UC 4 + 34A
LBL-18495
Preprint c.1

Lawrence Berkeley Laboratory
UNIVERSITY OF CALIFORNIA

Engineering & Technical Services Division

RECEIVED
LAWRENCE
BERKELEY LABORATORY

NOV 20 1984

LIBRARY AND
DOCUMENTS SECTION

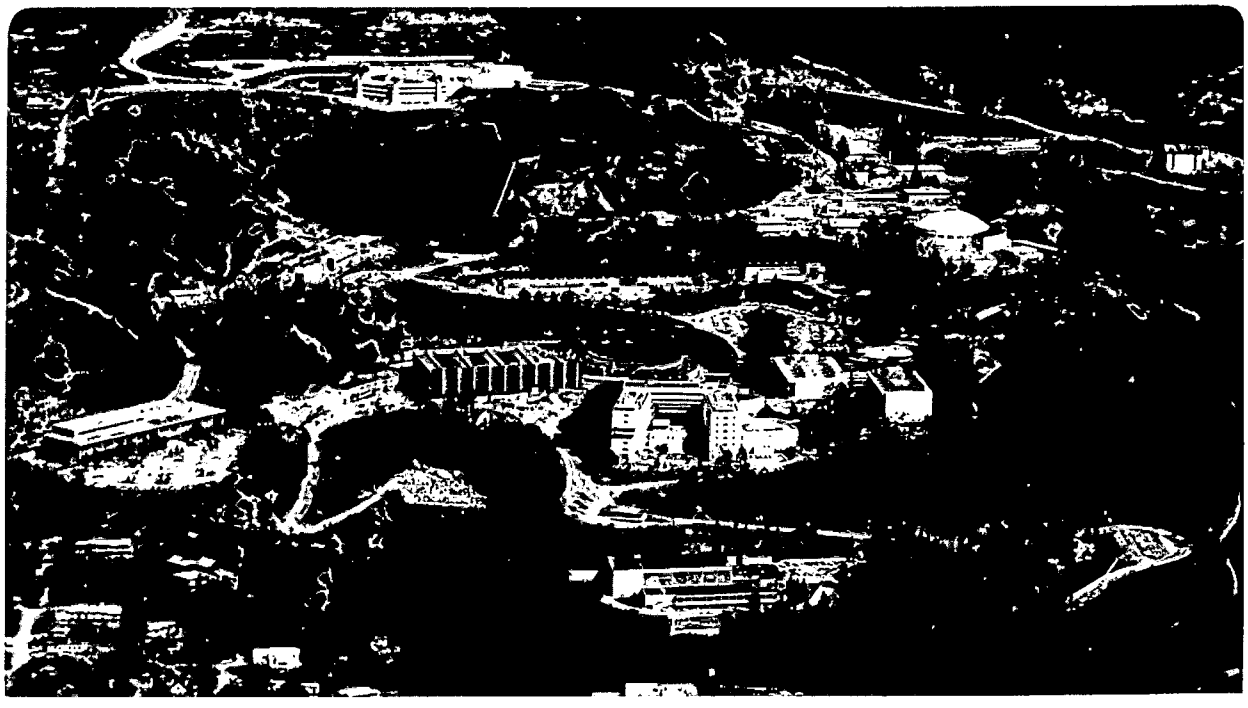
Submitted to the International Journal of X-Ray Spectrometry

STATISTICAL CONSIDERATIONS IN THE OVERLAP OF POWDER DIFFRACTION PATTERNS

J.M. Jaklevic

August 1984

For Reference
Not to be taken from this room



Prepared for the U.S. Department of Energy under Contract DE-AC03-76SF00098

LBL-18495 c.1

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

STATISTICAL CONSIDERATIONS IN THE OVERLAP OF POWDER DIFFRACTION PATTERNS

Joseph M. Jaklevic
Department of Instrument Science and Engineering
Lawrence Berkeley Laboratory
University of California
Berkeley, CA 94720

Abstract

A simple model for the calculation of the probability of random overlap of partial powder diffraction patterns has been constructed. The application of the model to the analysis of complex mixtures is discussed.

Introduction

The identification of chemical compounds using the x-ray powder diffraction (XRD) method involves the matching of a series of lines characteristic of the given compound with the lines observed in a diffractogram acquired from an unknown sample. In the case where the sample is a complex mixture of several compounds, the search-match procedure involves the matchup of the characteristic diffraction profile with a subset of lines observed in the more complex diffractogram. The ability to perform unambiguous compound identification is rapidly reduced as the number of chemical species in the sample is increased due to the inability of the instrumentation to resolve overlapping lines arising from the large number of patterns observed in the diffractogram of the unknown. Additional data such as relative line intensity and independent elemental analyses are often incorporated into the data interpretation in order to increase confidence in the compound assignments.

Recently the method of XRD has been applied to the analysis of ambient aerosol samples for the purpose of identifying the major chemical compounds present[1-4]. This class of samples encounters many problems which exacerbate

the problems of pattern matching previously described. The samples are small (typically less than 1 mg) and are deposited over a large area (8 - 10 cm^2) filter substrate. Diffraction data acquired for these samples exhibit a large diffuse scattered background intensity relative to the discrete diffraction peaks. For typical ambient aerosols, the concentration of any one component of interest is less than 10% of the total mass resulting in a further degradation in signal to background. As a result, a diffraction pattern observed for a given compound may not contain statistically significant peaks for all of the associated lines. Furthermore, because of the mechanics of the formation and collection of atmospheric particles, the sample may be inhomogeneous either because of preferential particle alignment or because the total number of particles is insufficient. These can affect the relative intensities of lines within a pattern to the point where certain major reflections may be unobservable.

For these and other reasons, it is seldom possible to identify all of the lines normally associated with a particular compound in the complex diffractogram acquired from an atmospheric aerosol sample or other complex samples. However, reasonable assignments can be made by resorting to a procedure in which the emphasis in the search-match procedure is placed on a small number of lines which are used as indicators of the presence of a particular compound. This removes the requirement that all lines in the diffractogram be identified but increases the probability of improper compound identification. The confidence in compound assignment is increased by comparing the elemental concentrations with those measured by other methods.

In order to obtain an independent estimate of the degree of confidence one can achieve using the above technique, a statistical model which estimates the probability of achieving a partial match-up of diffraction lines of a particular standard with those of an unknown diffractogram has been devised. Although

the model neglects the question of line intensities and the availability of other data to guide in identifying those compounds, it does give one some insight into the probability of spurious matchups in certain limiting cases.

Description of Model

The mathematical model can be described in two steps. The first is the calculation of the probability that a given line, characterized by an instrumental resolution function or half-width τ_1 will overlap with any one of a number of lines L with half-width τ_2, which are randomly distributed over a range of angles $\Delta\Theta$. This probability of overlap can be written as:

$$P = (\tau_1 + \tau_2) \frac{L}{\Delta\Theta} \tag{1}$$

A typical choice of values might be $\tau_1 = \tau_2 = 0.1$ degrees, L = 50 lines and $\Delta\Theta = 50$ degrees resulting in a probability of overlap of P = 0.2.

The combined probability that N overlaps will occur out of a possible M attempts can again be obtained from simple probability theory by considering the possible combinations in which the overlap conditions can be achieved, given that the probability of overlap in a single attempt is P. This can be written as[5]:

$$P_N^M = \frac{M!}{N!(M-N)!} P^N (1-P)^{M-N} \tag{2}$$

Some representative values calculated assuming P = 0.2 and some typical choices for N and M are given in Table 1. One application of these data is a calculation of the value of N for which there is a probability of 5% or less that the overlap could have occurred by chance. Those values are those above the bold line on the table. These results reinforce the intuitive feeling

that a match of, for instance, 2 out of 4 lines is not sufficient for identification (82% confidence) whereas 3 out of 4 represents a probable identification (97% confidence). It should be emphasized, however, that interpretation of these results in terms of confidence limits is not rigorous since other important factors such as intensity ratio and spatial coherence between the lines associated with various diffracting planes have been neglected.

A second illustrative example of the use of the model involves the reduction in the overlap probability as the unit probability P is reduced. A reduction in P is achieved either by reducing τ_1, that is improving the angular resolution of the instrument, or in reducing N which is achieved by working with a less complex mixture of substances. Table 2 illustrates the change in P_N^M for a few choices of N and M as a function of P. For instance, a reduction of P by a factor of two (from 0.2 to 0.1) brought about by a corresponding improvement in τ results in a dramatic reduction in the overlap probability for all cases. This emphasizes the importance of angular resolution in reducing ambiguity in compound identification.

The value of P is also proportional to the total number of lines in the unknown spectrum. An illustrative example can be generated by choosing $P = 0.04$ which corresponds to our earlier case but with the number of lines in the spectrum reduced to $L = 10$. If we now choose $M = 10$, the probability of overlap between two random patterns with 10 lines each can be calculated.

Summary

A simple, statistical model for calculating the probability of random overlap of portions of powder diffraction patterns has been described. The application of the formulae to certain cases can serve to establish minimal criteria for compound identification in complex mixtures where not all of the

lines in the diffractogram are clearly resolved. Important information normally considered in compound identification such as the relative intensity of lines and the spatial coherence of lines within a given compound which could reduce the assignment uncertainty has not been considered.

Acknowledgments

This work was supported by the Director's Office of Energy Research, Office of Health and Environmental Research, U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

References

1. P.D.E. Biggins and R.M. Harrison, Atmospheric Environment 13, 1213-1216 (1979).

2. B.L. Davis, "The use of x-ray diffraction quantitative analysis in air quality source studies". In: Electron Microscopy and X-Ray Applications to Environmental and Occupational Health Analysis, Vol. 2, ed. Russell (Ann Arbor Publishing, Ann Arbor, MI 1981) p. 131.

3. B.H. O'Connor and J.M. Jaklevic, Atmospheric Environment 15, 1681-1690 (1981).

4. A.C. Thompson, J.M. Jaklevic, B.H. O'Connor and C.M. Morris, Nucl. Instr. and Methods 198, 539 (1982).

5. W. Feller, An Introduction to Probability Theory and its Applications John Wiley, New York (1968).

TABLE 1. Probability of random overlap of N lines out of M possible lines assuming that the probability of a single overlap P = 0.2.

M \ N	0	1	2	3	4	5	6	7	8
1	.80	.20							
2	.64	.32	.03						
3	.51	.38	.10	.01	--				
4	.41	.41	.15	.03	--				
5	.33	.41	.20	.05	.01	--			
6	.26	.39	.25	.08	.02	--	--		
7	.21	.37	.28	.11	.03	--	--	--	
8	.17	.34	.29	.15	.05	.01	--	--	--
9	.13	.30	.30	.18	.07	.02	--	--	--
10	.11	.27	.30	.20	.09	.03	--	--	--
15	.03	.13	.23	.25	.19	.10	.04	.01	--
20	.01	.06	.14	.20	.22	.17	.11	.05	.02

TABLE 2. Calculation of P_N^M as a function of P for several N, M.

		P						
N	M	.02	.04	.06	.08	0.1	0.2	0.4
3	5	8×10^{-5}	6×10^{-4}	2×10^{-3}	4×10^{-3}	8×10^{-3}	0.051	0.23
3	6	2×10^{-4}	1×10^{-3}	4×10^{-3}	8×10^{-3}	0.014	0.082	0.28
4	7	5×10^{-6}	8×10^{-5}	4×10^{-4}	1×10^{-3}	3×10^{-3}	0.029	0.19
4	8	1×10^{-5}	1×10^{-4}	7×10^{-4}	2×10^{-3}	5×10^{-3}	0.046	0.23
5	9	*	1×10^{-5}	8×10^{-5}	3×10^{-4}	8×10^{-4}	0.016	0.17
5	10	*	2×10^{-5}	1×10^{-4}	5×10^{-4}	1×10^{-3}	0.026	0.20

This report was done with support from the Department of Energy. Any conclusions or opinions expressed in this report represent solely those of the author(s) and not necessarily those of The Regents of the University of California, the Lawrence Berkeley Laboratory or the Department of Energy.

Reference to a company or product name does not imply approval or recommendation of the product by the University of California or the U.S. Department of Energy to the exclusion of others that may be suitable.

TECHNICAL INFORMATION DEPARTMENT
LAWRENCE BERKELEY LABORATORY
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720