A Novel Approach of Gamma Spectrum Anomaly Detection Base on Machine Learning

Ri ZHAO, Liye LIU*

China Institute for Radiation Protection, Taiyuan , China, 030006

*Corresponding author's e-mail: liuliye@cirp.org.cn

Abstract. the traditional abnormal gamma energy spectrum recognition technology is based on the analysis of the structure of the full energy peaks in the gamma energy spectrum. The results are unreliable under the condition of poor energy resolution and low counting rate, and are vulnerable to environmental radiation background and other factors. in order to make up for the shortcomings of the traditional technology and improve the reliability of the identification results under adverse conditions, this paper studies a new abnormal gamma spectrum recognition technology based on principal component analysis and Mahalanobis distance under the framework of artificial intelligence. This technique is inspired by the phenomenon that the normal gamma energy spectrum without abnormal nuclide contamination in the same measuring environment has similar morphology. By making full use of the full spectrum information of gamma energy spectrum, the common morphological features are extracted by machine learning of a large number of measured normal gamma energy spectrum samples. On this basis, the coincidence degree between the identified samples and the morphological features is compared by mathematical means, and the intelligent identification of the samples is finally completed. In this paper, the main algorithm of this technology is studied first, and the correctness of the algorithm is verified by Monte Carlo simulation experiment, and the optimal setting of key parameters in the algorithm is obtained. Then, more than 20000 non-polluted and suspected contaminated gamma energy spectra measured by the whole body counter device in a nuclear power plant are used to identify the abnormal gamma energy spectrum. The results show that the method is reliable and effective. The work of this paper will provide a reference for domestic research on new technologies under the framework of artificial intelligence to break through the traditional model in the field of gamma ray measurement and analysis.

KEYWORDS: gamma spectrum anomaly recognition; artificial intelligence; principal component analysis; Mahalanobis distance

INTRODUCTION

Gamma energy spectrum is the count-energy statistical spectrum obtained by the deposition of energy in the detector by gamma rays emitted by radionuclides. Abnormal gamma spectrum recognition refers to the identification of gamma spectrum which may be contaminated by abnormal radionuclides in a specific measurement environment, so as to infer that the corresponding measuring object may contain dangerous radionuclides. In this paper, the gamma energy spectrum contaminated by abnormal radionuclides will be referred to as abnormal gamma energy spectrum, while the gamma energy spectrum not contaminated by abnormal radionuclides will be referred to as normal gamma energy spectrum. Abnormal gamma spectrum recognition is of great practical significance in many scenarios, such as port, cargo radioactivity screening at border, monitoring of

surface and in vivo radioactive contamination of workers in nuclear industry. At present, most of the abnormal gamma spectrum recognition in these cases still depends on the traditional energy spectrum analysis technique based on the omnipotent peak[1-3]To find all the peak structures in the spectrum which may be formed by the photoelectric effect of gamma rays in the detector and compare them with the normal gamma spectrum. However, this method is prone to misidentification and missing recognition in practical analysis, especially when the energy count rate is low, the relationship between channel address and energy is uncertain, and there is a false peak structure similar to the omnipotent peak in the energy spectrum. As a matter of fact, in practical scenarios such as cargo radioactivity screening and monitoring of radioactive contamination on the body surface, gamma detectors are often scintillation type detectors with high detection efficiency and no need for refrigeration, such as NaI detectors. The sites and energy obtained by these detectors vary according to the temperature of the measured environment and often contain false peak structures such as low energy scattering peaks and Compton edges, while the energy spectrum counting rate is very low due to rapid measurements. Therefore, in view of the disadvantages of the above traditional methods, a new abnormal gamma spectrum recognition technology suitable for various detector types and unfavorable measurement conditions needs to be studied.

In recent years, there have been pattern recognition techniques under the concept of artificial intelligence in the world. These technologies no longer focus on the detailed features of a single sample, but make full use of the rich information contained in a large number of sample sets, through learning and training. Because of the high intelligence and accuracy, these technologies have become the research hotspot in the international field, and have been developed rapidly in various fields. Preliminary theoretical and applied studies have been carried out abroad on the application of the idea of artificial intelligence to gamma spectrum analysis: for example, R.C.Runkle radiation detection for border vehicles, a large number of gamma energy spectra measured by mobile radiation monitors, and a machine-learning method to determine whether vehicles are loaded with illegal nuclear materials have been studied. The results show that the method is reliable and effective for radiation monitors of NaI type and plastic scintillation type[4]D.Boardman, on the basis of the above work, the classification of the materials contained in the vehicle cargo was further realized by principal component analysis[3]A large number of soil samples were classified according to geographical origin by using artificial intelligence algorithm based on gamma spectrum[5]At present, there are only a few rough works in China, such as Yang Yang's use of fuzzy clustering based on natural gamma spectrum to assist in the identification of formation lithology logging[6]。

Pattern recognition under the concept of artificial intelligence generally includes two main modules, namely feature extraction and classifier design. Because of the high dimension of gamma energy spectrum data, the principal component analysis algorithm commonly used in spatial dimension reduction and information extraction is selected in this paper. This algorithm realizes the low-dimensional projection of data points by looking for a set of special orthogonal bases in the sample space. At the same time, because abnormal gamma spectrum recognition only involves normal and abnormal dichotomy, the classification pattern is relatively simple, so this paper adopts a classifier based on Mahalanobis distance. Mahalanobis distance is a kind of generalized distance which is widely used in statistical analysis. It measures the possibility that a sample belongs to a sample set. The classifier based on it is simple and intuitive, especially suitable for dichotomy.

Firstly, the algorithm of abnormal gamma spectrum recognition based on principal component analysis and Mahalanobis distance is studied, and its correctness is verified by simulation

experiments. On this basis, more than 20000 non-polluted and suspected contaminated gamma energy spectra measured by the whole body counter of a nuclear power plant irradiation monitoring laboratory are used. The results show that the method is reliable and effective.

1. TECHNICAL PRINCIPLES

1.1. Principal component analysis

A principal component analysis method was invented by Karl Pearson in 1901. Since then, it has become the most commonly used unsupervised linear dimension reduction method in the field of statistical analysis and artificial intelligence[7]Combined with the background of gamma spectrum, the method can be explained briefly as follows.consider a sample set consisting of m normal gamma spectra $\Psi=\{x_i|x_i\in R^n;i=1,2,\ldots,m\}$, each gamma spectrum consists of n data, i.e. n dimensions. The covariance matrix of each channel is:

$$C_{jk}=\sum_{i=1}^{m}\frac{(x_{ij}-\tilde{x}_j)(x_{ik}-\tilde{x}_k)^T}{m-1}$$

Among them, i =1, 2,..., m; j=1, 2,..., n; k=1, 2,..., n; C_{jk} is the j line k column element of the covariance matrix,x_{ij} is the j count of the i spectrum, $\tilde{x}_j$ is the average of j channel

$$\tilde{x}_j=\frac{1}{n}\sum_{i=1}^{m}x_{ij}$$

Eigenvalue decomposition of C, that is

$$C=\zeta\lambda\zeta^T$$

The columns ζ are C eigenvectors, diagonal matrices, and diagonal elements are C eigenvalues.

The C eigenvector is the principal component of the set Ψ (PC)[8]The C eigenvectors are defined because they contain the main information Ψ the set: the eigenvectors are arranged in order of the size of the corresponding eigenvalues, and the eigenvectors corresponding to the largest eigenvalues are called the first principal components of the Ψ, that is, the PC_1. The eigenvector corresponding to the second eigenvalue is the second principal component of the Ψ, that is, the PC_2. The eigenvector of the second eigenvalue is the second principal component which is the direction of the maximum variance under the condition of the vertical first principal component and depicts the second largest information of the Ψ; and so on, up to the n principal component .These n principal components ultimately contain a complete description of the Ψ.

In order to reduce the dimension of the data and keep as much original data information as possible, the first k principal components can be selected as the basis of the subspace, and the new dimensionality reduced data can be obtained by projecting the original data into the subspace. For example: select the base of the subspace$\hat{\zeta}=[\zeta_1,\zeta_2,\ldots,\zeta_k]$, the new data is $\hat{x}_i=\hat{\zeta}^T x_i$. The new data obtained $\hat{x}_i$ from the projection retains most of the information of the original data, but the dimension is k

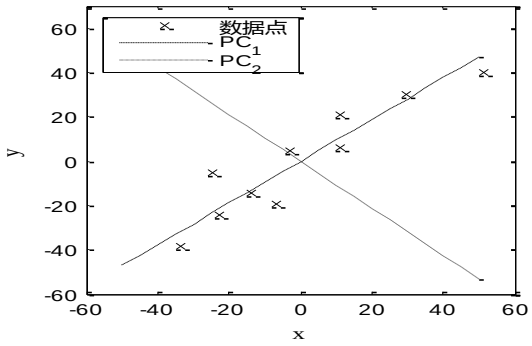

Fig .1 Schematic diagram of principal component

1.2. Mahalanobis distance

Mahalanobis distance is named after Indian mathematician P.C. Mahalanobis [7]. For the above original or reduced-dimensional set of gamma spectra analysed by principal components, the Mahalanobis distance from which a gamma spectrum y (belongs or does not belong this set) is d:

$$d = \sqrt{(y - \tilde{x})^T C^{-1} (y - \tilde{x})}$$

C is the same as the above definition, that is, the covariance matrix between the energy spectra in the set. $\tilde{x}$is the average energy spectrum.

Mahalanobis distance is a generalized distance, because the covariance between variables is considered. Compared with ordinary Euclidean distance, it can eliminate the influence of dimension and correlation between variables. In essence, it can be considered as a probability measure that a sample point belongs to a set. Taking the gamma spectrum as an example, the mechanism of the classifier based on Mahalanobis distance for anomaly recognition is to calculate the Mahalanobis distance from the center of the normal gamma spectrum set. The greater the probability that the spectrum belongs to the normal energy spectrum, and vice versa.

2. MATERIALS AND METHODS

2.1. Monte Carlo simulation experiment

Since the gamma detectors used in practical applications related to abnormal gamma spectrum recognition are often large volume scintillator detectors with high detection efficiency, the detector set in this paper is a 3*5*16 inch NaI detector. In order to absorb and scatter gamma rays, the water-filled cylinder with 60 cm bottom radius is about 15 cm. The simulation scenario is shown in figure 3.1.The simulation tool is Intel Core i5. for MCNPX 2.5.0, operating system for Windows 8, processors

Simulation of normal gamma energy spectrum: source energy is 1460 keV to simulate natural nuclides^{40}K. For simulating the uncertainty of source geometry, location and activity and the energy drift of NaI detector due to temperature change, For each simulation, The radius of the cylinder r, the distance from the detector d, the sampling number of the source n and the energy E of the source take random values in a certain range, Specifically, 0⩽r ⩽20, 10⩽d ⩽15, 1 e6⩽n ⩽2 e6, 1430 keV ⩽E ⩽1490 keV. The number of source samples is small, because the gamma energy spectrum in actual measurements is often low counting rate. Energy spectral broadening is calculated according to the measured broadening function of typical NaI detectors. When recording the spectrum, From 0 to 1800 keV take 512 equal spacing. Through simulation, Finally ,625 normal energy spectra were obtained, The spectral morphology is shown in figure 3.2.

Simulation of abnormal gamma energy spectrum: A source particle with 661 keV energy is added to the source sampling in the above scenario to simulate abnormal nuclides^{137}Cs, The abnormal energy spectra of 661 keV and 1460 keV γ photons are controlled. Finally, three abnormal gamma spectra were obtained, The sampling ratios of 661 keV photons corresponding to these three energy spectra and 1460 keV photons are :0.1:1,0.5:1 and 1:1, respectively. For ease of expression, As MC-1、MC-2、MC-3, below MC-1 the least abnormal, MC-3 anomaly is the most serious. The morphology of MC-1 and MC-3 is shown in figure 3.3.

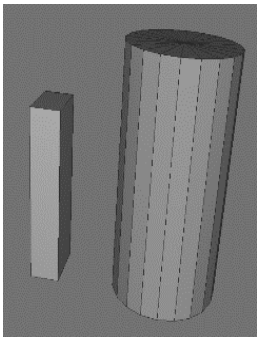

Figure 3.1 Schematic illustration of

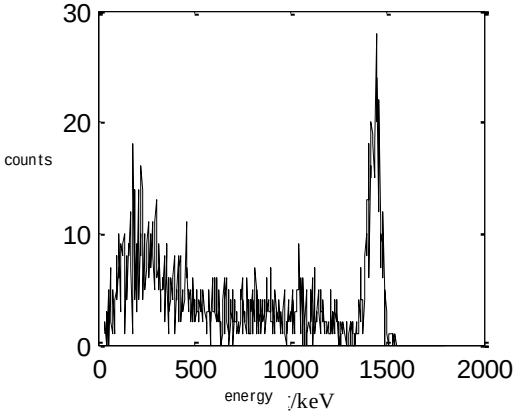

Figure 3.2 Normal gamma spectrum simulated

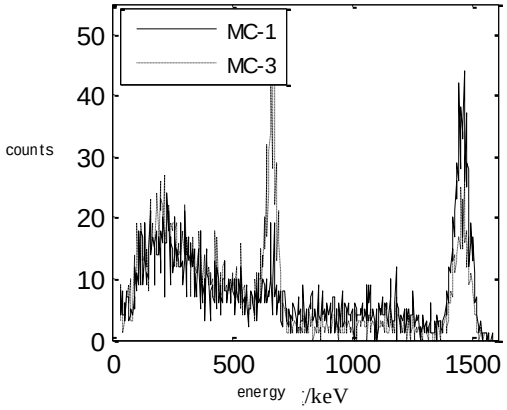

Figure 3.3 Abnormal gamma spectrum simulated

Figure 3.4 Can berra WBC of FASTSCAN company models

2.2. Description of measured gamma energy spectrum of nuclear power plant whole body counter

Normal gamma energy spectrum: this paper collects more than 20000 normal gamma energy spectra measured by the whole body counter of a radiation monitoring experiment in a running nuclear power plant. The whole body counter, or Whole Body Counter or abbreviated WBC,, is a device for directly measuring gamma rays emitted by radionuclides in the human body, which is used to determine whether the person under test is exposed to internal radiation due to the intake of radionuclides.WBC measurement is one of the important means of radiation protection in nuclear power plant, All operating nuclear power plants shall be equipped with corresponding measuring devices. The gamma spectrum collected in this paper was measured by a FASTSCAN WBC produced by a Canberra company (Figure 3.4), the WBC consists of two large volume NaI (tl) detectors of 4*4*16 inch size and necessary shielding and structural materials, A complete set of energy spectrum acquisition software is ABACOS. From 2010 to 2014, The spectral file format is automatically saved by the ABACOS. CNF format. The measured effective time (live time) of each energy spectrum is 60 s, All consist of 1024 counts, The corresponding energy and efficiency scales are carried out. The spectral morphology is shown in figure 3.5.

Abnormal gamma energy spectrum: due to the rare intake of radionuclides in nuclear power plants, only 2 abnormal energy spectra were collected from the WBC of the power plant. The file

format, measurement time and channel number of these two energy spectra are the same as the above normal energy spectra. For ease of description, the two spectra are recorded as WBC -1、WBC-2, shown in figure 3.6 below.

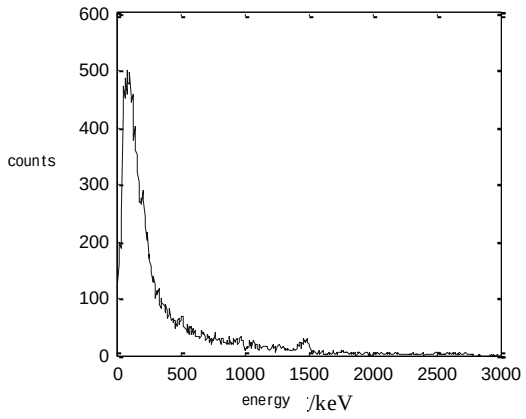

Figure 3.5 Normal WBC gamma spectrum

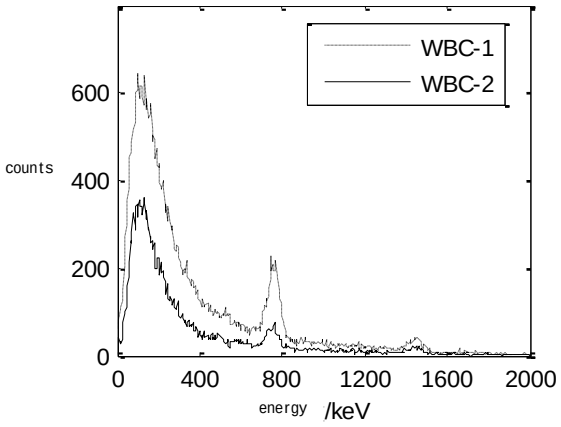

Figure 3.6 Gamma spectrum WBC anomaly

2.3. Spectral data preprocessing

a) Rebin of channel

Because the principal component analysis regards each channel as a random variable, that is, based on the channel, and the relationship between the channel and the energy in the actual measurement is determined by the scale experiment, there are differences in the calibration results between the spectra. In order to unify the relationship between Tao and energy, it is necessary to divide the channel address according to energy again, and merge the original gamma spectrum data into the new channel address by a certain method, which can be operated as follows.

Assuming E_{min}~E_{max}The energy range is divided into t intervals, that is, $[E_1,E_2)$,$[E_2,E_3)$,…, $[E_t,E_{t+1})$, of which$E_1=E_{min}$, $E_{t+1}=E_{max}$and energy spectrum x_I calibration curve is a f function relation f_i, the count of the new k can be calculated by the following formula:

$$x_{ik}^{rebin} = \sum_{j=1}^{n} x_{ij} \frac{p-q}{f_i(j+1)-f_i(j)} \quad \text{i=1,2,..., m;ki=1,2,..., t}$$

$x_{ik}^{rebin}$is the k count of the i spectrum after recombination.

In order to reduce the influence of statistical fluctuation, the new channel should be set wider, and the energy interval of each channel can be determined according to the criterion that the channel width is 3 times the half height and width of the peak.

b) Normalization

In order to highlight the morphological characteristics of gamma spectrum and weaken the effect of counting on morphology, it is necessary to normalize the spectral data after redividing the channel address.

The specific normalization treatment can be described as follows. For the gamma spectrum after re-location, it is normalized to:

$$x_{ij}^{norm} = x_{ij}^{rebin} / \max_{1\le j\le t} x_{ij}^{rebin} \quad \text{i=1,2,..., m;j i=1,2,..., t}$$

x_{ij}^{norm}which represents the j count of the i energy spectrum after normalization.

c) Centralization

In addition, because the principal component is the eigenvalue of the covariance matrix between channels, and the covariance matrix is related to the dispersion of each channel count and the correlation between each channel count, and is independent of the average value of each channel count itself. Therefore, in order to make the correspondence between principal component and energy spectrum data more obvious, the energy spectrum data should be centralized after normalization, that is, each energy spectrum minus the average energy spectrum, which can be expressed as follows:

$x_{ij}^{c} = x_{ij}^{norm} - \tilde{x}_{j}^{norm} = x_{ij}^{norm} - \frac{1}{m}\sum_{i=1}^{m} \tilde{x}_{ij}^{norm}$ i=1,2,..., m;j i=1,2,..., t

x_{ij}^{c}The normalized j count of the i spectrum.

It is worth noting that centralization only changes the energy spectrum data, has no effect on the covariance matrix between the channels, and the results of the subsequent principal components will not change.

2.4. Algorithm Flow

The method consists of two main steps: information extraction of normal gamma energy spectrum and anomaly recognition of unknown gamma energy spectrum.

The information extraction part of normal gamma spectrum samples first needs to preprocess the data of normal gamma spectrum sample set Ψ to obtain the preprocessed sample set Ψ'; Then the Ψ' is analyzed by principal component analysis to obtain all t principal components (the energy spectrum becomes t channel after the channel address is redefined) and sorted from large to small according to the corresponding eigenvalues. Select the first k principal components to form the k dimensional subspace Ω; Ψ' is projected on the Ω to obtain the new set after dimensionality reduction Ψ''; calculate the Mahalanobis distance from each point in the Ψ'' to the origin (the average point after centering is the coordinate origin), and obtain the 95% quantile d *. this distanced can be approximately regarded as 95% probability quantile of this distribution when the specific probability distribution function of the normal gamma spectrum Mahalanobis distance is unknown.

On this basis, the unknown energy spectrum is also preprocessed by the unknown energy spectrum y when the anomaly recognition of the unknown sample is carried out. Note that the average energy spectrum subtracted from the centralization processing is the normalized average energy spectrum of the Ψ, that is, $y'=y^{norm}-\tilde{x}^{norm}$. Then the y' is projected on the subspace Ω above to obtain y '', calculated Mahalanobis distance from the origin and compared with the d *. If the d is greater than the original, it indicates that the gamma spectrum is abnormal, otherwise it is normal.According to the definition of d *, the recognition confidence is 95%.

Based on the matlab, the algorithm of each module is implemented in this paper.

3. RESULTS AND DISCUSSION

3.1. Results of simulation experiments

During the pretreatment of gamma spectrum, in the energy range of 50 keV -1800keV ,13 channels were obtained according to the criterion that the channel width is 3 times the half height and width of the omnipotent peak. The morphology of the energy spectrum before and after division is shown in figure 4.1.

In order to reveal the relationship between principal components and raw data in terms of information content, With the increase of the k of subspace Ω dimension, The percentage of

variance of projection Ψ" on Ω to the original variance before projection, that is ,Ψ' variance, And var (Ψ")/var(Ψ'), where var represents variance. The results are shown in figure 4.2. As you can see, The first phase of the curve, As k increases, var (Ψ")/var(Ψ') rising fast, Soon to reach and exceed 90%, And then the curve rises slowly, And slowly approaching 100%, Finally reached 100% at k =13. This result illustrates, The top ranking PC contains most of Ψ' information, As the number increases, And later PC contain less and less information, but ultimately all PC contain Ψ' full information. Besides, As the number increases, PC information content trends can be more clearly shown in figure 4.3.

The next stage of abnormal gamma spectrum recognition test: first select k =2, that is, select PC_1， PC_2 as the basis of subspace Ω, figure 4.4 shows the recognition effect at this time. Thus, the normal gamma spectrum ensemble Ψ' gather together in this two-dimensional Ω space (plane), and the projection point of the MC-2,MC-3 is far away from the cluster, showing a certain ability of anomaly recognition. However, the projection point of the MC-1 is still in the cluster and can not be correctly identified.Mahalanobis distance d corresponding to MC-1,MC-2,MC-3 gamma spectrum are 0.63,3.86,4.92, respectively, while the 95% quantile of the Mahalanobis distance for all sample points in the Ψ" is d *=2.33, namely $d_1<d^*<d_2<d_3$This calculation is consistent with the results shown in Figure 4.4.

To make the MC-1,MC-2,MC-3 can be correctly identified, it is necessary to find the dimension k value of the appropriate subspace Ω. Therefore, this paper studies the change of relationship of d*,d_1,d_2,d_3 with the increase of k. See figure 4.5,when k <6, $d_1<d^*<d_2<d_3$, MC-1 can not be correctly identified; at 6≤k ≤11, $d^*<d_1<d_2<d_3$ meet the identification requirements. when k =12,13, the covariance matrix of the projection Ψ"Ψ' on the Ω is highly ill-conditioned and can not be reversed, that is, the anomaly recognition can not be completed at this time.

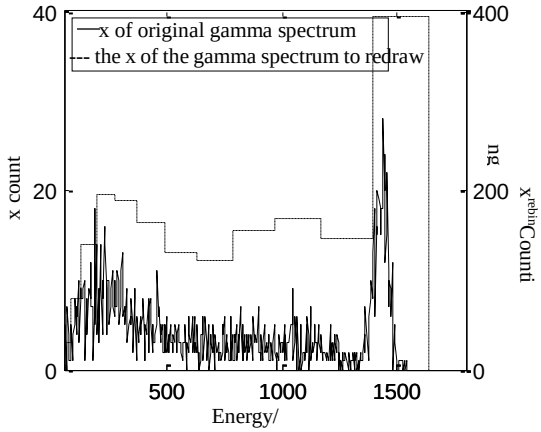

Fig .4.1 Gamma spectrum before and after the site reclassification

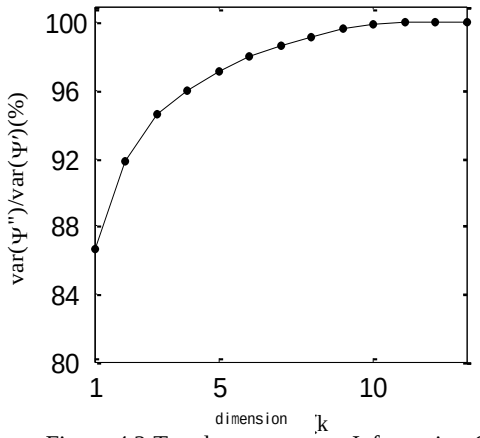

Figure 4.2 Trends of Subspace Information Content k with Dimension

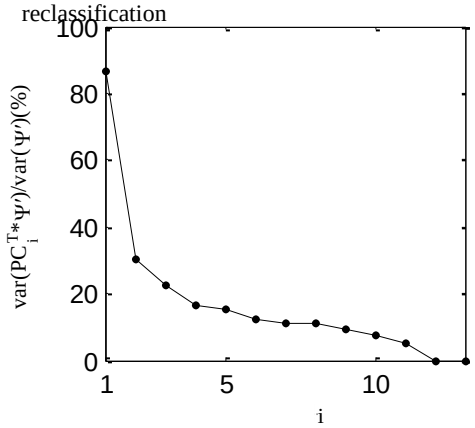

Figure 4.3 Principal component PC_iTrend of information content with sequence

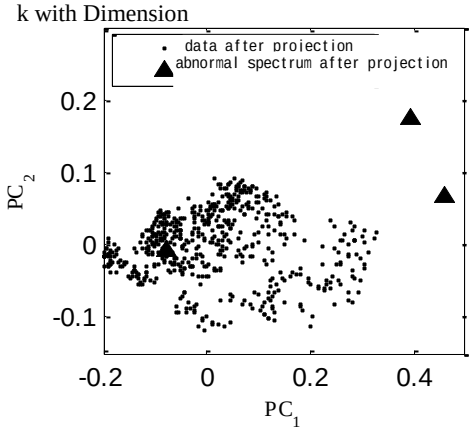

Fig .4.4 Recognition Effect of k=2 in Simulation Experiment

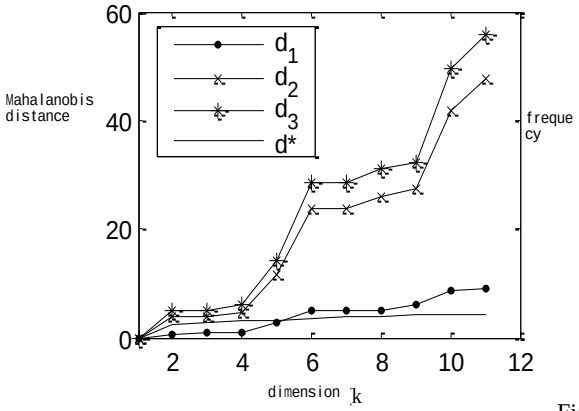

Figure 4.5 d_1d; and$_2$d; and$_3$d* with the k of subspace dimension

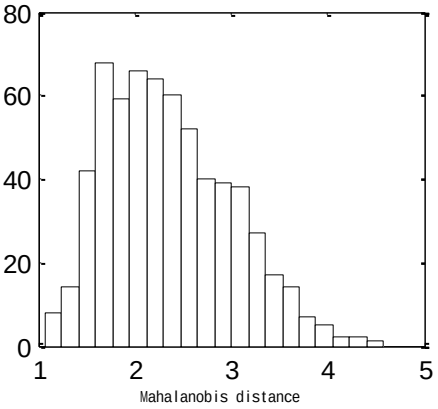

Fig .4.6 Mahalanobis distance distribution of normal gamma spectral subspace projection at 6 k = in simulated experiments

3.2. Experimental results of measured gamma energy spectra WBC nuclear power plants

At the stage of gamma spectrum pretreatment, the energy range of 50-3000 keV energy spectrum is selected to reclassify the channel address, and the criterion of division is the same as that of simulation experiment. After the division, the final 17.The energy spectrum morphology before and after division is shown in figure 4.7. Principal component analysis of more than 20000 preprocessed gamma spectra, A total of 17 principal components were obtained. Choose the first k principal components to form subspaces, And calculate the ratio of data variance before and after projection, var (Ψ")/var(Ψ'), The results are shown in figure 4.8.when k ≤11, var (Ψ")/var(Ψ')<99%, k =12, var (Ψ")/var(Ψ')=99.2 per cent, As a result, according to the k selection criteria, k takes 12.The Mahalanobis distance of each spectrum at k =12 was calculated and the d *.of 95% quantile was obtained.

The recognition spectrum WBC-1、WBC-2 pretreated by the same method and projected on the first 12 subspaces composed of the above principal components to calculate their Mahalanobis distance d; and$_2$。

WBC-1：d*=4.7, d_1=16.56, or d_1>d*, we know that this energy spectrum is abnormal energy spectrum;

WBC-2：d*=4.7, d_2=17.58, or d_2>d*,, we can see that this energy spectrum is also abnormal.

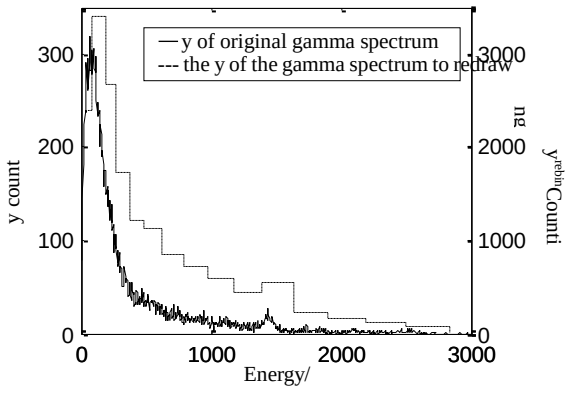

Figure 4.7 WBC gamma spectrum before and

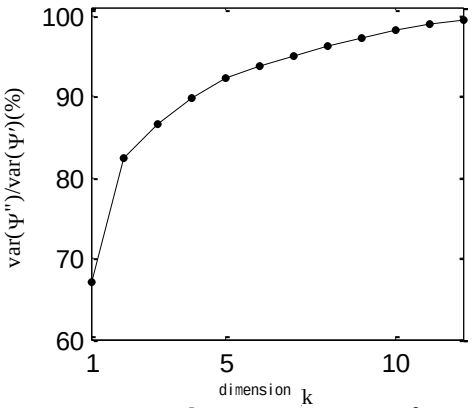

Figure 4.8 Trends of Subspace Information

3.3. Discussion

Above results show that the abnormal gamma spectrum recognition technology based on principal component analysis and Mahalanobis distance is effective and reliable on the premise of setting relevant parameters reasonably. At the same time, the implementation process also shows several obvious advantages of the method.First of all, the method is based on a large number of samples. When these samples are obtained, the measurement conditions such as ambient temperature, the size of the object under test, the relative position of the detector and the object under test, the background of environmental radiation and the energy scale curve are different, which makes the final recognition based on a more comprehensive and full grasp of the characteristics of the sample and avoids misjudgment by focusing only on a single energy spectrum; Secondly, because of the use of principal component analysis algorithm, the counting information of energy spectrum is weakened, and the morphological information related to the whole spectrum is extracted and highlighted. The amount of information used for screening is more specific than the peak structure concerned by traditional technology, which also improves the accuracy of recognition in low count rate cases; Moreover, it is because of its recognition based on the common characteristics of a large number of samples that it has a strong ability to resist false peak interference, which is evident in the analysis of the gamma energy spectrum of the WBC measured human body: the human body can be regarded as a mass source, and the proportion of the low energy part of the gamma ray after the full scattering of the human body increases, which makes the counting of the lower part of 300 keV of the human gamma energy spectrum measured by the detector much higher, and because the shell material of the detector has shielding effect on the low energy gamma ray, Finally, the gamma energy spectrum appears a low energy scattering peak in keV energy range of 90-120, which is close to the width of the omnipotent peak but much higher in amplitude. This false peak brings serious interference to the subsequent analysis, and it is easy to identify it as the omnipotent peak when using the traditional method, and then form the wrong judgment that the object under test contains abnormal nuclides. As a matter of fact, the false peak is often misidentified by commercial gamma spectrum software WBC a nuclear power plantThe^{144} Ce 133.5 keV the omnipotent peak of gamma ray formation, and once this situation occurs, it needs artificial auxiliary judgment of technicians to correct it. By using the new method studied in this paper, the characteristics of this low energy scattering peak have been extracted from a large number of samples and fully grasped, thus solving the problem of false recognition.

4. SUMMARY

In order to make up for the shortcomings of the traditional abnormal gamma spectrum recognition technology in the low energy spectrum counting rate, the uncertain relationship between channel address and energy, and the existence of false peaks, an anomaly gamma spectrum recognition technique based on principal component analysis and Mahalanobis distance under artificial intelligence framework is studied in this paper. Through theoretical research and experimental exploration, This paper establishes the basic algorithm flow of this technology, including gamma spectrum data preprocessing, principal component extraction, subspace projection, Mahalanobis distance calculation, etc, Two of them play a decisive role in the final effect of the technology. One is the preprocessing of gamma spectrum data, This is because the redividing of the path in the preprocessing step makes the relationship between the channel and the energy of all the

energy spectra consistent, The gamma spectrum with different energy scales is unified to the same scale, And when dividing the new road address, by setting the road width related to the half height and width of the omnipotent peak, The statistical fluctuation of each counting is reduced; The normalized energy spectrum with the highest count (i. e. amplitude) of 1 eliminates the effect of counting on the shape of the spectrum; These all lay the foundation for improving the accuracy of anomaly recognition. Second, when selecting the subspace Ω composed of k principal components before the normal gamma spectrum sample set, k shall be determined on the basis of var (Ψ")/var(Ψ')≥99 per cent, At the same time, we should avoid the matrix singularity in the calculation process caused by excessive k selection, Otherwise, the ideal recognition accuracy can not be achieved. Both the Monte Carlo simulation experiment and the measured gamma spectrum identification test WBC the nuclear power plant show good accuracy and reliability, This shows the possibility and potential of putting it into industrial application in the future.

REFERENCES

[1]Genie 2000Customization Tools Manual [M]Meriden ： Canberra Industries ,2006

[2]GammaVision V6Users Manual [M]U .S .A ： Advanced Measurement Technology,Inc.,2006

[3]David B oardman,M.R.,and Alison Flynn.Principal Component Analysis of Gamma-Ray Spectra for Radiation Portal Monitors [J]. IEEE Transa ctions on Nuclear Science,2012,59(1)：154-161.

[4]R .C.Runkle,M.F.Tardiff, K.K.Anderson, et al .Analysis of Spectroscopic Radiation Portal Monitor Data Usin g Principal Components Analysis [J]. IEEE Transacti ons on Nuclear Science,2006,53(3)：1418-1425.

[5]Snezana Dragovic,Antonije Onjia .Classification of soil samples according to their geographic origin using gamma-ray spectrometry a nd principal component analysis [J]. Journal of Environmental Radioactivity ,2006,89：150-158

[6] Yang Yang, Study and Development of Supporting Software for Log Identification of Stratigraphic Lithology Based on Fuzzy Clustering Chengdu: Southwest Petroleum University ,2014

[7] Zhou Zhihua C]. Machine Learning Beijing: Tsinghua University Press ,2016

[8]J.Han,M.Kamber and J.Pei,Data Mining：Concepts Techniques [C]. Singapore：Elsevier (Singapore) Pte Ltd.,2012