Classification Techniques in the Strong Intermittent Sequences of Vegetation Species and Bare Soil

M.I. Bitenbaev [a,*], B.V. Geldyev [b], A.A. Loktionov [a], N.P. Ogar [c], A.I. Polyakov [a]

[a] The Institute of Physics and Technology, 050032 Almaty 32, Kazakhstan – (bmarat@sci.kz) [b] Center for Remote Sensing and GIS "Terra", Almaty, [c] The Institute of Botany and Phytointroduction, Almaty.

Abstract - The arid zone of the south-eastern Kazakhstan is characterized with the strong heterogeneity of spatial and spectral distributions for neighbouring sites. The sites of bare soil and different species of vegetation form the strong intermittent pattern.
Applications of the signal theory methods and modern software allow one to solve the problem of classification in such intermittent environment. For a high dimensional space, as dimensionality increases, the volume of a hypercube concentrates in the corners. On this basis techniques have been developed to extract endmember spectra from remotely sensed data set.
The final outputs of this work are classification maps of site distributions for vegetation and bare soil, suitable for input into a GIS - project that will be able to provide the tools for the problems of effective ecosystem planning.

Keywords: arid zone, vegetation species, bare soil, heterogeneity, classification.

1. INTRODUCTION

Remote sensing data are used for precision farming, see, for example, site of American Society of Agricultural Engineers - ASAE (ASAE). Here it is important to point out the actual problems: assessing crop and soil condition, measuring crop parameters such as plant canopy (McNairn, 2001), water content, nitrogen, chlorophyll, and leaf area (Pathak, 2004). Classification techniques for these goals involve a broad variety of approaches that allow one to study a wide range of problems (Bonnell, 2004). However, the fields of cultural farming have the regular shape: rectangles, circles, etc.
The goal of this report is to perform the analysis for the strong intermittent sequences of different vegetation species and bare soil in the arid zone of the Shu river valley in Kazakhstan.
The IKONOS-2 satellite images with resolution 1 meter in the panchromatic band and 4 meters for the blue, green, red and near-infra-red bands are used.
Applications of the signal theory methods and modern software allow one to solve the problem of classification in the strong intermittent environment.

2. STUDY AREA AND MEASUREMENTS

The study was executed in the area of Shu river and its surroundings in the south-eastern of the Kazakhstan. The test sites consist of natural vegetation fields and bare soil sites. Ground measurements were acquired using the FieldSpec@HandHeld UV/VNIR Spectroradiometer.
Three – five measurements were taken at each sampling site in order to minimize errors and, thus, provide a representative average of the sample spectra. In order to reflect within-field variability, approximately seven sampling sites were selected per field.

3. DATA

In order to obtain reflectance spectra, spectral data of IKONOS-2 sensor (IKONOS) with four bands have been used. The first multispectral channel - MS-1 (Blue): 444.7 – 516.0 nm; center - 480.3 nm. The second - MS-2 (Green): 506.4 – 595.0 nm; center - 550.7 nm. The third - MS-3 (Red): 631.9 – 697.7 nm; center - 664.8 nm. Fourth - MS-4 (NIR = Near InfraRed): 757.3 - 852.7nm; center – 805.0 nm. Spatial resolution of this imagery is 4 m.

4. APPROACHES TO THE ANALYSES

4.1. Search for an Endmember.

The test sites – test points on a large geographic scale – cover only the very small part of the investigated area. What distribution of the principal components of an environment – bare soil, different species of vegetation - take place on the whole test region? It is obvious that in the natural environment all distributions of centers of vegetation or bare soil sites have the random character. For the quantitative study of the strong intermittent sequences of different objects we have to consider rather spectral structures of data than spatial ones: a spectral endmember is the spectral signature for a "pure" ground component. In other words an endmember is the set of data values defining one component in a mixture.
In the mathematical signal theory has been shown (Landgrebe, 2003) that for a high dimensional space, as dimensionality increases, the volume of a hypercube concentrates in the corners. Indeed, the fraction *FRACT* of the volume of a hypercube $(2r)^d$ in $[-r, r]^d$, contained in a hypersphere

$$\{2 r^d \pi^{d/2}\} / \{d\, \Gamma(d/2)\}$$

with the same radius r, inscribed in it, decreases as the dimensionality d increases:

$$FRACT = \pi^{d/2} / \{d\, 2^{d-1}\, \Gamma(d/2)\}$$

The limit of $FRACT\,(d \rightarrow \infty) = 0$ which implies that the volume of the hypercube is increasingly concentrated in the corners as d increases.

On this basis techniques have been developed to extract endmember spectra from remotely sensed data set. These in general involve the representation of the set of all scene pixels as a scatter plot in spectral space or some subspace thereof. Therefore, when pixel data are plotted in a scatter plot which uses image bands as plot axis, the spectrally purest pixels always occur in the corners of the data cloud, while spectrally mixed pixels always occur on the inside of this data cloud.

In general, the endmembers are vertices of the simplex. A simplex is a triangle in the plane – two dimensional scatter plot, a tetrahedron in 3-space – three dimensional scatter plot, etc.

The software system ENVI (ENVI 4.1, 2004) has excellent tools for 2D and nD analyses.

Ground measurements collected during summer field campaign 2003 and 2004 have shown that there are at least three endmembers on the test region. Therefore, experimentally search for the endmembers have been performed in the 2D scatter plot.

The example of the 2D scatter plot used to select the spectrally unique endmember surface of the test area, (without river) are shown in Fig. 1. Results confirm the analogous observations made in the field measurements. The broad right corner of the plot allows to recognize the additional type of vegetation - vegetation 3.

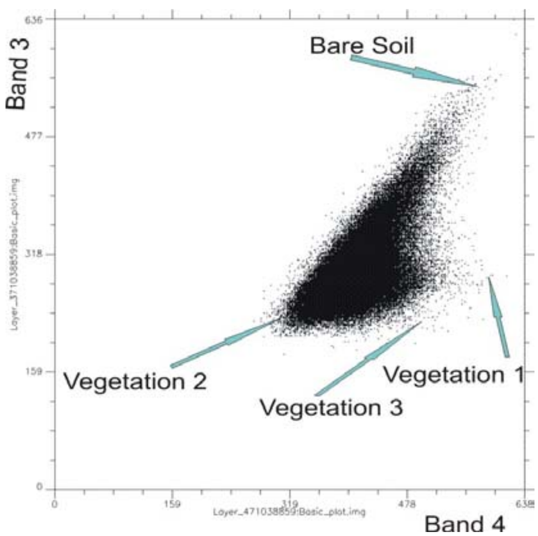

Figure 1. 2D scatter plot with the dominant spectral endmembers identified.

The reflection of points under the corners in the 2D simplex to the geographic map has shown that, on the one hand, all experimentally found endmembers have strong intermittent spatial distributions, and, on the other hand, all they correspond to the basic ground structure, collected during summer field campaign.

4.2. Image Analysis

Annotated, pseudo-color infrared image of the Shu river valley are presented on the Fig 2. The study area created from IKONOS bands 4, 3, 2, (RGB = 805.0 nm, 664.8 nm, 550.7 nm.).

It contains two test sites marked by arrows.

The bare soil sites show up as white areas because of gently sloping of the reflectance distributions at these bands. The red areas indicate vegetation, which reflects more in the near infrared than in green visible wavelengths. The river show up as dark area because near infrared wavelengths (757 – 853 nm) are absorbed in the very upper surface of water.

The characteristic peculiarity of the map on Fig. 2 consists in the strong intermittent structure of spatial distributions for sites of bare soil and different types of vegetation.

4.3. Unsupervised and Supervised Classification

In the problem of classification the best result are obtained on the basis of application of iterative sequence of unsupervised and supervised approaches (Lillesand, 2000). The quantitative study of remote sensing spectra in 2D scatter plot approximation has shown, that there are, at least, four area with different types of spectrally unique materials. These results are in the good coincidence with the field data observations.

Unsupervised classification – Cluster analysis – is performed when the analyst has no prior knowledge of the area of interest. In this type of classification an algorithm groups pixels with similar spectral characteristics together. In the report the ISODATA clustering algorithm with four classes established the initial group of centers and allocated each pixel to the nearest group according to its spectral characteristics. ISODATA makes several iterations through the data until the number of pixels that change groups during an iteration is less than a give threshold.

The result of the unsupervised classification can be presented as "unsupervised" map.

Supervised classification is carried out on an area with (partially) known features. This procedure involves the manual selection – field expedition – of some areas with unique spectral reflectance.

The first step in this approach is to find the genuine endmembers, which are the basic spectral constituents of the pixels within the vegetation and soil fields. Based on ground cover knowledge of these fields, four endmembers spectra were manually extracted from the reflectance image: bare soil and three species of vegetation, see Fig. 1.

Since the availability of pure pixels under natural field conditions is problematic, pure patches of soil and different species of vegetation were chosen on the fields. Endmember spectra were then extracted from these ten by ten meters patches. Although these patches were not exactly "pure", the selected endmember spectra were the "purest " spectra available from the field data. However, soil patches were 100% soil.

The classification algorithm have to find the desired classes. In this report the correlation classifier – the Spectral Angle Mapper (SAM) algorithm computes a spectral angle between each pixel spectrum and each target spectrum. The smaller spectral angle corresponds to the more similar target spectra. As increasing or decreasing of

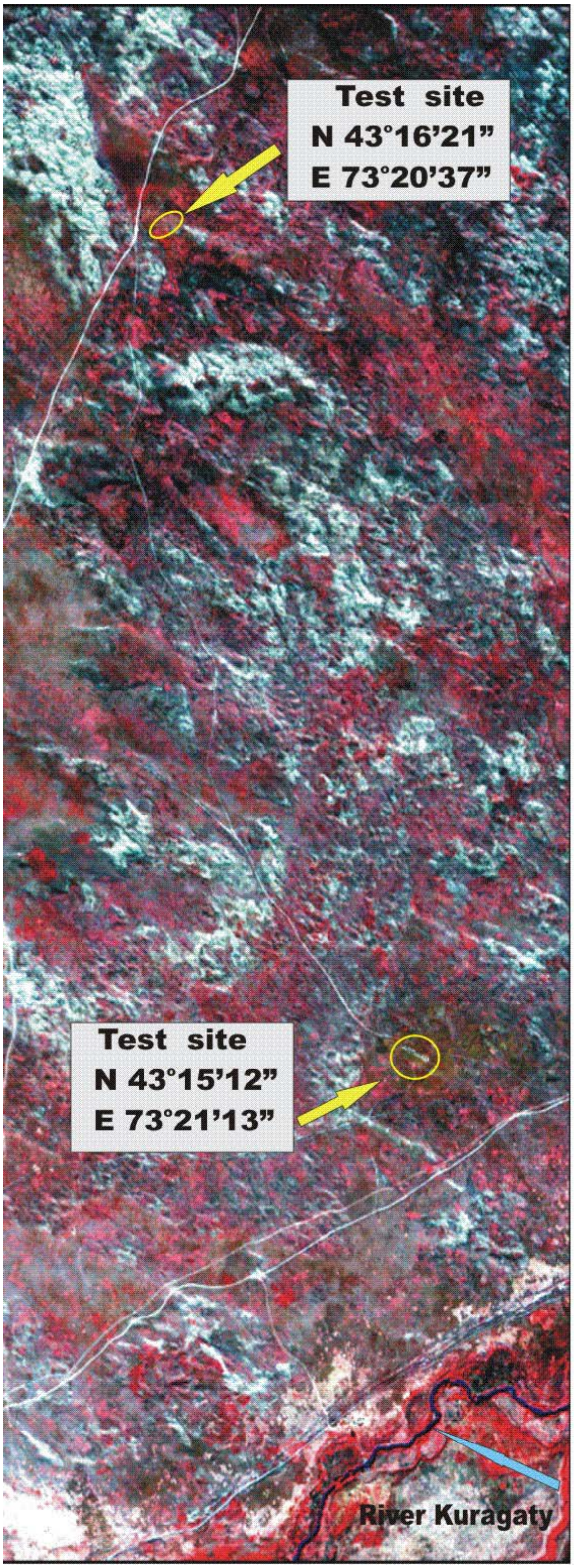

Figure 2. The map of the test region with two test sites.

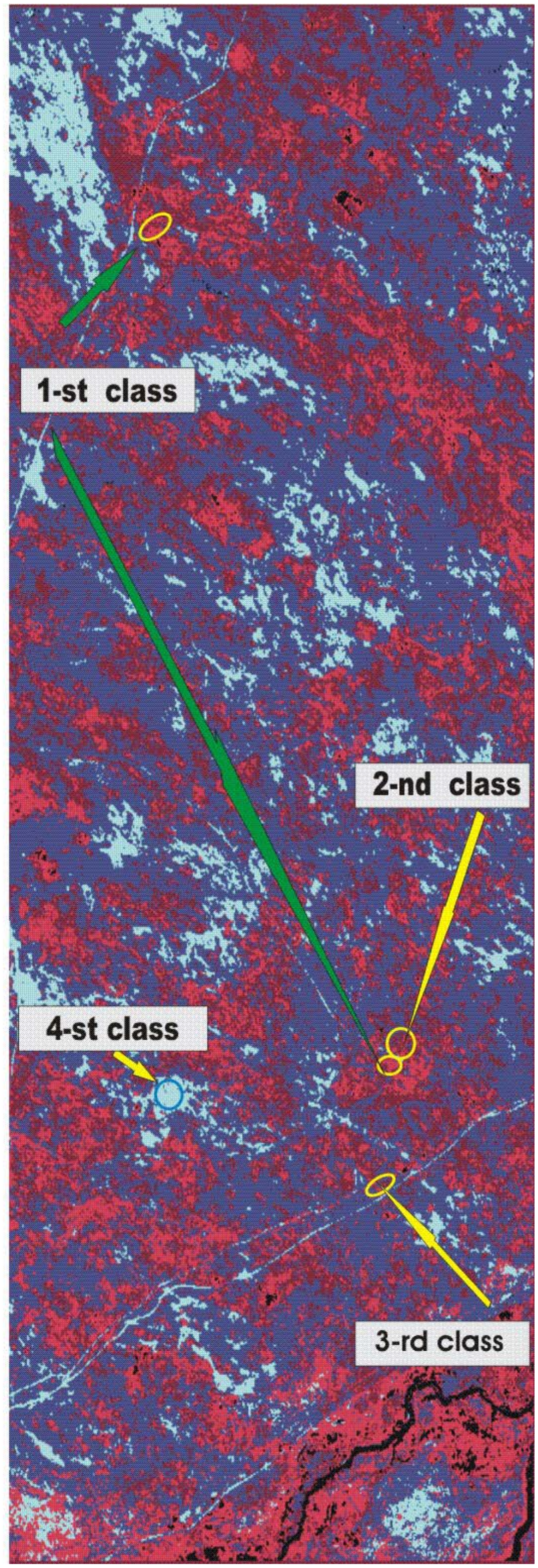

Figure 3. The map of the classification for the test region.

illumination doesn't change the direction of the vector, but only its magnitude, the spectral angle will be relatively insensitive to change in pixel illumination.

The result of the supervised classification can be presented as the "supervised" map.

The high degree of similarity of these supervised and unsupervised maps tell us about right results of the classification: the supervised and unsupervised analyses have led to the close conclusions. Therefore, the unique single map of the classification can be obtained. Such result of supervised classification is presented on the Figure 3. Each class is shown with corresponding color as the totality of spatially distributed sites.

5. SUMMARY

1. The arid zone of the south-eastern Kazakhstan is characterized with the strong heterogeneity of spatial and spectral distributions for neighbouring sites. The sites of bare soil and different species of vegetation form the strong intermittent pattern.

2. Data from sensors with high spatial resolution and modern mathematical and software tools allow one to classify these strong intermittent distributions.

3. This result might open new avenues ultimately leading towards deeper insight into long-distance properties of arid zones and the unsolved problems of identifying, mapping, and monitoring natural and anthropogenic effects for these ecosystems.

6. REFERENCES

ASAE – American Society of Agricultural Engineers, http://www.asae.org,

ASAE Technical Library, http://asae.frymulti.com

R.B. Bonnell, P.Dutillenl, Y.K. Zindashty et al, Classification accuracy of discriminant analysis artifical neural networks and decision trees for weed and nitrogen detection stress in corn, ASAE International Meeting, Aug. 1-4, 2004, Ottawa, Canada.

ENVI software, http://www.rsinc.com

IKONOS, http://www.spaceimaging.com

D.A. Landgrebe, Signal Theory Methods in Multispectral Remote Sensing, John Wiley& Sons Inc., 2003.

T.M. Lillesand, R.W. Kiffer, Remote Sensing and Image Interpretation, John Wiley& Sons Inc., Fourth Edition, 2000.

H. McNairn, J. C. Deguise et al., Estimating of crop cover and chlorophyll from hyperspectral remote sensing, Proc. of the 23rd Canadian Symp. on Remote Sensing, Aug., 2001, Quebec, Canada.

T.B. Pathak, Generating the relationship between vegetation index and crop yield and validation of existing crop yield model, ASAE International Meeting, Aug. 1-4, 2004, Ottawa, Canada.