

# A Statistical Approach to the Automated AES/XPS Qualitative Analysis

Y. Nagatsuka, K. Yoshida, Y. Nagasawa and Y. Ono  
JEOL Ltd. 3-1-2 Musashino, Akishima, Tokyo 196-8558

(Received September 30 1998; accepted November 20 1998)

Using spectral data processing techniques based mainly on the statistical theory, some attempts are made to develop a practical AES/XPS qualitative analysis software including the precise peak detection and the adequate assignment of transition, and finally reasonable elements identification. The present software is applied to some typical spectra data as well as the public VAMAS format data downloaded through the network from the Common Data Processing System (COMPRO).

## 1. Introduction

We have been studying the automated AES/XPS qualitative analysis for a few years[1] and applying it to some available data at hand as well as downloaded ones from the COMPRO through the public network. To cope with the various types of the COMPRO data which are measured using not only pulse counting detectors but also analog detectors, we have made our theory also applicable for the data using the analog detectors.

## 2. Peak Detection

In order to obtain excellent qualitative analysis results, it is necessary to make a reasonable peak detection for a given AES/XPS spectrum. Methods for automated peak detection have been reported so far, but there are very few papers written on the practical point of view.

We present here a simple method of detecting peaks for AES/XPS spectra, and check the validity of this method.

We assume that the spectrum is acquired by using a pulse counting measurement system in

which the variance of a measured count is roughly equal to the same measured count. Using this property, we can easily evaluate the variance curve for the second derivative of an AES/XPS spectrum.

If  $y_i$  represents the count accumulated in the  $i$ -th position of the spectrum, the second derivative curve  $d_i$  of the spectrum is obtained with the use of the Savitzky-Golay least-square method as follows:

$$d_i = \sum_{k=-n}^{k=n} g_k y_{i+k}, \quad (1)$$

where  $g_k$  represent weighting coefficients realizing the least-square fitting and the number of coefficients is  $2n+1$  which roughly cover the full width of a peak. The standard deviation of the second derivative,  $\sigma_i$  is then calculated utilizing the propagation of errors[2]:

$$\sigma_i^2 = \sum_{k=-n}^{k=n} g_k^2 (Var)_{i+k} + 2 \sum_{k=-n}^{k=n} \sum_{l=n-k}^{l=n} g_k g_l (Cov)_{i+k, i+l}, \quad (2)$$

where  $(Var)_{i+k}$  is the variance of  $y_{i+k}$  and  $(Cov)_{i+k, i+l}$  is the covariance of  $y_{i+k}$  and  $y_{i+l}$ . It is very difficult to evaluate Eq.(2) exactly.

However, it is possible to assume

$$(Var)_{i+k} \doteq y_{i+k}, \quad (3)$$

by the counting mechanism and to neglect the covariance terms by the randomness of noise.

Thus, we obtain

$$\sigma_i^2 \doteq \sum_{k=-n}^n g_k^2 y_{i+k}. \quad (4)$$

The criterion for detecting peak is given at the local minimum point of the 2<sup>nd</sup> derivative satisfying the following conditions:

$$d_i < -f\sigma_i, \quad (5)$$

where  $f\sigma_i$  is the noise threshold value, and at

the valley point  $i=j$ ,  $d_j$  takes a local minimum (valley) value and  $y_j$  a local maximum (peak) value respectively. At this point, we define relative standard deviation value  $s_j$  as

$$d_j = -s_j\sigma_j, \quad (6)$$

where  $s_j$  may indicate the 'plausibility' of peak.

We usually set  $f$  a constant of 2~3 for well defined data to obtain good results. However, for some data which are downloaded from COMPRO, we must set  $f$  much different values as 10~200 for obtaining reasonable results. Such data are not taken with the pulse counting system, but converted from the analog data using a technique such as voltage frequency conversion (VFC). In such a case, we cannot simply apply this counting statistics. However, it is possible to assume at least a linearity between the data obtained with analog and pulse counting systems. Then, we can assume in stead of Eq.(3)

$$(Var)_{i+k} \doteq Cy_{i+k}, \quad (7)$$

where  $C$  is an unknown constant for the analog counting system and  $C=1$  for the pulse counting

system. This equation is the generalized version of Eq.(3).

From purely statistical point of view, if all the data are taken with a pulse counting detecting system ( $C=1$ ), we can assure the detected peak is a genuine one for  $f=2$  with a probability of more than 95.4%, and for  $f=3$  with a probability of more than 99.7%. Inversely speaking, if a peak is detected with a similar probability, it may probably be a genuine peak even if  $f$  is greatly different from 2~3 for the analog detecting system.

### 3. Assignment of Transitions

The next problem is how to properly assign the transitions for the detected peaks. In order to assign the appropriate transitions for each detected peak with energy  $E_p$ , we extract transitions with energy  $E$  from the energy ordered transition table satisfying the following condition:

$$E_j - w(E_j) \leq E \leq E_j + w(E_j), \quad (8)$$

where  $w(E_j)$  is the energy dependent energy width for searching the transitions. Ideally,  $w(E_j)$  should be as narrow as possible. However, since there is an energy shift on the chemical state of sample, or misalignment at the energy axis of instrument, it is recommended to use an appropriate value somewhat larger than the peak width.

### 4. Element Identification

In the final stage of element identification, we must decide the element contained in the sample from information on the transitions selected for the detected peaks. At present, we are using the following information from the detected peaks of a spectrum:

①Peak energy  $E_j$ , relative standard deviation value  $s_j$ , peak area  $S_j(\%)$  from the second derivative spectrum, and local maximum  $P_{j+}$  and local minimum  $P_{j-}$  values from the first derivative spectrum.

- ②Transitions assigned to the detected peak.
- ③Energy domain of the measured spectrum.

From the above information, it is possible to make the qualitative analysis according to the following step:

- ①Select element names from the transitions.

It is possible to impose some restrictions on the selection of elements (e.g. setting the minimum relative intensity factor for selecting elements with major transitions).

- ②Apply the identification rule to each selected element.

There is an identification rule for each element. Each element identification rule which has several sub-rules to confirm the existence of the specified element includes the following types of rules:

- ①Positive rules to admit the existence of the element.

These rules support the existence of the element by evaluating the probability of existence for its major transitions. If some element that has only one major transition is assigned to a peak, the probability of existence for the element is equal to the probability of existence  $P$  of the peak with relative standard deviation  $s$ , and it is given by,

$$P = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^s e^{-x^2} dx. \tag{9}$$

Furthermore, if a certain element that has two major transitions is assigned to two different peaks with the probabilities of existence  $P_1$  and  $P_2$ , the resultant probability of existence  $P$  of the

element is given by

$$P = P_1 + P_2 - P_1 P_2. \tag{10}$$

- ②Negative rules to negate the existence of the element.

These rules oppose the existence of the element by considering the juxtaposition of the above major transition with the other transitions of the different elements. If the probability of existence for the different elements grow larger, and other inconsistencies are also found among the intensities of the transitions or the shape factors of the first derivative curve on the element concerned, the probability of existence for the element will become smaller.

In reality, due to the extremely difficult calculations, we also use a more ambiguous notion, the confidence of presence in stead of the probability of existence to allow the approximate or heuristic evaluations.

### 5. Analysis Example

Salt (NaCl) is a simple material but is liable to be identified incorrectly——Na is often identified as Zn, and Cl is sometimes identified as B. Na is easily distinguished from Zn by taking into consideration the existence of transition NaKLL-948, or by measuring the energy difference between the transitions NaKLL-979 and NaKLL-915. On the other hand, Cl is distinguished from B by comparing each local maximum and local minimum value ratio  $|P_{j+}/P_{j-}|$  of the first derivative curve, which is roughly 1 for Cl as compared 3 for B. Figure 1 shows the spectrum of NaCl and its 2<sup>nd</sup> derivative. The analysis results are shown in Table1.

### 6. References

- [1] Y. Nagatsuka, K. Yoshida, Y. Nagasawa, and Y. Ono, Hyomen Kagaku (J. Surf. Sci. Soc. Japan), 19, 7, 446(1988).  
 [2] P.Gans: "Data Fitting in the Chemical Sciences" (John Wiley & Sons, Ltd., Chichester, 1992).

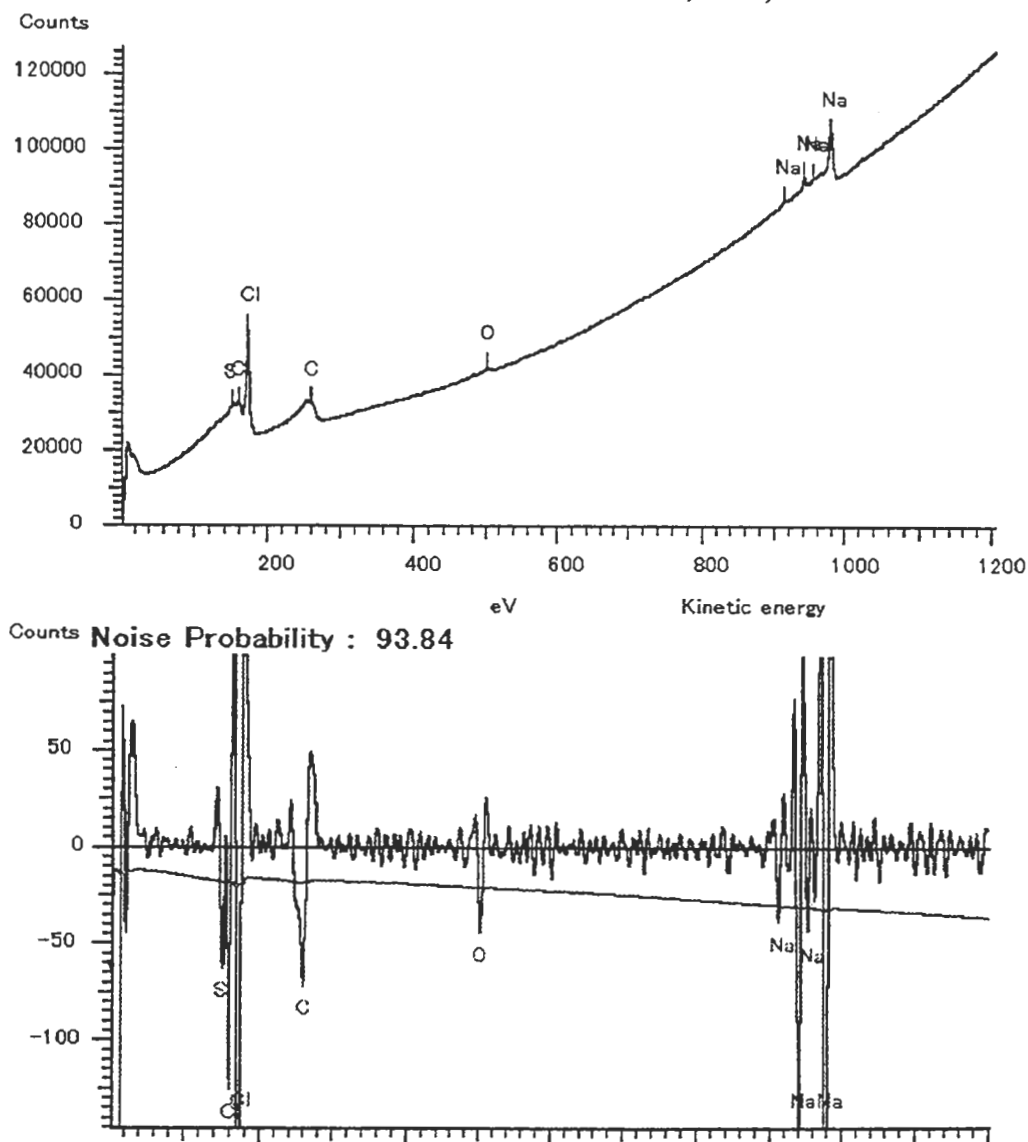


Fig. 1 AES spectrum of NaCl and its 2<sup>nd</sup> derivative with noise threshold curve.

PkEng (eV)	PkArea (%)	Sigma	PkLeft	PkRight	Element	Transition	Energy (eV)	Frac (%)	Result
151	2.6	10.3	318.8	78	S	LW00146	146	100	80
160	3.9	20.8	-67.1	-298.2	Cl	LW00165	165	5	80
172	52.5	210.7	2115.3	-2789.4	Cl	LW00176	176	100	80
260	6.5	11.6	271.7	-431.1	C	KLL00263	263	100	80
503	2.1	6.3	152.8	-63.8	O	KLL00503	503	100	80
912	1.6	3.8	233.5	74.8	Na	KLL00915	915	10	80
941	5.9	15.2	413.3	-135.2	Na	KLL00948	948	5	80
953	1.5	4.2	230.6	116	Na	KLL00948	948	5	80
979	23.4	55.6	1022.4	-1220.4	Na	KLL00979	979	100	80

Table 1 Qualitative analysis results of NaCl.