

Optimal Sampling Schemes applied in Geology

Pravesh Debba

CSIR, South Africa

UP 2010 Presentation

Outline

- 1 Introduction to hyperspectral remote sensing
- 2 Objective of Study 1
- 3 Study Area
- 4 Data used
- 5 Methodology
- 6 Results
- 7 Background and Research Question for Study 2
- 8 Study Area and Data
- 9 Methodology
- 10 Results
- 11 Conclusions

Outline

- 1 Introduction to hyperspectral remote sensing
- 2 Objective of Study 1
- 3 Study Area
- 4 Data used
- 5 Methodology
- 6 Results
- 7 Background and Research Question for Study 2
- 8 Study Area and Data
- 9 Methodology
- 10 Results
- 11 Conclusions

Outline

- 1 Introduction to hyperspectral remote sensing
- 2 Objective of Study 1
- 3 Study Area
- 4 Data used
- 5 Methodology
- 6 Results
- 7 Background and Research Question for Study 2
- 8 Study Area and Data
- 9 Methodology
- 10 Results
- 11 Conclusions

Outline

- 1 Introduction to hyperspectral remote sensing
- 2 Objective of Study 1
- 3 Study Area
- 4 Data used
- 5 Methodology
- 6 Results
- 7 Background and Research Question for Study 2
- 8 Study Area and Data
- 9 Methodology
- 10 Results
- 11 Conclusions

Outline

- 1 Introduction to hyperspectral remote sensing
- 2 Objective of Study 1
- 3 Study Area
- 4 Data used
- 5 Methodology
- 6 Results
- 7 Background and Research Question for Study 2
- 8 Study Area and Data
- 9 Methodology
- 10 Results
- 11 Conclusions

Outline

- 1 Introduction to hyperspectral remote sensing
- 2 Objective of Study 1
- 3 Study Area
- 4 Data used
- 5 Methodology
- 6 Results
- 7 Background and Research Question for Study 2
- 8 Study Area and Data
- 9 Methodology
- 10 Results
- 11 Conclusions

Outline

- 1 Introduction to hyperspectral remote sensing
- 2 Objective of Study 1
- 3 Study Area
- 4 Data used
- 5 Methodology
- 6 Results
- 7 Background and Research Question for Study 2
- 8 Study Area and Data
- 9 Methodology
- 10 Results
- 11 Conclusions

Outline

- 1 Introduction to hyperspectral remote sensing
- 2 Objective of Study 1
- 3 Study Area
- 4 Data used
- 5 Methodology
- 6 Results
- 7 Background and Research Question for Study 2
- 8 Study Area and Data
- 9 Methodology
- 10 Results
- 11 Conclusions

Outline

- 1 Introduction to hyperspectral remote sensing
- 2 Objective of Study 1
- 3 Study Area
- 4 Data used
- 5 Methodology
- 6 Results
- 7 Background and Research Question for Study 2
- 8 Study Area and Data
- 9 Methodology
- 10 Results
- 11 Conclusions

Outline

- 1 Introduction to hyperspectral remote sensing
- 2 Objective of Study 1
- 3 Study Area
- 4 Data used
- 5 Methodology
- 6 Results
- 7 Background and Research Question for Study 2
- 8 Study Area and Data
- 9 Methodology
- 10 Results
- 11 Conclusions

Outline

- 1 Introduction to hyperspectral remote sensing
- 2 Objective of Study 1
- 3 Study Area
- 4 Data used
- 5 Methodology
- 6 Results
- 7 Background and Research Question for Study 2
- 8 Study Area and Data
- 9 Methodology
- 10 Results
- 11 Conclusions

OVERVIEW OF HYPERSPECTRAL REMOTE SENSING

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

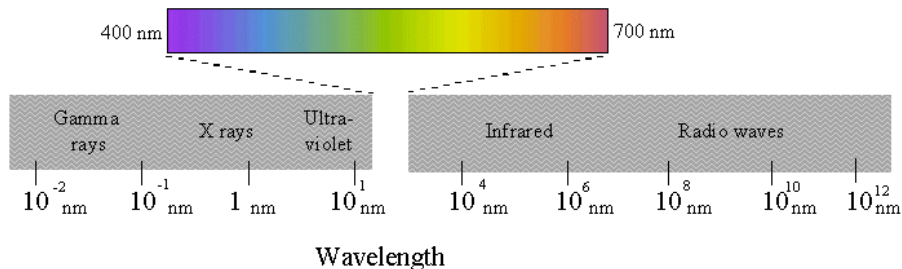


Figure: Spectral Range

OVERVIEW OF HYPERSPECTRAL REMOTE SENSING (cont...)

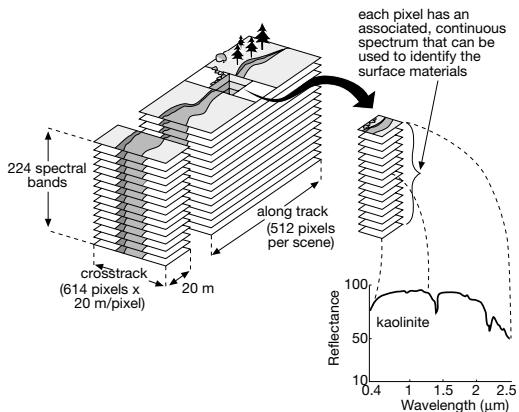


Figure: Hyperspectral cube

OVERVIEW OF HYPERSPECTRAL REMOTE SENSING (cont. . .)

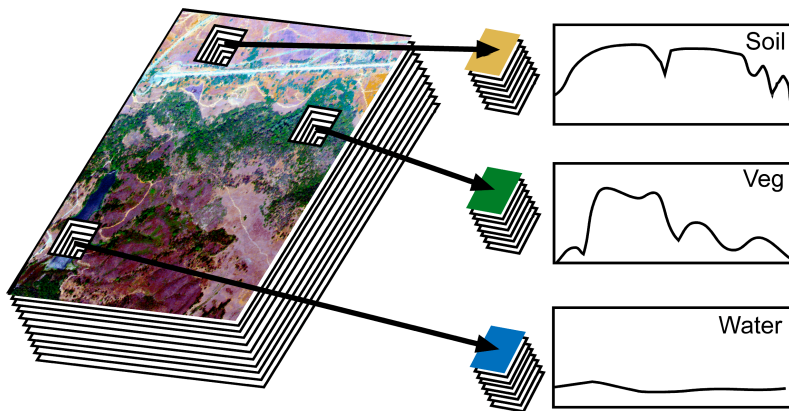


Figure: Pixels in hyperspectral image

OVERVIEW OF HYPERSPECTRAL REMOTE SENSING (cont. . .)

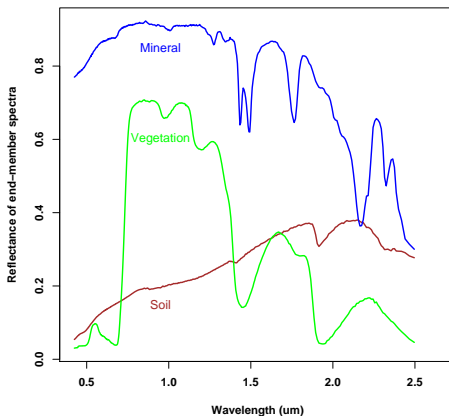


Figure: Example of 3 different spectral signatures

OBJECTIVE OF STUDY 1

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

STUDY SITE

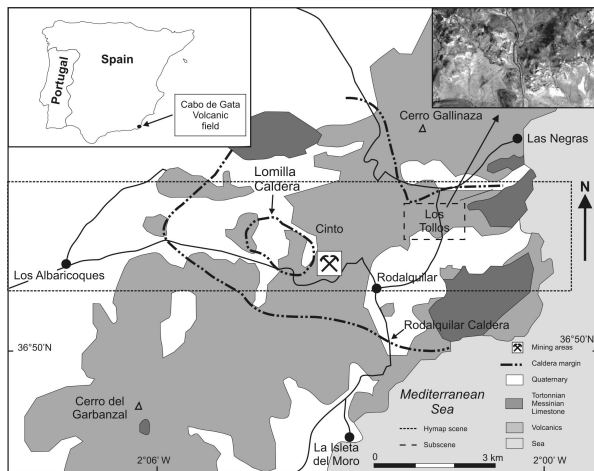


Figure: A generalized geological map of the Rodalquilar study area showing the flight line and the hyperspectral data

DATA USED

- HyMap: 126 bands – 0.4–2.5 μm
- Geology: 30 bands – 1.95–2.48 μm
- Distinctive absorption features at wavelengths near 2.2 μm
- We collected field spectra during the over-flight using the Analytical Spectral Device (ASD) fieldspec-pro spectrometer – 0.35–2.50 μm

ENDMEMBER SPECTRA

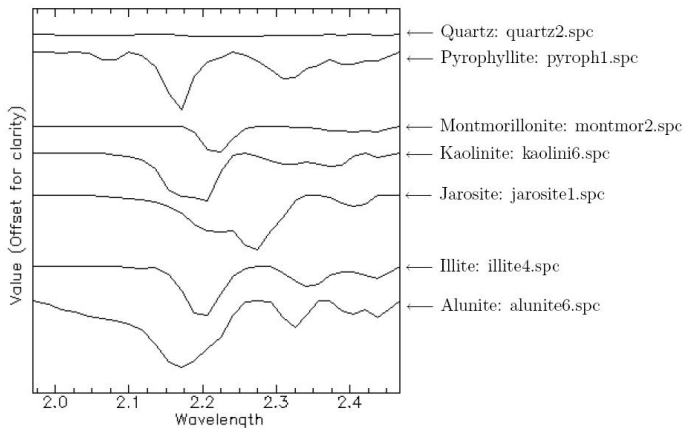


Figure: Plot of 7 endmembers from USGS spectral library for the 30 selected bands, enhanced by continuum removal.

CONTINUUM REMOVAL

Spectra are normalized to a common reference using a continuum formed by defining high points of the spectrum (local maxima) and fitting straight line segments between these points. The continuum is removed by dividing it into the original spectrum.

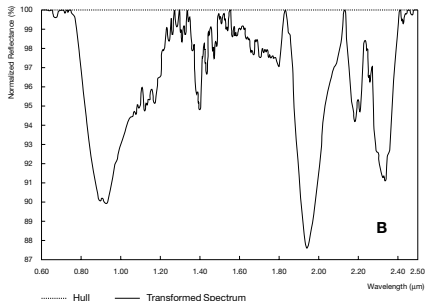
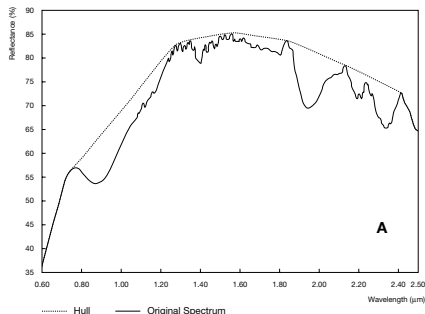



Figure: Concept of the convex hull transform; (A) a hull fitted over the original  spectrum; (B) the transformed spectrum.

METHODS: Spectral Angle Mapper (SAM) Classifier

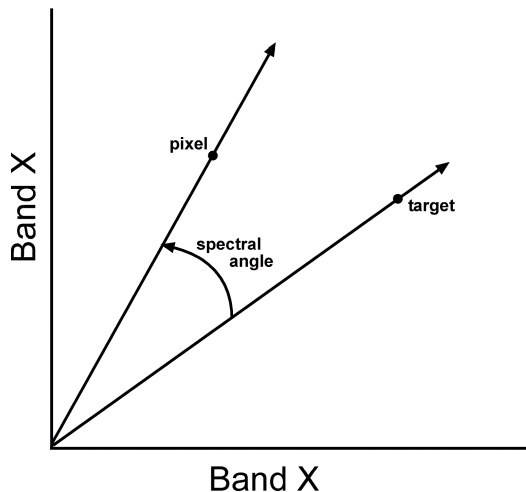
- SAM – pixel based supervised classification technique
- Measures the similarity of an image pixel reflectance spectrum to a reference spectrum
- Spectral angle (in radians) between the two spectra

$$\theta(\vec{\mathbf{x}}) = \cos^{-1} \left(\frac{f(\lambda) \cdot e(\lambda)}{\|f(\lambda)\| \cdot \|e(\lambda)\|} \right), \quad (1)$$

$f(\lambda)$ – image reflectance spectrum and $e(\lambda)$ – reference spectrum.

- Results in a gray-scale rule image – values are the angles

METHODS (cont. . .): Spectral Angle Mapper (SAM) Classifier



METHODS (cont. . .): SAM Rule Image for Alunite

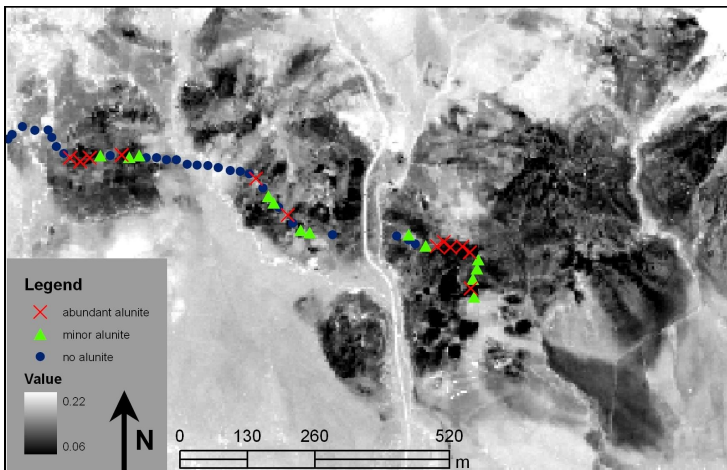


Figure: SAM classification rule image for alunite. Dark areas indicate smaller angles, hence, greater similarity to alunite.

METHODS (cont. . .): Spectral Feature Fitting (SFF)

- SFF – pixel based classification technique.
- Remove the continuum from both the reference and unknown spectra.
- SFF produces a scale image for each endmember selected for analysis by first subtracting the continuum-removed spectra from one (inverting it), and making the continuum zero.
- SFF determines a single multiplicative scaling factor that makes the reference spectrum match the unknown spectrum.

METHODS (cont. . .): Spectral Feature Fitting (SFF)

- SFF then calculates a least-squares-fit, band-by-band, between each reference endmember and the unknown spectrum.
- The total root-mean-square (RMS) error is used to form an RMS error image for each endmember.
- Scale/RMS provides a fit image that is a measure of how well the unknown spectrum matches the reference spectrum on a pixel-by-pixel basis.

METHODS (cont. . .): SFF Rule Image for Alunite

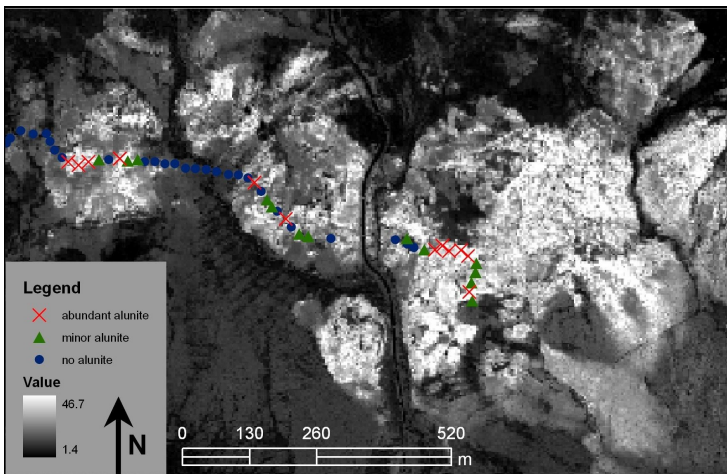


Figure: SFF fit image for alunite. Lighter areas indicate better fit values between pixel reflectance spectra and the alunite reference spectrum.

METHODS (cont...): Fitness Function

SAM values scaled to $[0, 1]$

$$w_1(\theta(\vec{\mathbf{x}})) = \begin{cases} 0, & \text{if } \theta(\vec{\mathbf{x}}) > \theta^t \\ \frac{\theta^t - \theta(\vec{\mathbf{x}})}{\theta^t - \theta_{\min}}, & \text{if } \theta(\vec{\mathbf{x}}) \leq \theta^t \end{cases}, \quad (2)$$

SFF values scaled to $[0, 1]$

$$w_2(\tau_F(\vec{\mathbf{x}})) = \begin{cases} 0, & \text{if } \tau_F(\vec{\mathbf{x}}) < \tau_F^t \\ \frac{\tau_F(\vec{\mathbf{x}}) - \tau_F^t}{\tau_{F,\max} - \tau_F^t}, & \text{if } \tau_F(\vec{\mathbf{x}}) \geq \tau_F^t \end{cases}, \quad (3)$$

METHODS (cont...): Fitness Function

Combination of SAM and SFF scaled to $[0, 1]$ is defined as

$$w(\theta(\vec{\mathbf{x}}), \tau_F(\vec{\mathbf{x}})) = \begin{cases} \kappa_1 w_1(\theta(\vec{\mathbf{x}})) + \kappa_2 w_2(\tau_F(\vec{\mathbf{x}})), & \text{if } \theta(\vec{\mathbf{x}}) \leq \theta^t \text{ and } \tau_F(\vec{\mathbf{x}}) \geq \tau_F^t \\ 0, & \text{if otherwise} \end{cases} \quad (4)$$

$$\phi_{\text{WMSD}}(\mathbf{S}^n) = \frac{1}{N} \sum_{\vec{\mathbf{x}} \in \mathbf{I}} w(\vec{\mathbf{x}}) \|\vec{\mathbf{x}} - W_{\mathbf{S}^n}(\vec{\mathbf{x}})\|, \quad (5)$$

METHODS (cont...): Fitness Function

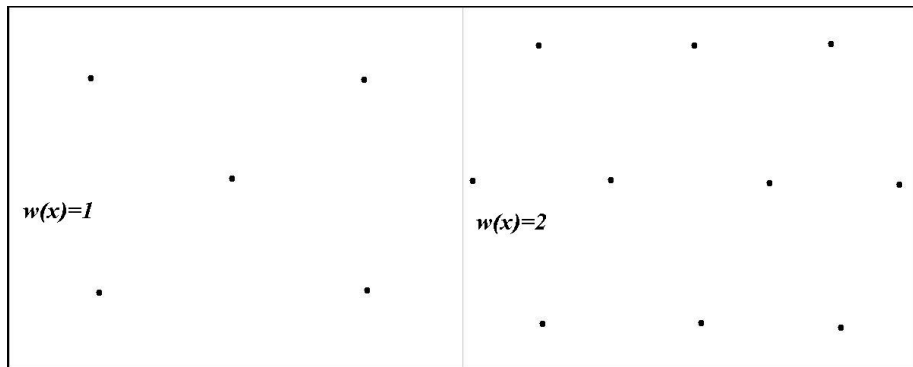


Figure: Fitness function with different weights for $N = 15$.

RESULTS OF THE OPTIMIZED SAMPLING SCHEME

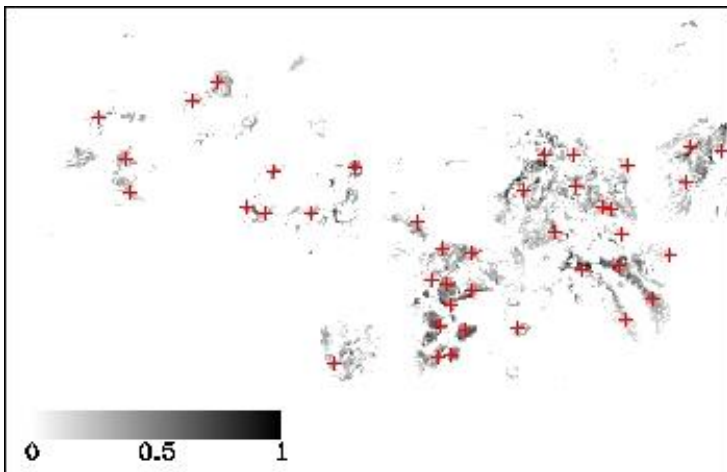


Figure: Optimized sampling scheme.

RESULTS (cont...): Distribution of 40 optimized sampling scheme

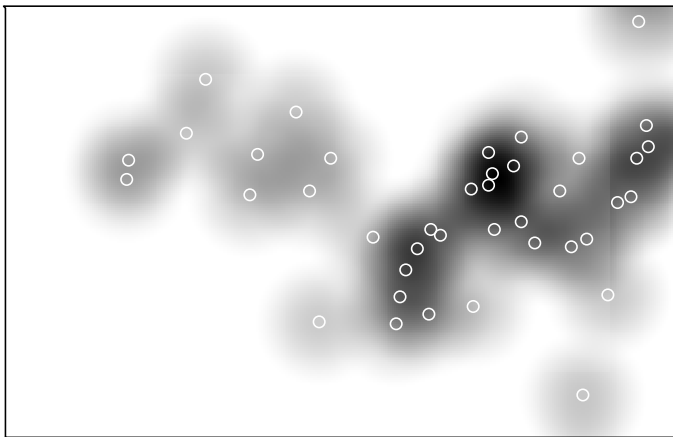


Figure: Distribution of 40 optimized sampling scheme

RESULTS (cont...): Distribution of 40 highest values

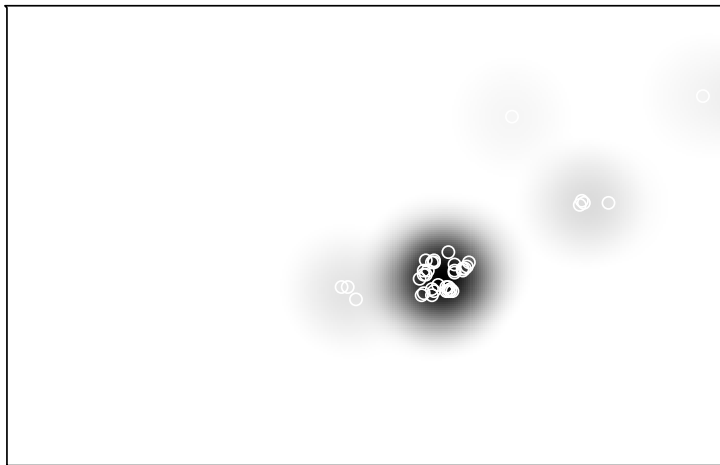
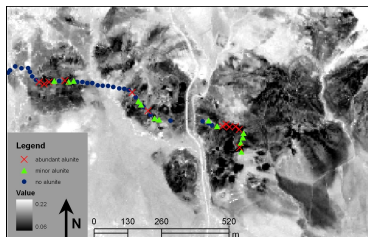
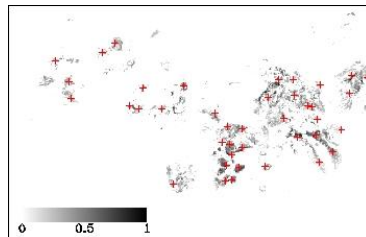


Figure: Sampling scheme: 40 highest values

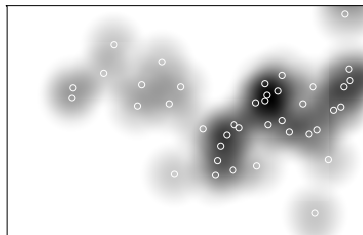
RESULTS (cont...): SUMMARY COMPARISON



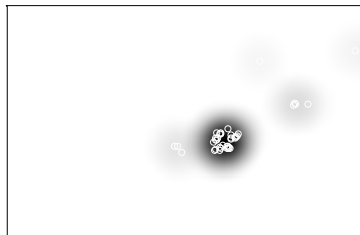
(a) SAM Classification



(b) 40 Optimized points



(c) Distribution sampling pts



(d) Distribution highest points

STUDY 2

- Mine wastes contains high concentrations of metals.
- Metals — leached from mine wastes — then released to and contaminating nearby ecosystems.
- Geochemical characterization of mine waste impoundments – important for rehabilitation; remediation; protect the surrounding environment and ecosystems.
- Effective geochemical characterization – entails surface (to subsurface) sampling – labor or cost intensive.
- Metals in mine waste impoundments – hosted by acid-generating sulphide-rich minerals (pyrite, pyrrhotite), or adsorbed onto surfaces of weathering products of such sulphide-rich minerals.
- Such minerals are difficult to detect or identify by using current remote sensing techniques including multispectral or even hyperspectral data.

- Certain sulphide-rich minerals, particularly pyrite, weathers to a series of iron-bearing sulfates, hydroxides and oxides (shown by Swayze et al., 2000).
- Such secondary iron-bearing sulfates/hydroxides/oxides have diagnostic spectral features – enables their detection or identification with analytical techniques using hyperspectral data (Crowley et al., 2003).
- Debba et al. (2005) showed the potential of using hyperspectral data to estimate abundances of spectrally similar iron-bearing sulfates/hydroxides/oxides.
- Kemper & Sommer (2002) showed that heavy metal contamination in soils can be quantified using reflectance spectroscopy.

- Remote sensing provides an indirect tool for surface characterization of mine waste impoundments with oxidizing sulphide-rich materials; namely, for mapping spatial distributions of secondary iron-bearing sulfates/hydroxides/oxides and heavy metals.

Hence, given a model of spatial distribution of secondary iron-bearing oxides/hydroxides, the problem is how to design a sampling scheme that would adequately capture the spatial distribution of certain groups of metals.

- A prospective sampling scheme is derived for nearby unsampled areas based on the variogram model of the adjacent sampled area.

STUDY AREA

- The present case study area is in the Recsk-Lahóca copper mining area in Hungary. The Recsk-Lahóca mining area is situated in the Mátra Mountains, about 110 km northeast of Budapest, Hungary.
- The Lahóca hill was mined for copper between 1852 and 1979.
- Mining of ore deposits in the Recsk-Lahóca area resulted in the exposure of sulphide bearing-rocks to surface water and atmospheric oxygen, which accelerate oxidation, leaching and release of metals and acidity.
- This study pertains to the tailings dumps northwest of Lahóca mine, which consist actually of two dumps referred to as “East Tails” and “West Tails” .

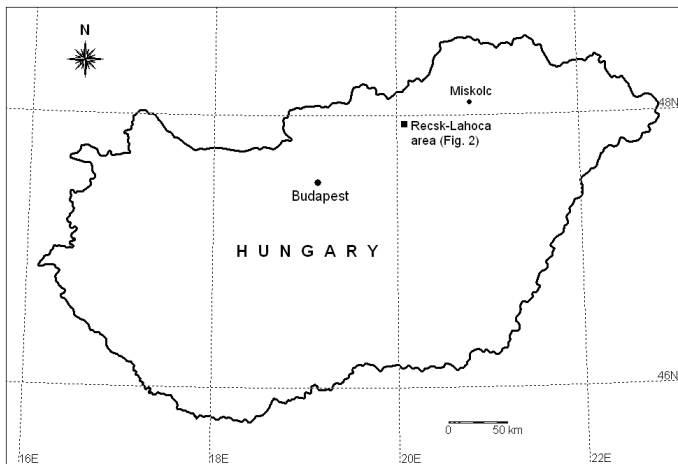


Figure: Study area: Recsk: Hungary.

THE DATA – HYPERSENSPECTRAL

- A subset of the Digital Airborne Imaging Spectrometer (DAIS-7915) is used.
- The resulting data is a 79 channel hyperspectral image, acquired over the Recsk. DAIS-7915 is a whisk broom sensor, covering a spectral range from visible ($0.4 \mu\text{m}$) to thermal infrared ($12.3 \mu\text{m}$) at variable spatial resolution from 3–20 m depending on the carrier aircraft altitude.
- Not all 79 channels were useful as many channels were too noisy and could not be corrected efficiently. Fortunately, the first 32 channels, spectral range 406–1035 nm, where iron-bearing oxides/hydroxides/sulphates have diagnostic features were found useful for this study.
- Samples from the tailings – collected shortly after collection of the DAIS hyperspectral data.

THE DATA – FIELD

- 53 samples were collected in the East Tails and 44 in the West tails – 10m×10m grid points.
- Concentrations of As, Cd, Cu, Fe, Mn, Ni, Pb, Sb and Zn in the decomposed samples were determined using the ICP-AES analyzer.

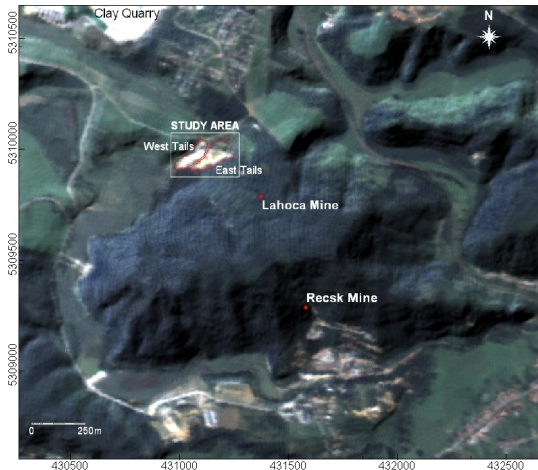


Figure: The Recsk-Lahóca area shown in pseudo-natural color composite image using DAIS data (red = ch10, green = ch5, blue = ch1) fused with a digital elevation model. Map coordinates are in meters (UTM projection, zone 34N).

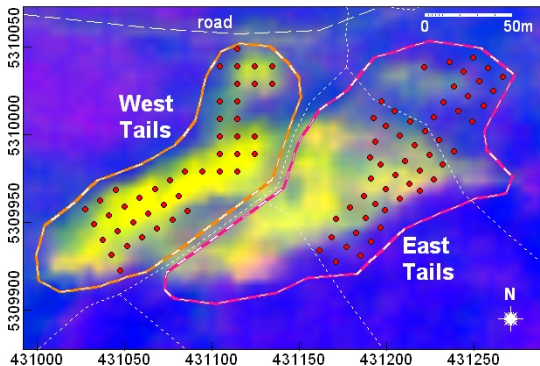


Figure: The “East Tails” and the “West Tails” shown in a color composite image of the DAIS data. Ratios of ch17 to ch28 (representing ferrihydrite reflectance and absorption peaks) was used as red band, ch13 to ch25 (representing jarosite reflectance and absorption peaks) was used as green band and ch32 to ch1 (representing non-iron-bearing minerals) was used as blue band. Red dots are locations of mine tailings samples. Short dashed lines indicates drainage lines of either active or non-active streams.

SPLITTING THE DATA

- The East and West Tails have different geochemical characteristics — split the data into two sets.
- Data from either sub-area are used to model a relationship between heavy metal associations and relative abundances of secondary iron-bearing minerals.
- The latter data are derived from spectral unmixing of hyperspectral data. See: Debba *et. al.* (2006). Abundance estimation of spectrally similar materials by using derivatives in simulated annealing, *IEEE Geoscience and Remote Sensing*, vol. 44, no. 12, 3649–3658.
- A model relationship between heavy metal associations and mineral abundances in one sub-area is then used as basis for optimal sampling design in the other sub-area.
- Division of the area and the data thus provides calibration analysis and prediction/validation analysis for optimal sampling design.

MODELING OF HEAVY METAL ASSOCIATIONS

- A factor component analysis with varimax rotation was performed on the logarithmic-transformed heavy metal concentrations to obtain the heavy metal association of interest.
- The scores of FA2E and FA2W — linearly transformed to $[0, 1]$ (for numerical compatibility with the mineral abundance estimates) — FA2ET and FA2WT.

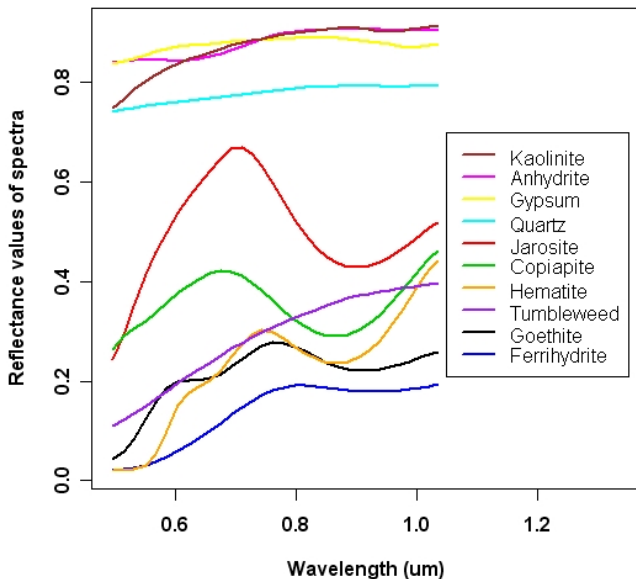


Figure: Reflectances of minerals which are common in contaminated areas.

KRIGING WITH EXTERNAL DRIFT

- Kriging with external drift is applicable to estimate primary variables of interest, which are practically measurable at only few sample sites, based on linearly related ancillary variables, which are measurable at much higher sampling density than the primary variables.
- Kriging with external drift is ideal if a primary variable could be measured more precisely and practically at a few locations (factor scores of heavy metal associations), whereas possibly less accurate measurements of linearly related ancillary variables are available everywhere in the spatial domain (relative abundances of metal-scavenging iron-bearing minerals – hyperspectral image).

KRIGING WITH EXTERNAL DRIFT

- The experimental semi-variogram $\gamma^*(h)$, where h is a fixed lag vector in both distance and direction, may be obtained from $\kappa = 1, 2, \dots, P(h)$ pairs of observations $\{z(\mathbf{x}_\kappa), z(\mathbf{x}_\kappa + h)\}$ at locations $\{\mathbf{x}_\kappa, \mathbf{x}_\kappa + h\}$, as:

$$\gamma^*(h) = \frac{1}{2 \cdot P(h)} \sum_{\kappa=1}^{P(h)} [z(\mathbf{x}_\kappa) - z(\mathbf{x}_\kappa + h)]^2 . \quad (6)$$

KRIGING WITH EXTERNAL DRIFT

- The k ancillary variables represented as regionalized variables $y_i(\mathbf{x})$, $i = 1, \dots, k$ with n_A observations, are less accurate measurements covering the whole domain \mathbf{A} at small scale and are considered as deterministic. The values $\{y_i(\mathbf{x})\}$ needs to be known at all locations \mathbf{x}_α of the samples as well as at the nodes of the estimation grid.
- Since $Z(\mathbf{x})$ and the set of $\{y_i(\mathbf{x})\}$ are two ways of expressing the same phenomenon, assume that $Z(\mathbf{x})$ is an average equal to a linear function of the set of $\{y_i(\mathbf{x})\}$ up to a constant b_0 and coefficients b_i , $i = 1, \dots, k$,

$$E[Z(\mathbf{x})] = b_0 + \sum_{i=1}^k b_i \cdot y_i(\mathbf{x}) = \sum_{i=0}^k b_i \cdot y_i(\mathbf{x}) , \quad (7)$$

where $y_0(\mathbf{x}) = 1$.

KRIGING WITH EXTERNAL DRIFT

- Assuming $Z(\mathbf{x})$ is a second order stationary random function, then

$$Z^*(\mathbf{x}_0) = \sum_{\alpha=1}^{n_A} \lambda_{\alpha} Z(\mathbf{x}_{\alpha}) \quad (8)$$

where λ_{α} denotes the weight of the α th observation and is constraint to unit sum.

- The kriging variance can then be written as

$$\sigma_{\text{KED}}^2(\mathbf{x}_0) = \text{Var}[Z(\mathbf{x}_0) - Z^*(\mathbf{x}_0)] . \quad (9)$$

KRIGING WITH EXTERNAL DRIFT

- The only factor influencing the kriging variance are the variogram $\gamma(h)$, the number of observations n_A , the sampling locations \mathbf{x}_α and the location \mathbf{x}_0 . This means that the kriging variance does not depend on the observations themselves, but rather only on their relative spacing. The advantage is that it can be used to optimize sampling schemes in advance of data collection.

PROSPECTIVE SAMPLING: SIMULATED ANNEALING

- The optimization procedure by simulated annealing is then performed by application of a criterion called the Mean Kriging Variance with External Drift (MKVED), the fitness function of which is defined as

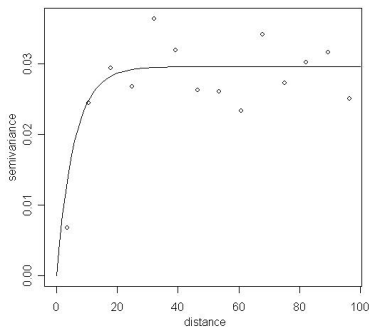
$$\phi_{\text{MKVED}}(\mathbf{S}) = \frac{1}{n_{\mathbf{A}}} \sum_{j=1}^{n_{\mathbf{A}}} \sigma_{\text{KED}}^2(\mathbf{x}_{\mathbf{A},j} | \mathbf{S}) , \quad (10)$$

where $n_{\mathbf{A}}$ is the number of raster nodes for which data for each of the covariates are available.

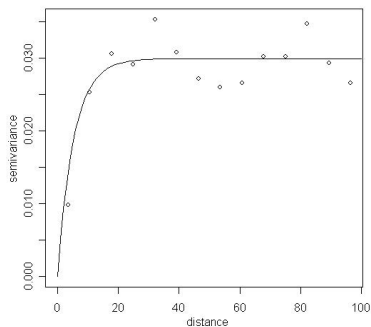
- The MKVED-criterion is proposed to derive the optimal prospective sampling scheme in an unvisited area based on a relevant model from a previously sampled area.

A PROSPECTIVE SAMPLING SCHEME

- A prospective sampling scheme for the West Tails is derived based on a model for the East Tails.
- As an illustration, it was decided to derive a prospective sampling scheme having 30 samples in the West Tails using the 53 samples from the East Tails.
- The exponential variogram was estimated with the data from the East Tails.
- To verify that this variogram is also appropriate for the West Tails, the East and West Tails data were combined.
- The similarity of the two variograms indicate that the variogram for the East Tails could be appropriate for modeling the West Tails.



(a) East Tails



(b) East and West Tails combined

Figure: The exponential variogram for the East Tails, combined East & West Tails.

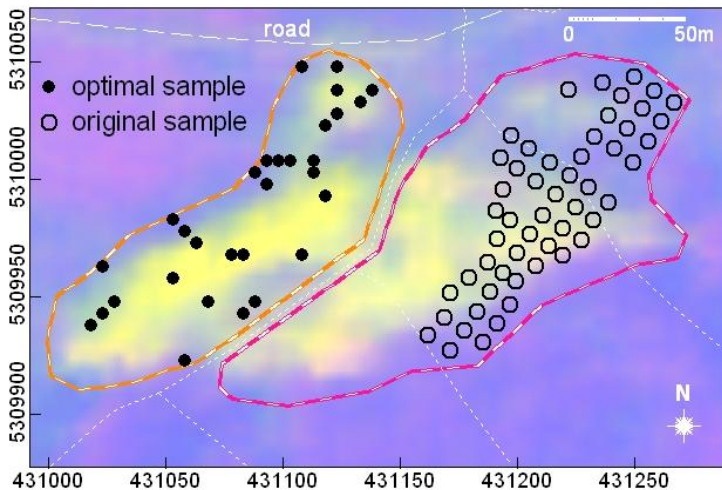


Figure: Prospective optimal sampling scheme in the West Tails using East Tails samples.

A PROSPECTIVE SAMPLING SCHEME

- The optimal sampling scheme constructed using the kriging external drift variance approach are spread over the West Tails region while retaining some close pairs of samples.
- These close pair samples are to improve the estimation of the variogram model.
- The mean kriging with external drift variance for the West Tails, using the combined East and West Tails sampling data, is 6.8×10^{-4} for the West Tails.
- This mean kriging variance was approximately the same when either of the two variograms was used.
- The optimal sampling scheme resulted in a mean kriging with external drift variance for the West Tails of 3.3×10^{-4} using the variogram derived from the East Tails data.
- This indicates that the optimal sampling scheme contains samples that reduces the mean kriging with external drift variance for the previously designed grid sampling scheme in the West Tails.

WHAT HAVE WE LEARNT?

- The spatial relationships between heavy metals and metal-scavenging minerals can be modeled adequately by kriging with external drift.
- The use of secondary information in designing optimal sampling schemes was also illustrated. Often these secondary information can be achieved at a relatively low cost and available over a greater region. These are the primary reasons for incorporating this information into the sampling design.
- Optimized sampling schemes using the mean kriging with external drift variance will result in sampling schemes that explicitly take into account the nature of spatial dependency of the data and together with hyperspectral data can be used to design sampling schemes in nearby unexplored areas.