# Quality and consistency of digitized data

N.Otuka[1] and S. Dunaeva[2]

[1]*Nuclear Data Center, JAERI, 319-1195 Japan*
[2]*Nuclear Data Section, IAEA, A-1400 Vienna, Austria*

(6 October 2005)

## 1. Background

At present, compilers in several centres use different software programs for digitizing numerical data from figures in (mostly old) papers. We notice that these programs, which employ different digitizing methods, get the different numerical data from the same image. We think it is important to check the quality and consistency of the digitized data going to EXFOR. For this purpose, a small comparison exercise was proposed (Memo CP-D/415, 1 April, 2005).

## 2. Method

Comparison and checking have been done by the following steps:

1. Digitize fig.2 (=Fig.1 of this paper) from D. Bucurescu *et al*., Nucl. Phys. **A674**, 11, 2000.
2. Send these data in the form of a short entry to the NDS for comparison
3. Compare digitized data with authors' tabulated data
    3.1. Plot digitized data with authors' data in same panels
    3.2. Evaluate "accuracy" of digitization

## 3. Results and discussion

We received entries from 5 centres (Table I). Data were mostly received in April, 2005. CNPD and CAJaD included accuracy of digitization under data headings `ANG-ERR-D` and `ERR-DIG` fields. CDFE digitized author's error bars as asymmetric, but we adopt upper errors (which were digitized almost larger than lower errors) in the present analysis.

Table I. Summary of centres and programs involved in the present analysis

| Centre | Digitizer | Developer | Format of digitized data[*)] | | |
|---|---|---|---|---|---|
| | | | $x$ (deg) | $y$ (mb/sr) | $\Delta y$ (mb/sr) |
| JCPRG | GSYS | JCPRG | 3 digits | 3 digits | 3 digits |
| CNPD | Graf_new | CNPD | 3 decimal | 5 digits | (given in %) |
| NNDC | GSYS | JCPRG | 3 digits | 3 digits | 3 digits |
| CAJaD | CAJaD | CAJaD | 2 decimal | 3 digits | (given in %) |
| CDFE | CDFE | CDFE | 2 decimal | 3-5 decimal | 3-5 decimal |

[*)] Example of data format: "*1.23*E+45": 3 digits, "12.*345*": 3 decimal.

### 3.1 Brief comparison between digitized data and authors' data (Fig. 2)

In Fig. 2, we plot digitized data and authors' data in same panels. Authors' data are well reproduced by digitized data in general. Some large discrepancy can be seen in the y values of the following data sets:

- NNDC at 993 keV: Underestimate authors' data at all angles.
- CAJaD at 2405 keV: Overestimate authors' data at all angles.
- CNPD at 814 keV: Over- and underestimate authors' data at small and large angles.

These were probably caused by trivial mistakes in setting of y-axis.

## 3.2 Evaluation of digitization accuracy (Figs. 3-5)

We try to evaluate the digitization accuracy in a quantitative method by defining an accuracy testing parameter. It is the ratio of digitized data from curve (*C*) to authors' tabulated data (*T*) minus 1 "*C*/*T*-1", which takes zero if digitization is perfect.

In Figs. 3 to 5, the values of *C*/*T*-1 are shown for x, y and y-error component of digitized data, respectively. "1277", "1720"… give level energies of corresponding panels in original figure (c.f. Fig. 2). Additional curves given by CNPD and CAJaD show digitization error coded under `ANG-ERR-D` (for x-values) and `ERR-DIG` (for y-values). These would be regarded as "allowable deviation" in the present checking.

Fig. 3 (accuracy for x values):

This component is plotted in linear scale in authors' figure. Therefore accuracy of digitization is expected to be constant in absolute unit (deg). Consequently, relative deviation (per-cent) takes smaller value at larger x-values. Deviations of CNPD and CAJaD data from authors' data are well covered by digitization errors given under `ANG-ERR-D`. It is worth studying their digitizer accuracy estimation.

Fig. 4 (accuracy for y values):

This component is plotted in logarithmic scale in authors' figure. Therefore accuracy of digitization is expected to be constant in relative unit (%). Deviations of CAJaD are well covered by their digitization errors under `ERR-D` when it is given.

There are large discrepancies in NNDC data at 993 keV, CAJaD data at 2405 keV, and CNPD data at 814 keV. These are easily understood by the discussion of Fig. 2. JCPRG and NNDC used same digitizer system, but accuracy is different in two centres. This would be caused by various sources (e.g. size of display, size of image, etc.).

Fig.5 (accuracy for y-error values):

Error bars in the original figure are very small. JCPRG, CNPD, NNDC and CAJaD therefore skip digitization of error bars for many data points. It is very difficult to digitize such a small error bar depicted in logarithmic scale in good quality. CAJaD data of 2405 keV are involved in the trouble as mentioned in the discussion of Fig. 2.

## 4. Conclusion

This experience has shown that all programs approximately work with identical accuracy. Accuracy depends on the size of the figure and slightly from the algorithm. The main problem is so called "human factor". To reduce this we have to check manually a few points on their compatibility with that which we have on the figure and also to see graphically received digitized curve.

We appreciate submission of entries from NNDC, CAJaD and CDFE.

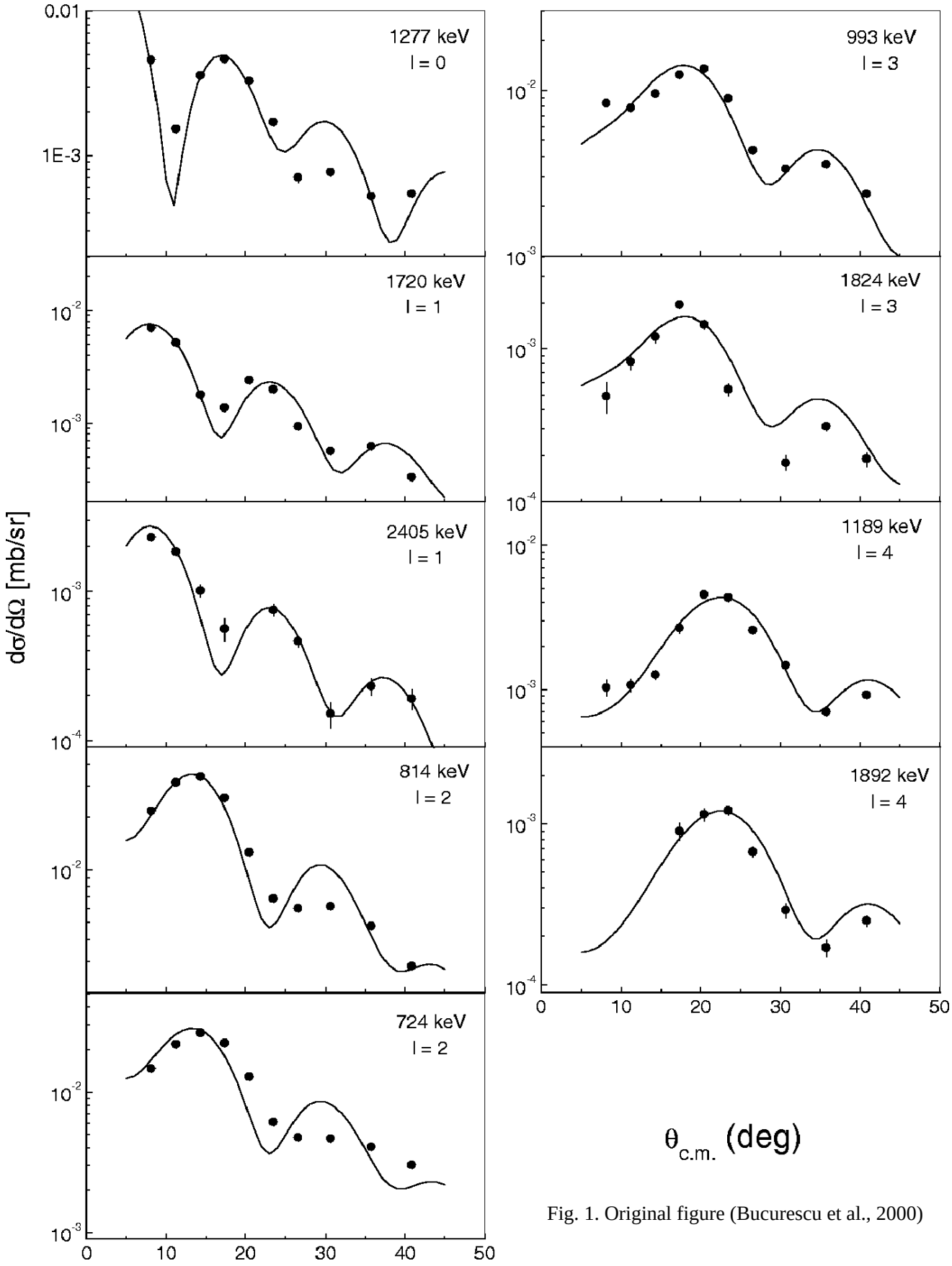


Fig. 1. Original figure (Bucurescu et al., 2000)

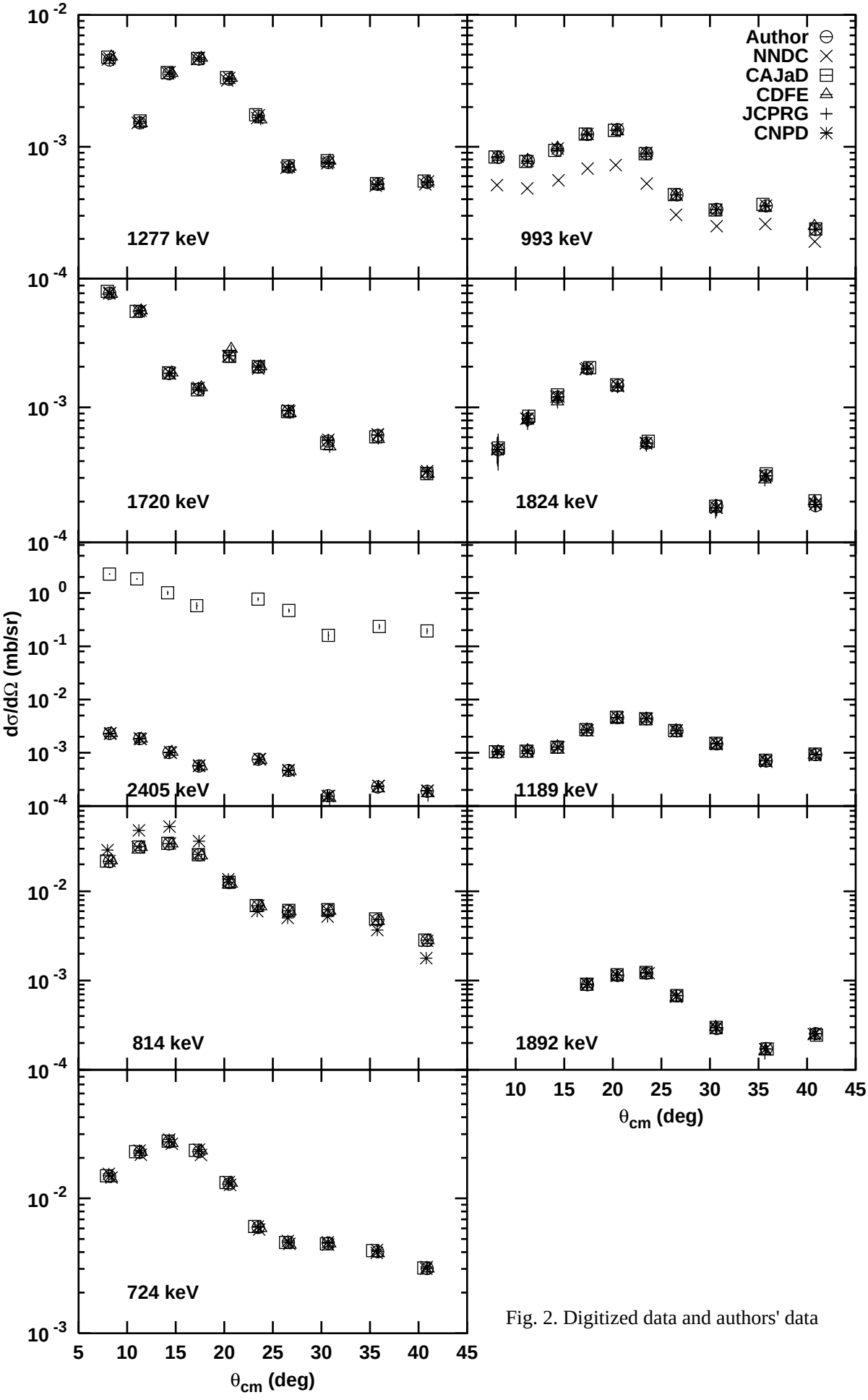


Fig. 2. Digitized data and authors' data

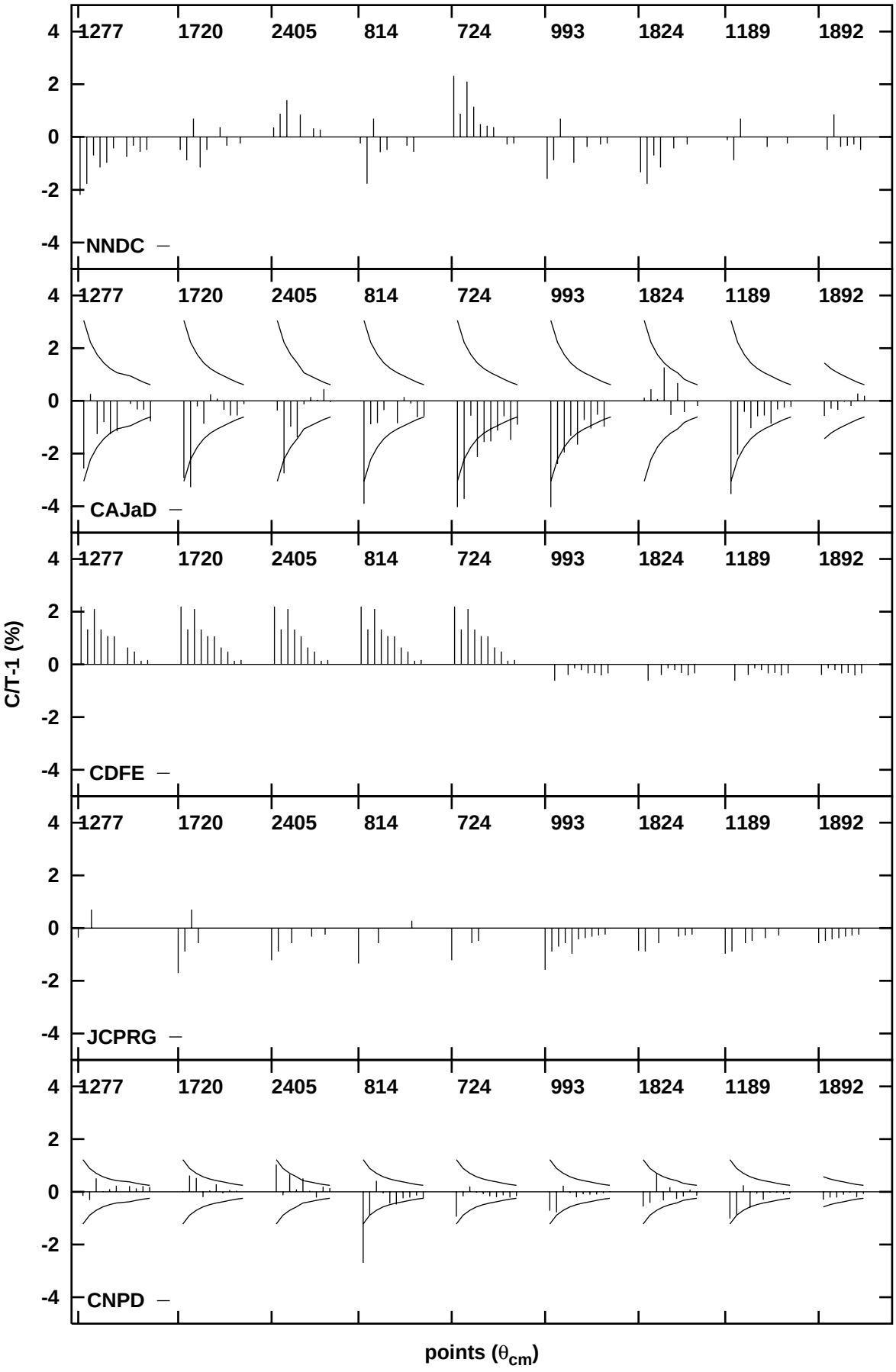


Fig. 3. Accuracy of digitization for x value at each data point

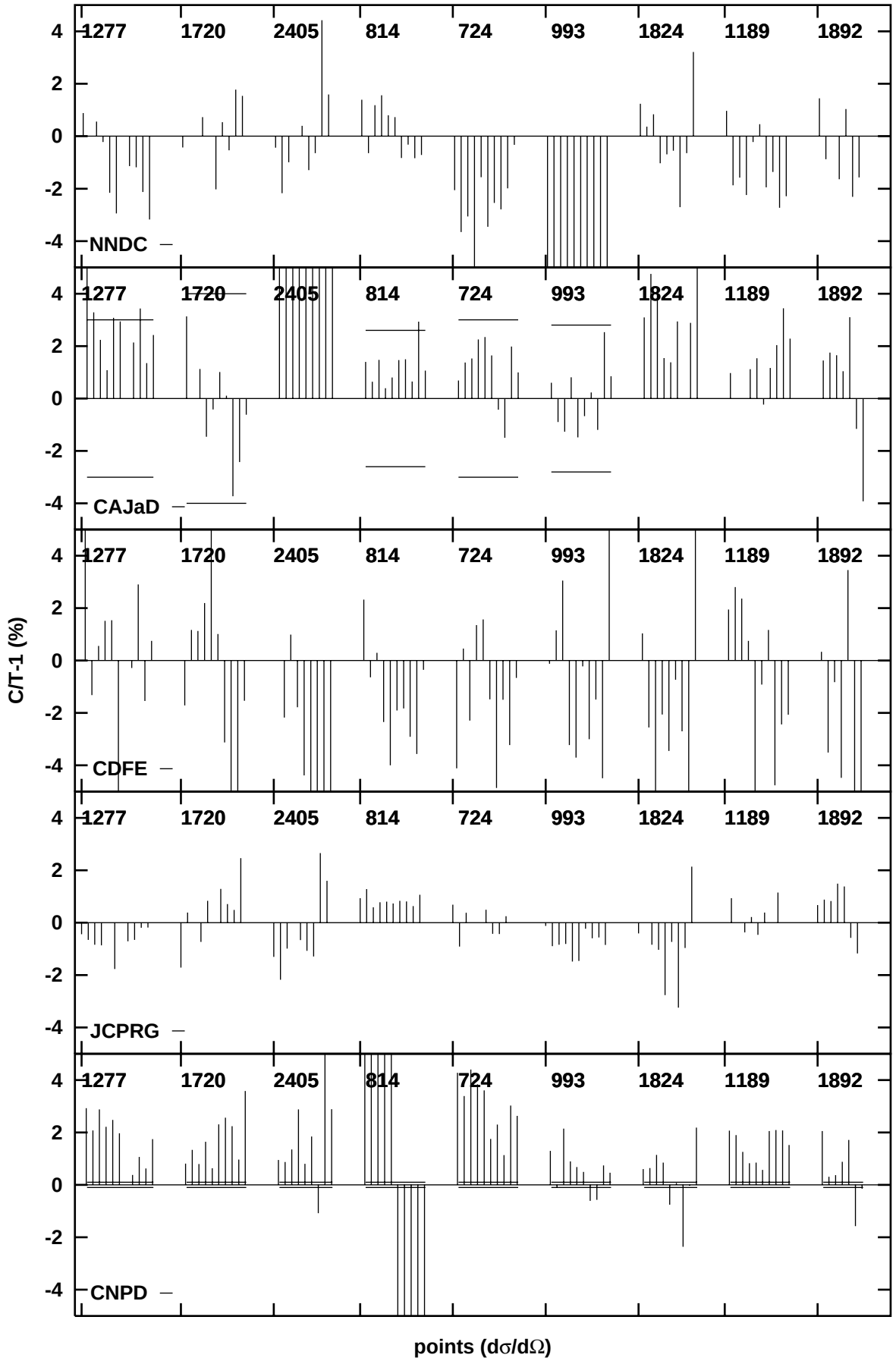


Fig. 4. Accuracy of digitization for y value at each point

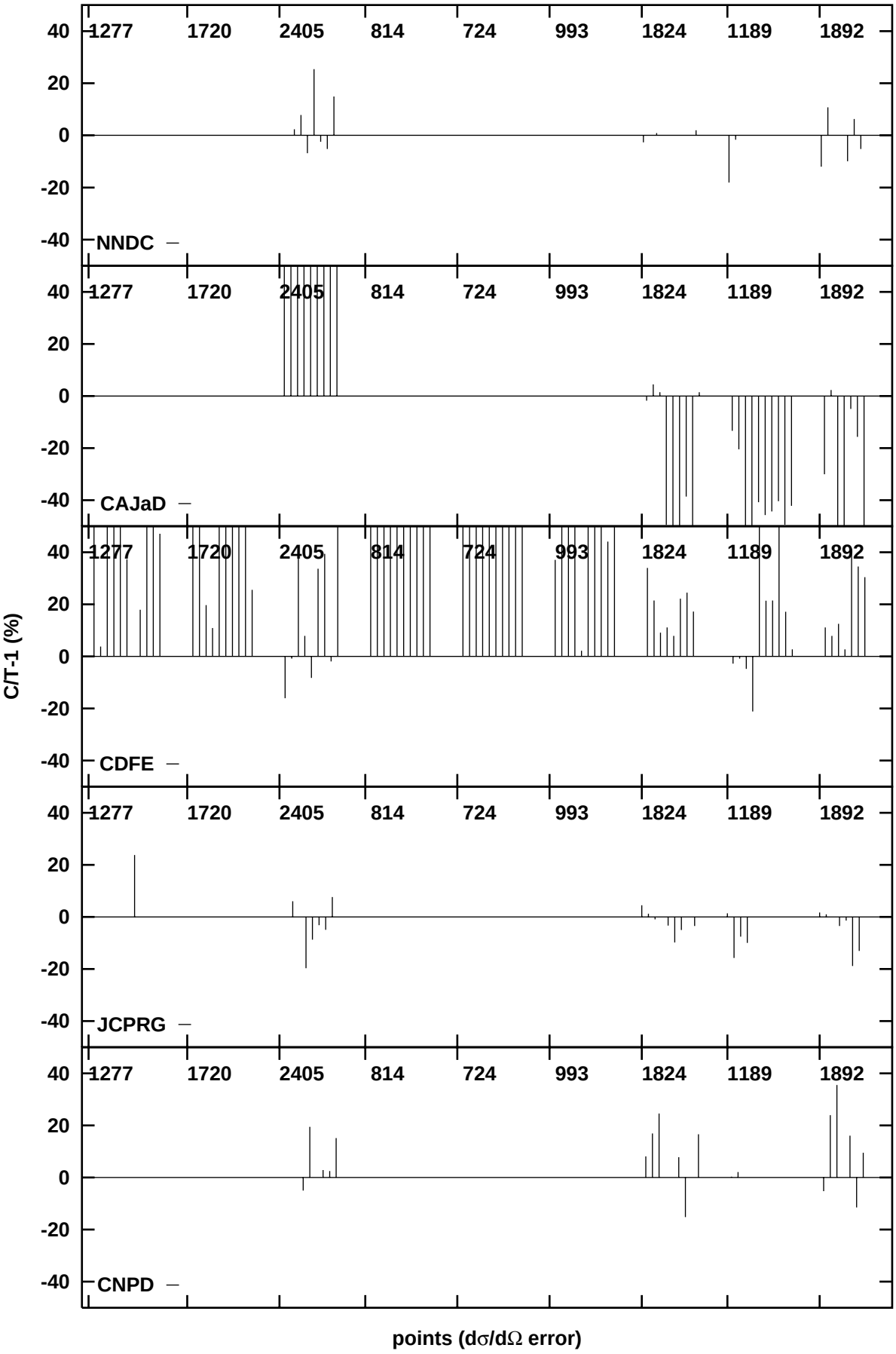


Fig. 5. Accuracy of digitization for y-error value at each data point