Study on application of entropy averaging method in airborne gamma-ray spectrum for local disturbance elimination

Chao Xiong 1, Jiankun Zhao 1, Hexi Wu 1, Liangquan Ge 2,*

1 School of Nuclear Science and Engineering, East China University of Technology, Nanchang 330013
2 Applied Nuclear Techniques in Geosciences Key Laboratory of Sichuan Province, Chengdu 610059

*Corresponding author and e-mail: Liangquan Ge, glq@cdut.edu.cn

Abstract. This paper briefly describes the basic principle and the entropy averaging method for local disturbance elimination in airborne gamma-ray spectrum. This method was employed to process the boundary of different lithologies and the identification of metal ore points. The results showed that applying the entropy averaging method to eliminate the interference of the airborne gamma-ray spectrum data, and selecting a larger window width, achieve more significant results compared with the ground gamma-ray spectrum.

1. Introduction
Airborne gamma-ray spectrum measurement that installed on an aircraft by an airborne gamma-ray detector is used to measure gamma ray from radioactive material in the surface medium and atmospheric radioactive material. We can explore shallow surface radioactive minerals and metal minerals through this method. Due to the randomness of the decay of radioactive materials, the radioactive element content has statistical fluctuation characteristics. In addition, formation of geological bodies are also random in the long process of geological evolution. At the same time, airborne gamma-ray spectrum measurements are susceptible to interference from factors such as air humidity, radon concentration in the air, topography, and flying height, which makes the measured data a random superposition of the real information of the radioactive level of the surface medium and the interference noise information. How to reduce the measurement statistical noise and improve the measurement accuracy is a new development direction in data processing in recent years.

Many scholars have tried various ways to eliminate the errors caused by statistical fluctuations. Eliminating local interference of gamma spectral data has been proven to be a necessary and effective method, and it is proved by experiments that the entropy averaging method is better than the multi-point slewing method. (Chen X, 2011), Liu et al. (Liu J, 2010) used the entropy averaging method to eliminate the local interference of the ground gamma-ray spectrum profile data of geological mapping in shallow coverage area. The results show that using the method can eliminate the random interference of surface vegetation and local soil enrichment. Xu (Xu S, 2011) compared a variety of methods to eliminate random interference of ground gamma-ray spectrum data. However, there are some significant differences between airborne gamma-ray spectrum and ground gamma spectrum survey. For instance, the number of measuring points on the airborne gamma-ray spectrum line is large; the single line covers a wide area and the point distance is fixed; due to the influence of terrain and other factors, the ground gamma-ray spectrum cannot be evenly distributed on the line, or the point distances are completely equal. In summary, the method of using the entropy averaging method to eliminate random interference of airborne gamma-ray spectrum data has not been reported.
2. The principle of entropy

The concept of entropy averaging is derived from the development of physics, which was originally used to
measure the disorder degree of a thermodynamic system. Claude Elwood Shannon first applied the concept
of entropy to information theory in 1948. So entropy is also called Shannon entropy. In information theory,
entropy is the concept of measuring uncertainty. That is, the higher the entropy, the larger the amount of
information, and the smaller the amount of information. In the airborne gamma-ray spectrum measurement,
the random variable X is set as the information source, where \( X \in \{ a_1, a_2, \ldots, a_k \} \). The frequency \( P_k \) at
which the variable X appears is as follows:

\[
P_k = P(a_k) (j = 1, 2, \ldots, K)
\]

Hence each line can be recorded as a finite number of independent repeat tests.

\[
X = \begin{bmatrix}
a_1 & a_2 & \cdots & a_k \\
P_1 & P_2 & \cdots & P_k
\end{bmatrix}
\]

This process can be expressed as a equation between X and \( P_k \):

\[
H(X) = -\sum_{k=1}^{K} P_k \log P_k
\]

Where \( H(X) \) is called information amount or information entropy about the source. In general,
information entropy can be calculated if the probability of a source is known. But because people often have
difficulty knowing the probability of random variables
appearing \( p_k = p(ak)(i=1,2,\ldots,K) \), therefore entropy is actually a mathematical expectation of the sum of
the bit quantities of random variables and the probability of sequential occurrence multiplied. If the
mathematical expectation of the function \( f(x) \) of \( x \) is:

\[
\sum_{k=1}^{K} p(a_k) f_m(a_k) = c_m, \quad m = 1, 2, \ldots, M
\]

thus we can estimate the distribution pattern \( (p_1, p_2, \ldots, p_k) \) of \( X \), and then find the solution of the \( X \)
distribution to calculate the information entropy.

In 1957, E.T.Jaynes proposed the principle of maximum entropy. The core idea of it is to make full use of
only a small amount of information to judge the distribution of unknown information without adding
speculation (Jaynes E.T, 1957). It can be known from the principle of maximum entropy that the estimation
problem of probability distribution has in fact been transformed into the optimal solution problem of the
planning scheme:

\[
H(X) = -\sum_{k=1}^{K} P(a_k) \log P(a_k)
\]

that is, the probability distribution of \( P = (p_1, p_2, \ldots, p_k) \) when the entropy is maximized. At the
same time, the following constraints are met:

\[
\sum_{k=1}^{K} p(a_k) f_m(a_k) = c_m, \quad (m = 1, 2, \ldots, M) \\
\sum_{k=1}^{K} p(a_k) = 1, \quad p(a_k) \geq 0
\]

The optimal solution obtained under this constraint is the probability distribution that maximizes the
system information entropy.

3. Entropy averaging function of airborne gamma-ray spectrum data

Since the airborne gamma-ray spectrum measurement has statistical fluctuations, and the measured data is
easily superimposed with the interference information, thus uncertainty appears. Such uncertain attributes
can be described by information entropy as follows:

\[
h_j = -k q_j \log q_j
\]

where \( q_i \) is the probability of describing the discreteness of a certain point data. The measurement process
of a single airborne gamma-ray spectrum point can be considered as an independent repeat test. Independent
repeat trials are uniformly distributed and can be considered as equal probability events. Since the
uncertainty of the equal probability event is the largest, we can assume that the entropy of the measured values of the airborne gamma-ray spectrum is the largest. At this time, the larger the value of the sharp change point in the window, the smaller the influence on the entropy averaging. Here the expression of $q_i$ can be written in the following form:

$$q_i = \frac{|A_i - A_0|}{\sum_{j=1}^{n} |A_j - A_0|}$$

where $A_0$ represents the actual measured value of the airborne gamma-ray spectrum, and $A_1$ represents the measured value when the entropy is maximum.

The steps of extracting anomaly information of airborne gamma-ray spectrum data based on entropy averaging method are as follows:

1. Select $n$ consecutive airborne gamma-ray spectrum measurement points from a certain measurement line to form a measurement window. The actual values of the measurement points are set to $A_1, A_2, A_3,...A_n$;

2. Select the point in the data set of the window that can make $H(A_0)$ obtain the maximum value. Since the equal probability event has the largest uncertainty, it has the largest entropy. To make $H(A_0)$ take the maximum value, $A_0$ should be the most equal distance from other observations falling within the window;

3. Solve the following equations

$$A_{ji} = A_j + \sum_{i=1}^{n} (A_j - A_i)h_i$$

$$k = \frac{1}{\log(n-1)}$$

$$q_i = \frac{|A_i - A_0|}{\sum_{j=1}^{n} |A_j - A_0|}$$

where $A_{ji}$ is the entropy averaging value of $n$ observation points in the current window, and $A_{ji}$ is recorded in the center of the window;

4. Move the window with one measuring point as the reference in the direction of the measuring line, and repeat steps (1)-(3) to complete the calculation of the single measuring line;

5. Repeat steps (1)-(4) with the line as the calculation unit to complete the calculation of all the measurement areas.

In order to visually reflect the effect of entropy averaging on raw data processing, the coefficient of variation is introduced into the statistics of airborne gamma-ray spectrum line data, that is, the coefficient of variation (CV) is used to characterize the degree of line change after different parameters are processed.

Because the ore-bearing areas are in the region of high coefficient of variation (Sun, 2009), the coefficient of variation of gamma energy spectrum can reflect the changing trend of uranium, thorium and potassium in the survey area, so it can be used to study the spatial distribution law of radioactive elements, general exploration and prospecting. The calculation formula is as follows:

$$CV = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2}$$

where $\bar{X}$ is the average value of a certain element in the survey area; $n$ is the total number of points in the survey area; $X_i$ is the element content value of a certain measuring point in the survey area.

The practical significance of the coefficient of variation in the measurement data of airborne gamma-ray spectrum is: Taking the average of the arithmetic points of the measured points in a single line as the reference, obtain the coefficient of variation of the measuring points, and further compare the degree of dispersion of the line data. The larger the coefficient of variation, the more intense the data jitter and the greater the instability.

Application to airborne gamma-ray spectrum data
The flight data of 1:5000 scale in a metal mine survey area in Inner Mongolia was selected as the data source. The data was measured by the AGS863 airborne gamma-ray spectrometer developed by Chengdu University of Technology. The geological map of the survey area is shown in Figure 1. The area is mainly composed of granite, sandstone and sedimentary rocks, and there are metal ore points (lead-zinc ore, iron ore) in the southeast of the survey area. The measured full spectrum of the gamma-ray spectrometer is shown in Figure 2, and the area of the survey area is about 300 square kilometers, with 21 lines and 6705 points. The measured nuclide statistics are detailed in Table 1.

![Figure 1. Geological map of the survey area](image1)

![Figure 2. Measured airborne gamma-ray spectrum by AGS863](image2)

Table 1. Measurement data statistics table

| Statistical parameter | Indicator name | K (%) | U(\(\mu g/g\)) | Th(\(\mu g/g\)) | TC(cps) |
|-----------------------|----------------|-------|----------------|----------------|--------|
| average value         |                | 3.59  | 3.97           | 14.33          | 20.26  |
| Median                |                | 3.23  | 3.32           | 11.76          | 17.52  |
| Standard deviation    |                | 1.47  | 3.19           | 7.79           | 9.46   |
| variance              |                | 2.15  | 10.15          | 60.74          | 89.56  |
| Minimum value         |                | 0.04  | 0.049          | 0.26           | 0.41   |
| Maximum value         |                | 19.87 | 37.20          | 57.19          | 107.32 |

Select a line in the survey area, and use the 5-point, 9-point, 13-point, 17-point and 21-point entropy averaging methods to process the line and calculate the coefficient of variation after processing. The comparison of the total count rate and coefficient of variation of a section of a survey line is shown in Figure 3.

![Figure 3. Comparison of different entropy mean and coefficient of variation for the total count rate of a survey line](image3)
It can be seen from the upper part of Figure 3 that the raw data can maintain the general trend of the content of the line elements after being processed by the entropy averaging method of different width windows, and the coefficient of variation is mostly below 0.05, and has the following characteristics:

In the region where the coefficient of variation is gentle (D3, D4), the results obtained by the entropy averaging method are basically consistent with the original data of the line, and the corresponding coefficient of variation is also less than 0.25, which is a weak variation. In the low-valued area D2, the data after the five-point entropy averaging is basically consistent with the original data form, and as the window increases, the low-value trend of D2 gradually decreases, and the lowest point increases significantly. The coefficient of variation here is the maximum of the whole line, reaching 0.12, which is a strong variation, indicating that the window size of the entropy averaging method has a greater impact here. The overall value of the D1 region is roughly the same as the original data, but with the increase of the entropy averaging window, it is controlled by the sliding average algorithm, and the peaks are suppressed by overlapping; while the coefficient of variation is weakly less than 0.6.

As can be seen from the above, the size of the window width directly determines the effect of the entropy averaging algorithm. According to the experience of ground gamma spectrum measurement, we select 7-point, 15-point, 30-point entropy averaging method to process the raw data, and extract the uranium, strontium and potassium anomaly information of the airborne gamma-ray spectrum. The results are shown in Figure 4 to Figure 6:

![Figure 4. 7-point entropy averaging method (a, uranium; b, thorium; c, potassium)](image)

![Figure 5. 15-point entropy averaging method (a, uranium; b, thorium; c, potassium)](image)

![Figure 6. 30-point entropy averaging method (a, uranium; b, thorium; c, potassium)](image)

It can be seen from Figure 4 to Figure 6 that when n=7 and n=15, the entropy averaging does not substantially eliminate the interference effect on the high, medium and low value regions of the airborne
gamma-ray spectrum profile. Liu Jinghua (Liu, 2010). studied the ground gamma spectrum measurement and found that using the entropy averaging method can eliminate local interference well and has less influence on anomaly information when n=5. The reason for this difference is the difference in the measurement carrier. Since the airborne gamma-ray spectrum system records a full spectrum of data per second, and the aircraft is flying at a high speed (speed 70-120 m/s), the area covered per unit time is much larger than the area of ground gamma spectrum measurement, which results in the distribution density of the sampling points of the airborne gamma-ray spectrum measurement is much higher than density of the ground gamma ray spectrum. Therefore, airborne data reflects far more radioactive information than ground data. However, the average gamma-ray spectrum entropy window width is much negligible compared to the airborne gamma-ray spectrum data, which requires increasing the entropy averaging window width to increase the data capacity within the window. When the window width n is chosen to be 30, it is found that the anomaly range and lithology distinction are the most consistent with the geological background shown in Figure 1. Therefore, it is determined that n=30 is used as the entropy averaging method to extract the window width value of the airborne gamma-ray spectrum anomaly information.

4. Conclusion
The entropy averaging of airborne gamma-ray spectrum is an effective interference cancellation method, for which data can maintain the general trend of the content of the line elements after being processed by the entropy averaging method of different width windows. After repeated tests, it is determined that n=30 is used as the entropy averaging method to extract the window width value of the airborne gamma-ray spectrum anomaly information.

Acknowledgement
This work is supported by the National Natural Science Foundation of China (No. 41804114), National Key R&D Project (No.2017YFC0602100) and Scientific research project of jiangxi education department (GJJ170451)

References
[1] Chen Xin.2011.Application research of entropy average method in radioactive data processing[J], (02):13-15.
[2] Liu Jing-Hua et al. 2010. Study on the application of entropy average method in ground gamma spectrometry for local disturbance elimination, Journal of Geophysics –Chinese edition[J], 53(8). 1961-1967
[3] Xu Shi. 2011. Research on Ground Gamma Spectroscopy Measurement Random Interference Cancellation Method[D],21-30
[4] Jaynes ET. 1957. Information theory and statistical mechanics. Physical Review[J], 106:620–630.
[5] Sun Zhongren.2009. Study on the coefficient of variation for the statistical work of specimen magnetic parameters[J],45(1).65-69