182
8-25-78

26.398

UCRL-52505

AN ANALYSIS OF THE POWDER DIFFRACTION FILE

R. L. Snyder, A. C. Johnson, E. Kahara,
G. S. Smith, and M. C. Nichols

June 22, 1978

Work performed under the auspices of the U.S. Department of Energy by the UCLLL under contract number W-7405-ENG-48.

Distribution Category
UC-4

LAWRENCE LIVERMORE LABORATORY
University of California/Livermore, California/94550

UCRL-52505

AN ANALYSIS OF THE POWDER DIFFRACTION FILE

By

R. L. Snyder*,
Q. C. Johnson, E. Kahara, G. S. Smith

And
M. C. Nichols**

MS. date: June 22, 1978

*With NYS College of Ceramics, Alfred University, Alfred, N.Y. 14802
**With Sandia Laboratory, Livermore, CA 94550

AN ANALYSIS OF THE POWDER DIFFRACTION FILE

ABSTRACT

The International Centre for Powder Diffraction Data has compiled an extensive powder diffraction database available on magnetic tape. We evaluated the database to determine the type and quality of the information it contains, including an analysis of d-I pairs per pattern, symmetry, quality marks, Q values, and intensities for subfiles of known crystal class. In addition, we evaluated nine functions to assess pattern quality for cubic and triclinic data. We present here a minimum amount of data analysis, but rather concentrate on the distribution and statistics. We also describe how these data were obtained.

INTRODUCTION

The purpose of this study is to establish the characteristics and quality of the Powder Diffraction File (PDF) as compiled and distributed on computer tape by the JCPDS (International Centre for Powder Diffraction Data).[1] Although the PDF has previously been studied,[2,3] we have investigated several new search and identification procedures.

For our study we created or used three different powder diffraction databases. All computer analyses were conducted on Control Data Corporation 7600 computers at Lawrence Livermore Laboratory. In addition, all of the figures were generated using the general plotting program CHARTIT.[4] The plots presented here in black and white were originally generated in color on a DICOMED D48 photographic plotting device. All programs referred to in this report were written in FORTRAN IV.

ANALYSIS OF THE COMPLETE POWDER DIFFRACTION DATABASE

We used sets 1 through 26 of the PDF as received on magnetic tape from the JCPDS to analyze the number and some characteristics of the patterns. The computer program that prepared this data is called DISTAT. Tables 1 and 2 show the breakdown of totals by set number for inorganic and organic patterns, respectively. Deleted patterns refer to those replaced by JCPDS editors in a later set. The old pattern is often left on the tape but marked as deleted.

The crystal-class indicator has been placed on only 16.2% of the file patterns on the magnetic tape. Patterns marked with an asterisk have been judged by the editors to be of high quality; patterns marked with a "c" are calculated from structural data; patterns marked with an "I" are indexed. The final three columns of Tables 1 and 2 give the number of patterns that have quality marks but no symmetry indication. To obtain the number of patterns with both a quality mark and symmetry indicator, subtract the proper columns (i.e. Col. 10–Col. 13). Table 3 gives a further breakdown of symmetry vs quality mark for the inorganic file only. This table contains the actual symbols found on the tape, indicating not only crystal class, but Bravais lattice type. The symbols have the following meaning:

- C – Primitive cubic
- F – Face-centered cubic
- B – Body-centered cubic
- T – Primitive tetragonal
- U – Base-centered tetragonal
- R – Rhombohedral
- H – Hexagonal
- O – Primitive orthorhombic
- P – Body-centered orthorhombic

Table 1. Statistics on the inorganic powder diffraction data base, sets 1-26.

SET	TOT	DEL	CUB	TET	HEX	ORT	MON	TRI	AST	CAL	IND	NS*	NSC	NSI
1	358	59	0	0	0	0	0	0	0	0	1	0	0	0
2	362	43	0	0	1	0	0	0	0	0	15	0	0	15
3	284	44	0	0	1	0	0	0	0	0	12	0	0	12
4	228	23	0	0	0	0	0	0	37	0	21	37	0	21
5	212	22	0	0	1	1	2	0	59	0	13	59	0	13
6	448	32	65	33	57	17	4	0	150	0	30	119	0	30
7	331	17	42	26	32	15	0	0	128	0	21	122	0	21
8	356	26	36	15	24	3	1	0	101	0	58	95	0	55
9	400	17	30	21	30	16	3	0	128	0	51	121	0	48
10	385	23	38	8	20	9	0	0	77	0	41	72	0	38
11	533	43	28	9	11	10	2	0	142	1	72	131	1	69
12	661	35	41	13	37	14	2	0	112	0	90	105	0	86
13	484	34	2	17	22	10	3	0	90	0	63	84	0	63
14	683	65	10	16	19	15	8	0	98	0	77	95	0	73
15	783	37	107	37	57	41	5	0	142	66	50	132	11	47
16	811	36	30	23	25	17	8	0	156	0	44	155	0	43
17	856	25	138	33	88	34	16	0	131	0	67	129	0	67
18	1395	52	132	98	144	37	19	0	296	13	77	283	4	75
19	1419	67	161	83	98	99	7	1	268	1	64	259	1	63
20	1446	60	40	23	61	39	6	0	383	52	66	382	36	66
21	1481	68	11	12	21	8	2	0	261	15	48	261	15	47
22	1498	86	21	16	32	13	10	1	324	49	449	308	49	388
23	1496	31	39	30	50	36	5	0	349	19	502	320	19	388
24	1496	22	35	40	49	21	4	0	377	121	587	341	109	497
25	1499	8	48	57	36	17	4	0	739	125	536	669	70	502
26	1498	0	66	56	91	71	21	1	233	56	1114	209	38	859
TOT	20428	975	1120	667	1007	543	132	3	4781	518	4169	4488	353	3587
%	(95.4	4.6)	5.2	3.1	4.7	2.5	0.6	0.0	22.3	2.4	19.5	21.0	1.6	16.8

Total number of patterns on file including those marked deleted = 21,403

SET = PDF Set No.
TOT = No. of patterns
DEL = No. marked deleted
CUB = Cubic
TET = Tetragonal
HEX = Hexagonal
ORT = Orthorombic
MON = Monoclinic
TRI = Triclinic
AST = No. with * and symmetry mark
CAL = No. with C and symmetry mark
IND = No. with I and symmetry mark
NS* = * but no symmetry mark
NSC = C but no symmetry mark
NSI = I but no symmetry mark

Q - Base-centered orthorhombic
S - Face-centered orthorhombic
M - Primitive monoclinic
N - Base-centered monoclinic
Z - Triclinic

The miscellaneous column refers to patterns that have some other symbol in the appropriate column, usually, if not always, a blank. Table 3 was prepared from the inorganic database by program ALLSTAT.

Program ALLSTAT also produced the d-I distribution given in Table 4 and plotted in Fig. 1, which shows the number of d-I pairs reported on each inorganic pattern in the PDF vs the number of patterns with this number of d-I pairs. The prominant spikes in this distribution at lines 39 and 40 are not a true phenomenon. A PDF card image will hold either 40 d-I pairs, or 39 with the message "N lines not included". JCPDS editorial policy over the years has apparently been not to continue a pattern on a second card unless it was of high interest or of high quality. These spikes can also be seen in the distributions given in the next section. Similar spikes near 20, 60, and 80 also reflect the 20-line-per-column layout of a PDF card image.

Table 2. Organic powder diffraction data base, sets 1-26.

SET	TOT	DEL	CUB	TET	HEX	ORT	MON	TRI	AST	CAL	IND	NS*	NSC	NSI
1	140	15	0	0	0	0	0	0	0	0	0	0	0	0
2	76	5	0	0	0	0	0	0	0	0	0	0	0	0
3	109	8	0	0	0	0	0	0	0	0	0	0	0	0
4	280	23	0	0	0	0	0	0	0	0	0	0	0	0
5	346	32	0	0	0	0	0	0	0	0	0	0	0	0
6	125	7	0	0	0	0	0	0	6	0	0	6	0	0
7	239	7	0	0	0	0	0	0	43	0	0	43	0	0
8	305	12	0	0	0	0	0	0	25	0	0	25	0	0
9	346	4	0	0	0	0	0	0	36	0	0	35	0	0
10	324	3	0	0	0	0	0	0	59	0	0	59	0	0
11	276	0	0	0	0	0	0	0	128	0	0	128	0	0
12	186	1	0	0	0	0	0	0	32	0	1	32	0	1
13	386	9	0	0	0	0	0	0	27	0	0	27	0	0
14	256	16	0	0	0	0	0	0	63	0	0	63	0	0
15	292	7	0	0	0	0	0	0	23	0	0	23	0	0
16	307	2	0	0	1	0	0	0	22	0	0	22	0	0
17	253	3	0	0	0	0	0	0	27	0	1	27	0	1
18	507	4	0	0	0	2	0	0	49	0	2	49	0	2
19	527	6	0	0	0	0	0	0	78	0	1	78	0	1
20	561	2	0	0	0	0	0	0	76	0	0	76	0	0
21	554	10	0	0	0	0	0	0	53	0	0	53	0	0
22	573	20	0	0	0	0	0	0	65	14	18	65	14	18
23	770	8	0	0	0	0	0	0	85	2	57	85	2	57
24	526	4	0	0	0	0	0	0	64	3	84	64	3	84
25	521	0	0	0	0	0	0	0	67	3	180	67	3	180
26	525	0	0	0	0	0	0	0	34	0	92	34	0	92
TOT	8902	208	0	0	1	3	0	0	1062	22	436	1061	22	436
%	(97.7	2.3)	0.0	0.0	0.0	0.0	0.0	0.0	11.7	0.2	4.8	11.6	0.2	4.8

Total number of patterns on file including those marked deleted = 9,110

SET = PDF Set No.
TOT = No. of patterns
DEL = No. marked deleted
CUB = Cubic
TET = Tetragonal
HEX = Hexagonal
ORT = Orthorombic
MON = Monoclinic
TRI = Triclinic
AST = No. with * and symmetry mark
CAL = No. with C and symmetry mark
IND = No. with I and symmetry mark
NS* = * but no symmetry mark
NSC = C but no symmetry mark
NSI = I but no symmetry mark

Table 3. Symmetry vs quality mark distribution for sets 1-26, inorganic powder diffraction data base.

Mark	Misc	C	F	B	T	U	R	H	D	P	Q	S	M	N	Z	Total
*	4488	33	55	15	25	51	16	61	21	2	5	0	4	5	0	4781
C	353	17	85	4	9	0	4	32	4	3	2	0	3	2	0	518
I	3587	30	71	20	48	65	34	145	83	25	22	5	11	22	1	4169
None	8595	298	374	96	205	244	103	600	212	25	100	24	47	35	2	10960
Total	17023	378	585	135	287	360	157	838	320	55	129	29	65	64	3	20428

Table 4. Pair distribution for sets 1–26 of the inorganic powder diffraction data base file.

d-I Pairs on a pattern	Patterns with the No. of d-I pairs	d-I pairs on a pattern	Patterns with this No. of d-I pairs
1	0	51	69
2	0	52	61
3	5	53	64
4	29	54	75
5	66	55	47
6	135	56	49
7	181	57	59
8	243	58	70
9	277	59	175
10	376	60	118
11	392	61	30
12	469	62	39
13	523	63	28
14	610	64	27
15	547	65	25
16	625	66	32
17	565	67	25
18	602	68	24
19	648	69	27
20	790	70	27
21	689	71	16
22	632	72	26
23	605	73	21
24	536	74	19
25	536	75	24
26	563	76	28
27	517	77	17
28	520	78	36
29	509	79	114
30	522	80	23
31	431	81	6
32	471	82	5
33	400	83	7
34	400	84	7
35	355	85	3
36	357	86	9
37	309	87	15
38	374	88	9
39	1205	89	17
40	593	90	2
41	135	91	2
42	143	92	0
43	135	93	3
44	157	94	0
45	134	95	0
46	119	96	2
47	99	97	0
48	114	98	1
49	85	99	12
50	107	100	0

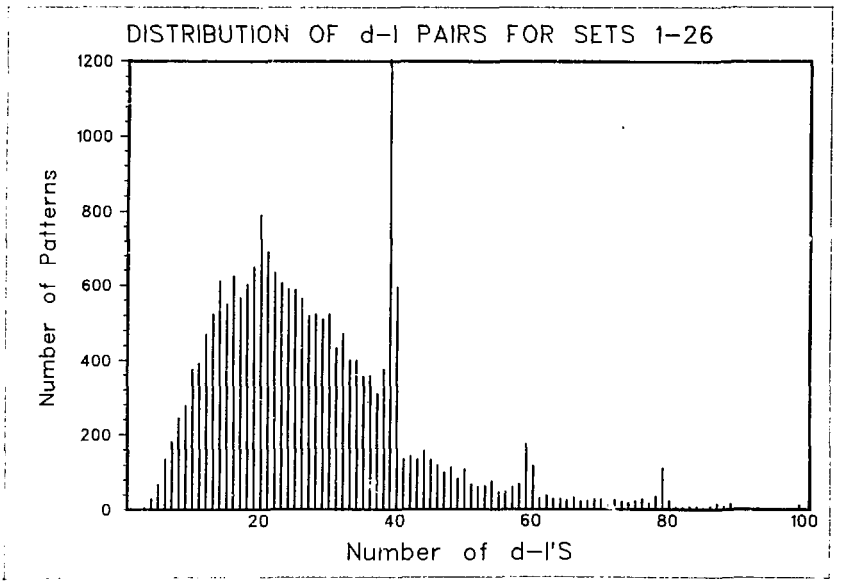

Fig. 1. Distribution of the number of d-I pairs per pattern in sets 1 through 26 of the inorganic PDF.

ANALYSIS OF PATTERNS WITH KNOWN CRYSTAL CLASS

Several years ago, G. G. Johnson, Jr. of Materials Research Laboratory at Pennsylvania State University was responsible for keypunching all the space groups and lattice parameters given on the JCPDS pattern cards. Since then, M. C. Nichols has continued this coding and when we began this study, the known space groups and lattice parameters up through set 24 were available. A program FETCHALL was written to accept these data along with the PDF inorganic master file and to produce eight new files: Cubic (CUB), Tetragonal (TET), Hexagonal and Rhombohedral (HEX), Orthohombic (ORTHO), Monoclinic (MONO), Triclinic (TRI), and two files labeled X and Y.

The first six files contain all the inorganic patterns known to have a particular symmetry, either because of the editorial symmetry mark found on the tape, or because this information had been included on the PDF card and subsequently coded. The amount of information on each of these patterns varies from known space group and lattice parameters, to nothing more than a crystal-class indication. Table 5 summarizes the amount of information known about the patterns on the files. The file labeled "X" contains all those inorganic patterns from sets 1 through 26 with no known quality or symmetry information. The Y file contains all the inorganic patterns with the quality marks "*", "C", or "I", but no further information. Note that because the space group, cell parameter data were only available through set 24, the X and Y files contain many patterns from sets 25 and 26. Many of these patterns have known space group and lattice parameters, but the crystal-class data mark was not coded onto the master tape. In fact, patterns from sets 25 and 26 on the six crystal-class files contain only symmetry marks with no space group or lattice

Table 5. Symmetry class files.

FILE	TOT	NOT F	NO CEL	NC*	NCC	NCI	CELL*	CELLC	CELLI	No. with known parm
CUD	3181	205	121	20	30	61	841	117	290	3060
TET	1973	160	120	43	9	59	491	17	333	1853
HEX	3125	263	149	17	29	81	784	68	540	2976
ORTHO	2554	271	99	11	6	68	807	56	579	2455
MONO	1829	286	38	3	2	21	724	57	543	1741
TRI	283	41	2	0	0	1	105	14	95	281
Y	2446			935	113	1398				
X	5037									
										12,416

TOT = No. on file
NOT F = No. of patterns not found
NO CEL = No. of patterns with symmetry mark but no lattice parameters
NC* = No. with * but no parameters
NCC = No. with "C" but no parameters
NCI = No. with "I" but no parameters
CELL* = No. with * and known parameters
CELLC = No. with C and known parameters
CELLI = No. with "I" and known parameters
No. with known parms = total No. with known parameters

parameters. These points are reflected in Table 6 which gives a breakdown of the number of patterns in each file by PDF set number.

The patterns referred to in Table 5 as "Not Found" represent patterns marked deleted on the master tape that were skipped in creating the eight files. Thus, a number of patterns that had their space groups and lattice parameters keypunched have since been marked deleted. In addition, some of the patterns not found were in the organic file that was not searched.

Table 7 gives the results of symmetry mark vs quality mark analysis for each of the eight files. These data were obtained by program RSTAT. In addition, RSTAT collected and prepared the data for the d-I pair-distribution analysis of each file, shown in Fig. 2.

To investigate relationships among the d and I values, a program RATIO was written to convert the eight files just described from d values to what we have called Q values. For each pattern, all d values were squared and then divided by the first d^2 value to expose geometric relationships within the various symmetry classes. These ratio or Q files are identical to the symmetry class files just described, except that d's have been replaced by the d^2/d_1^2 ratio. The various analyses of these files to be presented refer only to the nondeleted patterns from the inorganic PDF sets 1 through 26. The figures were prepared by program RPLOT.

Figure 3 shows the number of patterns with a particular Q for their second ratio (the first ratio in all patterns is, of course, 1.0, in that all d^2 values were divided by the first one). The prominent spikes in these figures are the result of two reflections, with Miller indices in the same order, commonly appearing as the first and second reflections in a pattern. For example, a pattern with d_{100} and d_{110} as the first two reflections will have its second ratio equal to 0.5. Similarly, Figs. 4 and 5 show the distributions for the 3rd and 4th ratios, respectively.

The next sets of figures refer to the intensities of the lines on the Q files. Figure 6 shows the distribution of the value of the intensity of the first three sequential lines on a scale of 1 to 8 for each pattern on the files. The next three sets of figures (7, 8, and 9) show the distribution of the values of the d^2 ratios for the strongest, second strongest, and third strongest lines in each of the patterns on the files.

Table 6. Number of patterns vs PDF set number for symmetry class files.

Q File															
Cubic		Tet		Hex		Ortho		Mono		Tri		Y		X	
Set	No.	Set	No.	Set	No.	Set	No.	Set	No.	Set	No.	Set	No.	Set	No.
1	28	1	8	1	18	1	14	1	6	1	1	1	1	1	231
2	162	2	26	2	50	2	14	2	10	2	0	2	2	2	62
3	77	3	24	3	41	3	20	3	8	3	0	3	2	3	75
4	71	4	29	4	56	4	18	4	12	4	3	4	2	4	22
5	48	5	24	5	35	5	41	5	15	5	3	5	3	5	23
6	123	6	53	6	109	6	52	6	24	6	7	6	4	6	52
7	90	7	47	7	74	7	42	7	28	7	8	7	5	7	29
8	72	8	57	8	64	8	42	8	45	8	8	8	5	8	45
9	78	9	48	9	86	9	73	9	33	9	11	9	5	9	57
10	100	10	34	10	59	10	45	10	32	10	6	10	4	10	90
11	86	11	47	11	56	11	83	11	73	11	8	11	7	11	139
12	102	12	68	12	123	12	81	12	55	12	22	12	8	12	175
13	61	13	48	13	85	13	69	13	48	13	6	13	2	13	139
14	60	14	70	14	95	14	94	14	69	14	22	14	4	14	212
15	157	15	103	15	119	15	115	15	58	15	12	15	2	15	183
16	131	16	117	16	88	16	89	16	68	16	10	16	4	16	276
17	204	17	84	17	158	17	106	17	78	17	12	17	2	17	185
18	235	18	182	18	319	18	153	18	113	18	21	18	12	18	316
19	223	19	144	19	257	19	233	19	119	19	18	19	2	19	364
20	171	20	128	20	220	20	236	20	154	20	31	20	4	20	450
21	153	21	90	21	243	21	186	21	149	21	19	21	7	21	574
22	181	22	121	22	207	22	188	22	231	22	21	22	11	22	460
23	217	23	149	23	223	23	234	23	187	23	24	23	7	23	432
24	261	24	173	24	237	24	262	24	213	24	31	24	13	24	287
25	49	25	58	25	37	25	18	25	5	25	0	25	1242	25	89
26	66	26	56	26	91	26	71	26	71	26	1	26	1106	26	86
TOT	3181	TOT	1973	TOT	3125	TOT	2554	TOT	1829	TOT	283	TOT	2446	TOT	5037

Table 7. Symmetry vs quality mark distribution for symmetry class files.

	Mark	Misc	C	F	B	T	U	R	H	O	P	Q	S	M	N	Z	TOT
Cubic	*	759	33	55	14	0	0	0	0	0	0	0	0	0	0	0	861
	C	41	17	85	4	0	0	0	0	0	0	0	0	0	0	0	147
	I	332	29	68	20	1	0	0	0	1	0	0	0	0	0	0	451
	Other	956	298	373	94	0	0	0	0	0	1	0	0	0	0	0	1722
	TOT	2088	377	581	132	1	0	0	0	1	1	0	0	0	0	0	3181
Tetra-	*	457	0	0	1	25	51	0	0	0	0	0	0	0	0	0	537
gonal	C	17	0	0	0	9	0	0	0	0	0	0	0	0	0	0	26
	I	278	1	0	0	47	65	0	0	0	0	1	0	0	0	0	392
	Other	570	0	0	1	205	244	0	0	1	0	0	0	0	0	0	1021
	TOT	1322	1	0	2	286	360	0	0	1	0	1	0	0	0	0	1973
Hex and	*	724	0	0	0	0	0	16	60	1	0	0	0	0	0	0	801
rhombo-	C	61	0	0	0	0	0	4	32	0	0	0	0	0	0	0	97
hedral	I	450	0	0	0	0	0	34	135	1	0	0	0	0	1	0	621
	Other	904	0	0	1	0	0	103	598	0	0	0	0	0	0	0	1606
	TOT	2139	0	0	1	0	0	157	825	2	0	0	0	0	1	0	3125
Ortho-	*	791	0	0	0	0	0	0	1	19	2	5	0	0	0	0	818
rombic	C	53	0	0	0	0	0	0	0	4	3	2	0	0	0	0	62
	I	516	0	1	0	0	0	0	2	76	25	21	5	1	0	0	647
	Other	669	0	1	0	0	0	0	0	209	24	100	24	0	0	0	1027
	TOT	2029	0	2	0	0	0	0	3	308	54	128	29	1	0	0	2554
Monoclinic	*	717	0	0	0	0	0	0	0	1	0	0	0	4	5	0	727
	C	54	0	0	0	0	0	0	0	0	0	0	0	3	2	0	59
	I	521	0	2	0	0	0	0	7	3	0	0	0	10	21	0	564
	Other	393	0	0	0	0	0	0	2	2	0	0	0	47	35	0	479
	TOT	1685	0	2	0	0	0	0	9	6	0	0	0	64	63	0	1829
Triclinic	*	105	0	0	0	0	0	0	0	0	0	0	0	0	0	0	105
	C	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14
	I	92	0	0	0	0	0	0	1	2	0	0	0	0	0	1	96
	Other	66	0	0	0	0	0	0	0	0	0	0	0	0	0	2	68
	TOT	277	0	0	0	0	0	0	0	1	2	0	0	0	0	3	283
Y file	*	935	0	0	0	0	0	0	0	0	0	0	0	0	0	0	935
	C	113	0	0	0	0	0	0	0	0	0	0	0	0	0	0	113
	I	1398	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1398
	Other	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	TOT	2445	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2446
X file	*	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	I	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Other	5037	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5037
	TOT	5037	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5037

POWDER DIFFRACTION PATTERN QUALITY

In the last phase of this study we evaluated the quality of a number of the patterns in the Powder Diffraction File. To do this, we postulated several functions and obtained the de Wolff M_{20} function [5] from the literature. Many of these functions required a knowledge of systematic extinctions for each pattern, thus restricting us to those patterns with both lattice parameters and space groups on the files described in the previous section. We needed the space group symbols to generate systematic extinctions and the lattice parameters to calculate theoretical 2θ positions, and thus were faced with an extensive amount of editing. All of the original errors as well as those that resulted from the various stages of keypunching became painfully obvious. To reduce the size of this task we restricted our studies to only the highest (cubic) and lowest (triclinic) symmetry cases.

Programs SPCUB and SPTRI were written to merge the space group and cell parameter data with sets 1 through 24 of the inorganic PDF master tape. These programs selected the cubic and triclinic patterns, respectively, and also checked the space group symbols against all valid symbols. In addition, they checked, corrected, and added the space group number searched for any inconsistencies, and merged the d-I information into the binary files, PDFCUB and PDFTRI. When all editing and error correction was complete, the files contained 2080 cubic and 292 triclinic patterns with all needed information. Table 8 lists the frequency of occurrence of the various space groups encountered in the PDF cubic and triclinic files. This tabulation is similar to others for all crystal systems [6,7].

To compare the average data found in the PDF to known high-quality data, we encoded the PDF numbers for all the cubic patterns determined by the National Bureau of Standards [8]. The patterns that correspond to these numbers were extracted from the cubic file PDFCUB to form a file called PDFNBSCUB which contains 315 patterns. The NBS patterns as they appear on the PDF master tape were affected by editorial procedures. All digits in d more than three places after the decimal had been eliminated, causing $\Delta 2\theta$ errors of more than 0.5 deg at high angles. Another difficulty arises from different values of the x-ray wavelengths used over the years. Both of these problems can be avoided if the measured 2θ values instead of a functionally derived quantity like d are used in publications and computer databases.

Because we still needed a set of accurate, reliable powder patterns for comparison purposes, we keypunched the complete set of 336 cubic NBS patterns and 38 high-quality triclinic patterns. The NBS patterns in their original published form are accurate to the level of the measurements made and

Table 8. Distribution of cubic and triclinic space groups.

No. of patterns with this space group	Space group symbol	Space group number
Cubic		
1	P23	195
3	F23	196
1	I23	197
76	P2(1)3	198
11	I2(1)3	199
4	PM3	200
4	PN3	201
47	FM3	202
3	FD3	203
15	IM3	204
95	PA3	205
42	IA3	206
8	P4(2)32	208
1	F4(1)32	210
2	I432	211
3	P4(3)32	212
10	P4(1)32	213
3	I4(1)32	214
10	P-43M	215
73	F-43M	216
31	I-43M	217
14	P-43N	218
5	F-43C	219
49	I-43D	220
289	PM3M	221
3	PN3N	222
68	PM3N	223
10	PN3M	224
645	FM3M	225
27	FM3C	226
103	FD3M	227
4	FD3C	228
39	IM3M	229
81	IA3D	230
Triclinic		
149	P_1	1
115	P_1	2
1	C_1	
1	A_1	
1	F_1	
2	I_1	
23	C_i	

should be as accurate as any available powder diffraction data. These two files, NBSCUB and NBSTRI, complete the data sets analyzed in this section. The notations on Figs. 10 through 51 mean the following:

- *PDF CUBIC*. These are the 2080 cubic patterns obtained from the PDF master tape and contained in file PDFCUB.
- *PDF-NBS CUBIC*. These are the 254 cubic NBS patterns as extracted from the PDF master tape and contained in file PDFNBSCUB.
- *TRUE NBS CUBIC*. These are the 336 cubic NBS patterns as punched from the original publications and contained in file NBSCUB.
- *PDF TRICLINIC*. These are the 315 triclinic patterns as extracted from the PDF master tape and contained in file PDFTRI.
- *NBS TRICLINIC*. These are the 38 high-quality triclinic patterns obtained from NBS publications and de Wolff patterns from the PDF card file.

We wrote five programs to analyze these five data files. The programs generated all of the possible reflections for each pattern read, and then eliminated all systematic extinctions required by the space group symbol. It also rejected any pattern with lines that could not be matched to a calculated line within 0.50-deg 2θ. This relatively mild rejection criterion eliminated over 400 of the PDF CUBIC patterns. Table 9 gives the actual number of patterns that could be indexed and that were used to produce all the figures in this section.

With the exception of Fig. 10, the figures in this section fall into five types labeled (a) through (e). Figures 11 to 15 show plots characterizing four of the five data files and testing some ideas of the authors. For the functions tested in these plots, the PDF-NBS cubic file is identical to the true NBS cubic data file. Thus, there are only four illustrations in each of these figures. Figures 16 through 51 test the properties of the nine different functions along with their average values. The functions tested are:

1. $\Delta d \quad = |d_{exp} - d_{calc}|$.

2. $\Delta 2\theta \quad = |2\theta_{exp} - 2\theta\ calc|$.

3. $\Delta 2\theta_{rms} = \sqrt{|2\theta_{exp} - 2\theta_{calc}|^2}$.

4. N/N_{poss}. This is the ratio of the number of a particular diffraction line (called N or N_{obs}), starting from the lowest angle reflection, to the number of possible lines, excluding systematically absent reflections. We refer to this quantity as "RATIO".

5. $$F_N = \frac{1}{\overline{|\Delta 2\theta|}} \frac{N}{N_{poss}}$$

The N in F_N refers to the number of the last diffraction line used in the calculation and is called AV. 2 THETA MERIT. F_N, which resulted from this work, is described in Ref. 9.

6. F_N (RMS) is similar to F_N, but uses the root mean square average $\Delta 2\theta$ in its definition, referred to as RMS 2θ MERIT.

7. $$R_{rms} = \sqrt{\frac{\sum |\Delta 2\theta|^2}{\sum 2\theta_{exp}^2}}$$

Table 9. Characteristics of the cubic and triclinic files and average values of functions evaluated for all lines of patterns to be indexed.

File	No. in file	No. with space group and cell	Calculated patterns	No. to be indexed	Evaluated					
					$\overline{\lvert\Delta 2\theta\rvert}$	N_{obs}/N_{poss}	F_N	M_N	M_{20}	R
PDF-cubic	3087	2080	117	1638	0.091	0.766	16.39	1109	40.1	0.156
PDF-NBS cubic	258	254	0	252	0.032	0.833	29.29	1447	77.5	0.057
NBS cubic	339	336	5	326	0.015	0.822	74.76	2396	124.9	0.025
PDF TRI	315	292	17	229	0.053	0.266	9.77	285	9.0	0.206
NBS TRI	38	38	0	37	0.019	0.490	33.00	624	14.8	0.072

$$8.\ M_{20} = \frac{Q_{20}}{2\overline{|\Delta Q|}\ N_{poss}}$$

where $Q_{20} = \frac{1}{d^2}$ of the 20th line.

This is a function derived by de Wolff and proposed for use in evaluating the reliability of a computer-indexed cell (see Ref. 5).

$$9.\ M_N = \frac{Q_N}{2\overline{|\Delta Q|}} \frac{N_{obs}}{N_{poss}} .$$

This function was our attempt to generalize the de Wolff M_{20} function so that it could be evaluated at any line number (N).

Table 9 lists average values for several of the above functions evaluated over all of the lines in each pattern. Using all lines in a pattern to evaluate any of the merit functions presents a problem for low-symmetry materials because the number of possible diffraction lines increases dramatically with diffraction angle. Thus, the matching of calculated and observed lines gets artificially better for high diffraction angles. This "easy matching" phenomenon causes an artificial lowering of $\Delta 2\theta$ (or ΔQ) for high line numbers. Figure 10 shows the average F_N figure of merit as a function of N for PDF triclinic data . The increasing value of F_N above line 30 results from the "easy matching" phenomenon. To avoid this artificial increase in the merit functions evaluated, we decided to limit the calculations to the first 30 lines of those patterns with more than 30 lines. Thus, in all of the figures that refer to an average function for a pattern, this function was evaluated up to the 30th or last line if the pattern contained less than 30 lines.

Figures 11 and 12 show the distribution of cubic lattice parameters and the ā parameters for triclinic at 0.2 and 0.1 Å resolution. The clustering of parameters for cubic materials is the result of a large number of common structure types in the file. Figure 13 shows the distribution of the intensities of all the diffraction lines on the four files. These distributions are of interest for various theoretical considerations of pattern searching procedures. Figure 14 shows the average | $\Delta 2\theta$ | of each line on the files as a function of the intensity of the line. These plots indicate, somewhat surprisingly, that the error in measuring the position of a diffraction line for average quality data is independent of the line's intensity. A correlation does appear to exist for the high-quality cubic data. However, because these patterns are uniformly reported to high angle where the rounding of d values has its most serious effect in | $\Delta 2\theta$ | and where reflections tend to be of linear intensity, this may be an artificial correlation. Because the intensity of the data were not coded onto the true NBS cubic file, this effect cannot be verified, although the lack of a correlation in Figure 14(d) for high-quality triclinic data appears to indicate that the cubic correlation is false. Figure 15 shows the F_N merit function vs PDF set number. Because these sets are released once per year, they indicate that average pattern quality has not significantly improved with time. The higher values of F_N in some of the earliest sets are not totally accurate because JCPDS editors have systematically replaced all but the best patterns over the years.

Figures 16 through 24 show average values for the nine functions evaluated up through line numbers 10, 20, 30, 40, and at the last line, whatever its number may be. These functions were evaluated for cubic primitive (P), face-centered (F), and body-centered (I) Bravais lattices as well as for triclinic primitive (P), any face-centered (F), and all faces or body-centered (I). In addition, these functions were evaluated for all the data on each file and for those patterns marked with an asterisk. Because we encoded the data for the true NBS file, no asterisks were attached; for the PDF-NBS file all patterns are marked with an asterisk. The existence of non-primitive triclinic patterns may appear unusual, but several triclinic cells have been reported in this manner. The slashed bars on top of the average bars represent the standard deviations of the averages.

Figures 25 through 33 show distributions of the average values for the nine functions evaluated at the 30th or last line, whichever comes first. To obtain these distributions, the average values were rounded off to the nearest ordinate value. For these figures, the values of the functions were divided into 100 intervals.

Figures 34 to 42 show average values for the nine functions evaluated at the line numbers from 5 to 55. The prominant increase in the merit functions and decrease in the error functions at line number 40 apparently results from the editorial policy previously described, to only include complete data for patterns of better-than-average quality. Another possibility is that authors publishing complete patterns may systematically have taken more care in

measuring the pattern. The single line in the M_{20} plots reflects the fact that this function is only defined at the 20th line.

Figures 43 to 51 show average values of the nine functions evaluated at different d values, starting with the largest. The generally increasing error functions and decreasing merit functions as d gets smaller is the result of the same editorial procedure that forced the independent encoding of the original high-quality cubic and triclinic patterns. The rounding off of d values on the magnetic tape database has its most severe effect at high angles and low d values. Note that pattens throughout this study were only evaluated up to the limit of the Cu Kα sphere of reflection. Thus, these plots do not extend lower than d=0.78.

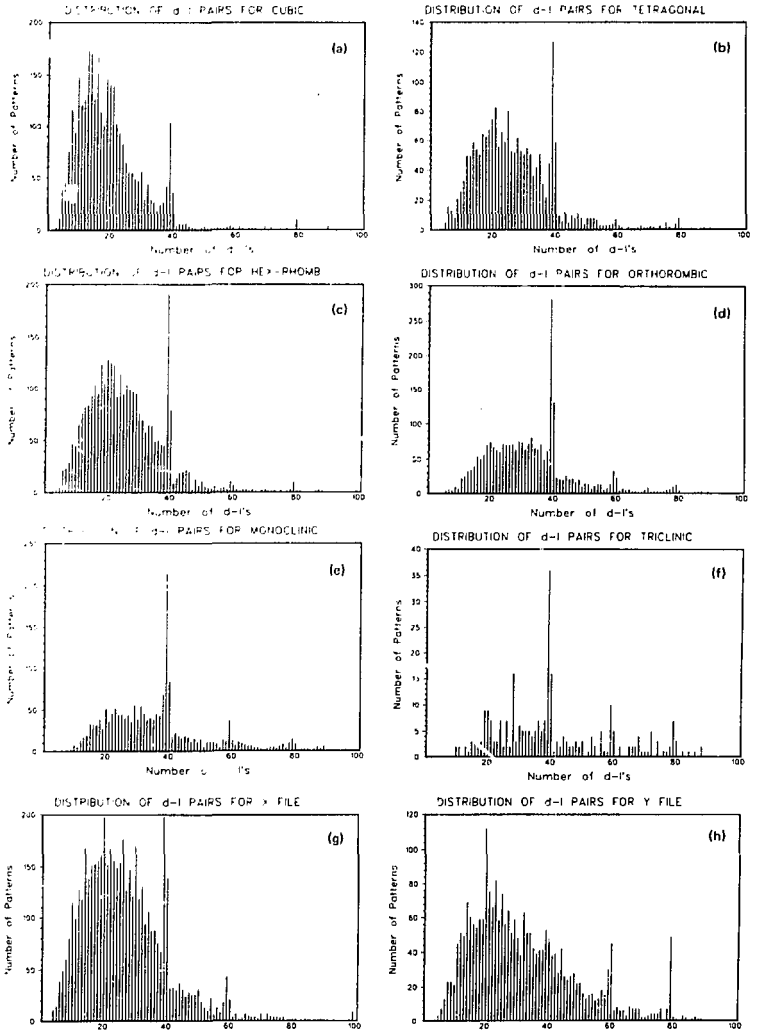

Fig. 2. Distribution of the number of d-I pairs per pattern on the eight crystal-class files.

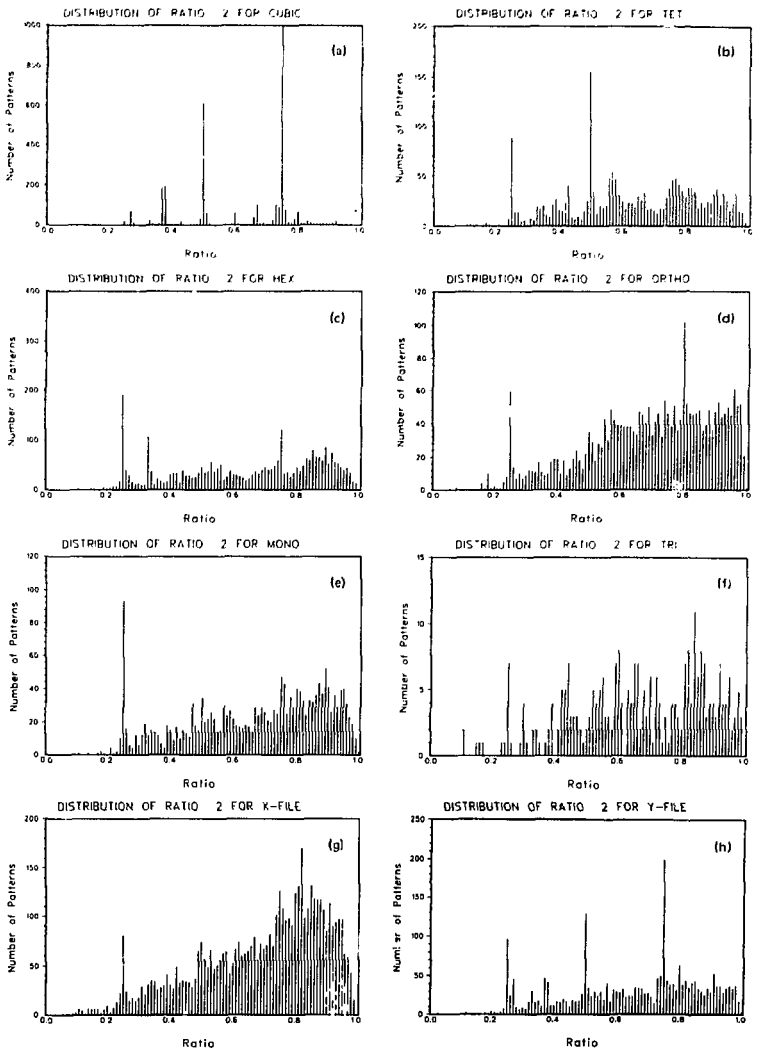

Fig. 3. Distribution of the value of d_2^2/d_1^2 for the eight crystal-class files.

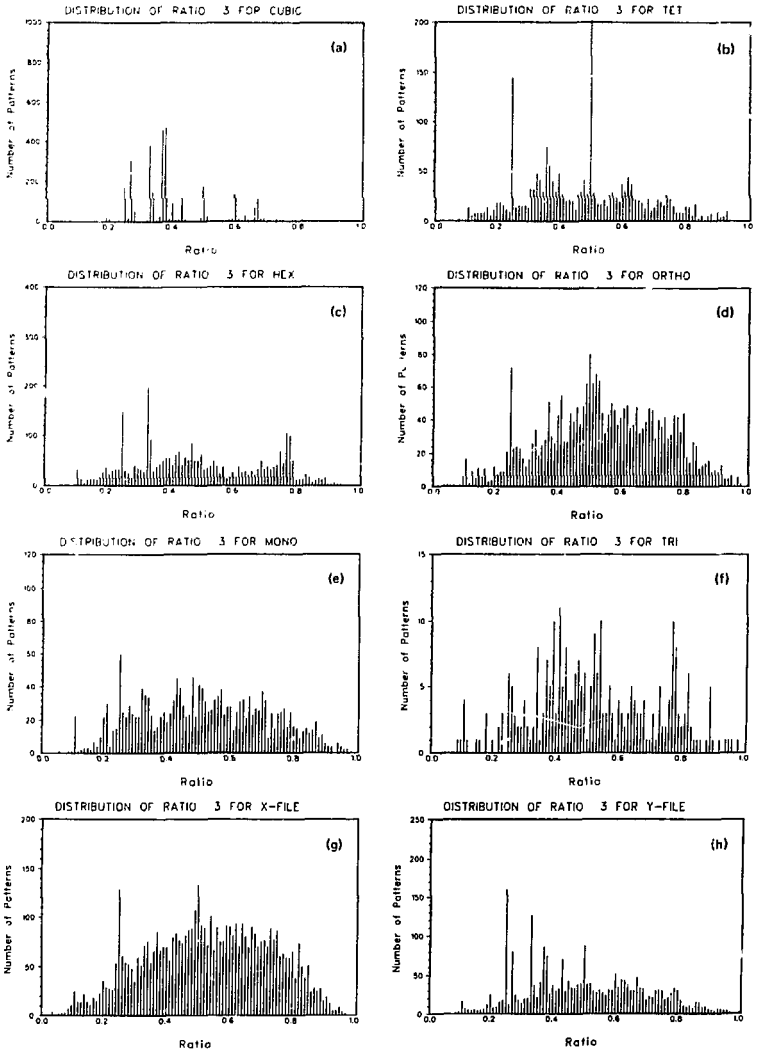

Fig. 4. Distribution of the value of d_3^2/d_1^2 for the eight crystal-class files.

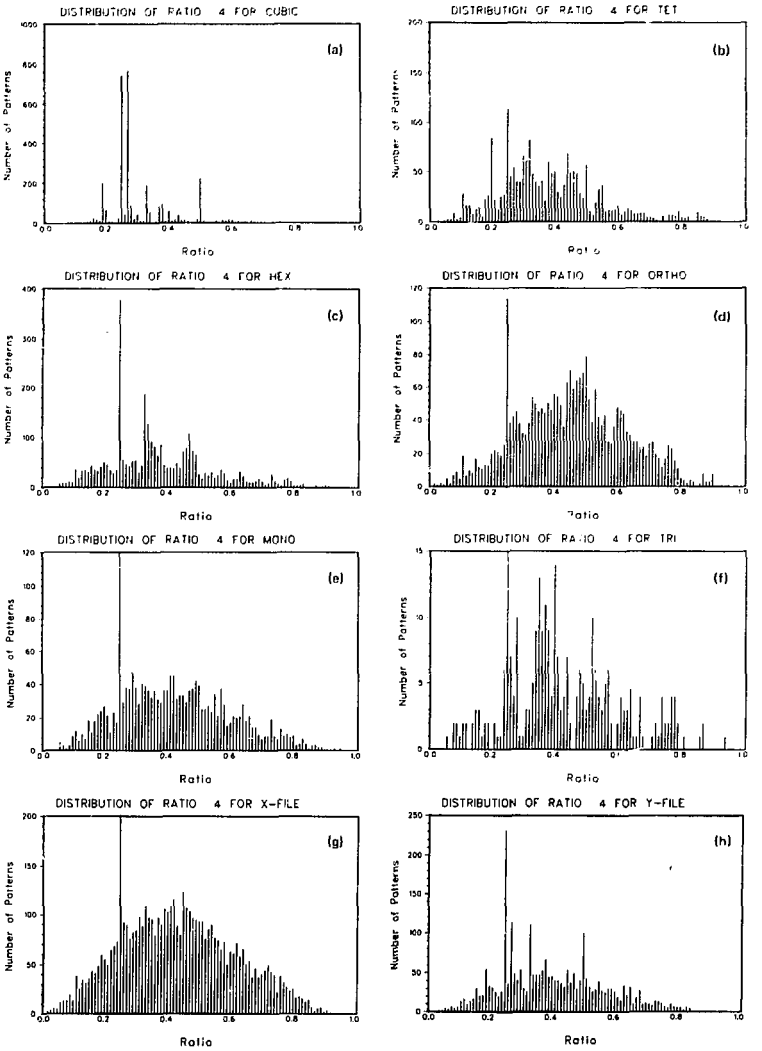

Fig. 5. Distribution of the values of d_4^2/d_1^2 for the eight crystal-class files.

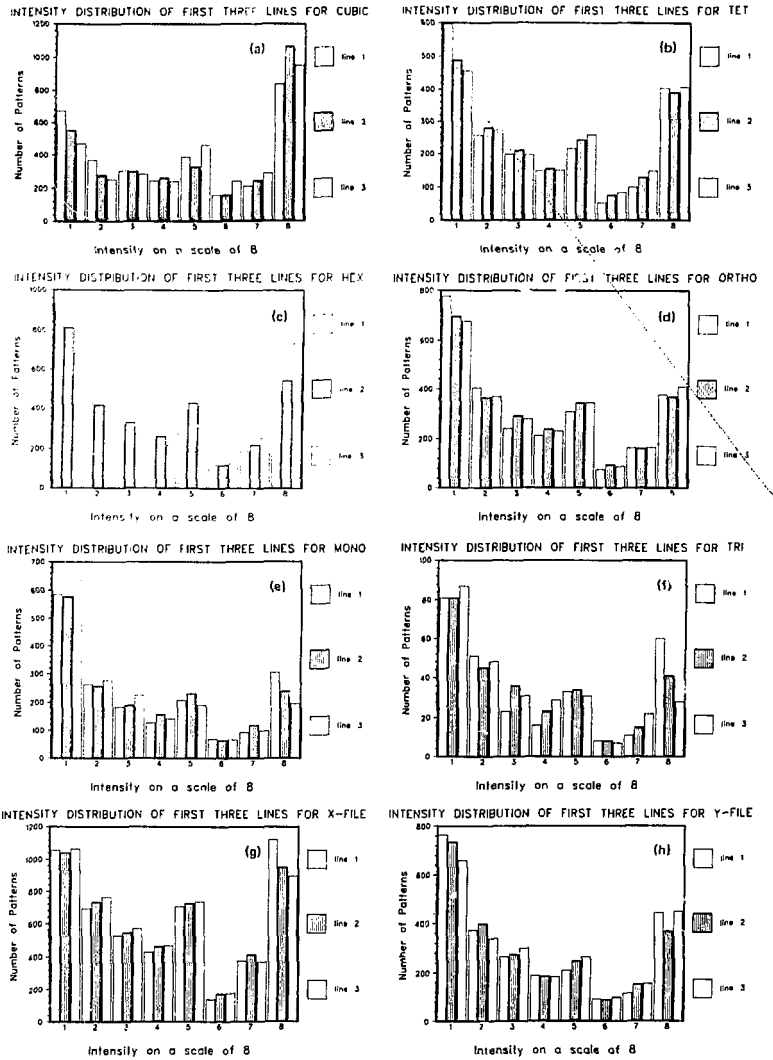

Fig. 6. Distribution of the intensities of the first three lines in each pattern for the eight crystal-class files. The intensity of line 1 is represented by the bar farthest to the left in each group of three.

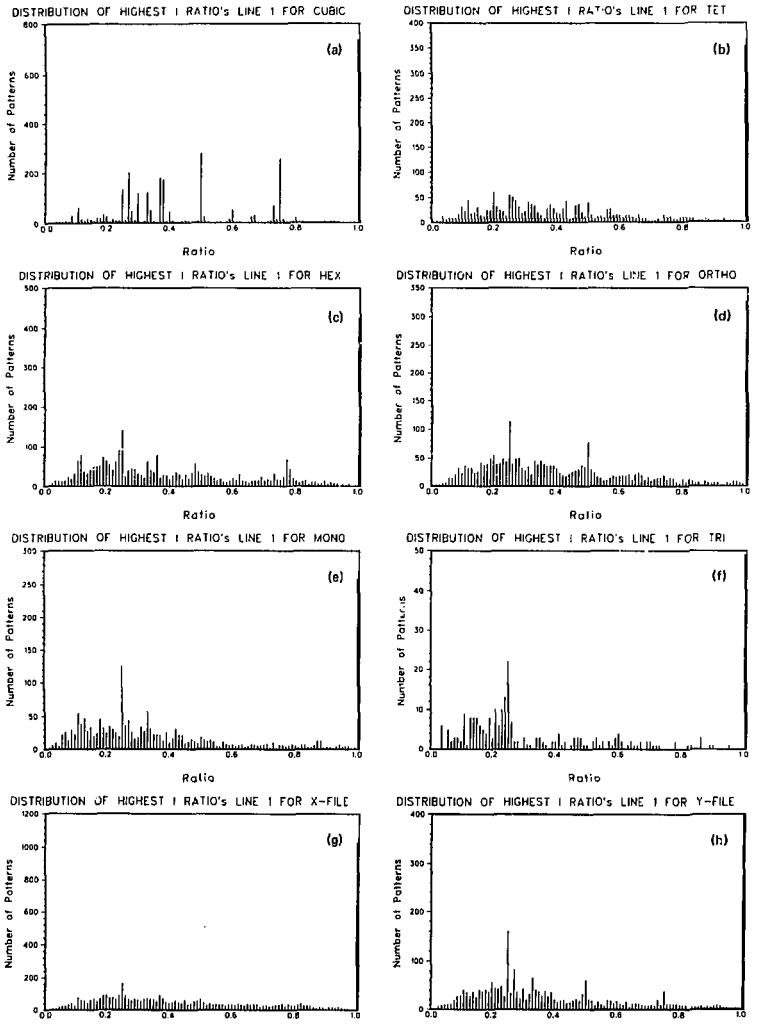

Fig. 7. Distribution of the d^2/d_1^2 values for the most intense line in the patterns on the eight crystal-class files.

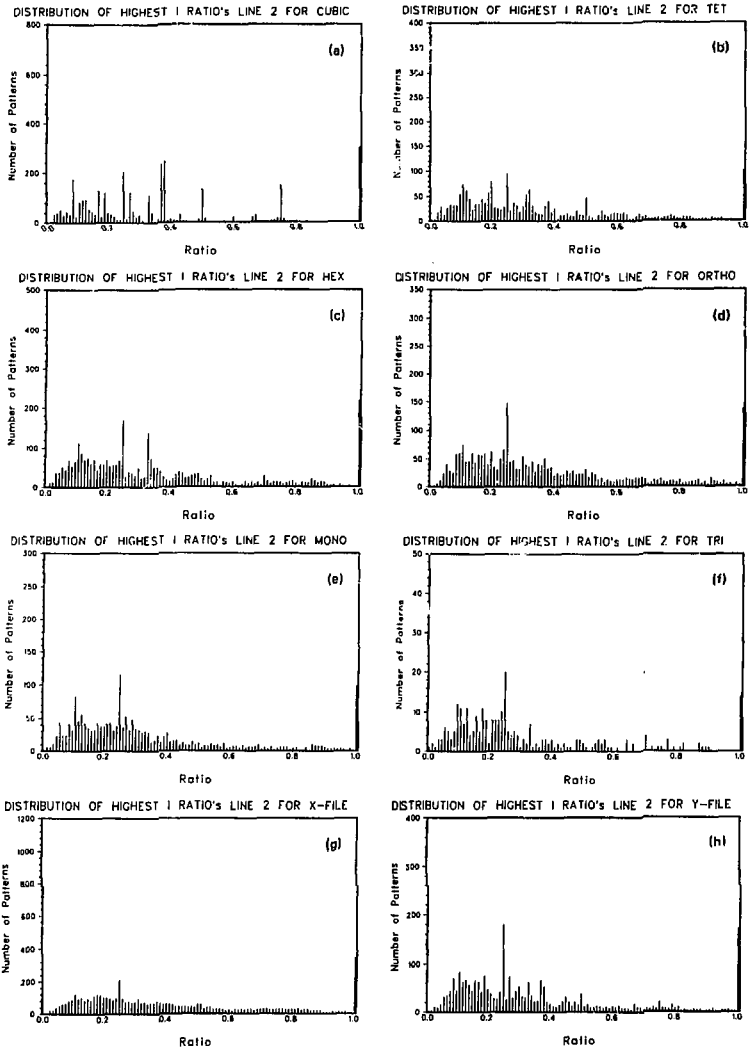

Fig. 8. Distribution of the d_2^2/d_1^2 values for the second most intense line in the patterns on the eight crystal-class files.[1]

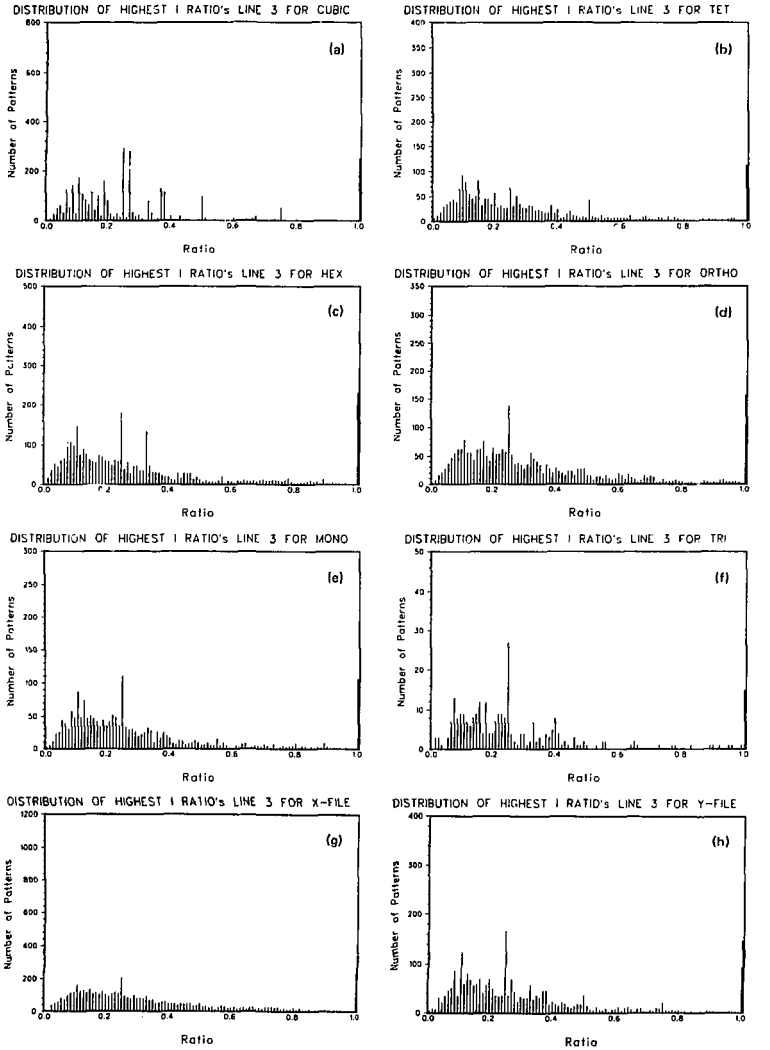

Fig. 9. Distribution of the d^2/d_1^2 values for the third most intense line in the patterns on the eight crystal-class files.

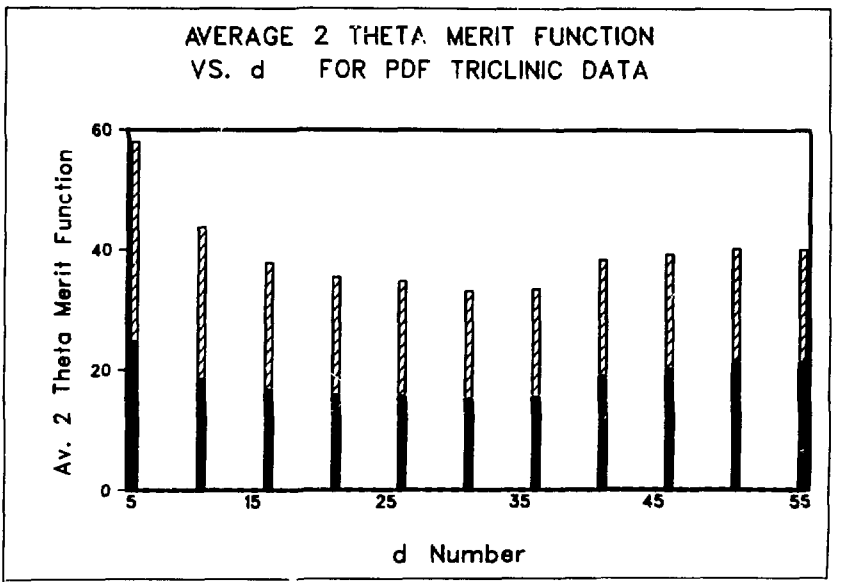

Fig. 10. Distribution of average values of the merit function F_N vs line number in each pattern. Slashed bars represent the standard deviation of average F_N values.

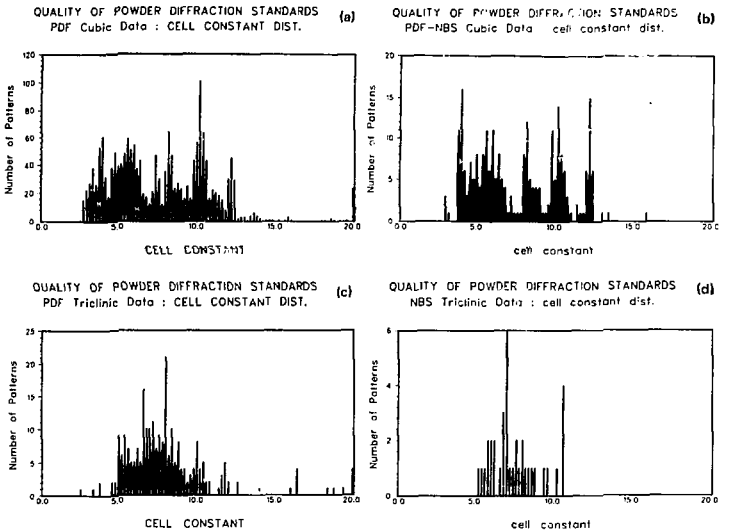

Fig. 11. Distribution of cubic lattice parameters and triclinic a values at 0.2 $\AA$ resolution.

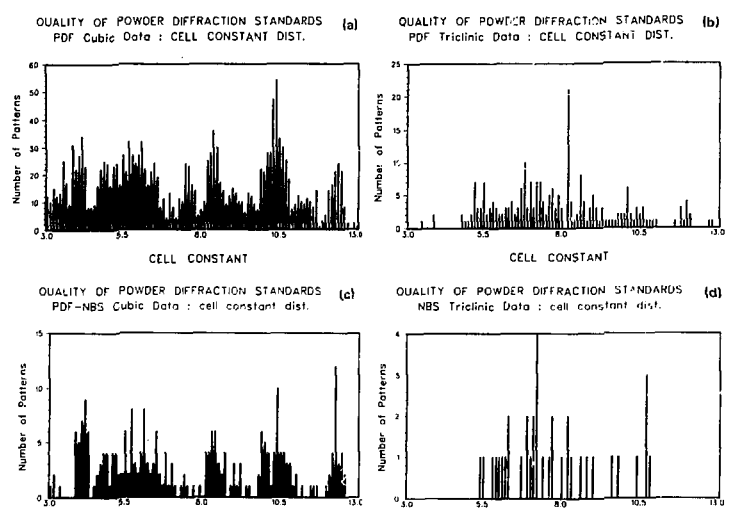

Fig. 12. Distribution of cubic lattice parametes and triclinic a values at 0.1 $\AA$ resolution.

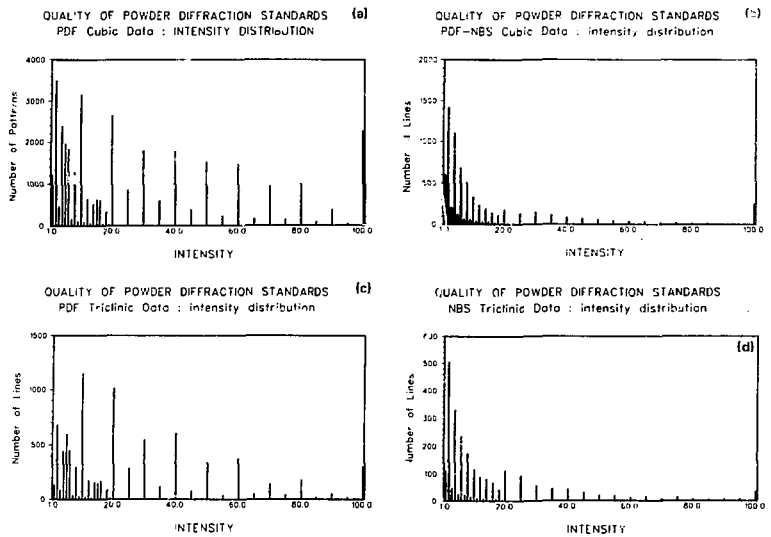

Fig. 13. Distribution of the intensities of the diffraction lines on the two cubic and two triclinic files.

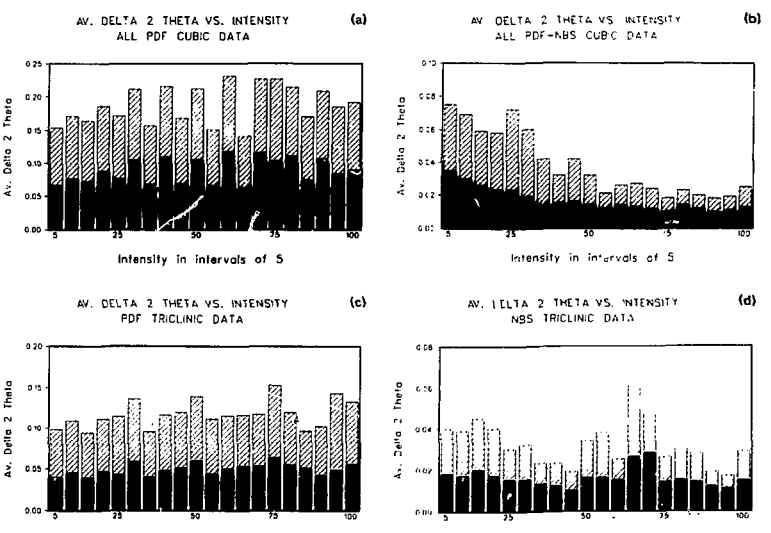

Fig. 14. Average 2θ error as a function of line intensity. Stashed bars represent standard deviations.

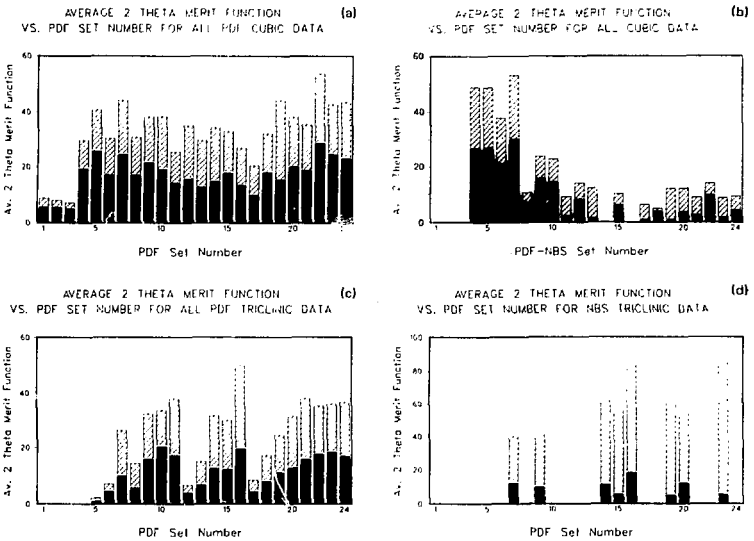

Fig. 15. Average values for the figure of merit F_N as a function of the PDF set number. Slashed bars represent standard deviations.

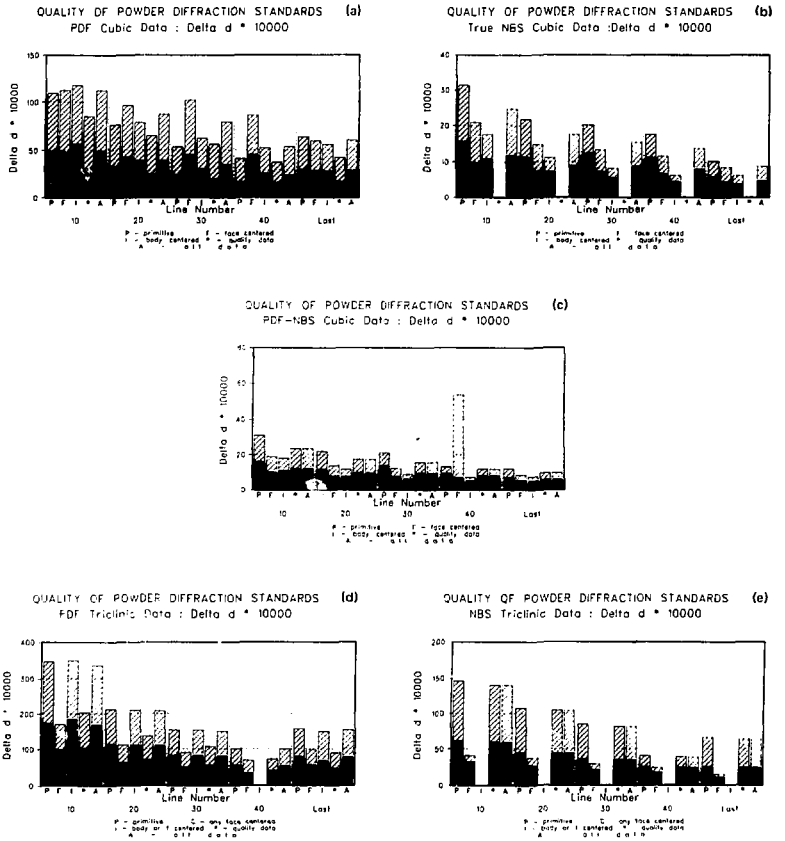

Fig. 16. Averagae value of | Δd | as a function of line number, lattice type, and PDF quality mark. Slashed bars represent standard deviations.

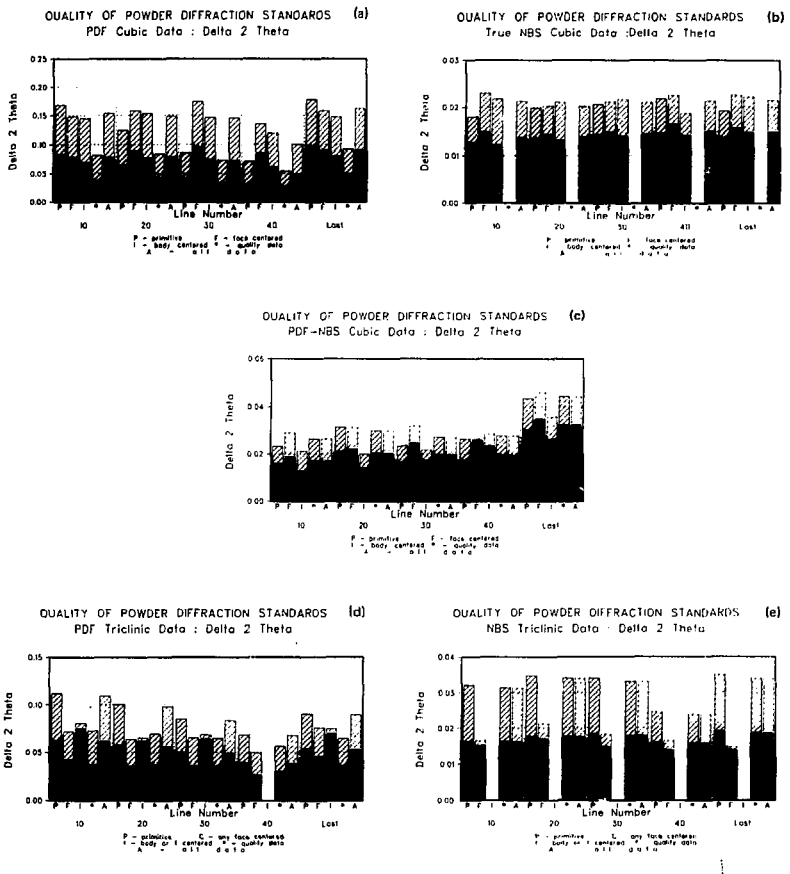

Fig. 17. Average value of $|\Delta 2\theta|$ as a function of line number, lattice type, and PDF quality mark. Slashed bars represent standard deviations.

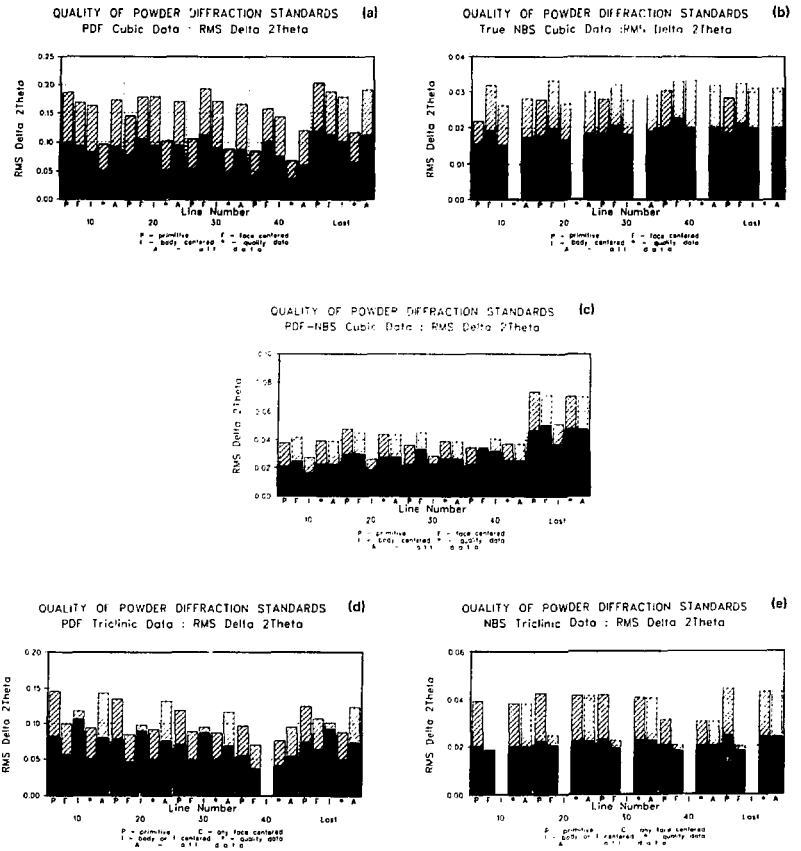

Fig. 18. Average value of root mean square $|\Delta 2\theta|$ as a function of line number, lattice type, and PDF quality mark. Slashed bars represent standard deviations.

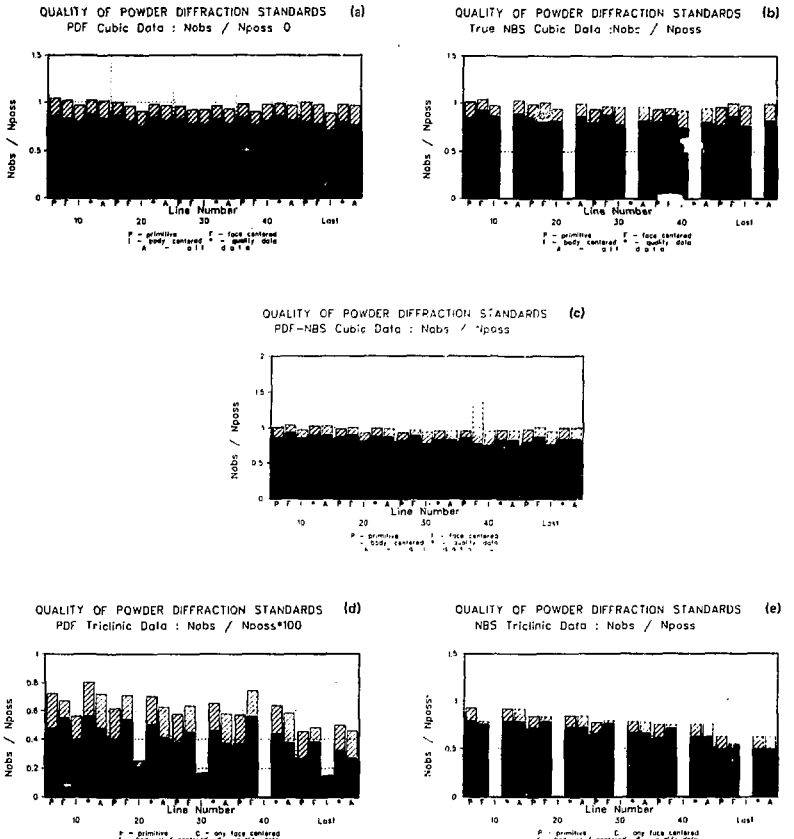

Fig. 19. Average value of the coverage factor N/N_{poss} as a function of line number, lattice type, and PDF quality mark. Slashed bars represent standard deviations.

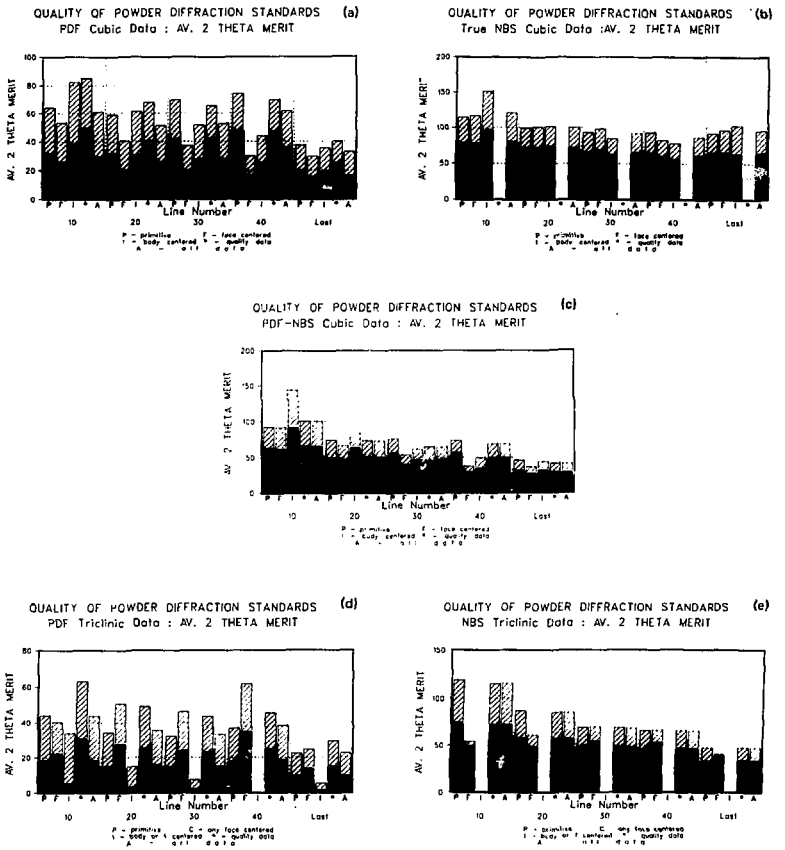

Fig. 20. Average value of F_N as a function of line number, lattice type, and PDF quality mark. Slashed bars represent standard deviations.

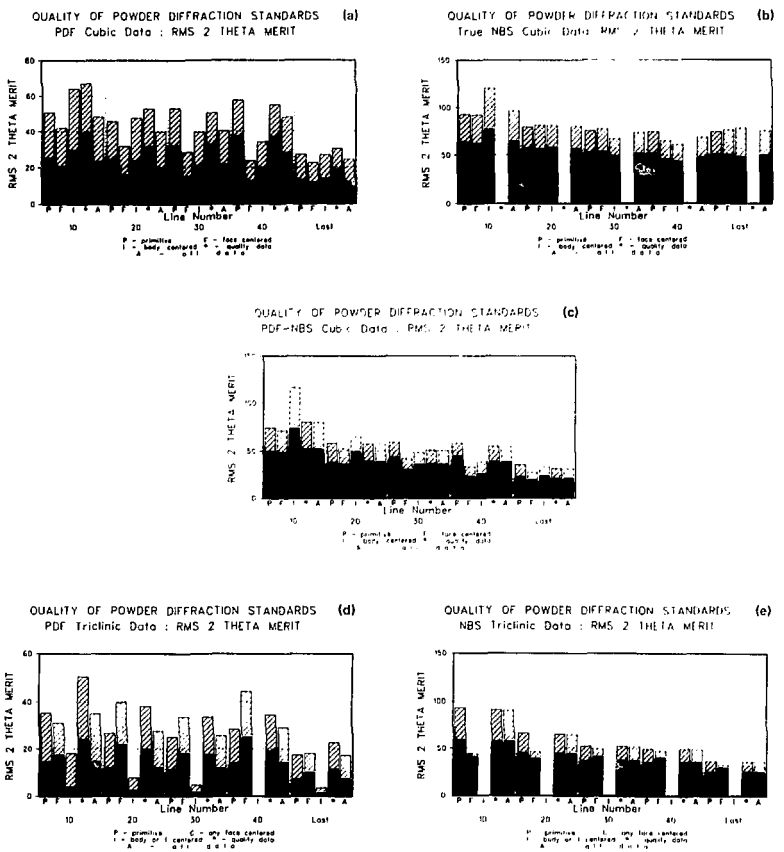

Fig. 21. Average value of root mean square F_N as a function of line number, lattice type, and PDF quality mark. Slashed bars represent standard deviations.

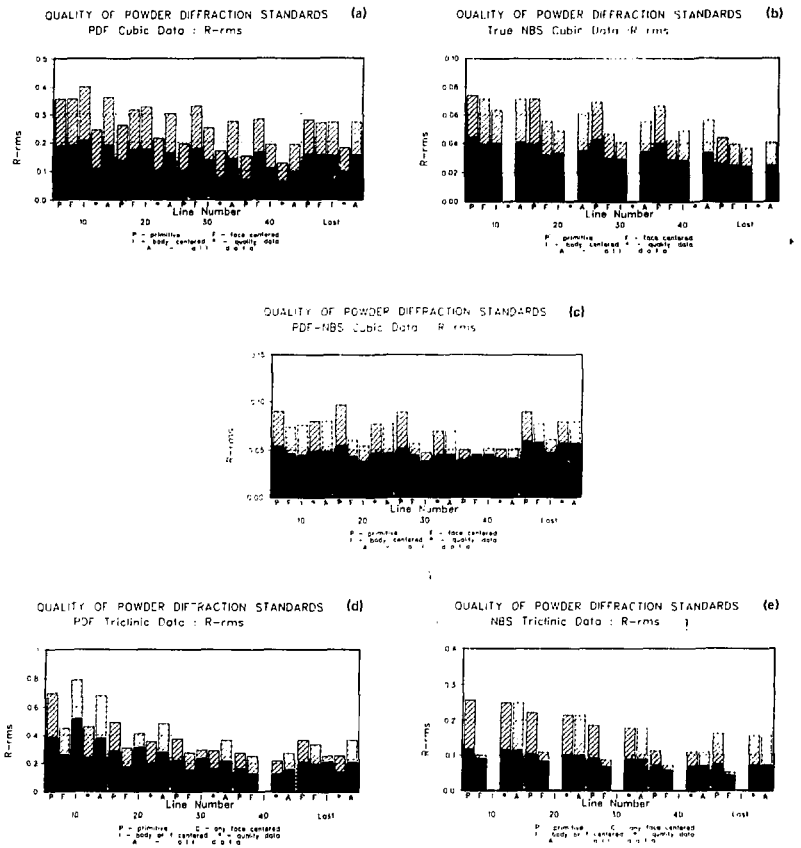

Fig. 22. Average value of R as a function of line number, lattice type, and PDF quality mark. Slashed bars represent standard deviations.

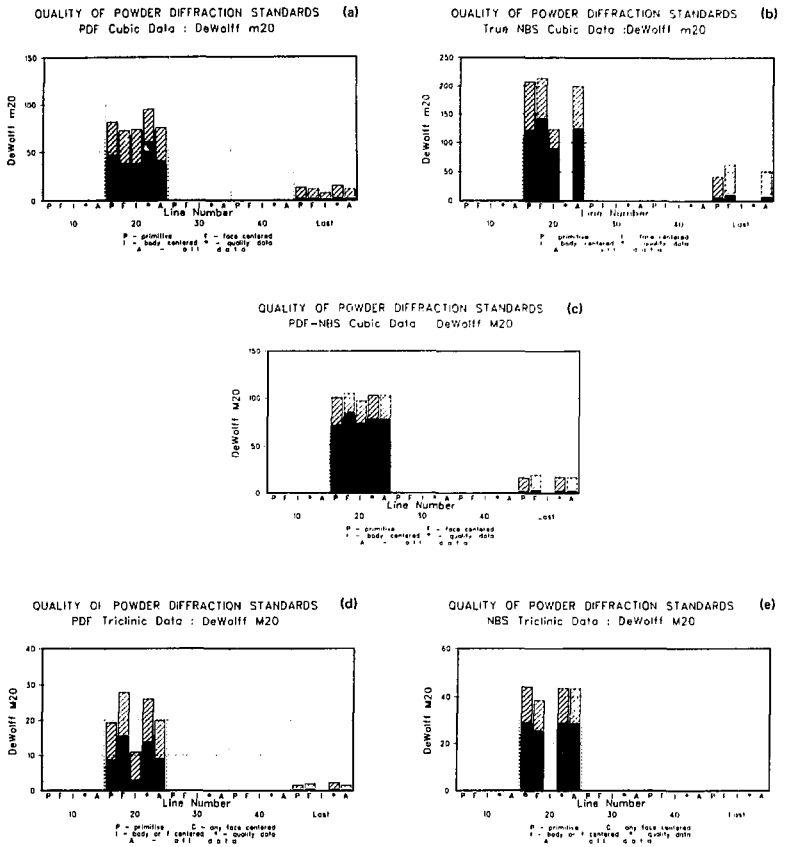

Fig. 23. Average value of M_{20} as a function of line number, lattice type, and PDF quality mark. Slashed bars represent standard deviations.

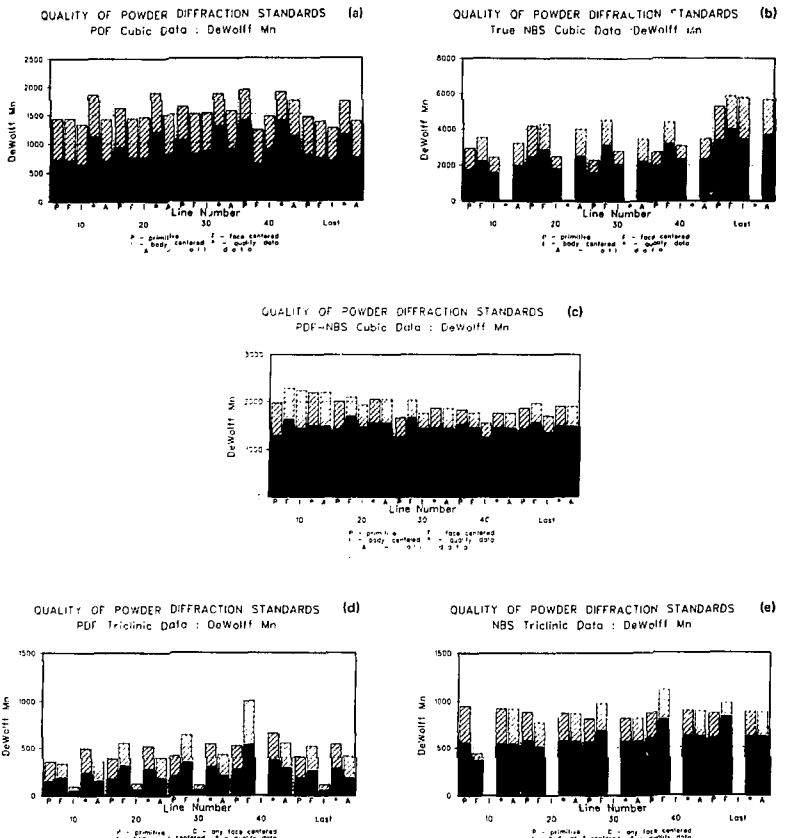

Fig. 24. Average value of M_N as a function of line number, lattice type, and PDF quality mark. Slashed bars represent standard deviations.

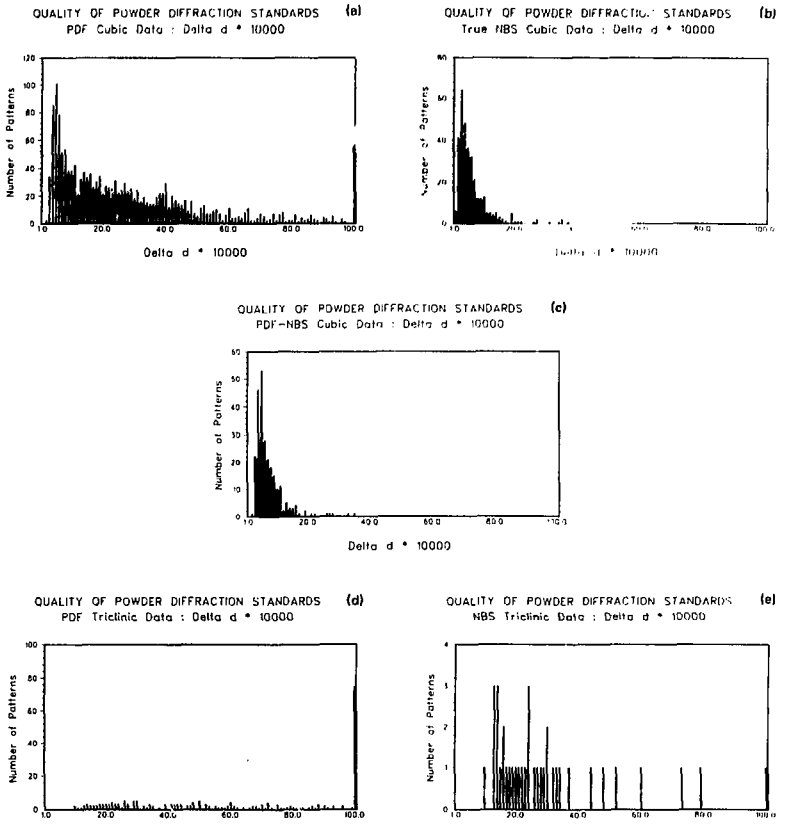

Fig. 25. Distribution of the average value of $|\Delta d|$ for cubic and triclinic patterns.

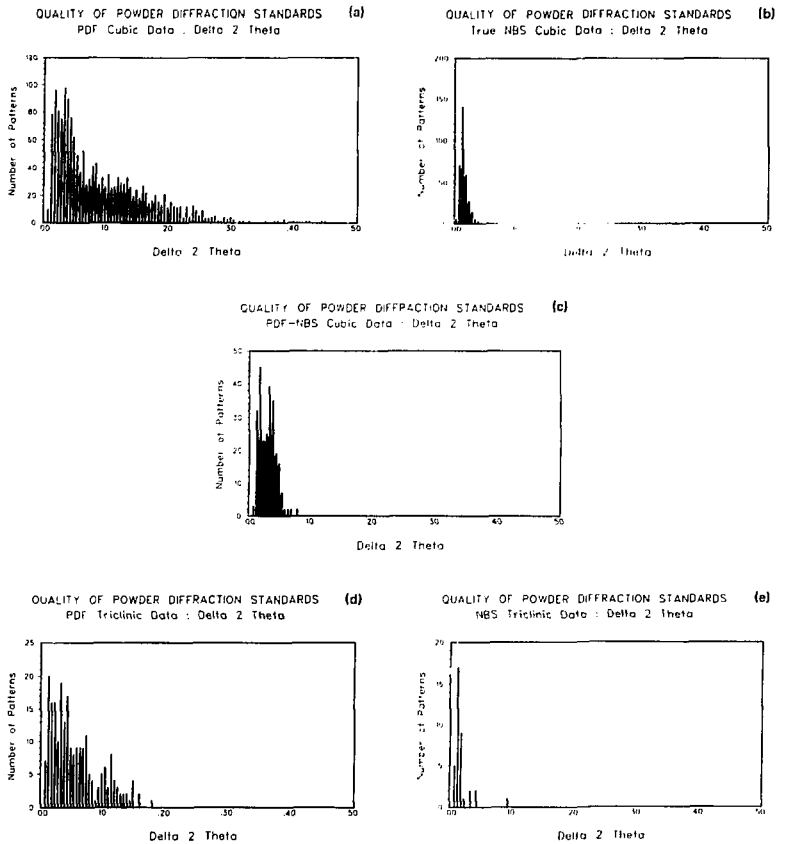

Fig. 26. Distribution of the average value of $|\Delta 2\theta|$ for cubic and triclinic patterns.

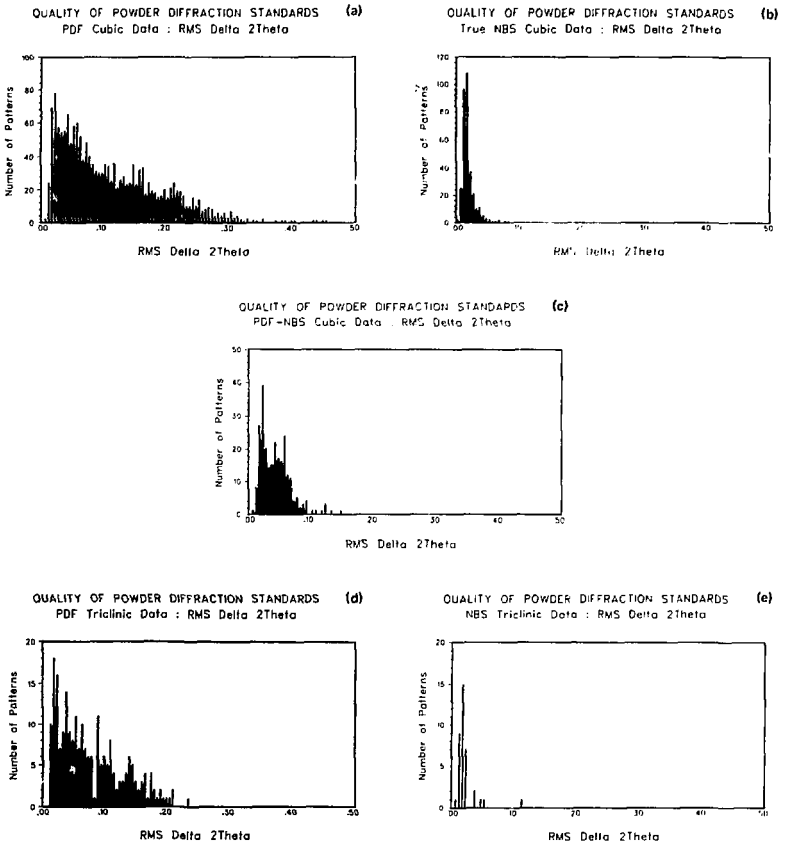

Fig. 27. Distribution of the average value of root mean square $| \Delta 2\theta |$ for cubic and triclinic patterns.

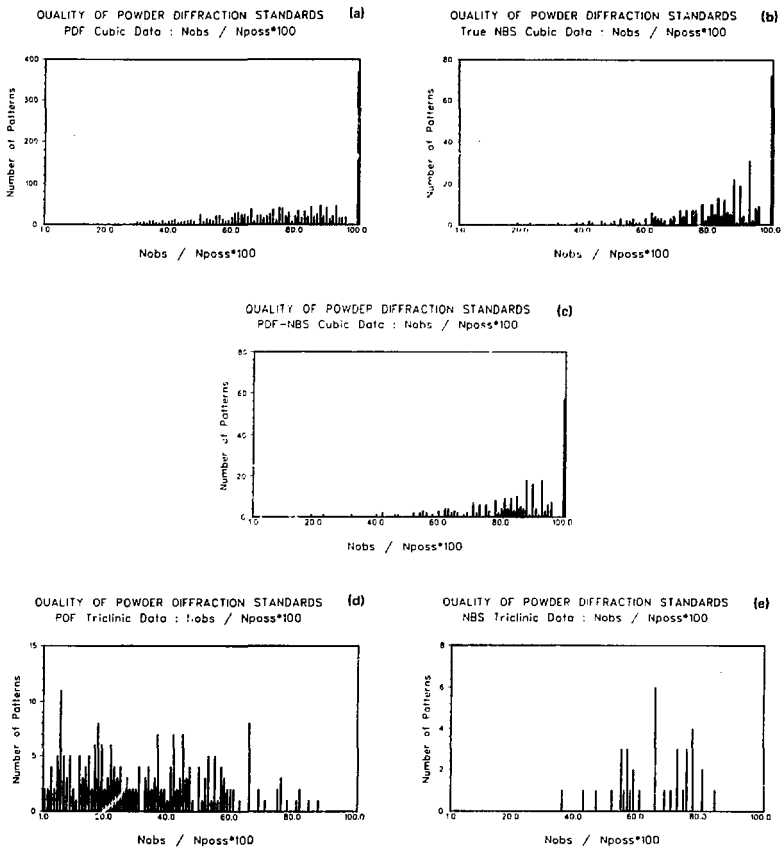

Fig. 28. Distribution of the average value of the coverage factor N/N_{poss} for cubic and triclinic patterns.

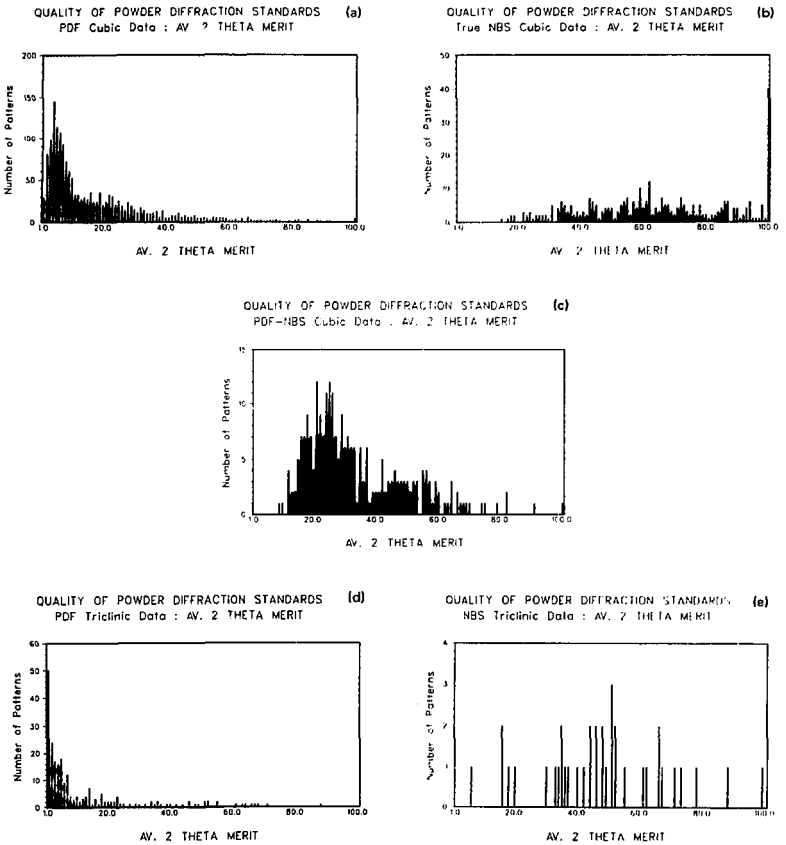

Fig. 29. Distribution of the average value of F_N for cubic and triclinic patterns.

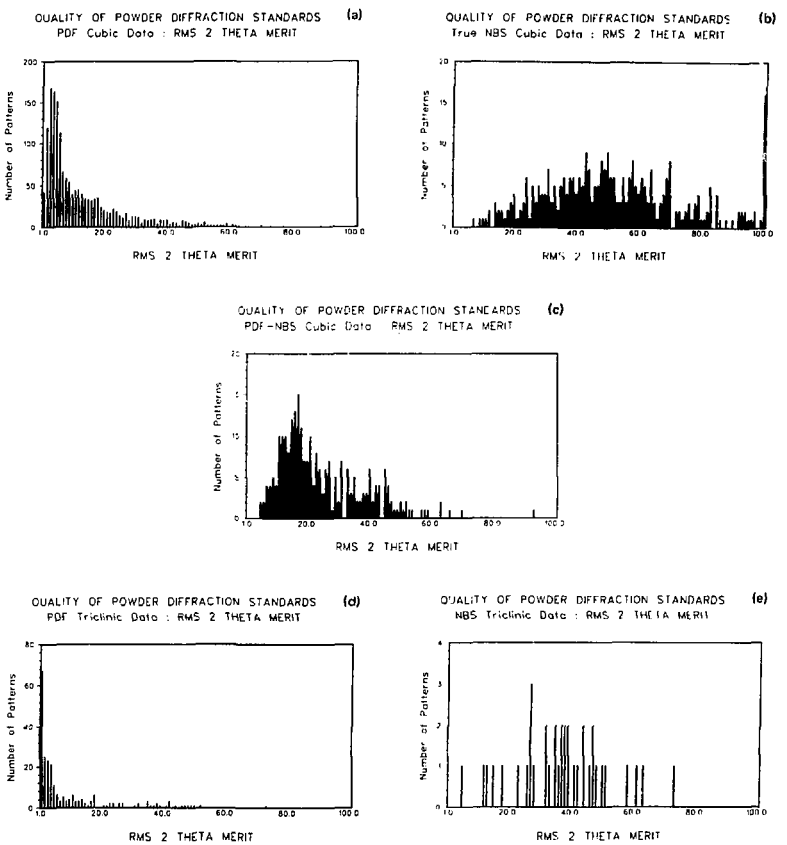

Fig. 30. Distribution of the average value of root mean square F_N for cubic and triclinic patterns.

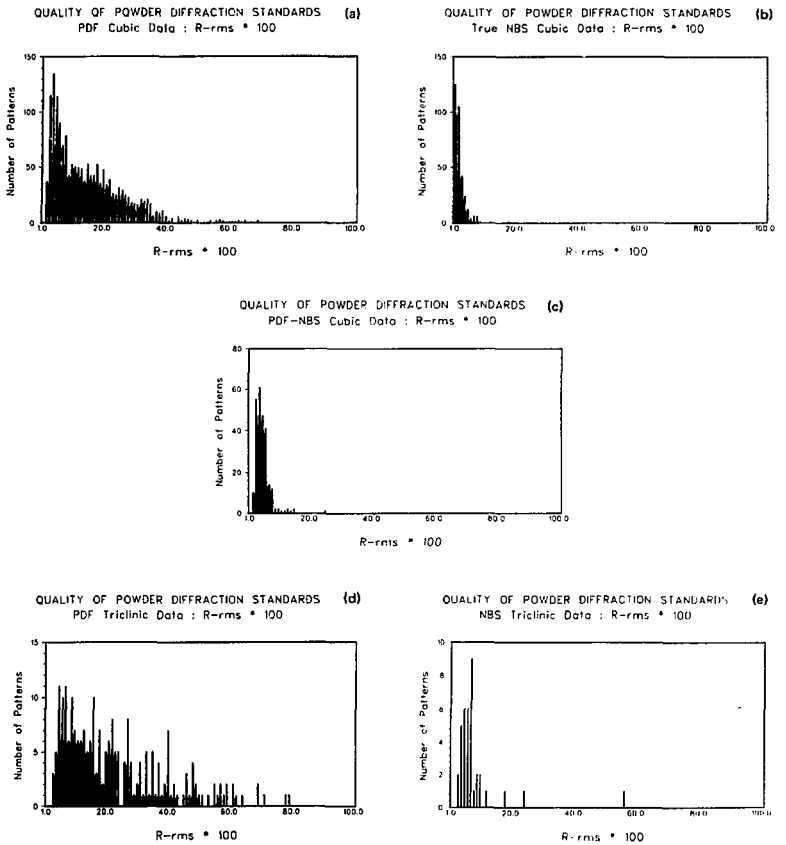

Fig. 31. Distribution of the average value of R for cubic and triclinic patterns.

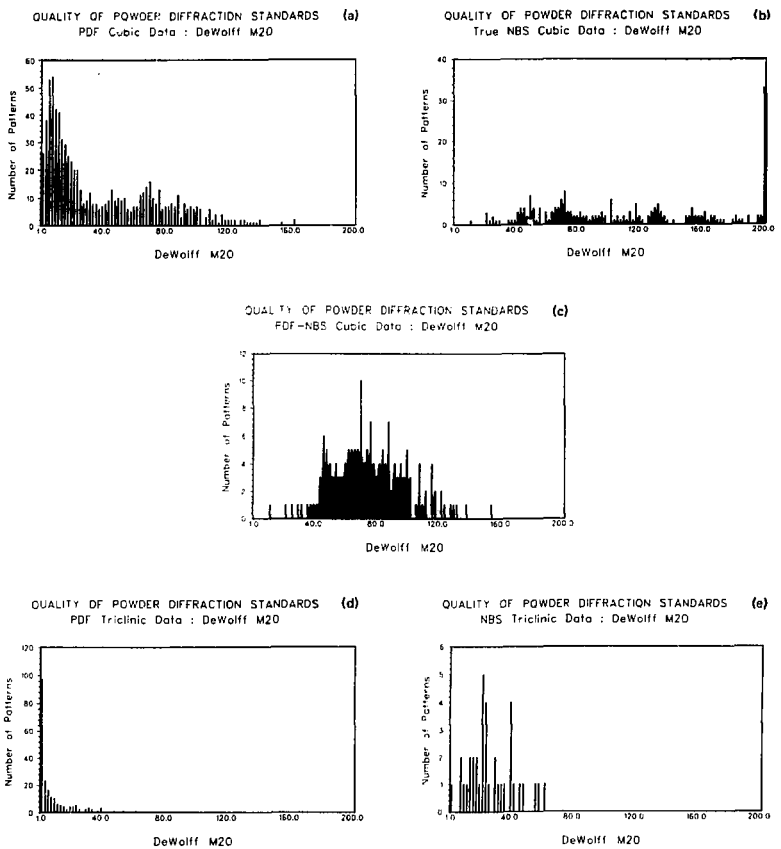

Fig. 32. Distribution of the average value of M_{20} for cubic and triclinic patterns.

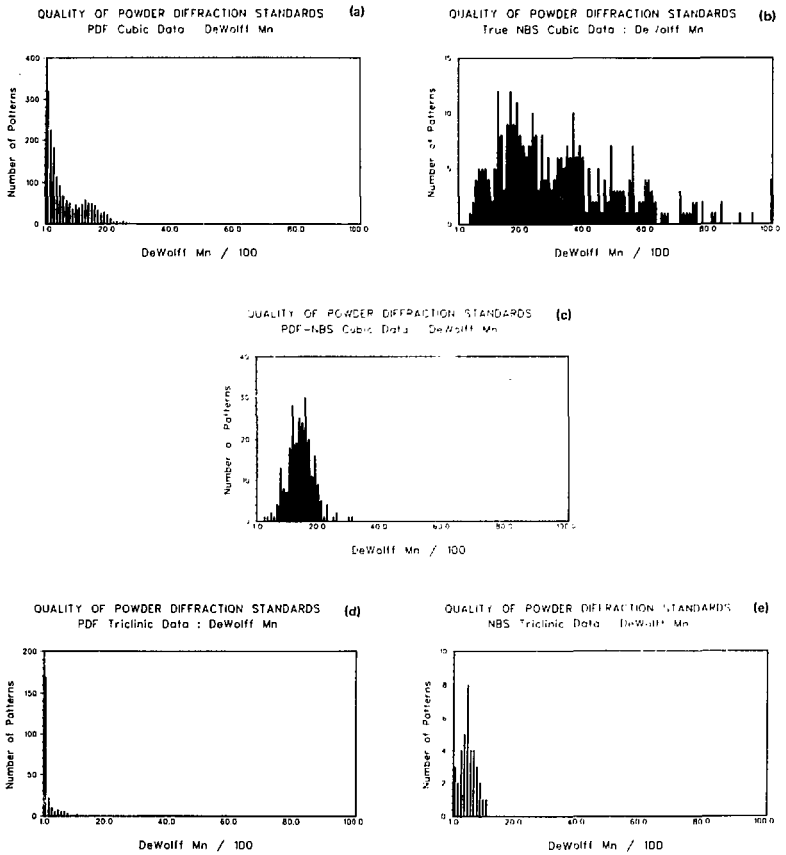

Fig. 33. Distribution of the average value of M_N for cubic and triclinic patterns.

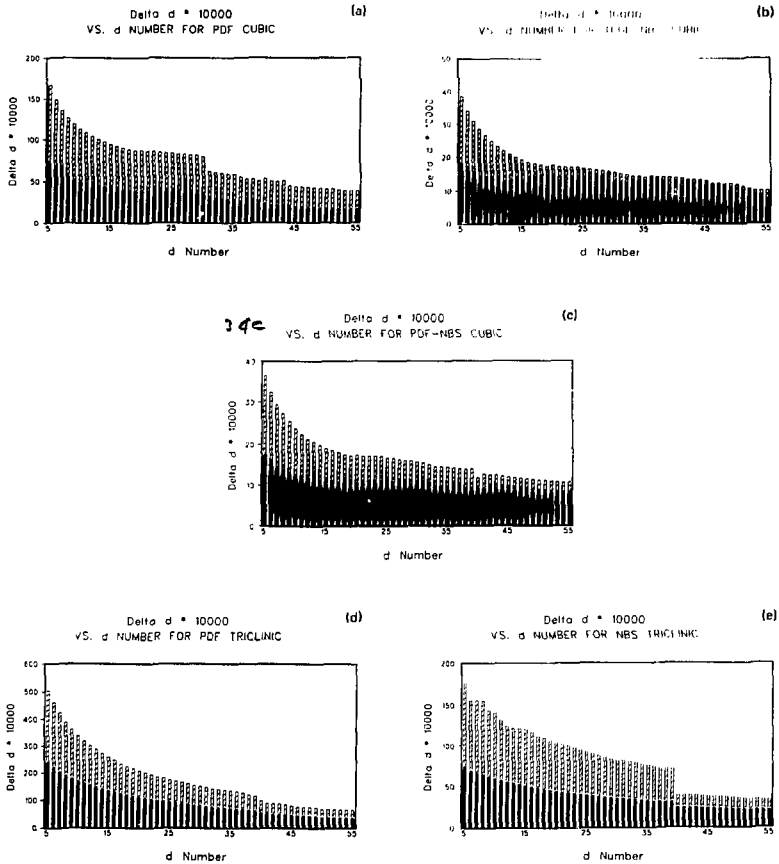

Fig. 34. | Δd | as a function of diffraction line number for cubic and triclinic patterns. Slashed bars represent standard deviations.

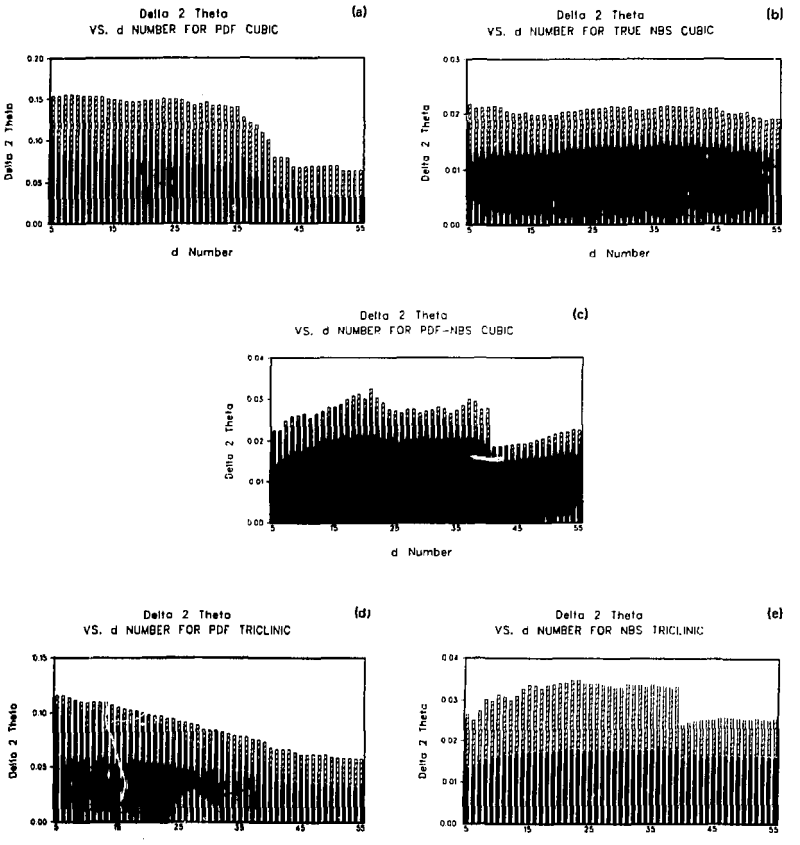

Fig. 35. $|\Delta 2\theta|$ as a function of diffraction line number for cubic and triclinic patterns. Slashed bars represent standard deviations.

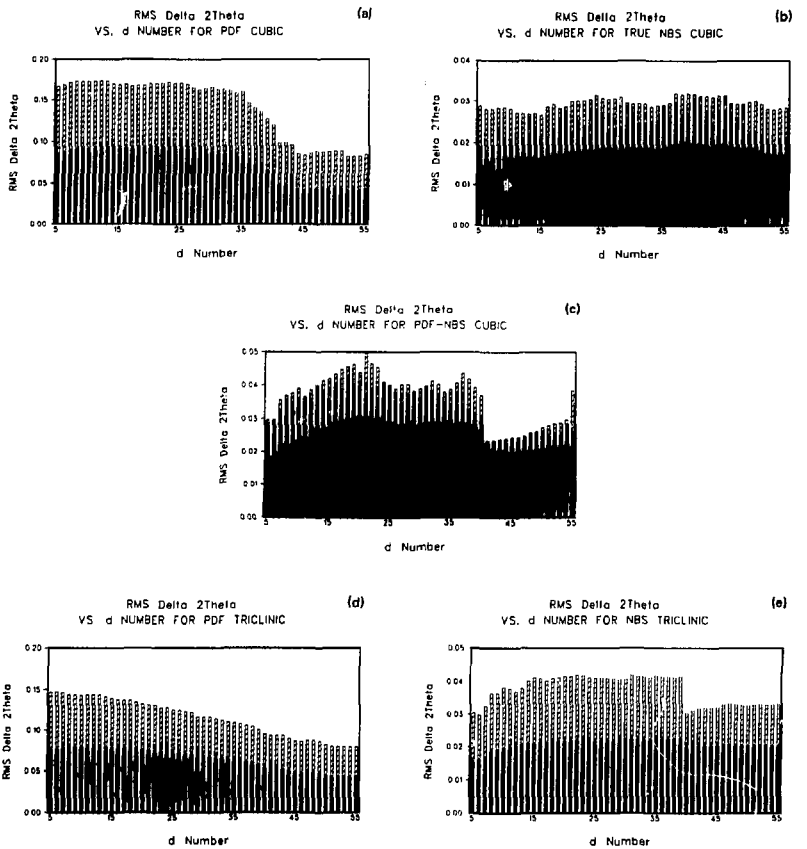

Fig. 36. Root mean square $|\Delta 2\theta|$ as a function of diffraction line number for cubic and triclinic patterns. Slashed bars represent standard deviations.

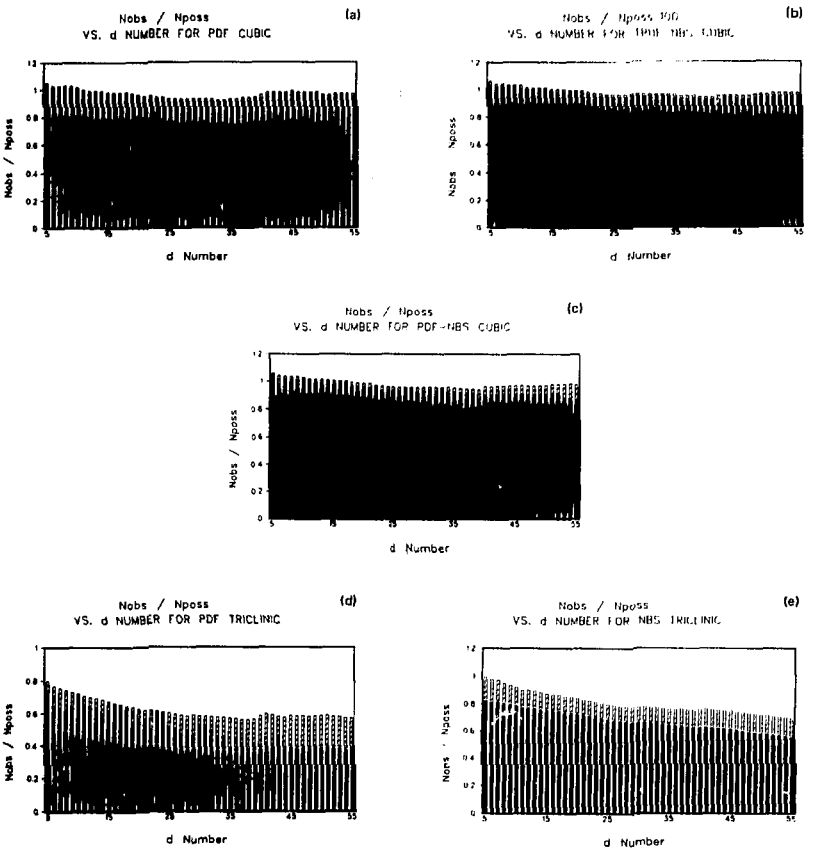

Fig. 37. The coverage factor N/N_{poss} as a function of diffraction line number for cubic and triclinic patterns. Slashed bars represent standard deviations.

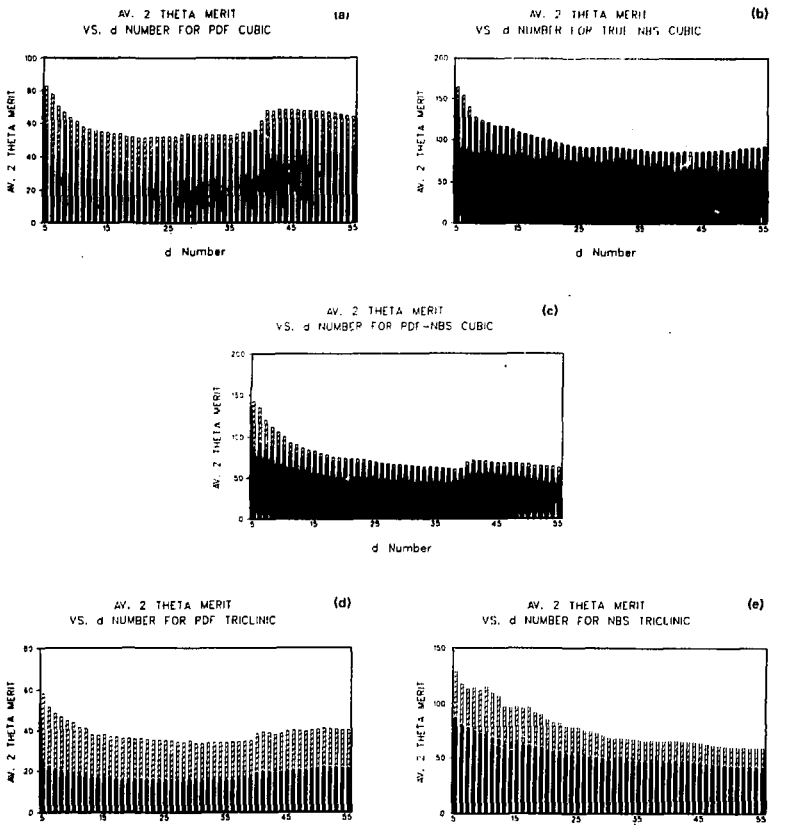

Fig. 38. F_N as a function of diffraction line number for cubic and triclinic patterns. Slashed bars represent standard deviations.

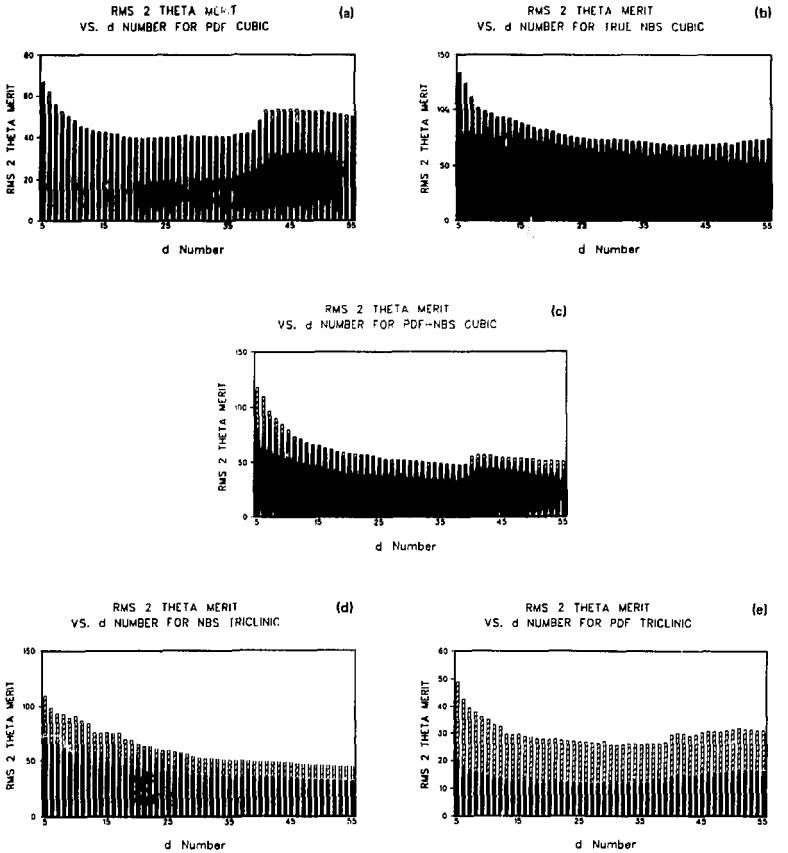

Fig. 39. Root mean square F_N as a function of diffraction line number for cubic and triclinic patterns. Slashed bars represent standard deviations.

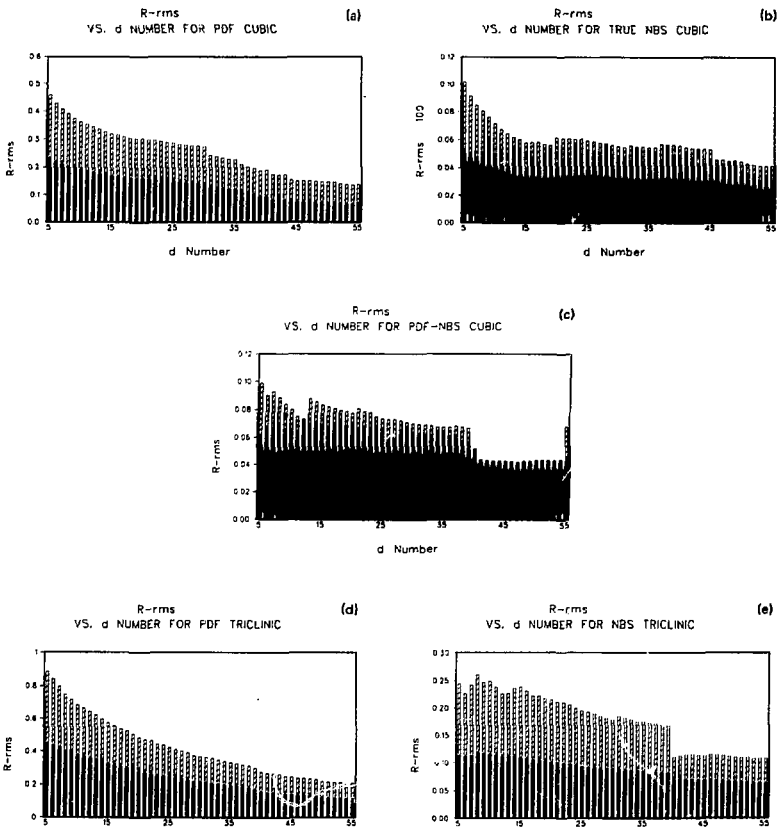

Fig. 40. R as a function of diffraction line number for cubic and triclinic patterns. Slashed bars represent standard deviations.

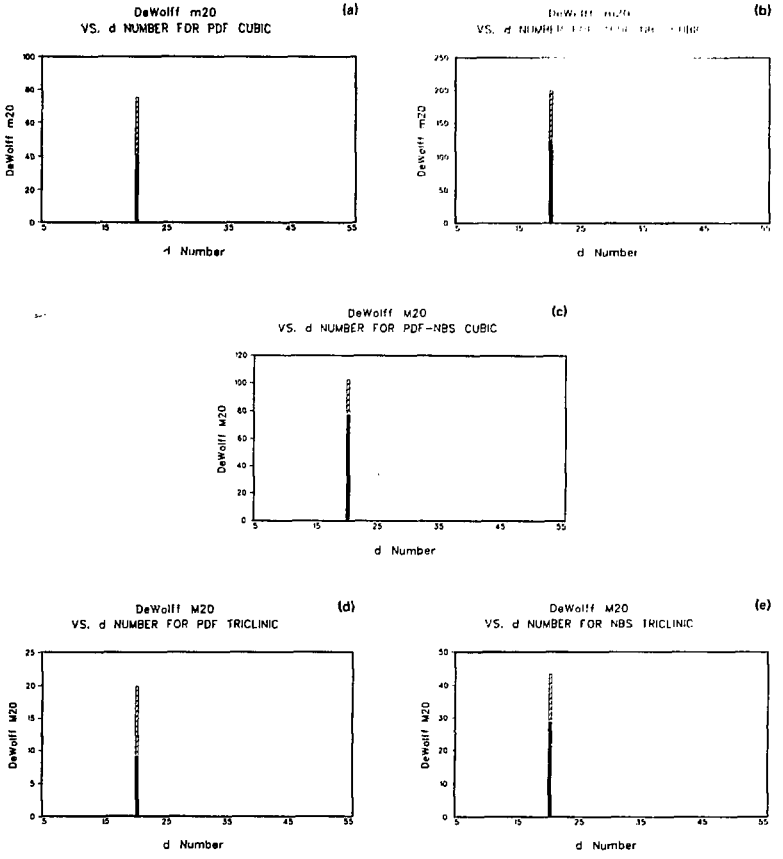

Fig. 41. M_{20} as a function of diffraction line number for cubic and triclinic patterns. Slashed bars represent standard deviations.

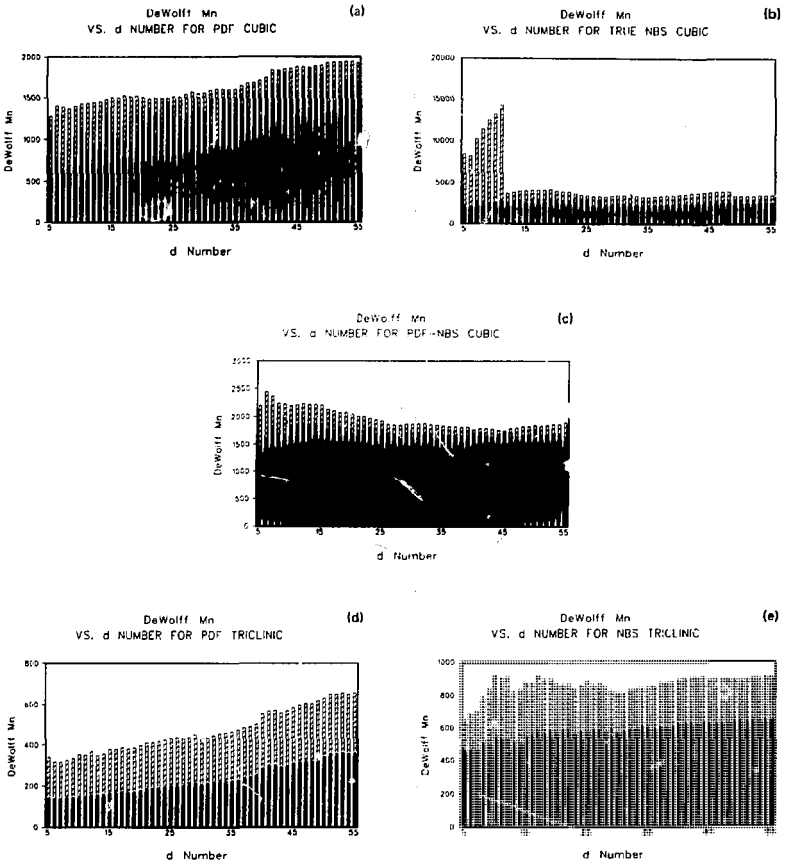

Fig. 42. M_N as a function of diffraction line number for cubic and triclinic patterns. Slashed bars represent standard deviations.

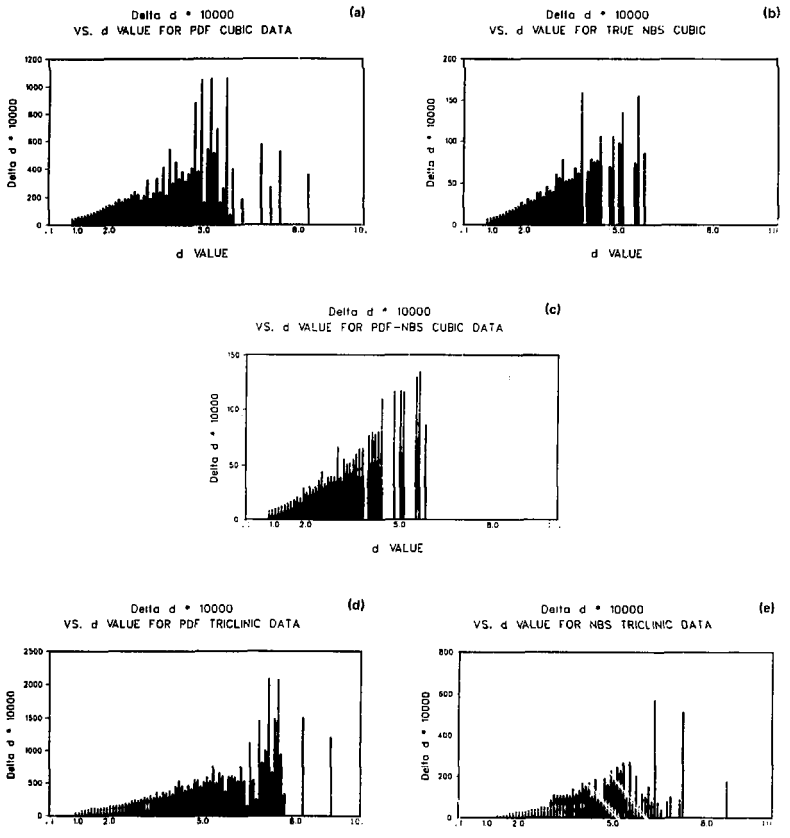

Fig. 43. | Δd | as a function of the value of d for cubic and triclinic patterns. Bars above the break represent standard deviations.

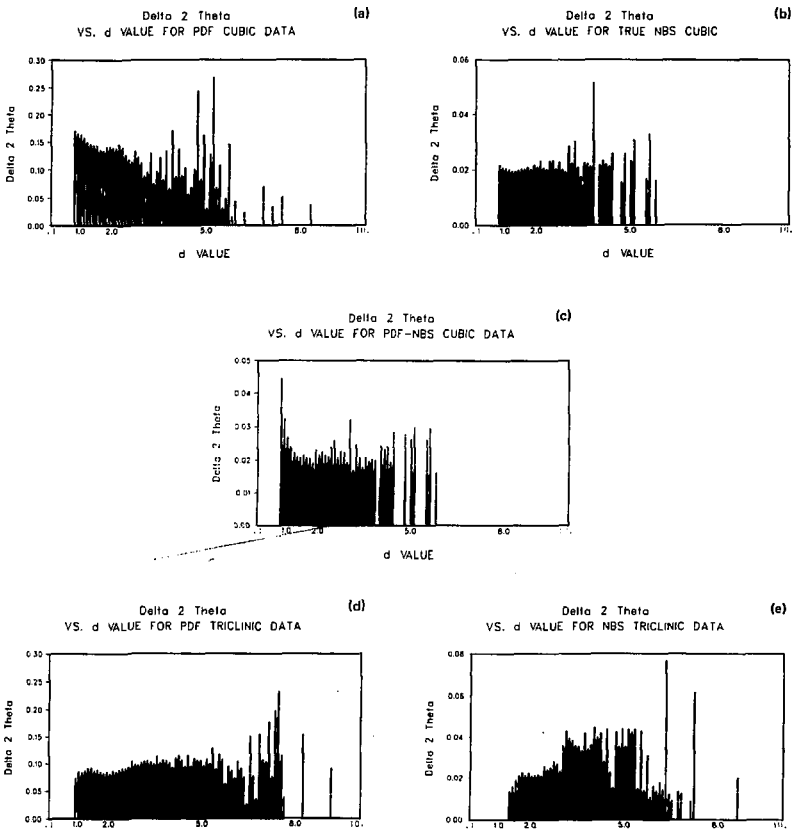

Fig. 44. $|\ \Delta 2\theta\ |$ as a function of the value of d for cubic and triclinic patterns. Bars above the break represent standard deviations.

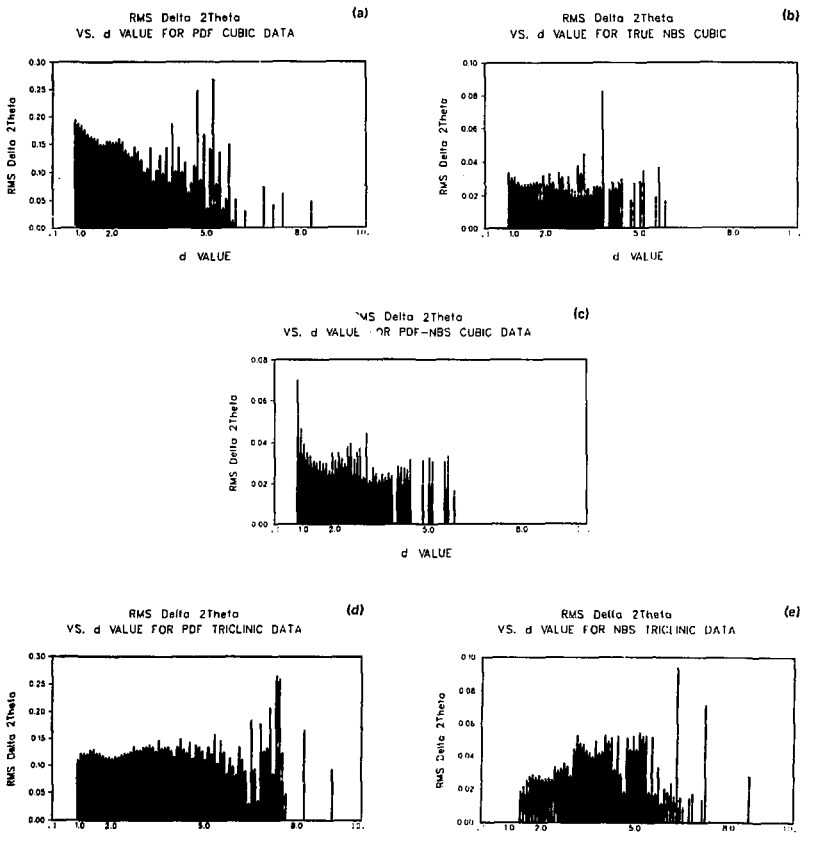

Fig. 45. Root mean square $|\Delta 2\theta|$ as a function of the value of d for cubic and triclinic patterns. Bars above the break represent standard deviations.

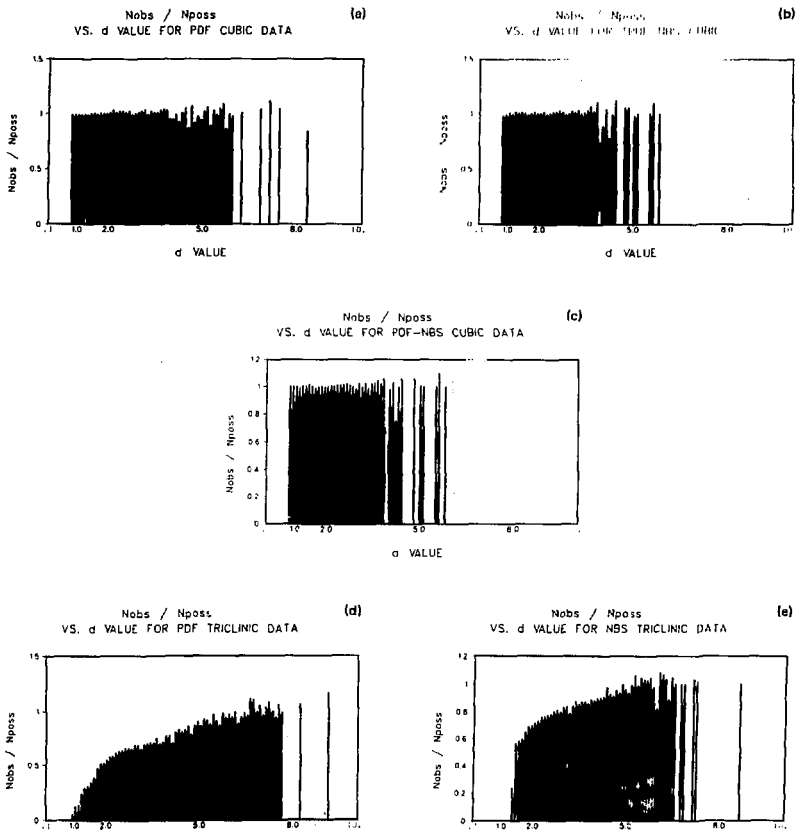

Fig. 46. Coverage factor N/N_{poss} as a function of the value of d for cubic and triclinic patterns. Bars above the break represent standard deviations.

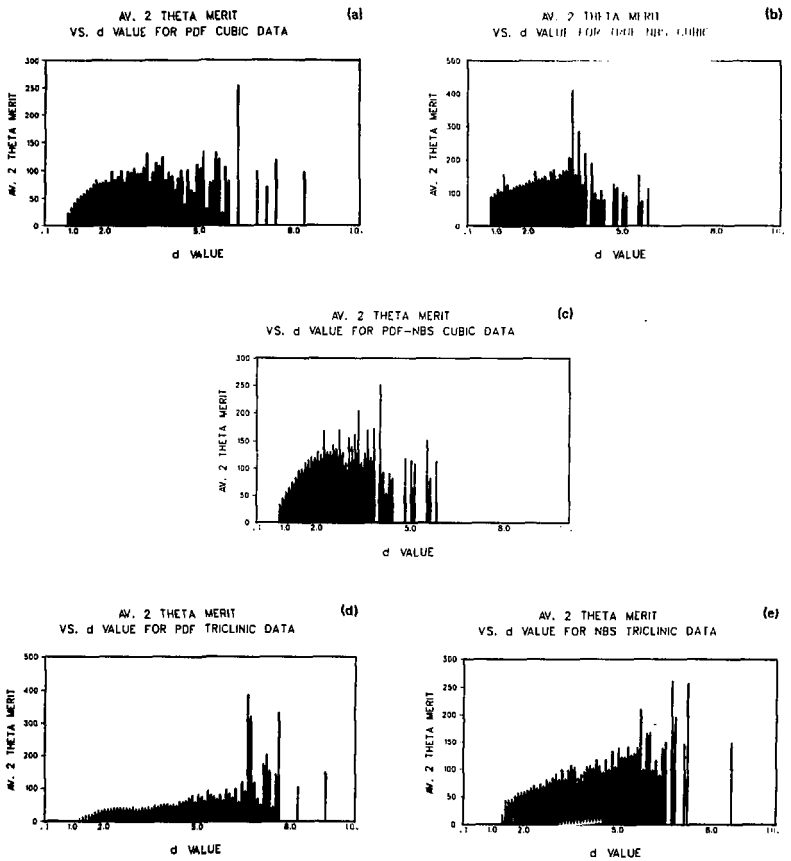

Fig. 47. F_N as a function of the value of d for cubic and triclinic patterns. Bars above the break represent standard deviations.

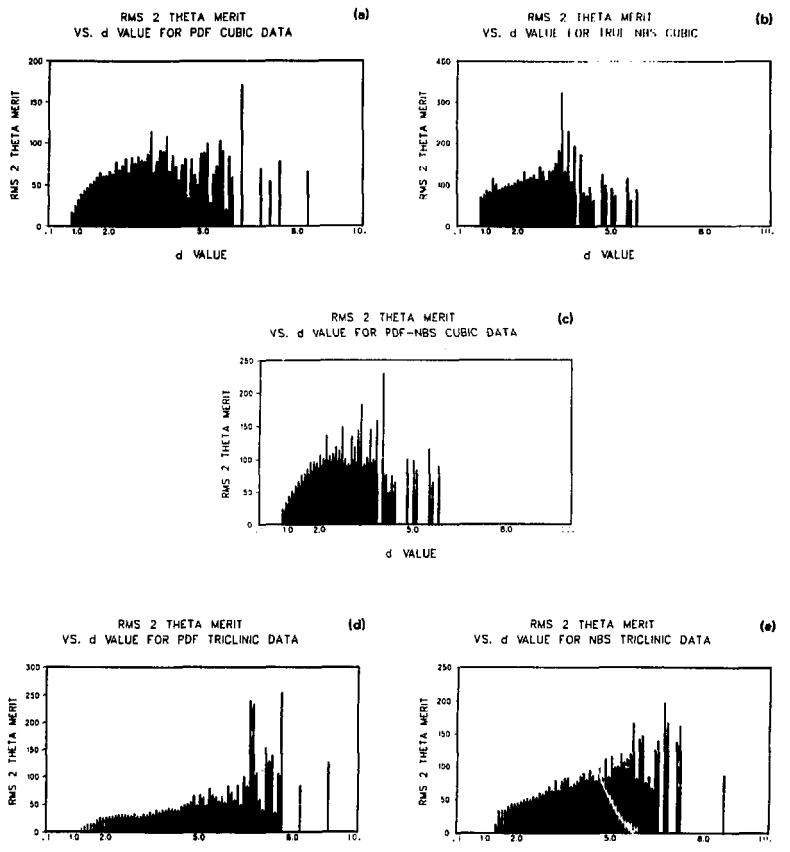

Fig. 48. Root mean square F_N as a function of the value of d for cubic and triclinic patterns. Bars above the break represent standard deviations.

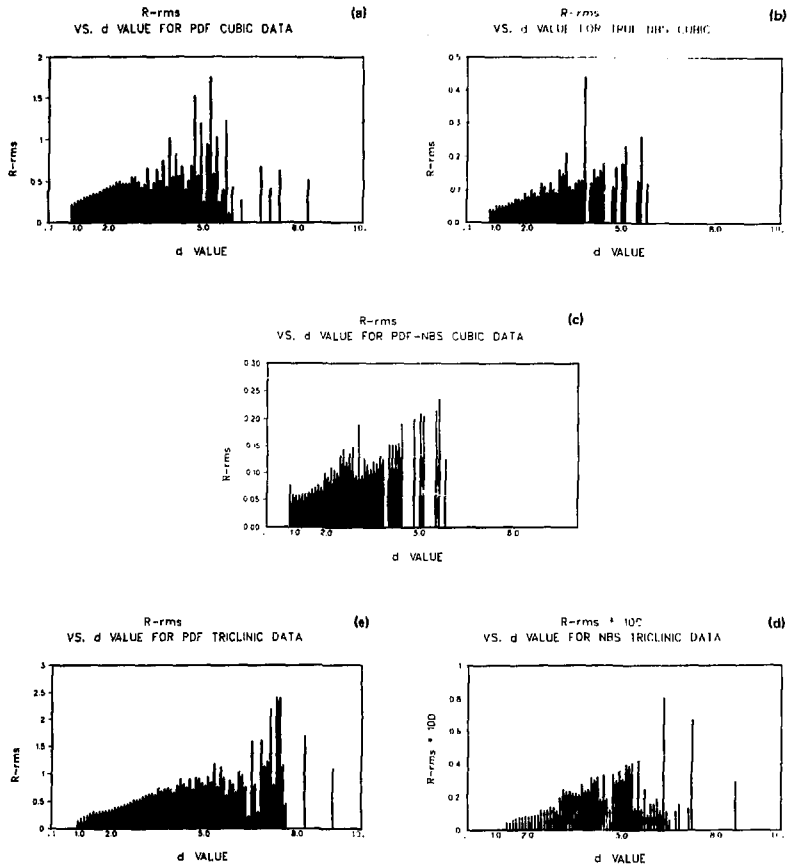

Fig. 49. R as a function of the value of d for cubic and triclinic patterns. Bars above the break represent standard deviations.

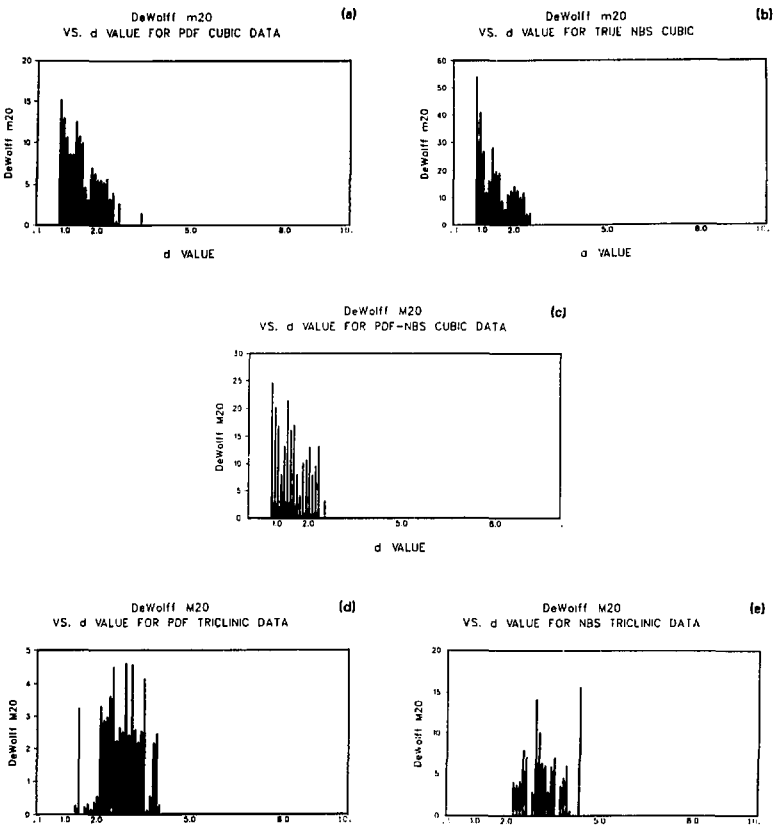

Fig. 50. M_{20} as a function of the value of d for cubic and triclinic patterns. Bars above the break represent standard deviations.

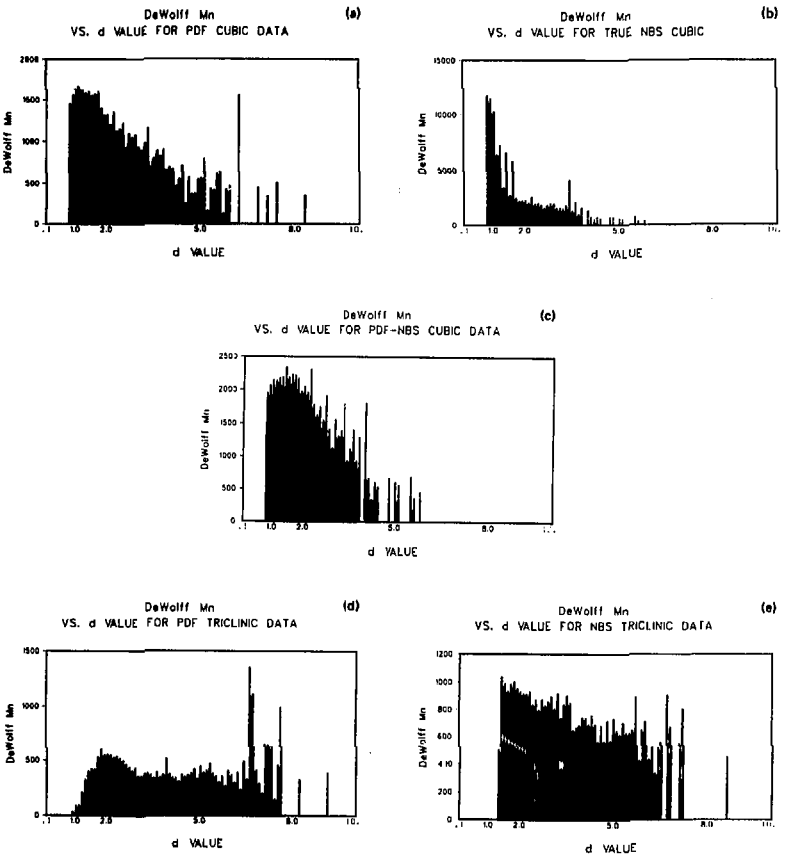

Fig. 51. M_N as a function of the value of d for cubic and triclinic patterns. Bars above the break represent standard deviations.

REFERENCES

1. JCPDS International Centre for Powder Diffraction Standards, 1601 Park Lane, Swarthmore, PA.
2. L. K. FREVEL, "Chemical Analysis by Powder diffraction," *Ind. and Eng. Chem. and Ed.* **16**, 209-218 (1944).
3. M. C. Nichols, *Graphical Representations of Selected Functions of Interplaner Spacings for 5184 Inorganic x-ray Powder Diffraction Patterns*, Lawrence Livermore Laboratory, Rept. UCRL-50115 (1966).
4. Peter Keller, *CHARTIT: A Computer Program That Constructs Bar Graphs in Color*, Lawrence Livermore Laboratory, Rept., UCRL-52310 (1977).
5. P. M. de Wolff, "A Simplified Criterion for the Reliability of Powder Pattern Indexing," *J. Appl. Cryst.* **1**, 108-113 (1968).
6. W. Nowacki, "Crystal Data Systematic Tables," *American Crystallographic Association Monograph* **6** (1967).
7. A. D. Mighell, "Symmetry and Reduced Cells" in *Proc. of American Crystallographic Association Meeting, Asilomar, Calif.*, **1977**, Rept. JN3.
8. "Standard X-Ray Diffraction Powder Patterns," United States Department of Commerce, National Bureau of Standards, Circular 539, Vols. 1-10 (1953-1960), Monographs 25, Sections 1-14, (1962-1977).
9. G. S. Smith and R. L. Snyder, "F_N: A Criterion for Rating Powder Diffraction Patterns and Evaluating the Reliability," of Powder Pattern Indexing *J. Appl. Cryst.*, to be published.

MLD