for point of interest selection but is not shown. A hallmark difference between mean year classes and segment mean classes was the appearance of spatial smoothing similar to the effect of a low-pass filter. A reduced number of speckled classes (lone pixels or small groups of pixels of differing class membership within other classes) were seen in the segment mean scheme. Extracted class membership and proximate classes were used to guide analysis of class temporal trajectory.

Point of Interest	Mean Year MODIS Class	Segment Mean MODIS Class	Imagery Description
1	56	38	Active agricultural field; active in historical imagery; additional CORONA image included for reference
2	56	38	Active ag. field; active historically
3	46	35	Inactive ag. field; N/A
4	41	34	Rangeland or inactive ag. field; inactive historically
5	63	38	Active ag. field; inactive historically

Table 4 Cluster membership of select point of interest locations.

Figure 18 The first point of interest is an active agricultural field near the confluence of the lower Balikh and Euphrates rivers. This area is dominated by irrigated agriculture.

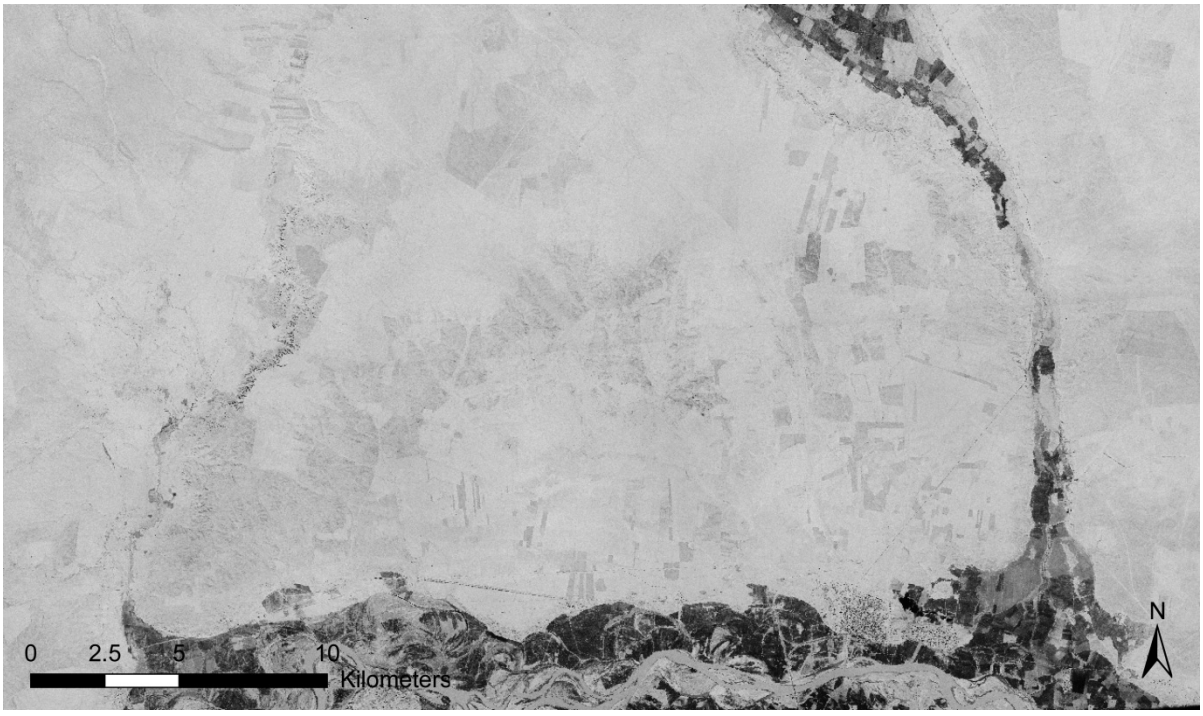


Figure 19 Historical satellite imagery from the CORONA program acquired November 1968 compared to modern imagery shows massive agricultural intensification (Casana et al. 2012).

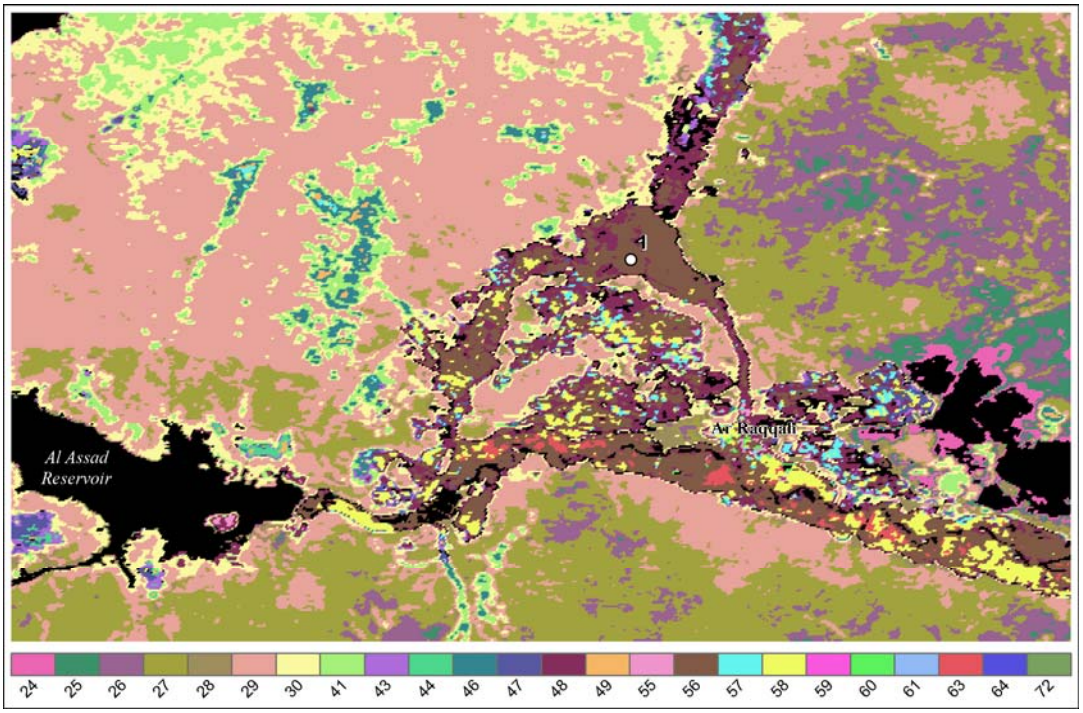


Figure 20 Point of interest 1 intersects class 56 of the mean year classification scheme. Class 56 appears to generally correspond to cropped areas in figure 18.

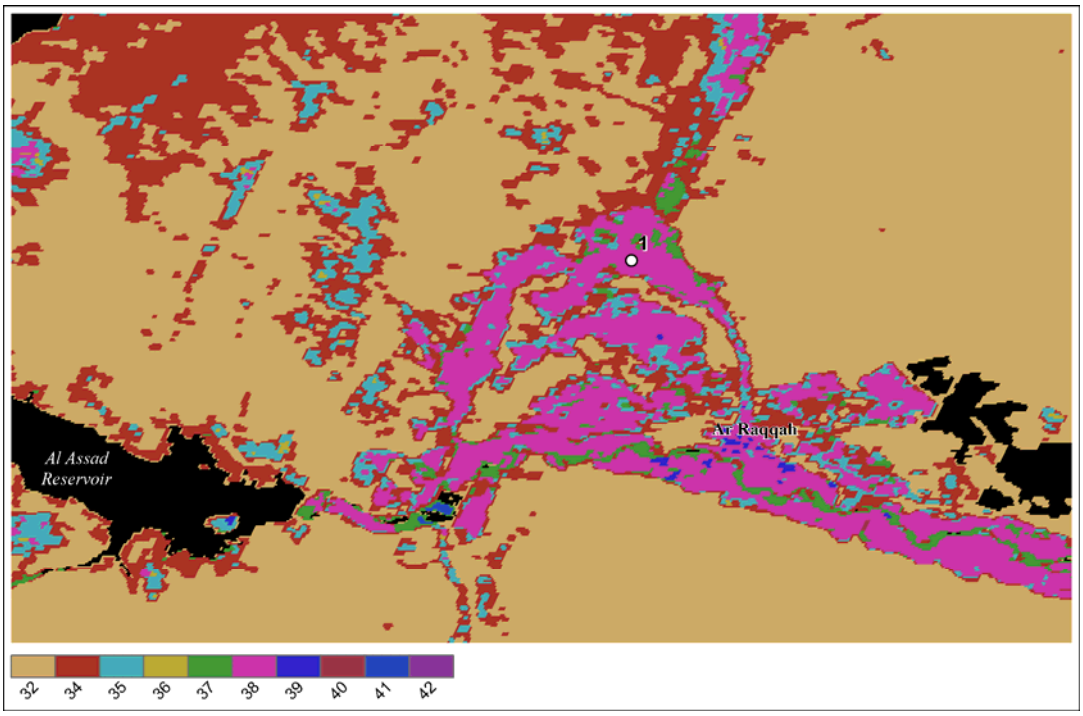


Figure 21 Point of interest 1 intersects with the main agricultural class of the segment mean classification scheme, class 38. Class 38 is less fragmented than class 56 of figure 19.

Figure 22 Points of interest 2 and 3 displayed over natural color imagery. Point 2 is over an active field in the Harran plain while point 3 is over an inactive field.

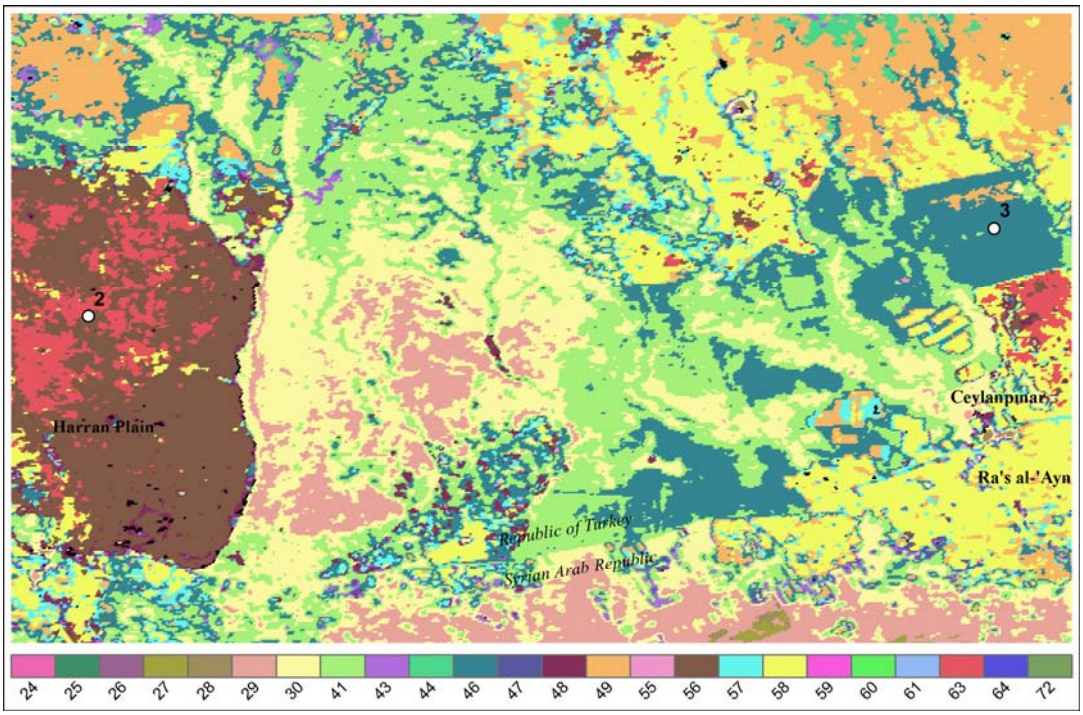


Figure 23 Points of interest 2 and 3 correspond to classes 56 and 46, respectively. A high level of class fragmentation is apparent between the points.

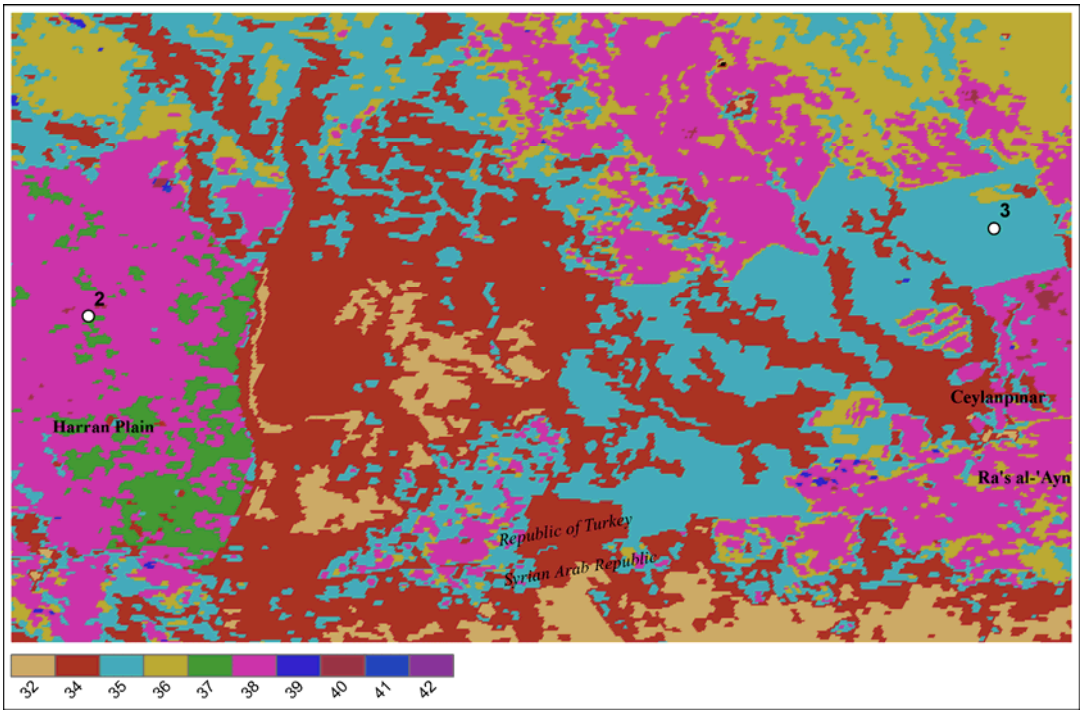


Figure 24 Points of interest 2 and 3 correspond with classes 38 and 35, respectively, in the segment mean classification map. Visible clusters are relatively un-fragmented.

Figure 25 Points of interest 4 and 5 are west of Al Assad Lake and correspond with rangeland and active fields, respectively.

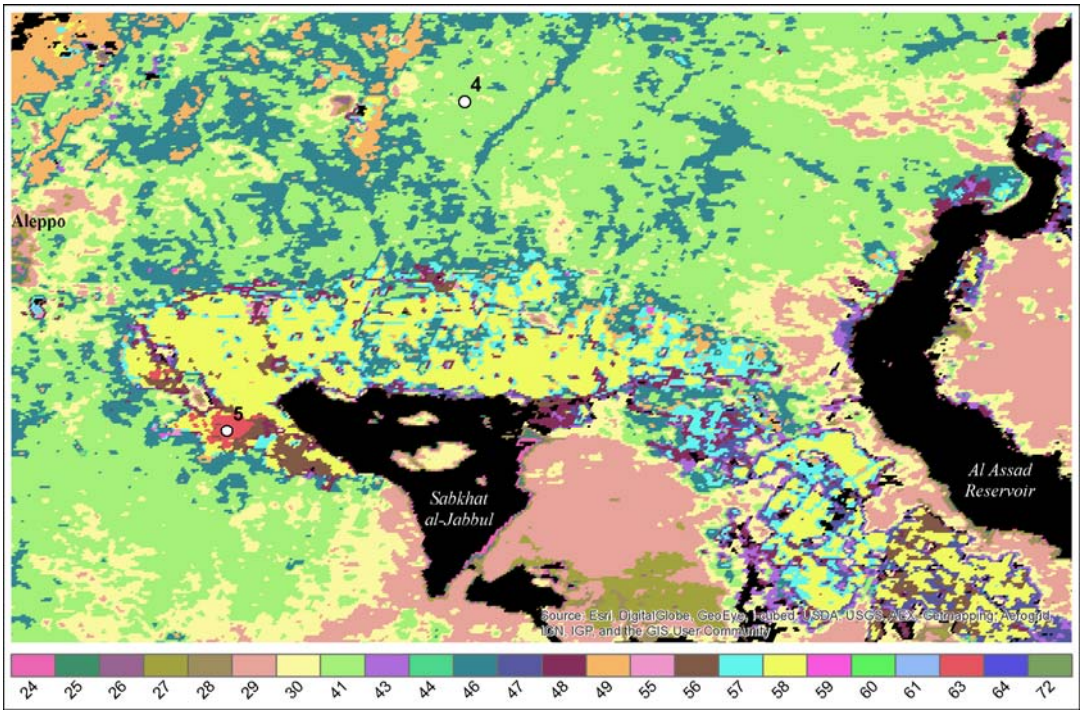


Figure 26 Point 4 is contained within class 41 which covers most of the rangeland but is peppered with classes 46 and 30. Point 5 intersects class 63 which is limited in spatial extent.

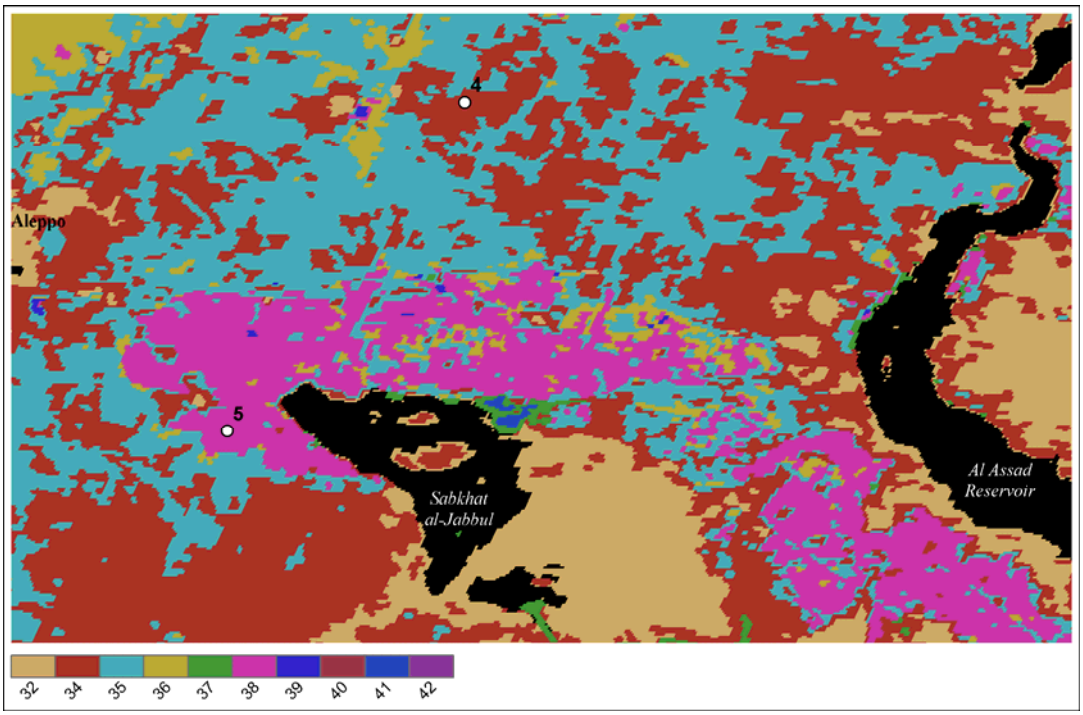


Figure 27 In the segment mean classification map, point 4 is now associated with class 40 which covers most of the rangeland (along with class 35). Point 5 is in class 38, the main agricultural class of the scheme.

4.3.2 TEMPORAL TRAJECTORY OF SELECTED CLASSES

Class dynamics over time cannot be obtained from the imagery or class map alone. Temporal trajectory was used to indicate the mean NDVI values throughout the study period. Graphs of the temporal trajectory were produced to provide insight into class structure and processes undergone over time. Not all graphs are reproduced herein. However, the graphs corresponding to the selected classes/clusters in figure 20 are included.

4.3.2.1 POINT OF INTEREST 1 & 2

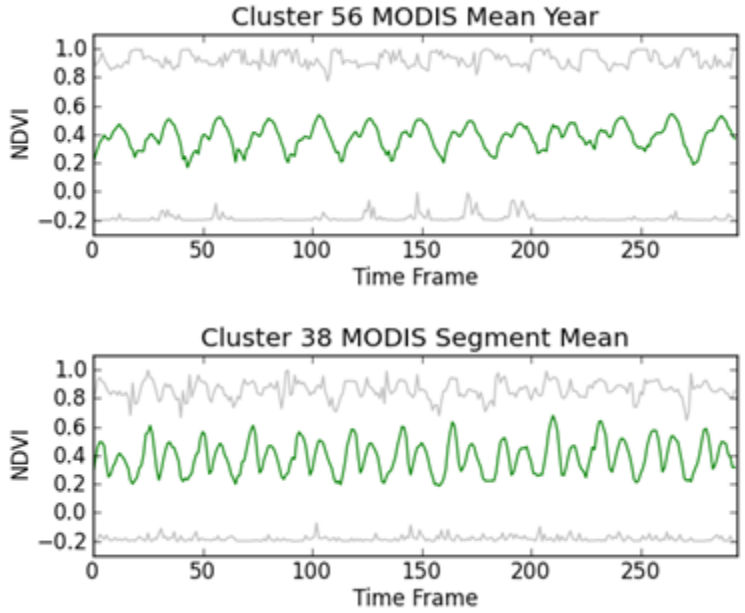


Figure 28 Cluster minimum, maximum, and mean (green) NDVI for clusters intersecting point of interest 1 and 2 for the length of the original time series.

At the first and second point of interest (figure 28), both class schemes produced a cluster with double-cropping peaks. The period between crop cycles, however, was smoothed in the mean year cluster by comparison to the more well-defined cropping periods of the segment mean cluster. Additionally, the peak NDVI cropping period was switched between the two classes. For the mean year class, the peak NDVI period came after the first cropping period of each year, while the segment mean class showed a peak NDVI during the first cropping period of each year.

4.3.2.2 POINTS OF INTEREST 3 & 4

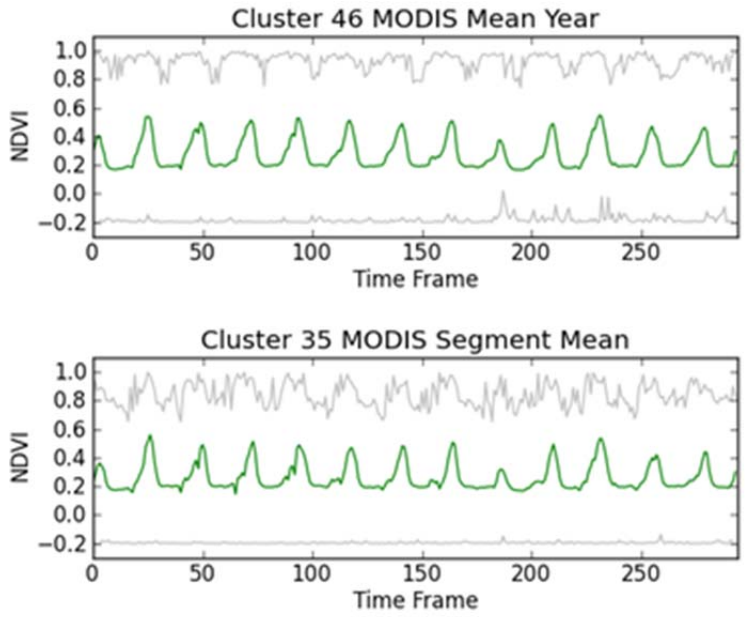


Figure 29 Cluster minimum, maximum, and mean (green) NDVI for clusters intersecting point of interest 3 for the length of the original time series.

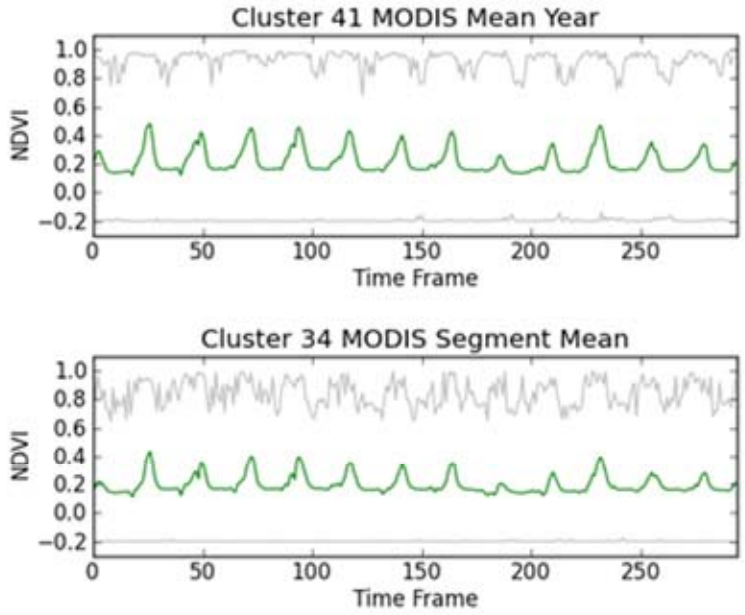


Figure 30 Cluster minimum, maximum, and mean (green) NDVI for clusters intersecting point of interest 4 for the length of the original time series.

A comparison of the classification time series chart for point of interest 3 (figure 29) and 4 (figure 30) was similar. Yearly crop cycle apparent in NDVI for each class showed a dormant period followed by a quick green-up and single harvest. The cycle shape for each year was similar between all compared classes indicating general agreement in class structure. Also, time frames above 190 in classes covering point 4 displayed a generally lower peak NDVI. An inspection of the situation of point 4 on the maps indicated that it was in a rangeland/dry-farmed area. The reduction of peak NDVI could indicate dryland degradation, prolonged drought, or changing water availability within the

region. In this instance, the segmentation mean method was favorable because it did not average NDVI values over time and maintained the magnitude of true NDVI peak. This property is seen in figure 31 (and less obviously in figure 30) with the segment mean NDVI peaks lower than the mean year peaks.

4.3.2.3 POINT OF INTEREST 5

Point of interest 5 (figure 31) resembled point of interest 1 and 2. Both points of interest were members of the same segment mean class but different mean year classes. Again, a double-cropped pattern was seen in both classes, and the first crop period was lower in mean year class and higher in the segment mean class. Point 5's depth of the dip between crop periods within a year for the mean year class differed from point 1's dip depth.

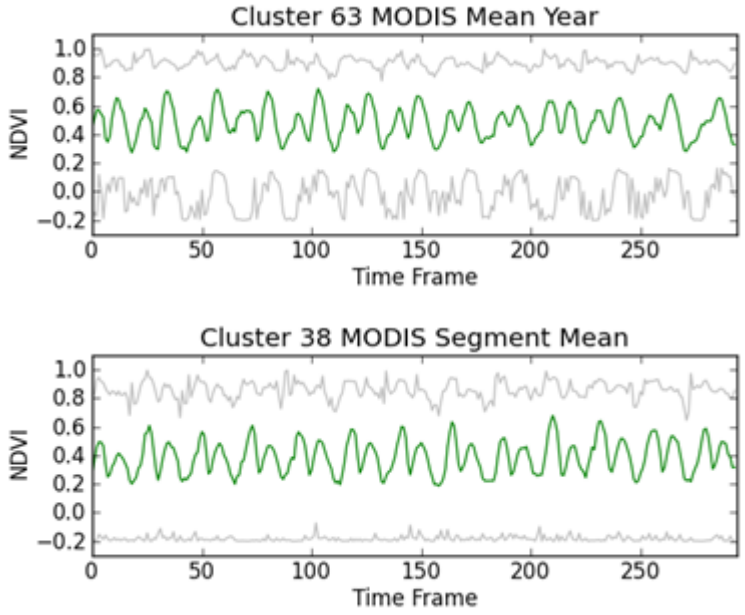


Figure 31 Cluster minimum, maximum, and mean (green) NDVI for clusters intersecting point of interest 5 for the length of the original time series.

4.4 OUTCOME OF WORK

This project succeeded in preserving temporal information while reducing data size, which is a valuable contribution to future researchers attempting to use temporal information from a full-data time series. Based on qualitative spatial and time series analysis, the research hypothesis, "there is no spatial difference between unsupervised classifications based on mean year reduced-data vs. segment mean reduced-data time series", was rejected. The provided examples

qualitatively differentiated the compared classification methods. Each technique has advantages depending on the desired spatial or temporal granularity. The mean year reduction method most likely picks out average intra-annual phenological variation. Variation in temporal trajectory is contained in the segment mean method. The MODIS segment mean classification map should be used for maximum utilization of the temporal information in the classification scheme.

4.5 FUTURE DIRECTIONS

Additional work is required to refine the segment mean data reduction method in a couple of areas. First, work is needed to define guidelines for determination of optimal scale size for the spatial data reduction technique. In this study, optimal scale size was determined by "guided guessing"; no quantitative measure of "best" segmentation level was used. For a more reliable outcome, future work needs to incorporate a method for quantitative evaluation of each proposed segmentation scheme. Second, additional work is required to determine a best cluster validity index to use to select the optimal classification scheme. Although this work used a quantitative measure (the minimum Transformed Divergence) to guide classification scheme selection, work is needed to compare the effect of using different CVIs. A list and descriptions of potential CVIs is presented by Arbelaitz et al. (2013).

Future research is also needed to determine the best distance (similarity) measure for use in unsupervised classification to take advantage of the properties of the segment mean method. Simple ISODATA was used for this research. However, similarity calculated in ISODATA (Euclidean distance) is essentially an averaging operation on pixel vectors. More effective similarity measures, which use the temporal information retained by the segment mean data reduction method, should take account of the shape of the NDVI curve over time.

5. CONCLUSION

Four land surface phenology classification maps were produced from time series satellite imagery to compare classification of mean year data reduction method against classification of the segment mean method. Two of the four maps were produced using 8 × 8 km AVHRR data over the years 1981 through 2006. The other two maps were produced using 250 × 250 m MODIS data for the years 2000 through 2012. The segment mean method showed both spatial smoothing similar to a low-pass filter and a reduction of the number of total classes covering the study area. The classification map created from the MODIS segment mean data reduction method represents the multi-decadal vegetation phenology record with an optimal spatial resolution.

6. REFERENCES

Arbelaitz, O., I. Gurrutxaga, J. Muguerza, J. M. Pérez, and I. Perona. 2013. "An Extensive Comparative Study of Cluster Validity Indices." *Pattern Recognition* 46 (1): 243–256.

Baatz, Martin, and Arno Schäpe. 2000. "Multiresolution Segmentation: An Optimization Approach for High Quality Multi-scale Image Segmentation." In *Angewandte Geographische Informationsverarbeitung XI*. Salzburg: Herbert Wichmann Verlag.

Bajgiran, Parinaz, Ali A. Darvishsefat, Ali Khalili, and Majid F. Makhdoum. 2008. "Using AVHRR-based Vegetation Indices for Drought Monitoring in the Northwest of Iran." *Journal of Arid Environments* 72 (6): 1086–1096.

Al-Bakri, J.T., and J.C. Taylor. 2003. "Application of NOAA AVHRR for Monitoring Vegetation Conditions and Biomass in Jordan." *Journal of Arid Environments* 54 (3): 579–593.

Ball, G. H., and D. J. Hall. 1965. "ISODATA, a Novel Method of Data Analysis and Pattern Classification". DTIC Document. http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=AD0699616.

Beaumont, Peter. 1996. "Agricultural and Environmental Changes in the Upper Euphrates Catchment of Turkey and Syria and Their Political and Economic Implications." *Applied Geography* 16 (2): 137–157.

Beck, Hylke E., Tim R. McVicar, Albert I.J.M. van Dijk, Jaap Schellekens, Richard A.M. de Jeu, and L. Adrian Bruijnzeel. 2011. "Global Evaluation of Four AVHRR-NDVI Data Sets: Intercomparison and Assessment Against Landsat Imagery." *Remote Sensing of the Environment* 115 (10): 2547–2563.

Benedetti, Riccardo, Paolo Rossini, and R. Taddei. 1994. "Vegetation Classification in the Middle Mediterranean Area by Satellite Data." *International Journal of Remote Sensing* 15 (3): 583–596.

De Brichambaut, G., and C. C. Wallen. 1963. "A Study of Agro-climatology in Semi-arid and Arid Zones of the Near East". Technical Note No. 56. Geneva: World Meteorological Organisation.

Brown, M.E., J.E. Pinzon, K. Didan, J.T. Morisette, and C.J. Tucker. 2006. "Evaluation of the Consistency of Long-term NDVI Time Series Derived from AVHRR,SPOT-vegetation, SeaWiFS, MODIS, and Landsat ETM+ Sensors." *IEEE Transactions on Geoscience and Remote Sensing* 44 (July): 1787–1793. doi:10.1109/TGRS.2005.860205.

Casana, Jesse, Jackson Cothren, and Tuna Kalayci. 2012. "Swords into Ploughshares: Archaeological Applications of CORONA Satellite Imagery in the Near East." *Internet Archaeology* (32). http://dx.doi.org/10.11141/ia.32.2.

Clark, R. N., G. A. Swayze, R. Wise, K. E. Livo, T. M. Hoefen, R. F. Kokaly, and S. J. Sutley. 2007. "USGS Digital Spectral Library Splib06a". U.S. Geological Survey, Data Series 231.

Cohen, J. 1960. "A Coefficient of Agreement for Nominal Scales." *Educational and Psychological Measurement* 20 (1): 37–46.

Congalton, Russell G. 2010. “Remote Sensing: An Overview.” *GIScience & Remote Sensing* 47 (4): 443–459.

Corcoran, P., and A. Winstanley. 2007. “Segmentation Evaluation for Object Based Remotely Sensed Image Analysis”. Unpublished manuscript. http://eprints.nuim.ie/00001329.

Deering, D. W., J. W. Rouse, R. H. Haas, and J. A. Schall. 1975. “Measuring Forage Production of Grazing Units from Landsat MSS Data.” In *Proceedings of the 10th International Symposium on Remote Sensing of the Environment*, 1169. Ann Arbor, Michigan.

DeFries, R., M. Hansen, and J. Townshend. 1995. “Global Discrimination of Land Cover Types from Metrics Derived from AVHRR Pathfinder Data.” *Remote Sensing of Environment* 54 (3): 209–222.

ERDAS. 2011. *ERDAS IMAGINE 2011* (version 11). Norcross, GA: ERDAS, Inc. http://geospatial.intergraph.com.

ESRI. 2012. *ArcGIS* (version 10.1). Redlands, CA: Environmental Systems Research Institute, Inc. http://www.esri.com.

Evans, J., and R. Geerken. 2006. “Classifying Rangeland Vegetation Type and Coverage Using a Fourier Component Based Similarity Measure.” *Remote Sensing of Environment* 105 (1): 1–8. doi:10.1016/j.rse.2006.05.017.

Geerken, Roland A. 2009. “An Algorithm to Classify and Monitor Seasonal Variations in Vegetation Phenologies and Their Inter-annual Change.” *ISPRS Journal of Photogrammetry and Remote Sensing* 64 (4): 422–431.

Gu, Yingxin, Jesslyn F. Brown, Tomoaki Miura, Willem J. D. van Leeuwen, and Bradley C. Reed. 2010. “Phenological Classification of the United States: A Geographic Framework for Extending Multi-Sensor Time-Series Data.” *Remote Sensing* 2 (2) (February 11): 526–544. doi:10.3390/rs2020526.

Gutman, G., and J. G. Masek. 2012. “Long-term Time Series of the Earth’s Land-surface Observations from Space.” *International Journal of Remote Sensing* 33 (15): 4700–4719.

Hargrove, William W., Joseph P. Spruce, Gerald E. Gasser, and Forrest M. Hoffman. 2009. “Toward a National Early Warning System for Forest Disturbances Using Remotely Sensed Canopy Phenology.” *Photogrammetric Engineering & Remote Sensing* 75 (10): 1150–1156.

Hird, J. N., and G. J. McDermid. 2009. “Noise Reduction of NDVI Time Series: An Empirical Comparison of Selected Techniques.” *Remote Sensing of Environment* 113 (1): 248–258.

Holben, Brent N. 1986. “Characteristics of Maximum-value Composite Images from Temporal AVHRR Data.” *International Journal of Remote Sensing* 7 (11): 1417–1434.

Hole, F., and B. F. Zaitchik. 2007. “Policies, Plans, Practice, and Prospects: Irrigation in Northeastern Syria.” *Land Degradation & Development* 18 (2) (March): 133–152. doi:10.1002/ldr.772.

James, M. E., and S. N. V. Kalluri. 1994. “The Pathfinder AVHRR Land Data Set: An Improved Coarse Resolution Data Set for Terrestrial Monitoring.” *International Journal of Remote Sensing* 15 (17): 3347–3363.

Jensen, John R. 2005. *Introductory Digital Image Processing: A Remote Sensing Perspective*. Third. Prentice-Hall Series in Geographic Information Science. New Jersey: Pearson Prentice Hall.

———. 2007. *Remote Sensing of the Environment: An Earth Resource Perspective*. Second. Prentice-Hall Series in Geographic Information Science. New Jersey: Pearson Prentice Hall.

Kouchoukos, N., R. Smith, A. Gleason, P. Thenkabail, F. Hole, Y. Barkoudah, J. Albert, P. Gluhosky, and J. Foster. 1998. "Monitoring the Distribution, Use, and Regeneration of Natural Resources in Semi-arid Southwest Asia." *Transformations of the Middle Eastern Natural Resources: Legacies and Lessons, Bulletin Series* (103): 467–491.

Kouchoukos, Nicholas. 2001. "Satellite Images and Near Eastern Landscapes." *Near Eastern Archaeology* 64 (1-2): 80–91.

Lhermitte, S., J. Verbesselt, W.W. Verstraeten, and P. Coppin. 2011. "A Comparison of Time Series Similarity Measures for Classification and Change Detection of Ecosystem Dynamics." *Remote Sensing of Environment* 115 (12) (December): 3129–3152. doi:10.1016/j.rse.2011.06.020.

Liao, T. Warren. 2005. "Clustering of Time Series Data—a Survey." *Pattern Recognition* 38 (11): 1857–1874.

Los, S. O. 1998. "Estimation of the Ratio of Sensor Degradation Between NOAA AVHRR Channels 1 and 2 from Monthly NDVI Composites." *Geoscience and Remote Sensing, IEEE Transactions On* 36 (1): 206–213.

Los, S. O., G. J. Collatz, P. J. Sellers, C. M. Malmström, N. H. Pollack, R. S. DeFries, L. Bounoua, M. T. Parris, C. J. Tucker, and D. A. Dazlich. 2000. "A Global 9-yr Biophysical Land Surface Dataset from NOAA AVHRR Data." *Journal of Hydrometeorology* 1 (2): 183–199.

Nguyen, Thi Thu Ha, C. A. J. M. De Bie, Amjad Ali, E. M. A. Smaling, and Thai Hoanh Chu. 2012. "Mapping the Irrigated Rice Cropping Patterns of the Mekong Delta, Vietnam, Through Hyper-temporal SPOT NDVI Image Analysis." *International Journal of Remote Sensing* 33 (2) (January 20): 415–434. doi:10.1080/01431161.2010.532826.

Pedelty, J., S. Devadiga, E. Masuoka, M. Brown, J. Pinzon, C. Tucker, D. Roy, et al. 2007. "Generating a Long-term Land Data Record from the AVHRR and MODIS Instruments." In *IGARSS*. IEEE International.

Pinzon, J., M. E. Brown, and C. J. Tucker. 2005. "EMD Correction of Orbital Drift Artifacts in Satellite Data Stream." *The Hilbert-Huang Transform and Its Applications (eds Huang N, Shen S)*: 167–183.

Piwowar, J. M., and E. F. LeDrew. 1995. "Hypertemporal Analysis of Remotely Sensed Sea-ice Data for Climate Change Studies." *Progress in Physical Geography* 19 (2): 216–242.

Rogerson, Peter A. 2006. *Statistical Methods for Geography: A Student's Guide*. Second. Los Angeles: Sage.

Rousseeuw, P. J. 1987. "Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis." *Journal of Computational and Applied Mathematics* 20: 53–65.

Schmidt, Heike, and Arnon Karnieli. 2000. “Remote Sensing of the Seasonal Variability of Vegetation in a Semi-arid Environment.” *Journal of Arid Environments* 45 (1): 43–59.

Solano, Ramon, Kamel Didan, Andree Jacobson, and Alfredo Huete. 2010. “MODIS Vegetation Index User’s Guide (MOD13 Series)”. Version 2.00. Vegetation and Phenology Lab, The University of Arizona.

Swain, Philip H., and Shirley M. Davis. 1978. *Remote Sensing: The Quantitative Approach*. McGraw-Hill International Book Co.

Tou, Julius T., and Rafael C. Gonzalez. 1974. *Pattern Recognition Principles*. Applied Mathematics and Computation 7. Reading, Massachusetts: Addison-Wesley.

Trigo, Ricardo M., Célia M. Gouveia, and David Barriopedro. 2010. “The Intense 2007–2009 Drought in the Fertile Crescent: Impacts and Associated Atmospheric Circulation.” *Agricultural and Forest Meteorology* 150 (9) (August 15): 1245–1257. doi:10.1016/j.agrformet.2010.05.006.

Trimble. 2011. *eCognition* (version 8). Munich, Germany: Trimble Germany GmbH. http://www.ecognition.com/.

Tucker, Compton J. 1979. “Red and Photographic Infrared Linear Combinations for Monitoring Vegetation.” *Remote Sensing of the Environment* 8 (2): 127–150.

Tucker, Compton J., Denelle M. Grant, and Jon D. Dykstra. 2004. “NASA’s Global AVHRR Data Set.” *Photogrammetric Engineering & Remote Sensing* 70 (3): 313–322.

Tucker, Compton J., Jorge E. Pinzon, Molly E. Brown, Daniel A. Slayback, Edwin W. Pak, Robert Mahoney, Eric F. Vermote, and Nazmi El Saleous. 2005. “An Extended AVHRR 8-km NDVI Dataset Compatible with MODIS and SPOT Vegetation NDVI Data.” *International Journal of Remote Sensing* 26 (20): 4485–4498.

Vermote, E., N. El Saleous, Y. J. Kaufman, and E. Dutton. 1997. “Data Pre-processing: Stratospheric Aerosol Perturbing Effect on the Remote Sensing of Vegetation: Correction Method for the Composite NDVI After the Pinatubo Eruption.” *Remote Sensing Reviews* 15 (1-4): 7–21.

Vermote, E., and Y. J. Kaufman. 1995. “Absolute Calibration of AVHRR Visible and Near-infrared Channels Using Ocean and Cloud Views.” *International Journal of Remote Sensing* 16 (13): 2317–2340.

Voss, Katalyn A., James S. Famiglietti, MinHui Lo, Caroline de Linage, Matthew Rodell, and Sean C. Swenson. 2013. “Groundwater Depletion in the Middle East from GRACE with Implications for Transboundary Water Management in the Tigris-Euphrates-western Iran Region.” *Water Resources Research* 49. doi:10.1002/wrcr.20078.

Weiss, E., S. E. Marsh, and E. S. Pfirman. 2001. “Application of NOAA-AVHRR NDVI Time-series Data to Assess Changes in Saudi Arabia’s Rangelands.pdf.” *International Journal of Remote Sensing* 22 (6): 1005–1027.

White, M. A., F. Hoffman, W. W. Hargrove, and R. R. Nemani. 2005. “A Global Framework for Monitoring Phenological Responses to Climate Change.” *Geophysical Research Letters* 32 (4): L04705.

7. APPENDIX

The following are selected Python code scripts used in the course of this work.

7.1 CREATION OF MEAN YEAR RASTERS

```
import os
import arcpy
from arcpy.sa import *

arcpy.env.overwriteOutput = True
arcpy.CheckOutExtension("Spatial")

#create a dictionary to keep track of all the compositing period days and corresponding years
cmpst_prd_dict = {}
for filename in os.listdir(r'D:\GIS\bbunker\Py\modis_time_series'):
	if os.path.splitext(filename)[1] == r'.img':
		if filename[5:8] not in cmpst_prd_dict:
			cmpst_prd_dict[filename[5:8]] = [filename[1:5]]
		else:
			cmpst_prd_dict[filename[5:8]].append(filename[1:5])
for daynum in cmpst_prd_dict:
	print 'Start day number ',daynum
	mask_sum = ''
	ndvi_sum = ''
	for year in range(0,len(cmpst_prd_dict[daynum])):
		print '\tworking on year ',(year + 1),' of ',len(cmpst_prd_dict[daynum])
		basepath = r'D:\GIS\bbunker\Py\modis_time_series'
		filename = 'A' + cmpst_prd_dict[daynum][year] + daynum + '.img'
		ndvi_in = os.path.join(basepath,filename)
		qa_in = os.path.join(basepath,'QA',filename)
		mask_temp = Con(qa_in, 1, 32767, 'VALUE <= 2 AND VALUE >= 0')
		ndvi_temp = Raster(ndvi_in) * mask_temp
		outpath = r'D:\GIS\bbunker\Py\modis_masked'
		ndvi_outpath = os.path.join(outpath, 'A' + cmpst_prd_dict[daynum][year] +
daynum + '.img')
		if year == 0:
			mask_sum = mask_temp
			ndvi_sum = ndvi_temp
		else:
			mask_sum = mask_sum + mask_temp
			ndvi_sum = ndvi_sum + ndvi_temp
	#divide the NDVI sum by the binary sum for the same time frame
	mask_sum = Con(mask_sum, 1, mask_sum, 'VALUE = 0')
```

```
        ndvi_avg = ndvi_sum / mask_sum
        outfilepath = r'D:\GIS\bbunker\Py\modis_mean_year_composites'
        outfilename = os.path.join(outfilepath, daynum + '.img')
        outfilename_qa = os.path.join(outfilepath, 'QA', daynum + '_qa.img')
        arcpy.CopyRaster_management(ndvi_avg, outfilename, '', '', '32767',
                '','','16_BIT_SIGNED', '','')
        arcpy.CopyRaster_management(mask_sum, outfilename_qa, '', '', '',
                '','','16_BIT_SIGNED', '','')
        #Do not use ndvi_avg.save(outfilename) as.save() creates a 32bit raster by default!
        print 'Done with ',outfilename
```

7.2 CALCULATION OF SEGMENT MEANS

```
import arcpy
import os
import time

arcpy.env.overwriteOutput = True

ZoneRasterData = r'D:\GIS\bbunker\eCog\modis_scale30\raster_image_segments.img'

imagefile_base = r'D:\GIS\bbunker\Py\modis_time_series'
for imagefilename in os.listdir(imagefile_base):
        if os.path.splitext(imagefilename)[1] == r'.img':
                print 'Processing: ',os.path.splitext(imagefilename)[0]
                sTime = time.time()
                input_img = os.path.join(imagefile_base,imagefilename)
                outTablePath = r'D:\GIS\bbunker\Py\modis_segment_means\tables'
                outTableName = os.path.splitext(imagefilename)[0] + r'.dbf'

                arcpy.CheckOutExtension("Spatial")
                arcpy.sa.ZonalStatisticsAsTable(ZoneRasterData, 'Value', input_img,
                        r'in_memory\table', "DATA", "MEAN")
                arcpy.TableToTable_conversion(r'in_memory\table', outTablePath,
                        outTableName)
                print r'Minutes Elapsed: ',(time.time()-sTime)/60.0
#If 'error: too many unique values!', set arcmap options -> raster -> no. Unique values
# to render = {>number of objects}
```

7.2 CREATION OF SEGMENT RASTERS

```
import arcpy
from dbfpy import dbf
import os
import time
import multiprocessing
```

```
arcpy.env.overwriteOutput = False
arcpy.env.rasterStatistics = 'NONE'
arcpy.env.pyramid = 'NONE'

refTable_path = r'D:\GIS\bbunker\eCog\MODIS_scale30\raster_segments.img.vat.dbf'
refTable = dbf.Dbf(refTable_path)
n = refTable.recordCount #number of segments in current segment scheme
tablefile_base = r'D:\GIS\bbunker\Py\modis_segment_means'

def build_raster_from_table(tablefilename):
	#print 'Processing: ',os.path.splitext(tablefilename)[0]
	sTime = time.time()
	input_table_path = os.path.join(tablefile_base,r'tables',tablefilename)
	output_ascii_path = os.path.join(tablefile_base,os.path.splitext(tablefilename)[0] + r'.txt')
	output_img_path = os.path.join(tablefile_base,os.path.splitext(tablefilename)[0] + r'.img')

	table = dbf.Dbf(input_table_path)
	output_ascii = open(output_ascii_path, 'w')
	output_ascii.write('NCOLS 1\n')
	output_ascii.write('NROWS ' + str(n) + '\n')
	output_ascii.write('XLLCENTER 0.0\n')
	output_ascii.write('YLLCENTER 0.0\n')
	output_ascii.write('CELLSIZE 1.0\n')
	output_ascii.write('NODATA_VALUE 65535\n')

	lines = []
	trCount = 0
	asciiCount = 0
	noDataCount = 0
	while asciiCount < n:
		if trCount < table.recordCount: #if there are still rows to be added from the table
			if refTable[asciiCount]['VALUE'] < table[trCount]['VALUE']:
				while refTable[asciiCount]['VALUE'] < table[trCount]['VALUE']:
					lines.append('65535\n')
					asciiCount += 1
					noDataCount += 1
			else:
				lines.append(str(int(table[trCount]['MEAN'])+2000) + '\n')
				trCount += 1
				asciiCount += 1
		else:
			lines.append('65535\n')
			asciiCount += 1
			noDataCount += 1

	lines[len(lines)-1] = lines[len(lines)-1][:-1] #remove the last newline character
```

```
        output_ascii.writelines(lines)
        output_ascii.flush
        output_ascii.close()
        del output_ascii
        in_mem_name = r'in_memory\temp'+os.path.splitext(tablefilename)[0]
        arcpy.ASCIIToRaster_conversion(output_ascii_path, in_mem_name, 'INTEGER')
        arcpy.CopyRaster_management(in_mem_name, output_img_path, '', '', '65535',
                '','','16_BIT_UNSIGNED', '','')
        arcpy.Delete_management(in_mem_name)
        eTime = (time.time()-sTime)/60
        #write a Quality Assessment file
        qa_file_path = os.path.splitext(output_ascii_path)[0] + r'_qa.txt'
        qa_file = open(qa_file_path, 'w')
        qa_file.write('No. Rows in DBF: ' + str(table.recordCount) + '\n')
        qa_file.write('No. Rows from DBF included: ' + str(trCounter) + '\n')
        qa_file.write('No. Rows written to ASCII file: ' + str(asciiCounter) + '\n')
        qa_file.write('No. Rows of added NoData (32767): ' + str(noDataCounter) + '\n')
        qa_file.write('Time taken to create raster: %.2f minutes' % eTime)
        qa_file.flush()
        qa_file.close()
        print r'Minutes Elapsed for ',os.path.splitext(tablefilename)[0],r': ',eTime

if __name__ == '__main__':
        table_list = []
        for tablefilename in os.listdir(os.path.join(tablefile_base,r'tables')):
                if os.path.splitext(tablefilename)[1] == r'.dbf':
                        table_list.append(tablefilename)
        pool = multiprocessing.Pool(7)
        pool.map(build_raster_from_table, table_list)
        pool.close() #Synchronize the main process with the job processes to cleanup
        pool.join()
```