Optimal Sampling Schemes for Vegetation and Geological Field Visits

Pravesh Debba

CSIR, Spatial Planning Systems, Decision Support and Systems Analysis, CSIR Built Environment
Optimal Sampling Schemes for Vegetation and Geological Field Visits

1. Introduction

2. Classification

3. Optimized sampling schemes case studies
   - Optimized field sampling for improved estimates of vegetation indices
   - Optimized field sampling representing the overall distribution of a particular mineral
Sample - small subset of the population of interest.
Sample should represent the characteristics of the population (parameters / distribution).
Draw inferences about a population based on incomplete knowledge.
Distinguish between two general approaches
  - Design-based Methods
    - ‘Ignore’ the spatial structure
    - Use some form of random sampling
    - Use feature space to design sample
  - Model-based Methods
    - Explicitly model the spatial structure
    - Selection of sample based on optimisation criterion
    - Use geographic space to design sample
IMPORTANCE OF OPTIMAL SAMPLING SCHEMES

- Environmental studies:
  - where to sample?
  - what to sample?
  - and how many samples to obtain?

- Remote sensing as ancillary information in the design of optimal sampling schemes.

- Advantages of using remote sensing images:
  - Provides a synoptic overview of a large area
  - Wealth of information over the entire area
  - In these methods sampling avoids subjective judgement
  - Reduces costs and saves time on the field (fewer samples)
Hyperspectral sensors

- record the reflectance in many narrow contiguous bands
- various parts of the electromagnetic spectrum (visible - near infrared - short wave infrared)
- at each part of the electromagnetic spectrum results in an image
Imaging spectrometry for monitoring tree damage caused by volcanic activity in the Long Valley caldera, California. This area suffers from volcanic activity, including frequent earthquakes, hydrothermal activity, and gas emissions. The research area is situated around Mammoth Mountain, a volcanic cone rising 760,000 years ago. After a period of rest (the last eruption was about 500 years ago), the area has since 1980 been suffering from frequent earthquakes, hydrothermal activity, and uplift of approximately 60 cm since 1980. The U.S. Geological Survey has measured an uplift of approximately 60 cm since 1980.

The Long Valley caldera measures approximately 17 km x 26 km and was formed by a large eruption about 760,000 years ago. The caldera is inflating; the U.S. Geological Survey has measured an uplift of approximately 60 cm since 1980.

Drought conditions have persisted in the area, leading to the death of trees. In 1990, areas of dying forests were found on the flanks of Mammoth Mountain. At first, the cause of tree death was sought in the persisting drought of the preceding years. However, trees died regardless of age or species, as shown in Figure 3. Research [22] revealed that high concentrations of carbon dioxide (30 to 350 parts per million) were measured in the area, with the highest concentrations found in the Long Valley caldera and on the flanks of Mammoth Mountain.

The die-off of trees is believed to be due to root damage from elevated carbon dioxide levels. The increase in carbon dioxide concentration deprives the roots of oxygen, resulting in trees that are under stress and ultimately die. Imaging spectrometry yields important information on tree conditions and on the presence of dead vegetative material. The use of imaging spectrometry to map the spatial extent of dead and stressed tree areas was successful.

The Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) was used for monitoring tree damage by volcanic activity. AVIRIS acquires images at an altitude of 14 km and has a nominal bandwidth of 10 nm. Each pixel has an associated, continuous spectrum that can be used to identify the surface materials. The hyperspectral cube generated by imaging spectrometers has bandwidths of only 20 nm or less. Since the continuous spectrum cannot resolve diagnostic spectral features of terrestrial or soil properties because their bandwidth of 70 to 240 nm cannot resolve diagnostic spectral features, only airborne imaging spectrometers can be used. To date, only airborne imaging spectrometers are available, but several imaging spectrometers are planned for the next 5 years.

Figure: Hyperspectral cube

ABSTRACT

In the last decade, a new remote sensing technique was introduced: imaging spectrometry. Imaging spectrometers measure a continuous spectrum that can be used to identify the surface materials. The concept of imaging spectrometry is shown in Figure 1. Conventional broad-spectrum sensors, such as multispectral and hyperspectral sensors, cannot resolve diagnostic spectral features. To date, only airborne imaging spectrometers have been used to identify materials or their properties by diagnostic absorption features. To date, only airborne imaging spectrometers have been used to identify materials or their properties by diagnostic absorption features. Figure 1 shows the concept of imaging spectrometry. Conventional broad-spectrum sensors, such as Spot-XS, Landsat MSS, and MODIS, cannot resolve diagnostic spectral features. To date, only airborne imaging spectrometers have been used to identify materials or their properties by diagnostic absorption features.

Developments in detector technology have triggered a new remote sensing technology: imaging spectrometry. Imaging spectrometers measure a continuous spectrum that can be used to identify the surface materials. Each pixel has an associated, continuous spectrum that can be used to identify the surface materials. The hyperspectral cube generated by imaging spectrometers has bandwidths of only 20 nm or less. Since the continuous spectrum cannot resolve diagnostic spectral features of terrestrial or soil properties because their bandwidth of 70 to 240 nm cannot resolve diagnostic spectral features, only airborne imaging spectrometers can be used. To date, only airborne imaging spectrometers are available, but several imaging spectrometers are planned for the next 5 years.

Figure: Hyperspectral cube
Figure: Pixels in hyperspectral image
OVERVIEW OF HYPERSPECTRAL REMOTE SENSING (cont...)

Figure: Example of 3 different spectral signatures
Optimal Sampling Schemes for Vegetation and Geological Field Visits

Debba

Introduction

Classification

Optimized sampling schemes case studies

Optimized field sampling for improved estimates of vegetation indices

Optimized field sampling representing the overall distribution of a particular mineral

UNSUPERVISED CLASSIFICATION

- No previous knowledge assumed about data.
- Tries to spectrally separate the pixels.
- User has controls over:
  - Number of classes
  - Number of iterations
  - Convergence thresholds
- Two main algorithms: Isodata and k-means
A set number of cluster centres are positioned randomly through the spectral space.

Pixels are assigned to their nearest cluster.

The mean location is re-calculated for each cluster.

Repeat 2 and 3 until movement of cluster centres is below threshold.

Assign class types to spectral clusters.
K-MEANS CLUSTERING (cont. . .)

(a) 1st iteration. Cluster centres are set at random. Pixels assigned to the nearest centre.

(b) 2nd iteration. Centres move to the mean-centre of all pixels in this cluster.

(c) N-th iteration. Centres have stabilised.
Optimal Sampling Schemes for Vegetation and Geological Field Visits

Introduction
Classification

Optimized sampling schemes case studies

Optimized field sampling for improved estimates of vegetation indices

Optimized field sampling representing the overall distribution of a particular mineral

---

ISODATA

- Extends k-means. Also calculate standard deviation for clusters.
- After stage 3 we can either:
  - Combine clusters if centres are close.
  - Split clusters with large standard deviation in any dimension.
  - Delete clusters that are too small.
- Then reclassify each pixel and repeat.
- Stop on max iterations or convergence limit.
- Assign class types to spectral clusters.
ISODATA (cont. . . )

(d) Data is clustered but blue cluster is very stretched in band 1.

(e) Cyan and green clusters only have 2 or less pixels. So they will be removed.

(f) Either assign outliers to nearest cluster, or mark as unclassified.
Optimal Sampling Schemes for Vegetation and Geological Field Visits

Debba

Introduction

Classification

Optimized sampling schemes case studies

Optimized field sampling for improved estimates of vegetation indices

Optimized field sampling representing the overall distribution of a particular mineral

SUPERVISED CLASSIFICATION

- Start with knowledge of class types.
- Classes are chosen at start.
- Training regions are created for each class.
- Ground truth used to verify the training regions.
- Quite a few algorithms. Here we will look at:
  - Parallelepiped
  - Maximum likelihood
Training classes plotted in spectral space. In this example using 2 bands.
For each training region determine the range of values observed in each band.

These ranges form a spectral box (or parallelepiped) which is used to classify this class type.

Assign new image pixels to the parallelepiped which it fits into best.

Pixels outside all boxes can be unclassified or assigned to the closest one.

Problems with classes that exhibit high correlation between bands. This creates long ‘diagonal’ data-sets that do not fit well into a box.
For each training class the spectral variance and covariance is calculated.

The class can then be statistically modelled with a mean vector and covariance matrix.

This assumes the class is normally distributed. Which is generally okay for natural surfaces.

Unidentified pixels can then be given a probability of being in any one class.

Assign the new pixel to the class with the highest probability — or unclassified if all probabilities low.
Optimal Sampling Schemes for Vegetation and Geological Field Visits

Debba

MAXIMUM LIKELIHOOD (cont. . . )
Outline

1. Introduction

2. Classification

3. Optimized sampling schemes case studies
   - Optimized field sampling for improved estimates of vegetation indices
   - Optimized field sampling representing the overall distribution of a particular mineral
The design of the optimal prospective sampling scheme for field visits in an agricultural study, using a segmented hyperspectral image.

The optimal prospective sampling scheme will be representative of the whole study area for various parameters embedded by the segmentation and bands selected for the segmentation.
Study site – Tedej – Hungary.
Crops: barely, maize, sugar beet, sunflower, alfalfa.
Digital Imaging Spectrometer – DAIS-7915 – 79 channel hyperspectral image.
Spectral range from visible (0.4 \( \mu \text{m} \)) to thermal infrared (12.3 \( \mu \text{m} \)).
Spatial resolution 3–20 m depending on the carrier aircraft altitude.
Optimal Sampling Schemes for Vegetation and Geological Field Visits

Debba

Introduction

Classification

Optimized sampling schemes case studies

Optimized field sampling for improved estimates of vegetation indices

Optimized field sampling representing the overall distribution of a particular mineral

Figure: Study area in Tedej, Hajdu-Bihar area, Hungary.
Figure: Hyperspectral image of study area in Tedej, Hajdu-Bihar area, Hungary. Reflectance values for bands 29 (0.988 µm), 39 (1.727 µm) and 1 (0.496 µm).
Optimal Sampling Schemes for Vegetation and Geological Field Visits

Introduction

Classification

Optimized sampling schemes case studies

Optimized field sampling for improved estimates of vegetation indices

Optimized field sampling representing the overall distribution of a particular mineral

METHODS: ITERATED CONDITIONAL MODES (ICM) ALGORITHM

- Adequate image segmentation takes into account both spectral features and spatial information.
- Markov Random Fields (MRF) have been useful in this respect.
- For each category \( k = 1, 2, \ldots, K \), let
  - \( C_k^{(\alpha)} \) denote the set of pixels which belongs to the \( k \)th category and \( C^{(\alpha)} = \bigcup_{k=1}^{K} C_k^{(\alpha)} \) the segmented image at the \( \alpha \)th iteration, \( \alpha = 0, 1, 2, \ldots \),
  - \( N_k^{(\alpha)} \) denote the number of elements in \( C_k^{(\alpha)} \), i.e. the number of pixels in the \( k \)th category at the \( \alpha \)th iteration,
  - \( \mu_k^{(\alpha)} = \sum_{(i,j) \in C_k^{(\alpha)}} f_{ij} / N_k^{(\alpha)} \) be the \( m \)-dimensional mean vector of the \( k \)th category at the \( \alpha \)th iteration.
METHODS: ITERATED CONDITIONAL MODES (ICM) ALGORITHM (cont. . .)

\[
\text{arg min}_k \left\{ \left( f_{ij} - \mu_k^{(\alpha)} \right)^T \left( f_{ij} - \mu_k^{(\alpha)} \right) - \beta \nu^{(\alpha)} N_{ij}^{(\alpha)}(k) \right\}
\]  

(1)

\[
\nu^{(\alpha)} = \frac{1}{N} \sum_{k=1}^{K} \sum_{(i,j) \in C_k^{(\alpha)}} \left( f_{ij} - \mu_k^{(\alpha)} \right)^T \left( f_{ij} - \mu_k^{(\alpha)} \right) .
\]  

(2)
A second order MRF was applied in which the neighbors of each pixel consists of its eight adjacencies, with border pixels adjusted appropriately.

**Figure:** Calculation of \( N_{ij}^{(\alpha)}(k) \) for an arbitrary interior pixel \((i, j)\) belonging to category \(k\).
METHODS (cont. . . ): SAMPLE SIZE PER CATEGORY

For a pre-specified number of $n$ samples, the sample size for category $k$ equals

$$n_k = n_{(0)} + (n - K \cdot n_{(0)}) \cdot \frac{N_k^{(r)} \sqrt{\nu_k^{(r)}}}{K \sum_{t=1} N_t^{(r)} \sqrt{\nu_t^{(r)}}},$$

(3)

where $\nu_k^{(r)} = \frac{1}{N_k^{(r)}} \sum_{(i,j) \in C_k^{(r)}} (f_{ij} - \mu_k^{(r)})^T (f_{ij} - \mu_k^{(r)})$. 

For a pre-specified number of $n$ samples, the sample size for category $k$ equals
Simulated annealing — optimization method to find the global optimum of an objective function in the presence of local optima. A fitness function $\phi(S)$ has to be minimized. A probabilistic acceptance criterion decides whether $S_{i+1}$ is accepted or not:

$$P_c(S_i \rightarrow S_{i+1}) = \begin{cases} 1, & \text{if } \phi(S_{i+1}) \leq \phi(S_i) \\ \exp \left( \frac{\phi(S_i) - \phi(S_{i+1})}{c} \right), & \text{if } \phi(S_{i+1}) > \phi(S_i) \end{cases}$$
The initial sampling scheme for the $k$th category $S_k^{(0)}$ is a random selection of $n_k$ [see Equation 3] points from category $k$. For $S_k$, the fitness function equals

$$\phi_{\text{MMSD}}(S_k) = \frac{1}{N_k^{(r)}} \sum_{(i,j) \in C_k^{(r)}} \| c_k(ij) - W_{S_k}(c_k(ij)) \| , \quad (5)$$

where $c_k(ij) \in C_k^{(r)}$ is a location vector denoting the $(i, j)$th pixel belonging to category $k$ and $W_{S_k}(c_k(ij))$ denotes the location vector of the nearest sampling point in $S_k$. 
RESULTS: GENERATED SEGMENTED IMAGE

Figure: Generated segmented image.
RESULTS (cont. . .): OPTIMIZED SAMPLING SCHEME

**Figure:** Optimized sampling scheme.
Figure: Original hyperspectral image. Reflectance values for bands 29 (0.988 µm), 39 (1.727 µm) and 1 (0.496 µm).
RESULTS (cont. . . ): SEGMENTED IMAGE – 8 CATEGORIES

**Figure:** ICM Segmented image with eight categories.
RESULTS (cont. . .): SEGMENTED IMAGE – 4

ROI CATEGORIES

Figure: Segmented image confining sampling regions to the four categories.
RESULTS (cont. . .): OPTIMIZED SAMPLING SCHEME

**Figure:** Optimized sampling locations of 50 points distributed over 4 categories.
RESULTS (cont. . .): DIFFERENT VEGETATION INDICES

- Normalized Difference Vegetation Index (NDVI)
  \[
  \text{NDVI} = \frac{R_{0.886} - R_{0.675}}{R_{0.886} + R_{0.675}}
  \]  

- Renormalized Difference Vegetation Index (RDVI)
  \[
  \text{RDVI} = \frac{R_{0.886} - R_{0.675}}{\sqrt{R_{0.886} + R_{0.675}}}
  \]  

- Modified Simple Ratio (MSR)
  \[
  \text{MSR} = \left( \frac{R_{0.886}}{R_{0.675}} - 1 \right) \sqrt{\frac{R_{0.886}}{R_{0.675}} + 1}
  \]  

- Soil-Adjusted Vegetation Index (MSAVI)
  \[
  \text{MSAVI} = \frac{1}{2} \left[ (2R_{0.886} + 1)^2 - 8(R_{0.886} - R_{0.675}) \right]
  \]
## RESULTS (cont.): COMPARISON OF SAMPLING SCHEMES

<table>
<thead>
<tr>
<th>Sampling Scheme</th>
<th>Mean NDVI</th>
<th>Mean RDVI</th>
<th>Mean MSR</th>
<th>Mean MSAVI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image</td>
<td>0.59</td>
<td>8.8</td>
<td>1.34</td>
<td>1.24</td>
</tr>
<tr>
<td>Optimized sampling scheme</td>
<td>0.58</td>
<td>8.6</td>
<td>1.32</td>
<td>1.22</td>
</tr>
<tr>
<td>Random sampling scheme</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.49</td>
<td>7.9</td>
<td>1.18</td>
<td>1.09</td>
</tr>
<tr>
<td>2</td>
<td>0.38</td>
<td>6.1</td>
<td>0.94</td>
<td>0.89</td>
</tr>
<tr>
<td>3</td>
<td>0.45</td>
<td>7.0</td>
<td>1.11</td>
<td>1.06</td>
</tr>
<tr>
<td>Grid sampling scheme</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.49</td>
<td>7.8</td>
<td>1.14</td>
<td>1.13</td>
</tr>
<tr>
<td>2</td>
<td>0.53</td>
<td>8.2</td>
<td>1.25</td>
<td>1.13</td>
</tr>
<tr>
<td>3</td>
<td>0.53</td>
<td>8.3</td>
<td>1.26</td>
<td>1.15</td>
</tr>
</tbody>
</table>
Optimal Sampling Schemes for Vegetation and Geological Field Visits

Debba

Outline

1. Introduction

2. Classification

3. Optimized sampling schemes case studies
   - Optimized field sampling for improved estimates of vegetation indices
   - Optimized field sampling representing the overall distribution of a particular mineral
Using a hyperspectral image, to guide field sampling collection to those pixels with the highest likelihood for occurrence of a particular mineral, for example alunite, while representing the overall distribution of alunite.

Usefulness: To create a mineral alteration map
Optimal Sampling Schemes for Vegetation and Geological Field Visits

Introduction

Classification

Optimized sampling schemes case studies

Optimized field sampling for improved estimates of vegetation indices

Optimized field sampling representing the overall distribution of a particular mineral
Optimal Sampling Schemes for Vegetation and Geological Field Visits

Debba

Introduction
Classification
Optimized sampling schemes case studies
Optimized field sampling for improved estimates of vegetation indices
Optimized field sampling representing the overall distribution of a particular mineral

Legend
- abundant alunite
- minor alunite
- no alunite

Value

SPECTRAL ANGLE MAPPER (cont...
**Continuum Removal**

**Figure:** Concept of the convex hull transform; (A) a hull fitted over the original spectrum; (B) the transformed spectrum.

---

**Introduction**

**Classification**

**Optimized sampling schemes case studies**

**Optimal field sampling for improved estimates of vegetation indices**

**Optimized field sampling representing the overall distribution of a particular mineral**
Figure: SFF fit image for alunite. Lighter areas indicate better fit values between pixel reflectance spectra and the alunite reference spectrum.
Combination of SAM and SFF scaled to [0, 1] is defined as

$$w(\theta(\vec{x}), \tau_F(\vec{x})) = \begin{cases} \kappa_1 w_1(\theta(\vec{x})) + \kappa_2 w_2(\tau_F(\vec{x})), & \text{if } \theta(\vec{x}) \leq \theta^t \text{ and } \tau_F(\vec{x}) \geq \tau^t_F, \\ 0, & \text{if otherwise} \end{cases}$$

(10)

$$\phi_{\text{WMSD}}(S^n) = \frac{1}{N} \sum_{\vec{x} \in I} w(\vec{x}) \| \vec{x} - W_{S^n}(\vec{x}) \|_1$$

(11)
Optimal Sampling Schemes for Vegetation and Geological Field Visits

Debba

Introduction

Classification

Optimized sampling schemes case studies

Optimized field sampling for improved estimates of vegetation indices

Optimized field sampling representing the overall distribution of a particular mineral

METHODS (cont. . .): Fitness Function

Figure: Fitness function with different weights for $N = 15$. 

$$w(x)=1$$

$$w(x)=2$$
RESULTS (cont. . . ): OPTIMIZED SAMPLING SCHEME

Figure: Optimized sampling scheme.
RESULTS (cont. . . ): Distribution of 40 highest values

**Figure:** Sampling scheme: 40 highest values
RESULTS (cont. . .): Distribution of 40 optimized sampling scheme

Figure: Distribution of 40 optimized sampling scheme
This is a numerical measure of the quality of the sampling design. The most common are:

- Minimise the maximum kriging variance
- Minimise the average kriging variance
- Maximise the information in a sample variogram

Kriging variance does not depend on the observed values, but only on the spatial structure and the location of the sample points i.e. the only factors influencing the kriging variance are therefore the variogram, the number of observations and the location of the prediction point. This means that it is possible to calculate the kriging variance before actual sampling takes place, provided the variogram is known or can be assumed. This feature is used to optimise spatial sampling schemes for minimal kriging variance.
Example

\[ \phi_{\text{OK}}(S) = \frac{1}{N} \sum_{j=1}^{N} \sigma_{\text{OK}}^{2}(x_j | S) , \quad (12) \]

or

\[ \phi_{\text{MAX}}(S) = \max \left( \sigma_{\text{OK}}^{2}(x_j | S) \right) , \quad (13) \]

where

\[ \sigma_{\text{OK}}^{2}(x_0) = \sum_{i=1}^{N} \lambda_i \cdot \gamma (x_i - x_0) + \Phi , \quad (14) \]

where \( \lambda_i \) denotes the weight of the \( i \)th observation and \( \Phi \) a Lagrange multiplier.