Applied Radiation and Isotopes xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Applied Radiation and Isotopes

journal homepage: www.elsevier.com/locate/apradiso

Identification of weak peaks in X-ray fluorescence spectrum analysis based on the hybrid algorithm combining genetic and Levenberg Marquardt algorithm

Hua Du[a,b], Wuhui Chen[a,b], Qingjun Zhu[a,*], Songlin Liu[a], Jianbin Zhou[c]

[a] Institute of Plasma Physics, Chinese Academy of Sciences, Hefei, Anhui 230031, China
[b] University of Science and Technology of China, Hefei, Anhui 230027, China
[c] The College of Nuclear Technology and Automation Engineering, Chengdu University of Technology, Chengdu, Sichuan 610059, China

HIGHLIGHTS

- A novel algorithm is developed to solve weak peaks in X-ray fluorescence analysis.
- The hybrid algorithm has higher accuracy, better stability and faster convergence.
- This is a promising method that can be used to determine heavy metals in other grains.

ARTICLE INFO

Keywords:
Cadmium
Weak peak
Hybrid algorithm
Genetic algorithm
Levenberg-Marquardt algorithm

ABSTRACT

Accurate measurement of cadmium content in rice is of utmost importance to determine if the inspected rice product is safe to people. X-ray fluorescence analysis is frequently used for multi-element analysis because it has characteristics of fast, accurate and nondestructive. However, due to the low content of cadmium in rice, its corresponding characteristics energy peak is relatively weak and is sensitive to the background information in the X-ray energy spectrum. Thus, it is very tough to obtain the accurate values of cadmium content by utilizing traditional X-ray fluorescence analysis. In this paper, the identification of weak peaks of cadmium is much improved by proposing a hybrid algorithm combining genetic algorithm (GA) and Levenberg-Marquardt algorithm (LM). The hybrid algorithm not only takes full advantages of GA and LM respectively but also inhibits their unwanted properties: poor local search ability of GA and locally convergent of LM. The proposed hybrid algorithm is employed to identify weak peaks in X-ray spectra of six contaminated rice samples with different contents of cadmium. Two comparative experiments are conducted to compare the performance between GA, LM and the proposed hybrid algorithm. One of the comparative experiments has the relative error varying with the number of calculations, which aims to verify the accuracy and stability. The results show that the hybrid algorithm is a better option in terms of accuracy and stability. Another comparative experiment of which the average relative error varies with the number of iterations is conducted to verify the computing efficiency. The experiments show that the hybrid algorithm exhibits a faster convergence rate. Two numerical experiments demonstrate that the proposed algorithm can well resolve the identification issue of the cadmium in the X-ray spectra and significantly improve the content measurement accuracy of cadmium in the quality evaluation experiment of rice products.

1. Introduction

The detection for heavy metal cadmium in rice has been listed as an important indicator of the national routine monitoring. Metal cadmium and its compounds can penetrate the body through the digestive and respiratory tract, and they will accumulate in kidneys and liver, eventually lead to the lesions of an immune system, nervous system and reproductive system (Williams and David, 1976; Nogawa et al., 1983; Watanabe et al., 2002; Simmons et al., 2005; Kobayashi et al., 2009).

The determination of cadmium content in rice has attracted great attention of many scholars. The graphite furnace atomic absorption spectrometry (GFAAS) (Holcombe and Borges, 2004; Fang et al., 1994)

* Corresponding author.
E-mail address: zhuqingjun@ipp.ac.cn (Q. Zhu).

https://doi.org/10.1016/j.apradiso.2018.07.009
Received 7 June 2017; Received in revised form 20 June 2018; Accepted 10 July 2018

was one of the most commonly used methods for measuring the cadmium in rice. However, the general heating time was too long, which affected the speed of analysis. Other analysis methods such as electrothermal vaporization atomic fluorescence spectrometry (XiaoDong et al., 2009), UV–vis spectrophotometry (Keawkim and Chuanuwatanakul, 2013), were also applied to analyze the cadmium in rice. But there are so many problems in the above methods, such as time-consuming, complex structure of detection instruments, harsh conditions, and high cost.

X-ray fluorescence analysis, with the characteristics of non-pollution and non-destructive, is a rapid, accurate and economical method for multi-element analysis (He et al., 1993; Ao et al., 1997; Sbarato and Sánchez, 2001; Myint et al., 2003; Revenko, 2002; Hall et al., 2010). The elements in solid, liquid, powder and even gas samples can be qualitatively and quantitatively analyzed by this technology (Sbarato and Sánchez, 2001; Revenko, 2002; Potts et al., 1984). It has become one of the most popular methods of spectral analysis in China. However, due to the low content of cadmium in rice, its corresponding characteristics energy peak is relatively weak and is sensitive to the background information. Meanwhile, most weak peaks cannot be correctly identified by the traditional spectral method in the process of X-ray fluorescence analysis. In many cases, identifying weak peaks accurately is a key technology in X-ray fluorescence analysis.

In this paper, a pioneer study is reported to identify the weak peaks in the X-ray fluorescence analysis. In order to improve the content measurement accuracy of cadmium in rice products, a hybrid algorithm combining genetic algorithm (GA) and Levenberg-Marquardt algorithm (LM) is developed to resolve the identification issue of cadmium in X-ray spectra.

2. Description of the algorithm

In the process of traditional X-ray fluorescence spectrum analysis, the peak's centroid position and peak's boundaries are firstly obtained by the peak detection algorithm, such as the first derivative method, the symmetric zero-area conversion method, the convolution method, and so on. Then the peak's area can be calculated by the total peak area method (TPA), Wasson method or Covell method. However, these algorithms are often used to solve strong peaks. For weak peaks, its corresponding characteristics energy peak is sensitive to the background information. Thus, it is very tough to identify some weak peaks by the traditional spectral methods.

In order to solve this kind of problem and to avoid the influence of the noise, the fitting method has been proposed in the spectrum analysis. Because the measured spectrum can be fitted against a fitting model expressed as a linear combination of full-energy-peaks from several individual nuclides. Therefore, the best parameters of the fitting model can be obtained by minimizing the difference between a measured spectrum and the fitted spectrum. Thus, some optimization algorithms are usually used to minimize the difference between the measured spectrum and the fitted spectrum.

2.1. Parametric model

The observed X-ray fluorescence spectrum can be approximately modeled as a linear combination of full-energy-peaks from several individual nuclides and be superimposed on a background (Meier, 2005; Ying-jie et al., 2016). Each peak can be simply modeled by a Gaussian curve, which consists of the amplitude A_i, the centroid position p_i and the half width σ_i, so the energy spectrum can be expressed as:

$$\hat{f(X)} = \sum_{i=1}^{N} A_i e^{\frac{-(x-p_i)^2}{2\sigma_i^2}} + B(x) \tag{1}$$

where the independent variable x is the number of channels of the spectrum, $\hat{f(X)}$ represents the estimated counts on the corresponding channel, N is the number of Gauss peaks. B(x) stands for the function of the background, which is obtained by fitting in the form of Ying-jie et al. (2016):

$$B(x) = ax + b \tag{2}$$

When the estimated count $\hat{f(X)}$ on the corresponding channel is obtained, the corresponding error equation is acquired:

$$V = \hat{f(X)} - L \tag{3}$$

where L is the count in channel i of the measured spectrum, V is the difference between the measured spectrum and estimated spectrum, then the sum of squares of residuals R is:

$$R = V'V = (\hat{f(X)} - L)'(\hat{f(X)} - L) \tag{4}$$

It is a common computational optimization problem that to minimizes minimize R, The most common method for nonlinear least-squares minimization is the LM algorithm. However, the LM algorithm is locally convergent. The iterative divergence occurs when the initial guess is poor (Moré, 1978). In order to solve this kind of problem and to relax the constraints on initial guess, the GA is utilized to generate initial guess for LM algorithm because of its global optimization, strong adaptability, and strong robustness. Therefore, a hybrid algorithm combining GA and LM algorithm is proposed to improve its accuracy, stability and convergence speed.

2.2. Genetic algorithm

The genetic algorithm is probabilistic methods mimicking “real life”, which is a stochastic optimization algorithm based on natural selection and genetic law with characters of global optimization, strong adaptability, strong robustness, and so on. It is widely used in many fields such as function optimization, nonlinear parameter estimation, pattern recognition, image processing and so on (Houck et al., 1995; Fleming and Fonseca, 1993; Freeman et al., 1999; Rennard, 2000; Li, 2006). But it has the weakness of premature and poor ability of local searching. It may wastes much time for local optimum value when the population is near the optimization.

The genetic algorithm is formed by a population of individuals of randomly chosen parameter values. Potential solutions are referred to as individuals and are encoded into binary strings which represent chromosomes. For our specific application, individuals represent the adjustable parameters of the amplitude, the centroid position, the half width and background coefficients. Each parameter can be forced to assume values inside a predefined range according to the model. Several operators inspired by the evolutionary biology, such as inheritance, mutation, selection and crossover, act on in the population of evolutionary. The process of parent selection, mating, crossover, and mutation through successive generations is called evolution. With the evolution of solutions, increasingly stronger solutions are generated. The parameter values that minimize R are known as the ultimate solutions.

In the genetic algorithm, the probability of the individual genetic to the next group depends on the value of individual fitness. In our case the fitness function is:

$$fitiness = \frac{1}{\sum_i^N (f(\hat{X}_i) - L_i)^2} \tag{5}$$

The larger the fitness value of individual is, the smaller the difference between the measured spectrum and estimated spectrum is. The general step of the genetic algorithm as follows:

1. Initialize parameters.

a) Set the range of centroid position, amplitude, peak width and background coefficients.