Spectral data processing based on maximum noise fraction

Wentai Guo¹, Liangquan Ge¹,², Fei Li¹,² and Chuanhao Hu¹,²

¹College of Nuclear Technology and Automation Engineering, Chengdu University of Technology, China
²Applied Nuclear Technology in Geosciences Key Laboratory of Sichuan Province, China

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

Abstract. The processing of gamma spectrum measurement data has always been an important step in data interpretation of the survey area. It includes the correction of radioactivity interference caused by non-measuring targets, the correction of effects caused by different heights or topography in the survey environment, the smoothness and denoising of the measurement spectrum, etc. The accuracy of each step will affect the final data resources. The interpretation of the materials results in inaccurate regional radioactivity assessment. Maximum noise separation transformation is a similar method to principal component transformation, which is used to reduce dimensionality and noise of data. This paper takes the noise reduction process of gamma spectrum data processing as the research direction, uses SPSS data processing software, combines with MATLAB to denoise the data, and uses the maximum noise separation transformation as the denoising method, applies it to denoise the spectral data, and uses the principal component analysis method and the maximum noise fraction method to denoise the same set of data separately. After analysis, it is concluded that the maximum noise fraction method is superior to principal component analysis when there is a large noise variance for a set of data.

1. Introduction

Measuring gamma-ray intensity and energy is an important method for nuclear radiation detection. Today, gamma-ray spectroscopy measurement methods have become mature, and there are a considerable number of instruments that have been applied in many fields to choose from under different measurement needs and different conditions. Gamma-ray spectroscopy is a widely used method for radioactivity measurement. In noise reduction, many scholars have studied for decades, and many methods have been published, including least squares fitting, NASVD, MNF and wavelet transformation, and many other methods have been compared and improved. At present, the most commonly used methods are the maximum noise fraction method and the noise-adjusted singular value decomposition method. The MNF method, proposed by Green et al. in the late 1980s, is based on the improvement of principal component transformation [1]. In the field of gamma spectrum data denoising, Nielson studied the application of MNF to aeronautical data in 1994 [2]. In 2004, Dickson et al. compared various noise treatment methods under four different conditions, and concluded that the root mean square error of K, U, and thorium element content after maximum noise fraction treatment would be smaller [3-4]. Later, some scholars used it to process U-thorium ratio data, and found that it has excellent results. The standard deviation and coefficient of variation of the data result
processed by the maximum noise fraction method are both smaller than that by the noise-adjusted singular value decomposition method. Principal Component Analysis (PCA) is mainly used to reduce the dimension of data. It transforms a multivariate dataset into a set of irrelevant variables with decreasing variance. The maximum noise fraction is essentially a transformation of multiple principal components. One of the difficulties to deal with is to obtain the noise covariance estimates.

At present, few domestic journal literatures have published the content related to the processing of nuclear data using the maximum noise fraction method. Most of them are related to the direction of processing hyperspectral data, or some improvement and other content in the processing process [4-6]. Some domestic scholars related to the processing of nuclear data have chosen to publish on foreign journal websites and write in English. Therefore, there are still some vacancies in this area. This paper takes MNF method as the research object, uses MATLAB to program, and combines SPSS data processing software to remove noise from the data, and verifies its noise removal performance.

2. Principles and algorithms

2.1. Principal component analysis
Principal Component Analysis (PCA) is a basic data dimension reduction method. Orthogonal transformation transforms a set of possibly related multivariate data into a set of variables with decreasing variance and linear irrelevance, which is called the principal component. It is mainly used for data compression, de-correlation, de-noising and feature extraction. The idea of principal component analysis is widely used in the development evaluation of regional economy, the formulation of clothing standards, satisfaction evaluation, pattern recognition, image compression and many other fields.

For a dataset that contains two variables, a simple transformation step for principal component analysis is:

1. Centralize the data so that the mean of the data corresponding to each variable is zero. Simply put, subtracting the mean of each column from the original data will result in a zero mean for the new dataset.

2. Establish a two-dimensional coordinate system. Since all the data is centralized, we can treat each row of data as a point in the two-dimensional coordinate system.

3. Imagine that the coordinate system rotates with a fixed origin. The data is centralized, and the examples from each point to the origin do not change during the rotation process.

4. We need to find a coordinate system in which these points can express more information on an axis. Generally speaking, if the projection of these points on an axis in the current coordinate system is to be as dispersed as possible, the corresponding variance must be maximized. Since the mean of these data does not change with the rotation of the coordinate system, we can find that the variance is simplified to the maximum of the sum of all projected coordinates on an axis, where the corresponding data is the first principal component of the principal component analysis. The second principal component is data projected on the other axis.

2.2. Maximum noise fraction transformation
The maximum noise fraction transformation is essentially a multi-principal component transformation, but the components are ordered from variance to signal-to-noise ratio.

1. The most important problem in calculating MNF transformations is to estimate noise, which is to generate an approximate noise covariance matrix. There are currently five main ways to achieve this process:

2. Differential method. Noise is estimated directly from the difference between adjacent data. It is equivalent to using variance covariance instead of noise covariance.

3. Cause-and-effect synchronous autoregression. The residual of the synchronous autoregressive model is used to estimate the noise.
(4) The difference from the mean of the neighborhood. Noise is estimated using the difference between the neighborhood data and the population data.

(5) The difference from the median value in the neighborhood. Mean difference may affect details, which can be mitigated by this method.

(6) Quadric method. Noise estimation is based on the residuals of a quadric surface fitted within the neighborhood.

3. Experimental design
In this paper, principal component analysis and maximum noise fraction are used to denoise the same set of data. The following figure is the initial data for processing and the 30 spectral lines obtained after thirty consecutive measurements with the same instrument at the same location.

![Figure 1. Original dataset.](image)

4. Discussion and analysis

4.1. Principal component analysis processing
SPSS software has its own principal component analysis, so we only need to import all the data into SPSS, then select Analysis-Dimension Reduction-Factor Analysis, add all the variables to the list to be processed, and then select the analysis type, method and so on.

In the output list, we first focus on explaining the total variance of the data, as shown in the following table
**Table 1.** Explanation of total variance for principal component transformation.

| Component | Total Variance | Variance percentage | Cumulative% |
|-----------|----------------|---------------------|-------------|
| 1         | 29.454         | 98.181              | 98.181      |
| 2         | .372           | 1.240               | 99.422      |
| 3         | .017           | .057                | 99.479      |
| 4         | .016           | .053                | 99.532      |
| 5         | .010           | .033                | 99.565      |
| 6         | .010           | .032                | 99.597      |
| 7         | .008           | .028                | 99.625      |
| 8         | .008           | .026                | 99.652      |
| 9         | .007           | .025                | 99.677      |
| 10        | .007           | .023                | 99.700      |
| 11        | .007           | .022                | 99.722      |
| 12        | .006           | .021                | 99.743      |
| 13        | .006           | .020                | 99.764      |
| 14        | .006           | .020                | 99.784      |
| 15        | .006           | .019                | 99.803      |
| 16        | .005           | .018                | 99.821      |
| 17        | .005           | .017                | 99.838      |
| 18        | .005           | .017                | 99.855      |
| 19        | .005           | .016                | 99.871      |
| 20        | .004           | .015                | 99.886      |
| 21        | .004           | .014                | 99.900      |
| 22        | .004           | .014                | 99.913      |
| 23        | .004           | .013                | 99.926      |
| 24        | .004           | .012                | 99.938      |
| 25        | .004           | .012                | 99.950      |
| 26        | .003           | .011                | 99.961      |
| 27        | .003           | .011                | 99.972      |
| 28        | .003           | .011                | 99.982      |
| 29        | .003           | .010                | 99.992      |
| 30        | .002           | .008                | 100.000     |
Figure 2. Graphic lithotripsy with eigenvalues of each principal component.

The extracted eigenvalues of the first principal component accounted for 98% of the total variance, while the second principal component was small from the beginning, all below 0.5, indicating a high correlation of data. The SPSS software automatically extracts a principal component and can already interpret these thirty sets of data fairly well. The lithotripsy map is a line graph of the characteristic values of all the principal components, from which you can also see that the contribution of the first principal component is much greater than that of the other principal components.

Next, the transformation coefficients of the principal component transformation are needed to calculate the matrix data after the principal component transformation. Here we need to get the principal component load matrix $\Sigma_C$ and the eigenvalue $\lambda_i$ of the principal component.

Transformation coefficients of principal component $i$:

$$ A_i = \frac{\Sigma_C}{\sqrt{\lambda_i}} $$

Then the first principal component data after principal component transformation is:

$$ Y_i = X_S \ast A_i $$

where $X_S$ is the original standardized data.

The first transformed principal component polyline is shown in the following figure:
Figure 3. First principal component of SPSS processing.

It is significantly smoother and less wrinkle and distortion than the original data.

Below, the transformation is implemented using the code of principal component analysis of MATLAB written by ourselves. The code is not given in this article because it is limited to space. This paper chooses to output both the first and second principal components as pictures for comparison, as shown in the following figure:

Figure 4. The first (blue) and second (purple) principal components of MATLAB principal component transformation.

The smoothness of the first principal component is still very high, and it is very similar to the shape of the first principal component calculated manually in SPSS. In the second principal component, near the first peak, the second principal component has many folds, and it was once below zero. Therefore, the second principal component should be abandoned and only the first principal component should be retained.

4.2. Maximum noise fraction processing

The maximum noise fraction method is a kind of orthogonal linear transformation based on information, which is similar to principal component analysis. It uses a set of specific basis vectors to
change the original spatial coordinates, and maps the data to a new set of coordinates. Figure 5 shows the eigenvalues of the noise covariance matrix obtained after noise estimation. As mentioned above, with the increase of the number of components, the variance will still be arranged from large to small, but the signal-to-noise ratio is not necessarily. When this happens, we will find that the eigenvalues in the noise covariance matrix of the maximum noise fraction analysis will not be arranged from large to small. For example, in the case of the 11th and 12th components in Figure 5, the noise eigenvalue of the 12th component is slightly larger than that of the 11th component (in fact, it is less than), and sometimes we can find this situation slightly. In this case, the maximum noise fraction method can do better than the principal component analysis method.

Figure 5. Characteristic value of noise covariance matrix.

Figure 6 shows the eigenvalues of the adjusted covariance matrix. The eigenvalues of the first component account for 99% of the eigenvalues of all components, which can restore the initial data to the maximum extent. This extreme ratio of eigenvalues not only shows that the correlation of the data is very good, but also helps us to get the transformation data more accurately.
4.3. Comparative analysis of treatment results

As like as two peas of two and two in Figure 4 and 7, we can observe the first component of the two methods and observe the two waves together.

It is found that the principal component transform and the maximum noise fraction transform are equivalent in some cases. When the uncorrelated noise is uniformly distributed, the results of PCA and MFN are the same. Therefore, when there is a large noise variance in a group of data mentioned above, the maximum noise fraction method will be better than the principal component analysis in this case.
5. Conclusions
This paper comes to a very important part of gamma spectrum processing, noise reduction process. Because noise reduction affects the subsequent spectral analysis and also indirectly affects the most important data interpretation. Data interpretation is related to whether the measurement data can be correctly summarized, and the quality of noise reduction is related to whether the abnormal areas in the data can be detected and whether the correct data interpretation can be made. At present, the most commonly used noise reduction methods are the noise adjusted singular value decomposition method and the maximum noise fraction method. Because the maximum noise fraction method is similar to the principal component analysis, I choose the principal component analysis method and the maximum noise fraction method as the comparison method, and more do it yourself programming to complete the implementation of the two methods. The components of principal component analysis are sorted according to variance, and the sorted principal components are uncorrelated, which can be used to extract useful information. But it also ignores that if the variance of noise in a component is large, the data transformed by principal component will still have a lot of residual noise, which cannot be denoising. Therefore, we need to use another basis to sort the components, SNR being a good choice. The higher the signal-to-noise ratio is, the lower the noise is, which solves the problem of principal component analysis. Of course, in the future, it is very likely that there will be other situations, which will lead to the fact that the signal-to-noise ratio cannot accurately judge and remove the noise. Later scholars will have to continue to improve this method.

In addition, through the research of this paper, it is found that in the process of data processing, when principal component analysis and maximum noise fraction method are used to deal with three-dimensional data, that is, h×w×p data, it is not difficult to see that principal component analysis in most cases is equivalent to the maximum noise fraction, not just the similar processing method. However, when dealing with other multispectral data, such as the regional image distribution of U/Th ratio, the maximum noise fraction method is more often used.

Acknowledgement
This work was funded by Sichuan Science and Technology Program (2020YJ0334), (2021YJ0522) and the National Natural Science Foundation of China (No.11905020).

References
[1] ANDREW A. GREEN, MARK BERMAN, PAUL SWITZER, AND MAURICE D. CRAIG. 1988. A Transformation for Ordering Multispectral Data in Terms of Image Quality with Implications for Noise Removal [J]. IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, 26(1): 65-74.
[2] Allan Aasbjerg Nielsen. 1994. Analysis of Regularly and Irregularly Sampled Spatial, Multivariate, and Multi-temporal Data [D]. Institute of Mathematical Modelling Ph.D. Thesis.
[3] Bruce L. Dickson. 2004. Recent advances in aerial gamma-ray surveying [J]. Journal of Environmental Radioactivity, 76: 225-236.

[4] Li Fei, Ge Liangquan, Tang Zhuoyao, Chen Yi, Wang Jing. Recent developments on XRF spectra evaluation. APPLIED SPECTROSCOPY REVIEWS.2020, 55(4): 263-287.

[5] FEI LI, ZHIXING GU, LIANGQUAN GE, DI SUN, XUTAO DENG, SHUN WANG, BO HU, JINGRU XU. Application of artificial neural networks to X-ray fluorescence spectrum analysis. X-Ray Spectrometry. 2018, 48(2): 138-150.

[6] FEI LI, ZHI XING GU, LIANGQUAN GE, HUI LI, XINYU TANG, XINLE LANG, BO HU. Review of recent gamma spectrum unfolding algorithms and their application. Results in Physics. 2019, 13: 102211.