GEOCHEMICAL SAMPLING SCHEME OPTIMIZATION ON MINE WASTES BASED ON HYPERSPECTRAL DATA

Zhao Ting[1][2], Pravesh Debba[3][2], Alfred Stein[2]

[1] Institute of Mineral and Resource of China, Academy of Geological Science, 10037
[2] International Institute for Geo-Information Science and Earth Observation (ITC), Hengelosestraat 99, PO Box 6, 7500 AA Enschede, The Netherlands
[3] University of South Africa (UNISA), PO Box 392, Pretoria, Gauteng, South Africa, 0003

KEY WORDS: Hazards, Hyperspectral, Sampling, Pollution, GIS

ABSTRACT:

Spatial sampling optimization is an important issue for both geo-chemists and geo-statisticians. Many spatial sampling optimization methods have been previously developed. In this paper, we present a spatial simulated annealing method is presented using hyperspectral data.This sampling method was applied in a project concerning environment assessment of the Dexing Copper Mine. Mine waste contains high concentrations of metals, mostly of a non-economic value. Most of them are discharged without any decontamination, for example, acid-generating minerals. Acid rock drainage can adversely have an impact on the quality of drinking water and the health of riparian ecosystems. To assess or monitor environmental impact of mining, sampling of mine waste is required. Optimal geochemical sampling schemes, which focus on ground verification of mine wastes extracted from hyperspectral data, was derived automatic from a JAVA program.Hyperspectral data help to identify ground objects by a larger spectral range. Spectral angle mapper classification technique is carried out to obtain rule images. A rule image provides weights that are utilized in defining the objective function for the sampling scheme. These are optimized by means of simulated annealing. The simulated annealing uses the Weighted Means Shortest Distance (WMSD) criterion between sampling points. The scaled weight function intensively samples areas where an abundance of weathering mine waste occurs. A threshold is defined to constrain the sampling points to certain areas of interest.

1. OVERVIEW OF OPTIMIZATION OF SAMPLING SCHEMES

To obtain a better cost-effect spatial sampling method, geo-statistical research has been carried out. In comparing traditional sampling schemes Burgess found that a regular grid results in only slightly less precise estimates than a triangular grid, for the same sampling density. Different grid designs can reach the same precision with different sampling costs(Burgess, 1981) . Some specially designed grids can be sampled at a lower price than a regular grid, while maintaining the same precision (Debba, 2006).

We distinguish two kinds of sampling schemes. First, a retrospective sampling scheme that contains sample locations that are already planned. During assessment, some points will be removed from or added to the existing sampling scheme. Second, a prospective scheme contains sample locations that are pre-determined before actual sampling in the field. Both of these schemes require optimization. Optimization of a sampling scheme means a reduction in the numbers of sampling points with the same or even a higher accuracy of certain unknown parameters.

In this paper we employed the second sampling scheme to explore the abundance data from an area where no prior field work is done. The mine waste deposits normally cover large area. Some piles already exist for ages. Intensive sampling over the whole area is costly. With the aid of remote sensing data the design of an optimal sampling scheme is vital to provide the researcher with the relevant information about mine wastes.

2. METHODOLOGY

Figure 1 shows graphically the overview of the sampling process. The sampling process consists of three parts, of which each has several sub-processes.

First, we need to decide on the area to perform the sampling scheme. We need to know the size of the area and how many samples would be deployed in this area. Prior knowledge or background information can play a key role in arriving at a suitable answer to the above. For example, the geological setting in the area, the deposit type, mineral type, land cover, land user and weather are important traits in making such decisions. Such information helps our work through the whole period. For example, based on weather information we can predict the approximate time for satellite to acquire data as the quality of hyperspectral satellite image strongly relies on weather condition. If too much cloud gathers in the air, the satellite can not get an image of a sufficient high quality. After the acquisition of an image, a quality assessment should be performed before further analysis.

The second stage is the analysis of the image data. During this stage the ground weathering waste rock should be identified and mapped. Atmospheric correction is a prerequisite to most hyperspectral imagery data analysis approach. During the correction, the atmospheric disturbances should be removed. Normally, cloud areas should be masked and then further analysis will exclude these areas. After image preparation, the most important stage is ground weathering waste rock spectral discrimination and acquirement. We must ensure what we wish to identify from the image data. The ground weathering waste rock spectral data can be acquired from three different ways. First, the best option is by means of field observed spectral data. Second, requires the collection from a spectral library. Third, is to directly detect the spectrum from images if we know the exact position of these rocks. Then classification techniques, for example, Spectral Angle Mapper (SAM) can be implemented to match the image spectrum and reference spectrum. It results into a rule image which gives a quantitative representation of the ground object.

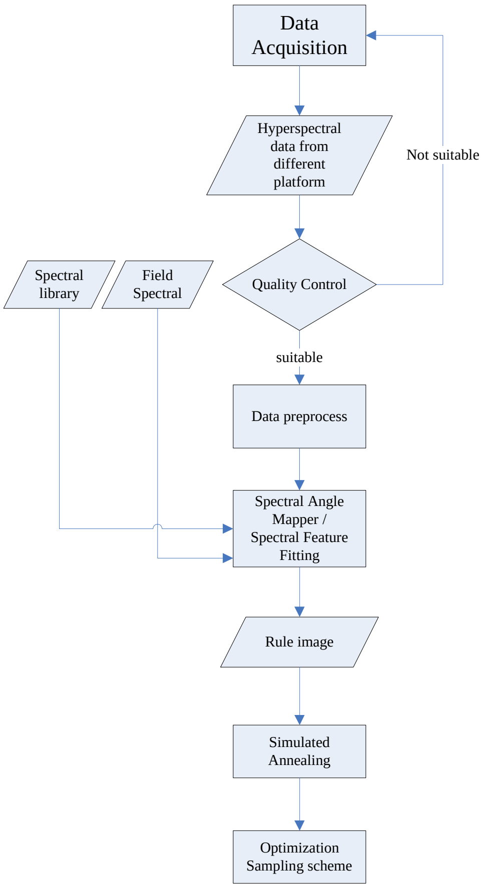

Figure 1. Workflow of the optimization of sampling scheme

The third stage is the sampling optimization, which is based upon an optimization algorithm. In this assignment, spatial simulated annealing (SSA) was selected. SSA is a very effective method to find the global optimum. A fitness function must be defined to support SSA algorithm. For different applications of SSA, different fitness functions can be defined. In this case we defined a Weighted Means Shortest Distance (WMSD) fitness function. In this case, weight refers to the angle value of rule image. Lower angle has a lower weight. Subsequently a software program is being developed based on SSA and WMSD. In the sampling stage the first input is a collection of randomly distributed sampling points. The SSA will calculate new samples to replace old sample points. Then it will calculate the fitness function value. When the fitness function value converges to a stable state, the final result scheme is considered as optimized.

3. BACKGROUND OF DEXING COPPER MINE

Dexing copper mine is one of the largest open-cast copper mine in the world(Figure 2) in Jiangxi province China. Along with the copper mine development, a city was built. At the same time the development of the mine results into local environment degradation. Its ore production in recent years is more than ten thousand tons per day. Big amounts of waste rock were dumped in nearby area. River water is acidified by the process of sulfide oxidation (mainly microbial oxidation of pyrite) and is then partly neutralized by hydrolysis reactions with aluminosilicates and other minerals present in the waste rock. This leads to the accumulation of Fe sulfates, oxyhydroxides and oxides in a spatial and temporal sequence.

Figure 2. Open-cast field

4. DERIVING OPTIMAL SAMPLING SCHEMES BASED ON SPATIAL SIMULATED ANNEALING

Spatial simulated annealing is presented to optimize spatial sampling schemes. The workflow is showed in figure3. Sampling schemes are optimized at the point-level, taking into account concentration of contaminant on the ground. The workflow indicates input parameters and key steps during optimization.

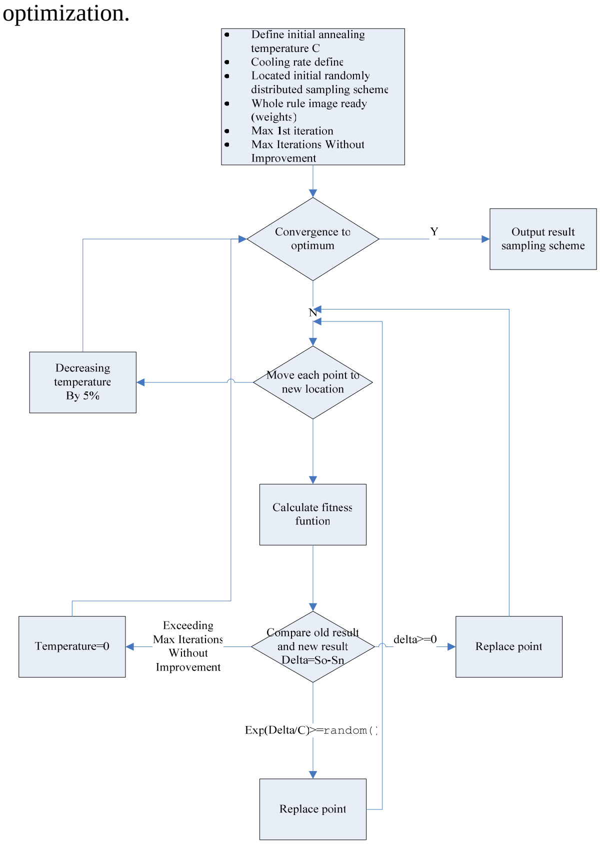

Figure 3. Schematic overview of Spatial Simulated Annealing sampling

4.1 Transition probabilities

Our goal of optimization is to get best sampling scheme with a fixed number of samples. In case of tailing sampling, the optimization means making transition from the randomly points to new points if the new point has lower angle value (new state). The probability of transition is a function (see equation 1) $P_c(S_i \rightarrow S_{i+1})$ of two states, where state 1 is denoted by $\phi(S_i)$ and state 2 is denoted by $\phi(S_{i+1})$.

$$P_c(S_i \rightarrow S_{i+1}) = \begin{cases} 1, & if\, \phi(S_{i+1}) \le \phi(S_i) \\ \exp(\frac{\phi(S_i) - \phi(S_{i+1})}{c}), & if\, \phi(S_{i+1}) > \phi(S_i) \end{cases} \ldots\ldots\ldots\ldots(1)$$

Figure4 shows a fitness function $\phi(S_i)$ should be defined which has to be minimized. ***S*** represents the sampling configuration. Starting with a random sampling schemes S_0, then S_i and S_{i+1} represent two solutions with fitness $\phi(S_i)$ and $\phi(S_{i+1})$, respectively. Sampling scheme S_{i+1} was derived from S_i . $P_c(S_i \rightarrow S_{i+1})$ is a probabilistic acceptance criterion that decides whether new sampling scheme S_{i+1} meet the fitness criteria.

In equation 1, *c* denotes a positive control parameter, and is usually called the temperature in simulated annealing problems. The parameter *c* is lowered according to a cooling schedule as the process evolves, to find the global minimum. If probabilistic acceptance of S_{i+1} is higher than the acceptance value which we predefine, a solution S_{i+2} will be derived from S_{i+1}. The $P_c(S_{i+1} \rightarrow S_{i+2})$ will be calculated (Debba, 2006).

4.2 Fitness function

SSA is employed to find global optimum sampling scheme. It starts by randomly selecting a collection of points in an image. In case of Dexing mine, 131 points were randomly located on the preparing stage based on a Hyperion image obtained from NASA. During the process of optimization, a new point in the image is then randomly selected and replaces a randomly selected old point from the current collection. This displacement occurs based on a probabilistic criterion. This criterion has to be translated into a fitness function, in order to be used in spatial simulated annealing. It this study the fitness function Weighted Means Shortest Distance (WMSD)–criterion has been used.

The WMSD–criterion is a weighted version of the Minimization of the Mean Shortest Distances (MMSD)–criterion (Groenigen, 1997). We first discuss the MMSD–criterion. In this case let ***S*** represented a sampling scheme, consisting of n_s sampling points. The q^{th} sampling point is denoted with $p_{s,q}$, with coordinates $x_{s,q}$ and $y_{s,q}$. Furthermore let A_R represent all points in the image.

$$\phi_{MMSD}(S) = E[\min\{\delta(\mathrm{P}_r, P_{s,i}), i \in 1,\ldots,n_s\}] \ldots\ldots(2)$$

where δ represents the Euclidean distance between two points. This expectation can be estimated by substituting the average value for the expectation:

$$\hat{\phi}_{MMSD}(S) = \sum_{j=1}^{n_e} \frac{\min\{\delta(P_{e,j}, P_{s,i}), i \in 1,\ldots,n_s\}}{n_e} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3)$$

Based on MMSD, WMSD take weights into account. These weights were extracted from the rule image. Then fitness function includes a location dependent weight function $\omega(\vec{X})$ that is scaled to [0, 1] .

$$\phi WMSD(S^n) = \frac{1}{N}\sum \omega(\vec{X}) \| \vec{X} - Ws^n(\vec{X}) \| \ldots\ldots\ldots\ldots (4)$$

where $\omega(\vec{X})$ play the weight function role. The $\| \vec{X} - Ws^n(\vec{X}) \|$ is the Euclidean distance between two points. $\vec{X}$ are the location vectors of each point in the image except sampling points. $Ws^n(\vec{X})$ is the location vector of sampling point in S^n nearest to $\vec{X}$. Two kinds of weight function were implemented in the process.

5. RESULT

sample number	10
start temperature	10
cooling rate	0.05
maximum iteration at each temperature	2000
MS: maximum number of iterations without improvement	2000
Thresholds	30

Table.1 Environment setting

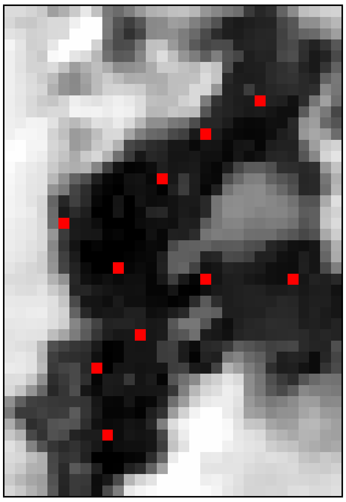

Figure 4. Result: 10 samples distribute in rule image based on a threshold of 30

Figure 4 shows the algorithm made a good result. Sampling points were spread over the area of interest, placing more samples in areas with smaller spectral angle.

Based on observation, environment parameters were defined in table 1. The threshold was defined as 30 degree. In these areas where weight values are between 0 to 30 degree have very high probability of weathering waste rock exist. The lower weight means the higher probability.

6. CONCLUSION

Overall we conclude that sampling optimization using hyperspectral data can be applied in thegeochemical environment assessment purpose. Optimization based on Spatial Simulated Annealing algorithm is a good way to find global optimal sampling position.

REFERENCE:

Debba, P. (2006). "Sampling scheme optimization from hyperspectral." ITC PHD Thesis 2006.

Groenigen, J. W. v. (1997). "Spatial simulated annealing for optimizing sampling." GeoENV I. *Geostatistics for environmental applications* pp 351-361.

Burgess, T. M., Webster, R., & McBratney, A. B. (1981). Optimal interpolationand isarithmic mapping of soil properties. IV Sampling strategy. *Journal of SoilScience*, 32(4), 643–660.