https://doi.org/10.61092/iaea.tdbx-syfr

IAEA
International Atomic Energy Agency

INDC(NDS)-0629
Distr. G+NC

INDC International Nuclear Data Committee

Summary Report of the Consultants' Meeting on

Benchmarking of Digitization Software

IAEA Headquarters, Vienna, Austria

12 – 14 November 2012

Prepared by

Naohiko Otsuka and Valentina Semkova
IAEA Nuclear Data Section, Vienna, Austria

January 2013

IAEA Nuclear Data Section, Vienna International Centre, A-1400 Vienna, Austria

Selected INDC documents may be downloaded in electronic form from
http://www-nds.iaea.org/publications/
or sent as an e-mail attachment.
Requests for hardcopy or e-mail transmittal should be directed to
nds.contact-point@iaea.org
or to:

Nuclear Data Section
International Atomic Energy Agency
Vienna International Centre
PO Box 100
A-1400 Vienna
Austria

Produced by the IAEA in Austria
January 2013

INDC (NDS)-0629
Distr. G+NC

Summary Report of the Consultants' Meeting on

Benchmarking of Digitization Software

IAEA Headquarters, Vienna, Austria

12 – 14 November 2012

Edited by

Naohiko Otsuka, Valentina Semkova
IAEA Nuclear Data Section, Vienna, Austria

January 2013

Abstract

This report summarizes the IAEA Consultant Meeting on Benchmarking of Digitization Software, held at IAEA Headquarters in Vienna, Austria from 12 to 14 November 2012. The meeting was attended by twelve participants from five Member States and IAEA. A summary of the meeting is given in this report along with the recommendations to the software developers, EXFOR compilers and NRDC coordinator. The papers presented by the participants are also appended.

TABLE OF CONTENTS

LIST OF ACRONYMS

ATOMKI	Nuclear Research Institute, Debrecen, Hungary
BARC	Bhabha Atomic Research Centre, Trombay, Mumbai, India
CAJAD	Centre for Nuclear Structure and Reaction Data, Kurchatov Institute, Moscow, Russia
CDFE	Centr Dannykh Fotojad. Eksp., Moscow State University, Russia
CHEX	EXFOR check program (originating from NNDC)
CIAE	Chinese Institute of Atomic Energy, Beijing, China
CJD	Russian Nuclear Data Centre, IPPE, Obninsk, Russia
CNDC	China Nuclear Data Centre, CIAE, Beijing, China
CNPD	Centre of Nuclear Physics Data, RFNC-VNIIEF, Sarov, Russia
CP...	Numbering code for memos exchanged within the NRDC
EXFOR	Format for the international exchange of nuclear reaction data
GDgraph	Digitizing software developed by CNDC
GSYS	Digitizing software developed by JCPRG
IAEA	International Atomic Energy Agency, Vienna, Austria
InpGraph	Digitizing software developed by CNPD
IPPE	Institute of Physics and Power Engineering, Obninsk, Russia
JCPRG	Japan Nuclear Reaction Data Centre, Sapporo, Japan
NDPCI	Nuclear Data Physics Centre of India
NDS	IAEA Nuclear Data Section, Vienna, Austria
NNDC	National Nuclear Data Center, Brookhaven National Laboratory, USA
NRDC	International Network of Nuclear Reaction Data Centres
NRDF	Japanese Nuclear Reaction Data File
ORDER	EXFOR program for addition of record identification and bookkeeping information
VNIIEF	Russian Federal Nuclear Centre, Sarov, Russia
ZCHEX	Current version of CHEX, updated and maintained by NDS

MEETING SUMMARY

1. INTRODUCTION

The IAEA Consultants' Meeting on "Benchmark of Digitization Software" was held at IAEA Headquarters in Vienna, Austria from 12 to 14 November 2012. The meeting was attended by twelve participants (Appendix A) from five Member States (China, India, Japan, Russia and USA) and IAEA.

About 40% of all entries included in the EXFOR library during the last decade contain digitized data, and it is important to ensure the correctness of these data. Since the last benchmark organized by the IAEA Nuclear Data Section (NDS) in 2005 (See Appendix E), there have been new developments in the digitization software used by EXFOR compilers. The purpose of the meeting was to discuss the quality and consistency of data obtained by the digitization software developed in three countries (China, Japan and Russia) between EXFOR compilers and software developers.

Totally eleven papers were presented and discussed at the meeting (see Appendix D). Some topics such as automatic digitization and digitized value rounding were intensively discussed. The outcomes from the meeting were summarized in eleven recommendations to software developers, fifteen recommendations to EXFOR compilers, and four recommendations to the NRDC coordinator.

2. BRIEF SUMMARY

2.1 Opening

R. Forrest, Head of the IAEA Nuclear Data Section, welcomed the participants of the meeting on behalf of the IAEA. The participants introduced themselves. B. Lalremruata (Mizoram Univ., India) was elected as Chairman, and N. Otsuka (NDS) was elected as the Rapporteur. The agenda was adopted without change (see Appendix B).

V. Semkova (Scientific Secretary, NDS) outlined the objectives of the meeting. She stressed importance of this meeting by reminding participants of the fact that about 40% of EXFOR entries compiled in the last 10 years contain digitized data, and explained the following four items to be discussed in the meeting: 1) current status of digitization codes; 2) quality assurance of digitized data; 3) recording of digitized data in EXFOR entries, and 4) digitization uncertainties.

2.2 Presentations by Software Developers

G. Pikulina (RFNC-VNIIEF, Russia) presented the Russian EXFOR Digitizer (InpGraph) Ver. 2.2. She explained two directions in further developments of the digitizing software: 1) implementation of important feedback; 2) major redesigning of the current user interface. In the new version, SRC (service) files can be checked by "Processing" function of the Russian EXFOR editor, and the SRC files can be edited on a special window when necessary. Examples of checking and editing were demonstrated.

S. Taova (RFNC-VNIIEF, Russia) reported the development of EXFOR Digitizer Wizard Ver. 3.0. This Wizard adopts "Parameter gathering technique". This technique asks users for a minimum amount of input while the rest is determined by the program itself. With this Wizard, manual input of coded words interpreted by the software is no longer necessary. Addition of headings and unit codes, curve information (*e.g.*, figure number) setting and scale setting on the Wizard were demonstrated.

R. Suzuki (Hokkaido Univ., Japan) introduced the Japanese digitization software GSYS. Following an introduction and demonstration of basic functions (*e.g.*, "drag and drop", "snapshot", "feedback function", "glass", "loupe"), he presented details about three subjects: "auto-point recognition", "feedback function", and "orthogonality of x- and y-axis". The algorithm of auto-point recognition adopted in GSYS is based on the template matching technique which is used for the real-time tumor-tracking radiotherapy developed by the Hokkaido University Hospital for automatic tracking of the surrogate gold markers implanted in/near a tumor. From his experience in medical physics, he also strongly recommended checking of digitization results with a check list.

Chen Guochang (CIAE, China) presented the Chinese digitization software GDgraph. There were requirements of digitization from evaluators and experimenters around the Chinese Nuclear Data Center (CNDC), and the Centre has developed GDgraph since 1997. Version 1.0 released in 2001 was written in Visual C++. It accepted only BMP image files as input and users could not digitize error-bars. These previous inconveniences are resolved in the current version (Ver. 4.4) written in Perl. A new function ("Project and Remarks") allows users to treat a set of files (*i.e.*, input image, output numerical data, various setting parameters, free text remark) as one package. He explained that CNDC has recently identified many old articles published in Chinese journals and missing in the EXFOR library, and digitization becomes more important for their EXFOR compilation.

G. Pikulina reported testing of two generic digitization tools (Graph2Digits, DAGRA) to review requirements for automatic or semi-automatic digitization. She concluded that 1) filtration for noises, 2) rotation of image, and 3) elimination of axis lines and captions are necessary for successful automatic digitization.

S. Taova also introduced an existing digitization tool (GETDATA Graph Digitizer) to study automatic or semi-automatic digitization. In this tool, users can choose two algorithms of automatic digitizing ("auto trace lines" and "digitize area" as well as manual digitizing. Two conditions for automatic digitizing were given: 1) curves are well separated from each other; 2) quality of the plot is high enough.

V. Zerkin (NDS) stressed that further automation of operations (*e.g.*, auto-focusing of symbols and scales) is the most promising direction, and such automation can be made more effective by editing (filling) defects on symbols and error-bars as well as calibration of scales by introducing an appropriate fitting function. He also emphasized that independent plotting of digitized values may find mistakes which cannot be found within a simple digitization procedure, and demonstrated some examples of such advanced checking with his ZVView software as an additional option.

2.3 Presentations by EXFOR Compilers

B. Lalremuruata (Mizoram Univ., India) introduced the current EXFOR compilation activity by NDPCI in India, and reported comments from three Indian EXFOR compilers (S. Badwar, R. Ghosh and R. Mandal). R. Ghosh regularly meets "I/O error 32" and "I/O error 103" as well as the "does not exist" error message when she uses the Russian digitizing software even if both AXS and SRC files are properly located. She also meets "java.lang.ArrayIndexOutOf Bounds Exception" when she uses GSYS, but she can continue digitization if she does not click "OK" at this stage. S. Badwar meets an extra data line at the end of outputs from the Russian digitizer. R. Mandal wishes that the Russian digitizer reports error messages at an earlier stage so that she does not need to repeat the whole digitization procedure again. She also proposes implementation of digitization error evaluation in GSYS.

M. Mikhaylyukova (IPPE, Russia) started her presentation from technical questions, comments and proposals for InpGraph and GSYS. Then she reported her own benchmark problem where $^{16}O(n,\alpha)^{13}C$ cross sections measured by Vitaly Khryachkov et al. (IPPE) and digitized by InpGraph and GSYS from figures (linear-linear) published in three conference, and compared digitized values with the original data from the author. She reported that the digitization errors estimated by InpGraph were smaller than the mean absolute deviations for both x-values (neutron energies) and y-values (cross sections). She also compared the results for the 2011 paper ("bad" quality plot) between InpGraph and GSYS, and found room for improvements in both digitizers.

N. Otsuka (NDS) discussed the expression of digitized values in EXFOR entries. He started his presentation from the expression of values and uncertainties used in scientific literature, and proposed to keep the same number of digits for values (*e.g.*, `DATA`), experimental uncertainties (*e.g.*, `DATA-ERR`), and digitization uncertainties (*e.g.*, `ERR-DIG`) in EXFOR entries by rounding digitized values to the same digits in the fixed decimal number expression. He also introduced a model of constant digitization uncertainty assuming that the uncertainty is corresponding to a fixed-size rectangle on the figure image, and concluded that 1) absolute (*e.g.*, mb) and relative (*e.g.*, %) expressions are appropriate for errors of digitization from linear and logarithmic scale, respectively; 2) the fixed decimal point number expression (*e.g.*, `12.345`) and floating decimal point number expression (*e.g.*, `1.2345E+02`) are suitable for values digitized from linear and logarithmic scale, respectively. Finally he proposed a guideline for the expression of digitized values in EXFOR entries for adoption in the meeting.

2.4 Benchmark Results

V. Semkova reported results from the Benchmark on Digitization Software. In the beginning she presented the objectives of the benchmark and the procedure that was applied. Different types of graphs were distributed by Memo CP-D/761 in order to test the quality and consistency of the digitization procedure. Digitized data in a form of EXFOR Entries were received from 14 participants. Plots with submitted digitized data and authors' data for each of the distributed graphs were presented. It was concluded that there is an overall consistency of the digitized data, but also systematic deviations in some datasets were present. The accuracy of the digitized data were also evaluated by the parameter (C/T)-1, where C (Curve) is the digitized value, T (Table) is the value from authors. Graphs with calculated (C/T)-1 value for each of the files submitted to the Benchmark were shown. The parameter (C/T)-1 was calculated for X and Y coordinates of each point as well as for the error bars ΔX and ΔY. It was shown that an accuracy below 1% for the digitized X and Y values was achieved in some digitized datasets. Such agreement was found in datasets digitized by different types of software, therefore the programs used in the Benchmark work with comparable accuracy. However the so called "human factor" also plays a role for the final results since a spread of the C/T-1 values for different datasets digitized with the same software exists. For the error bars ΔX and ΔY the calculated C/T-1 values were found to be higher than those for the X and Y values. The formats of the numerical data chosen by the compilers were compared with the format of the numerical data received from the authors. It was emphasized that if X and ΔX

values are given with less precision than the author's data this may cause repetition of the same values of the independent variable despite the fact that the points are not overlapping. Such a mistake may also lead to significant deviation of the digitized data from the real point in case of steep changes of the value of the variable (*e.g.*, energy in resonance structure). V. Semkova concluded that good knowledge of the software, accumulation of experience, and attentiveness are recommended in order to ensure precise digitization results. Further software development and implementation of the users' feedback will provide better performance.

2.5 Discussion

Recommendations to software developers, EXFOR compilers and the NRDC coordinator were extensively discussed based on presentations, and in total thirty recommendations were agreed among the participants. One recommendation proposed by Otsuka (rounding of the digitized values according to the estimated digitization uncertainty) led to intense discussion. Rounding by keeping only one significant digit in the digitization error may result in repetition of an independent variable (*e.g.*, incident energy in excitation function showing resonance structure) even if two data points can be distinguished by eyes on the figure image, and some participants proposed to recommend rounding to two or three significant digits in the digitization error. This follows the general (established) rules for the number of digits necessary to express the value of uncertainty, and also help to avoid coincidence for digitised values of argument or function. However, some other participants argued to keep all digits created by the digitization software. The participants could not build any consensus, and finally concluded not to include the issue into the list of recommendations.

2.6 Closing

V. Semkova thanked the Chairman and all participants for very productive contributions and discussions that resulted in a very successful meeting. She also thanked some EXFOR compilers (T. Ashizawa, S. Badwar, R. Ghosh, A. Makinaga, R. Mandal, S. Takàcs) who could not attend the meeting, but provided their digitization results for discussion in the meeting. She then closed the meeting.

RECOMMENDATIONS

To Software Developers

R1 (To GDGraph) Make an English version, and make it available for EXFOR compilers.

R2 (To GSYS) Implement a function to estimate accuracy of digitization.

R3 (To InpGraph) Resolve too low relative (%) digitization error evaluated by the code.

R4 Share useful features among developers.

R5 Consider feedback provided by the meeting participants and compilers.

R6 Indicate contact details of the developer in the software (*e.g.*, “help”, “about”).

R7 Implement automatic focusing procedure for symbols, axis and its legend (if it has not been implemented yet).

R8 Implement feedback function so that compilers can reproduce their digitized values on the images (if it has not been implemented yet).

R9 Introduce calibration procedure to digitizing programs (if it has not been implemented yet).

R10 Show digitized scales on the screen independently at the checking stage (if it has not been implemented yet).

R11 Display a warning message when repetition exists in digitized values of an independent variable.

To EXFOR Compilers

R12 Follow “Guideline for expression of digitized data in EXFOR” (Appendix C of this report) for digitized data included into EXFOR entries.

R13 Provide feedback to software developers for further improvement.

R14 Ask for numerical data from the authors. Explain to the authors that the original numerical data from the authors are preferable than values digitized from images.

R15 Do not change the number of digits for values provided by the authors without their permission.

R16 Avoid adoption of digitized values when the original values are explicitly given in the article (e.g., incident energy, detection angle).

R17 Consult with NDS if digitization and compilation are really necessary when the image quality is poor, for cases where there have been many experimental data sets in EXFOR.

R18 Ask for help of other centres when the compiler cannot prepare an image with sufficient quality (*e.g.*, due to restriction to electronic access).

R19 Enlarge the image as much as possible when the image file is created and digitized. Utilise functions available on the software for this purpose (e.g., “magnifying glass”).

R20 Digitize the beginning and end points of scales with special care to avoid systematic errors.

R21 Check the values of the beginning and end points (labels on scales) after digitization of the image.

R22 Investigate the reason of the repetition of a digitized value of an independent variable to avoid human errors such as digitization of the same symbol twice.

R23 Keep the field for the uncertainty blank when the error-bar is not clearly visible, and explain the situation in free text (with `FLAG` when appropriate).

R24 Keep text information associated to numerical data (*e.g.*, range of cross sections) in the EXFOR entry as free text when digitized values are adopted in the entry and such information is seen in the article.

R25 Use InpGraph when deformation of the image is serious.

R26 Correct defects on objects to be digitized (*e.g.*, by filling symbols or lines) when auto-focusing is applied to a poor image, in order to make auto-focusing more stable.

To NRDC Coordinator

R27 Add the contact details of software developers to the NRDC software web page.

R28 Introduce a set of benchmark problems for new EXFOR compilers.

R29 Introduce a set of benchmark images (*e.g.*, png) for software developers.

R30 Collect feedback on digitizing software, and list it on the NRDC web page.

LIST OF PARTICIPANTS

CHINA

Mr Guochang **Chen**
China Nuclear Data Centre
China Institute of Atomic Energy
P.O. Box 275 – 41
BEIJING 102413
Tel: +86 10 693 58604
Fax: +86 10 693 57008
E-mail: cgc@ciae.ac.cn

INDIA

Mr Lalremruata **Bawitlung**
Department of Physics
Mizoram University
AIZAWL 796004
Tel: +91 3892305572
Fax: +
E-mail: marema08@gmail.com

JAPAN

Mr Ryusuke **Suzuki**
Department of Medical Physics
Hokkaido University Hospital
Kita 14-jo Nishi 5-chome
SAPPORO 060-8648
Tel: +81 11 706 7638
Fax: +81 11 706 7639
E-mail: ryusuke@med.hokudai.ac.jp

RUSSIAN FEDERATION

Ms Marina **Mikhaylyukova**
Institute for Physics and Power Engineering
Bondarenko Sq. 1
OBNINSK 249033
Kaluga Region
Tel: +7 484 399 8779
Fax: +7 084 396 8225
E-mail: mmarina@ippe.ru

RUSSIAN FEDERATION

Ms Galina **Pikulina**
Russia Federal Nuclear Centre
All Russia Scientific Research
Institute of Experimental Physics
Mira 37
SAROV 607188
Nizhnii Novgorod Region
Tel: +7 831 302 7779
Fax: +7 831 304 5569
E-mail: pikulina@expd.vniief.ru

RUSSIAN FEDERATION

Ms Sophiya **Taova**
Russia Federal Nuclear Centre
All Russia Scientific Research
Institute of Experimental Physics
Mira 37
SAROV 607188
Nizhnii Novgorod Region
Tel: +7 831 302 7779
Fax: +7 831 304 5569
E-mail: taova@expd.vniief.ru

OBSERVER – USA

Mr Boris **Pritychenko**
National Nuclear Data Center
Brookhaven National Laboratory
Bldg. 197D, PO Box 5000
Upton NY 11973-5000
UNITED STATES OF AMERICA
Tel: + 1 631 344 5091
Fax:+ 1 631 344 2806
E-mail: pritychenko@bnl.gov

IAEA

Mr Robin A. **Forrest**
Head
Nuclear Data Section
Division of Physical and Chemical Sciences
Tel.: +43 1 2600 21709
Fax: +43 1 2600 7
E-mail: r.forrest@iaea.org

IAEA

Mr Stanislav **Simakov**
Head
Nuclear Data Service Unit
Nuclear Data Section
Division of Physical and Chemical Sciences
Tel.: +43 1 2600 21717
Fax: +43 1 2600 7 21717
E-mail: s.simakov@iaea.org

IAEA

Ms Valentina **Semkova** (Scientific Secretary)
Nuclear Data Service Unit
Nuclear Data Section
Division of Physical and Chemical Sciences
Tel.: +43 1 2600 21727
Fax: +43 1 2600 7
E-mail: v.semkova@iaea.org

IAEA

Mr Naohiko **Otsuka**
Nuclear Data Service Unit
Nuclear Data Section
Division of Physical and Chemical Sciences
Tel.: +43 1 2600 217175
Fax: +43 1 2600 7
E-mail: n.otsuka@iaea.org

IAEA

Mr Viktor **Zerkin**
Nuclear Data Service Unit
Nuclear Data Section
Division of Physical and Chemical Sciences
Tel.: +43 1 2600 21714
Fax: +43 1 2600 7 21714
E-mail: v.zerkin@iaea.org

Appendix B

AGENDA

Monday, 12 November 2012

09:30 - 10:30	**Opening Session**
	Welcome address **(Robin Forrest)**
	Self-introduction of Participants
	Selection of Chairperson and Rapporteur
	Approval of Agenda
	Administrative Announcements **(Alexander Oechs)**
	Objectives of the Meeting (**Valentina Semkova)**
10:30 - 10:50	Coffee break

Presentations

10:50 - 11:25	**Galina Pikulina, CNPD** “Development of EXFOR Digitizer on the basis of user’s feedback”
11:25 - 12:00	**Sophiya Taova, CNPD** “Development of EXFOR Digitizer Wizard - New interface”
12:00 - 13:30	Lunch break
13:30 - 14:30	**Ryusuke Suzuki, Hokkaido University** “Introduction to digitization software GSYS” “Design and implementation of digitization software GSYS”
14:30 - 15:30	**Guochang Chen, CNDC** “Introduction of the digitizing software GDgraph”
15:30 - 16:00	Coffee break
16:00 - 16:30	**Galina Pikulina, CNPD** “Proposals on the development of the software on data digitizing for data entering into the EXFOR library”
16:30 - 17:30	**Sophiya Taova, CNPD** “GetData digitizing program code: description, testing, training”

Tuesday, 13 November 2012

09:00 - 09:45	**Lalremruata Bawitlung, Mizoram Univ.** "An overview of Indian EXFOR compilation (Feedback and suggestions)"

09:45 - 10:30	**Marina Mikhaylyukova, CJD** "Digitizing process: comments, real error estimation, recommendations".
10:30 - 10:50	Coffee break
10:50 - 11:30	**Naohiko Otsuka, IAEA** "Expression of digitized values in EXFOR entries based on digitization accuracy"
11:30 - 12:30	**Valentina Semkova, IAEA** "Reporting the results from the digitization software benchmark exercise"
12:30 - 14:00	Lunch break
14:00 - 14:30	**Viktor Zerkin, IAEA** "Some requirements for digitizing software and using advanced plotting for checking results"
14:30 - 15:30	Discussions and drafting of Conclusions & Recommendations
15:30 - 15:50	Coffee break
15:50 - 17:30	Discussions and drafting of Conclusions & Recommendations
19:00 -	**Social event: Visit to Pürstner Gaststube** (http://www.puerstner.com/)

Wednesday, 14 November 2012

09:00 - 11:00	Discussions and Drafting of Conclusions & Recommendations
11:00 -	**Final Remarks and End of the Meeting**

Appendix C

GUIDELINE FOR EXPRESSION OF DIGITIZED DATA IN EXFOR

1. Keep consistency for the number of digits between the digitized values and uncertainties.

   Example:
```
DATA
EN         DATA       DATA-ERR
MEV        MB         MB
 14.1       12.34      2.34
 14.3       12.3       2.3
 14.5        1.234E+01  0.234E+01
 14.6        1.23 E+01  0.23 E+01
…
ENDDATA
```

2. Use the fixed and floating decimal point expression for the numbers digitized from linear and logarithmic scale, respectively.

   Example:

   `12.345` (a value digitized from linear scale)

   `1.2345E+02` (a value digitized from logarithmic scale)

3. Digitization accuracy may be given in the absolute unit (e.g., `ADEG`) or relative unit (*e.g.*, `PER-CENT`) for the numbers digitized from linear and logarithmic scale, respectively.

   Example:
```
COMMON
ANG-ERR-D  ERR-DIG
ADEG       PER-CENT
 0.12       1.2
ENDCOMMON
DATA
ANG-CM     DATA       DATA-ERR
ADEG       MB         MB
  5.67       3.456E+02  0.234E+02
 12.31       2.345E+02  0.123E+02
…
ENDDATA
```

4. Consider rounding of digitized values to integers if values are for atomic numbers, mass numbers etc., and digitized values are close to integers.

Appendix D

PAPERS PRESENTED IN THE MEETING

TITLE	Presented by
Development of EXFOR Digitizer on the basis of user's feedbacks	G. Pikulina
Development of EXFOR Digitizer Wizard. New interface	S. Taova
Introduction, design and implementation of digitization software GSYS	R. Suzuki
Introduction of the digitizing software GDgraph	Chen Guochang
Proposals on the development of software on data digitizing for data entering into the EXFOR library	G. Pikulina
GetData digitizing program code: description, testing, training	S. Taova
An overview of Indian EXFOR compilation (feedback and suggestions)	B. Lalremruata
Digitizing process: comments, real error estimation, recommendations	M. Mikhaylyukova
Expression of digitized values in EXFOR entries	N. Otsuka
Benchmarking of digitization software (Memo CP-D/761): Results and discussion	V. Semkova
Some requirements for digitizing software and using advanced plotting for checking results	V. Zerkin

Note: These presentations are available online: http://www-nds.iaea.org/digitization/ .

Development of EXFOR Digitizer on the Basis of User's Feedback

G. Pikulina
Russian Federal Nuclear Center – VNIIEF, Sarov 607188, Russia

We have got a lot of user feedbacks for our version of the EXFOR Digitizer during the last two years.

So we decided to develop our Digitizer in two directions. The first one is implementation of the most important feedbacks and issue of the next Digitizer version. The second way is the redesigning of the user interface for the Digitizer with quiet new possibilities.

The main difficulties of our program are connected with error searching in the service files with digitized information. So we have started to develop some sort of plug-in for checking and editing the service files with AXS and SRC extension.

There was a problem of running compiling program Graph_New from the EXFOR-Digitizer interface. Sometimes the user come across an error message 'The file does not exist'. Three required folders should be contained in the main directory where this program is installed:

- BMP – for source image files
- SRC – for service files
- EXE – for service program Graph_New.EXE.

In this case the menu item PROCESSING of the Digitizer Main Window runs the compiling process of EXFOR file without such bugs.

There is also an opportunity to look through the compiled EXFOR-file and make some changes in it by a simple Editor.

The results of service files checking will be presented in a special window. I am going to enumerate some functions that have been already implemented. Then I will briefly describe our plans.

There is a possibility to look through the texts of SRC and AXC files and make necessary changes with the help of Text Editor.

The presence of all required commands such as BEGW, BEGG, BEGC, all ENDs and so on is verified.

The User can also edit names and units of independent variables, change number of them and correct axes names and units.

We are planning to implement the following opportunities:

- checking axes linearity and correctness of axes scaling
- editing reference points
- data plotting of the compiled EXFOR file.

The last function will be also available for additional checking of digitization results. For its implementation we are going to use the ready Chart component from the EXFOR-Editor.

We shall try to prepare the final version of the EXFOR-digitizer for the EXFOR Workshop next year.

Development of EXFOR Digitizer Wizard – New Interface

S. Taova
Russian Federal Nuclear Center – VNIIEF, Sarov 607188, Russia

We have got a lot of user feedbacks for our version of the EXFOR Digitizer during the last two years. Development of a new version of InpGrapf was started in our centre. This decision was motivated by four serious reasons:

a) Now our digitizing program is used not only in our center
b) Different comments and remarks come from the users
c) It is not possible to take into account these comments performing the minor changes only
d) New techniques have been developed providing the improvement of user's interface.

The idea was to use a new approach - Parameter Gathering Techniques – which allows the user entering the minimum amount of information. The rest of the parameters are determined by the program itself.

We decided to use a Wizard which can lead the user through the interface step by step to do tasks in a prescribed order.

Some new possibilities of our Graph Digitizer are listed below:

1. There is no necessity to enter coded words any more. These steps are hidden from the user.
2. It is possible to use automatic scaling (boundary values and a number of reference points are set).
3. Editing of information is allowed at any digitizing stage.
4. Data checking is provided at the stage of entering information.

There are four new buttons on a panel (**Fig.1**):

- Add Curve: start new digitizing
- Save Data: save digitized data
- Check Data: check the order of input (definition of scale, reference points, independent variables and so on).
- Process Data: call subprogram to transfer plot coordinates to physical values.

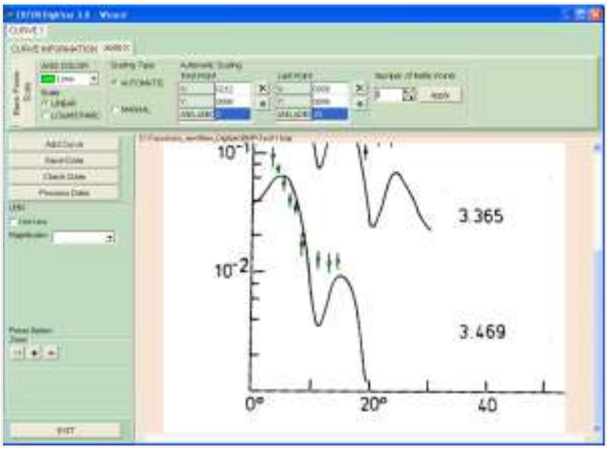


FIG. 1 (colour online). A new window of InpGraph.

Our future plans are:

1. Introduce a possibility of entering asymmetric errors
2. Provide an option for setting data precision (the number of significant figures in data values)
3. Connection with Exfor-Editor will be implemented. It will be possible to include digitized data to Exfor-Editor directly.

Introduction, Design and Implementation of Digitization Software GSYS

Ryusuke Suzuki
Department of Medical Physics, Hokkaido University Hospital, Sapporo 060-8648, Japan

1. Introduction

Japan Charged-Particle Reaction Data Group (JCPRG) has accumulated the charged-particle induced nuclear reaction data measured by accelerators in Japan for the Nuclear Reaction Data File (NRDF). Though it becomes possible to obtain experimental data directly from experimentalists, it is still necessary to convert the graphical data on the paper into numerical data in case that numerical data cannot be obtained from the authors.

In the past, an input device called ‘digitizer’ was used for reading numerical data from printed matters directly. In 1985, JCPRG started to develop a system for this type of device, GRADIS [1-4]. Then such a device was replaced by digitization software which digitizes numerical data from electric image files. The first trial at JCPRG is development of SyGRD [sigúrd] [5] (**Fig.1**), which is a macro program for an existing image analysis tool.

FIG. 1 (colour online). Screenshot of SyGRD [5].

In FY 2004, Dr. K. Arai developed the first version of GSYS [6] (now called as GSYS1**, Fig. 2(a)**), which is a cross-platform GUI application built from scratch. This system was adopted as a successor of SyGRD.

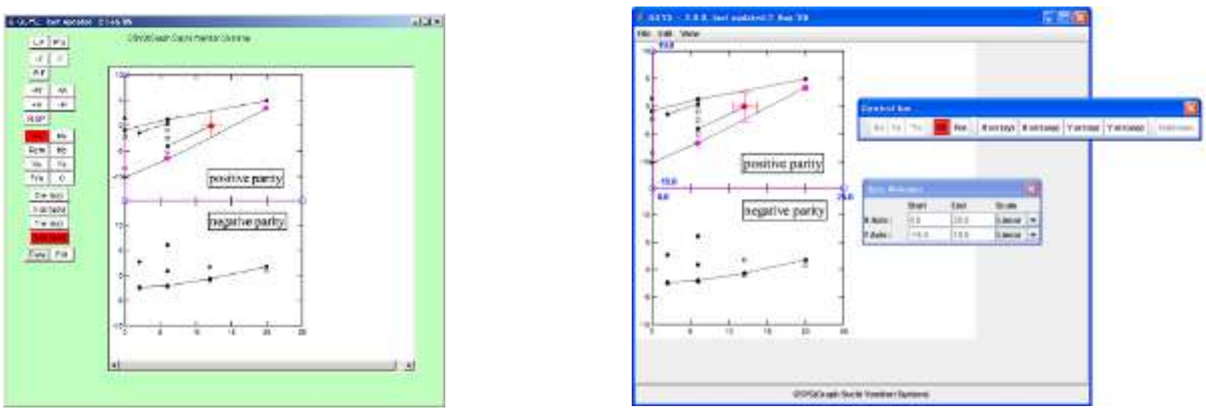


FIG. 2 (colour online). Screenshot of main window; GSYS1 (a) and GSYS2.0 (b).

In FY 2005, the author revised whole user interface, and added “feedback function” to GSYS in order to reuse the numerical data for checking digitization accuracy by plotting the numerical data on the image. Many ideas suggested by compilers were also incorporated. This new version was released as the second version of GSYS [7] (now called as GSYS2.0, **Fig. 2(b)**) in August, 2006.

After release of GSYS2.0, the automatic axis detection system, which automatically detects and sets the position of axis by easy operation, was included by S. Ito. It reduced operators’ work and the ambiguity of human judgment. In addition to this new feature, some usability improvements were performed. This new system was released as GSYS2.2 [8] on December 31, 2006.

On April 22, 2010, GSYS2.4 [9] was released as an update version of GSYS2.2. Main features added in this version are (1) “Automatically point recognition function”, which automatically recognize the point near the clicked position, and correct the marker position, and (2) “Magnifying glass function”, which enlarges the figure partially and enable us to perform usual operation on this enlarged figure. Many kind of new features were also included in this version.

After GSYS2.4 was released, further development of GSYS continues. GSYS2.6 is planned to release in the future.

GSYS has all features for digitizing process from beginning to end, and also has tools for quality assurance (QA) of digitized data. GSYS is cross-platform application which only requires the Java Runtime Environment (JRE) without any specific library. It can be downloaded from the JCPRG web site [10]. GSYS is Free, but there are terms of use

- PLEASE USE THIS SYSTEM AT YOUR OWN RISK.
- IT IS NOT ALLOWED TO USE THIS SYSTEM FOR ANY KIND OF COMMERCIAL BUSINESS PURPOSE.

In this paper, some features and technical details of GSYS are discussed. A short manual is also appended for people who want to use GSYS.

2. Features

GSYS is featured software. Features of GSYS are easy to use and many functions, such as “drag-and-drop”, “Loupe”, “automatic axis detection function”, “automatically point recognition function”, “feedback numerical data function”, and “magnifying glass functions” are implemented. The detailed explanations of these functions are explained in the GSYS manuals [7-9]. Some features are introduced in this subsection and technical detail is discussed in the next subsection.

a) Feedback function

Feedback function is a function to load the numerical data from files and plot them directly on the image displayed on the main panel. This function is implemented in GSYS2.0 [7]. Before GSYS2.0, digitizing processes were one way process that read the data from an image, then digitize it to the numerical value. Therefore, if there are some mistakes in the data, or the quality of the data is not very good, the user must reread the data from the beginning of digitizing process to improve the data. The feedback function enables us to compare the numerical data visually with the real data on the graph by plotting the compiled numerical data on the image as shown in **Fig. 3**. It is also possible to modify the data by moving or adding the data points. GSYS can read not only the numerical data produced by GSYS, but also general numerical data. Thus, the feedback function enables us to reuse the data easily and check the data accuracy in greater detail.

b) Automatic axis recognition function

Automatic axis recognition function was implemented by S. Ito in GSYS2.2 [8]. This function detects and sets the position of axis by easy operation. It reduced operators’ work and the ambiguity of human judgement. After clicking Auto button and enclosing X or Y axis by dragging as shown in **Fig. 4**, detected axis and tics are shown as green and blue lines as shown in **Fig. 5**, respectively. It was suggested that there is possibility to use this function for detecting the distortion of graph in this meeting.

c) Magnifying glass function

To digitize the numerical data from the graph precisely, it is essential to point the data position and axis position accurately. Magnifying glass function is implemented in GSYS2.4 [9] for this purpose by enlarging the figure partially as shown in **Fig. 6**. The author is inspired by InpGraph [11] to implement this function. Usual operations (add point, correct the marker position, and so on) can also be performed on this glass window. You can move this window by dragging or using cursor key.

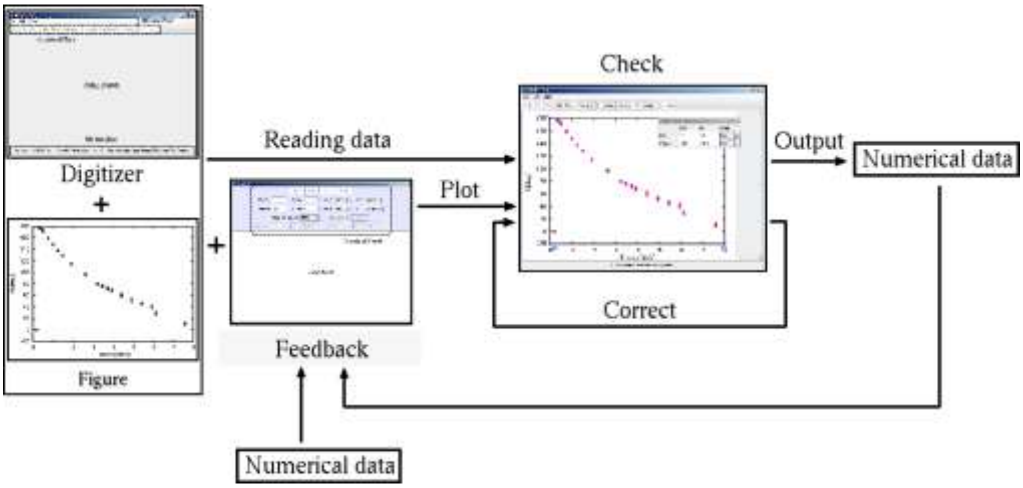


FIG. 3 (colour online). Data reading process and feedback function.

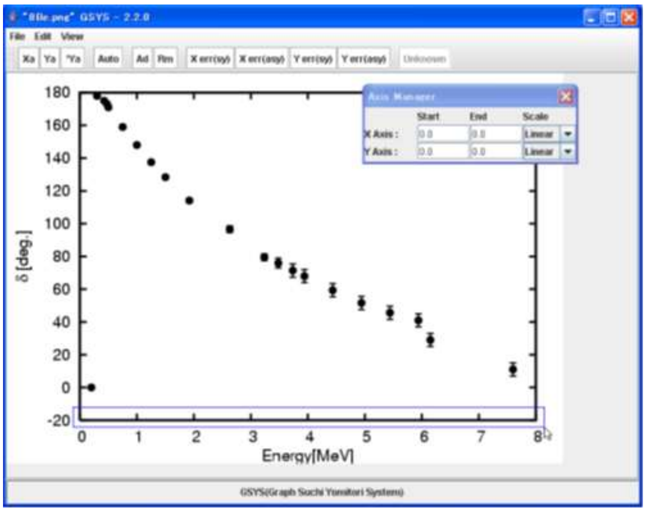


FIG. 4 (colour online). Window to select a starting point of an automatically detected axis.

FIG. 5 (colour online). Window to select a starting point of an automatically detected axis.

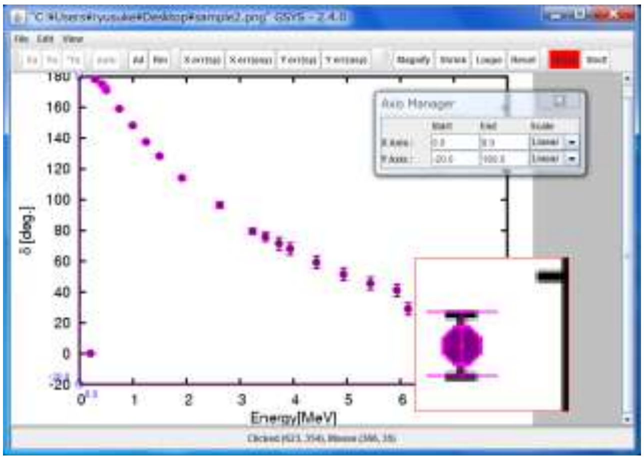


FIG. 6 (colour online). Magnify glass function.

3. Technical details

a) Automatically point recognition function

Recent version of GSYS (GSYS2.4) has the automatic point recognition function (as shown in **Fig. 7**). With this function, the maker on the figure is automatically recognized and the position is automatically corrected.

For recognizing the point from image, template matching technique is used. The author is inspired this technique from Real-time Tumour-tracking radiotherapy (RTRT), which is developed by the Hokkaido University hospital [12]. In the RTRT, a surrogate gold fiducial marker implanted in/near tumour is tracked using a pair of diagnostic fluoroscope X-ray units. For tracking the gold marker, the marker position is searched and recognized in the fluoroscopic images by using template matching technique. The same technique is used in GSYS.

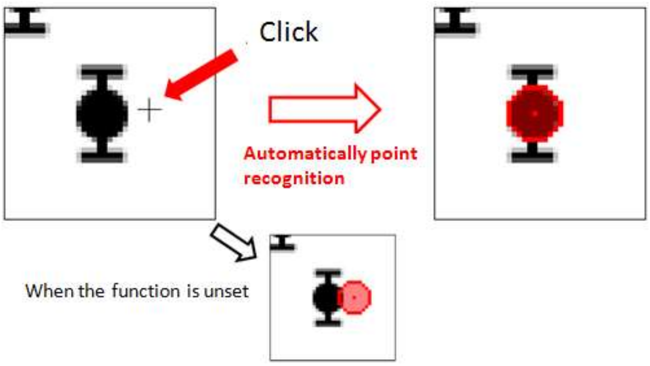


FIG. 7 (colour online). Automatic point recognition function. If you click the neighbourhood of point as shown in left hand side, GSYS2.4 recognize and correct the point position, and add the point. If you do not use this function, data point is added at just the point where you click.

In the template matching technique, template image and target image are prepared. Pattern recognition score (PRS) are calculated to determine the position of point on the template image. PRS is calculated as

$$r = \frac{N \sum_{i=1}^{N} I_i T_i - \left(\sum_{i=1}^{N} I_i \right) \sum_{i=1}^{N} T_i}{\sqrt{\left| N \sum_{i=1}^{N} I_i^2 - \left(\sum_{i=1}^{N} I_i \right)^2 \right| \left| N \sum_{i=1}^{N} T_i^2 - \left(\sum_{i=1}^{N} T_i \right)^2 \right|}}$$

$$PRS = 100 \times r^2 \quad (r \geq 0)$$
$$PRS = 0 \quad (r < 0)$$

where r is normalized Cross-Correlation coefficient, I_i is pixel value of target image, T_i is pixel value of template image and N is number of pixel.

In the automatic point detection function implemented in GSYS, small region centered on clicked position is used for target image and circles with various radiuses are prepared as template image. PRS are calculated all possible location of template image on target image and all prepared template image. The location which gave the largest PRS is assumed to be the detected point position. If the maximum value of PRS is not sufficiently large, it is judged that point is not detected. The implementation of this feature is not so mature, and further improvement of point recognition is planned.

b) Header of output file and feedback function

Some meta-data are included in the header part of output file generated by GSYS2. There are `Axis_X`, `Axis_Y` and `MD5` lines, such as

```
Axis_X : 3fba8106fc4d40833fd6f96f96f96f973fe40df371b3450f3fd6fb00e1ba0c58
Axis_Y : 3fba81be6e3668a23fd6fb00e1ba0c583fba81be6e3668a23fac5ce08d1447b7
MD5Fig : cc66d9f4df54b8f1888329fb2c130da7
```

This information are used for feedback function. The data of `Axis_X` and `Axis_Y` lines are hexadecimal number representation of coordinate of X axis and Y axis on the image file, respectively. These information are only meaningful for the image file, which is used for digitizing. If you use these axis information to different figure, mismatch problem will occur as shown in right-hand of **Fig. 8**. To identify the images (to avoid the mismatch problem), finger print of image file is used, since finger prints of different image are different. In this matter, GSYS can restore the axis information correctly from the header information when we use the feedback function.

MD5: cc66d9f4df54b8f1888329fb2c130da7 MD5: 41ba3c52b6cff490a2a1236057ca463c

FIG. 8 (colour online). Both figures are same graph, but not the same image. Axis information is used for corresponding image in left figure, and axis information is used for different figure (right figure).

c) Coordinate, orthogonality condition

It is assumed that GSYS is used only for graph in orthogonal coordinate system. In the case of graph image is distorted, the graph cannot keep the orthogonality of X-axis and Y-axis. GSYS can used even for such a graph by disabling the orthogonality condition on the property dialog and set the X-axis and Y-axis as coordinate of graph. In such a case, there are two possibilities to calculate the coordinate of point a) by drawing perpendicular lines toward X and Y-axis or b) by using oblique coordinate system. GSYS adapt method a) and does not use oblique coordinate system for calculating coordinate of point under the assumption that nuclear data are plotted only on the orthogonal coordinate system, because of continuity of Japanese digitizer and simplicity.

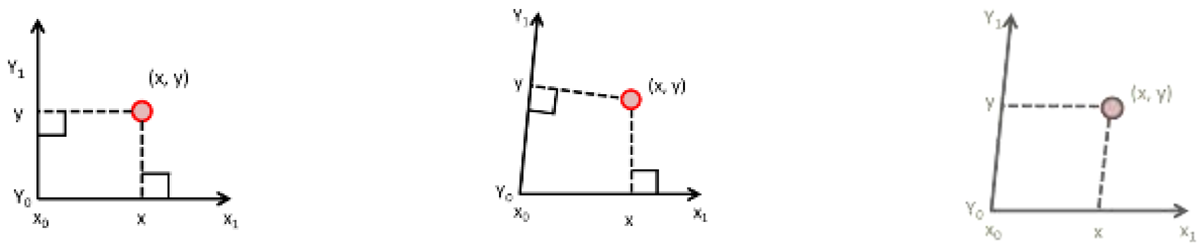


FIG. 9 (colour online). Graph in orthogonal coordinate system (left) and Graphs in non-orthogonal coordinate system (middle and right). Point position is calculated by drawing perpendicular lines towards X and Y-axis (middle), and oblique coordinate system (right) is not used.

4. Software development (version numbering)

After GSYS2.0 was released, GSYS is developed by using trunk and branch scheme as shown in **Fig. 10**. The each releases of GSYS have the version number, e.g, Gsys2.4.2. The 2nd number ("4" in this example) means the major version number and the 3rd number ("2" in this example) means minor version number. GSYS with even major version number is released branch and odd version is used for the trunk (main development line). The minor version number used in the released branch is increased if bug is fixed, or minor function improvement is performed.

The new features are included in the trunk branch (which have odd major version). If the features are included sufficiently and the application become stable, the major version number is incremented by 1 and released. After releasing the stable release, the major version number of trunk is incremented one more to odd number and the development of GSYS continues in this odd numbered branch.

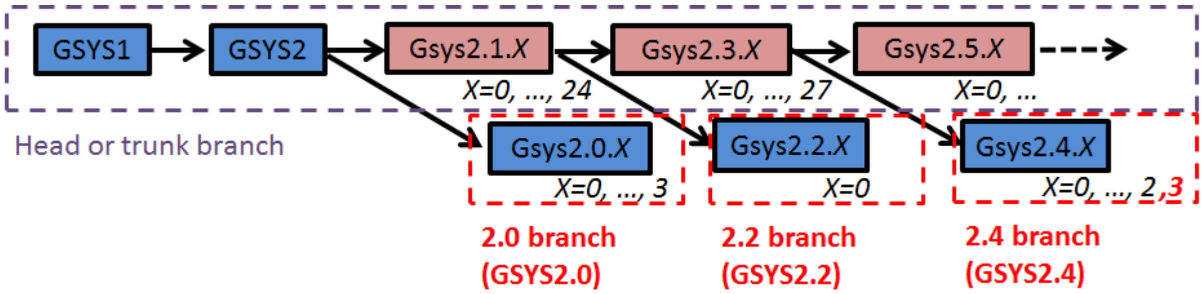


FIG. 10 (colour online). Simplified diagram of GSYS history.

5. Quality of digitized value

Regarding to the quality of the digitized value, the author presented that the quality of digitized numerical values is influenced by many kind of sources or factors (Image, Digitizing software and digitizer (individual or group)) as shown in **Fig. 11**. Though the digitizing software can be developed to improve the quality of digitized numerical value, it is difficult to take care of all possible sources which reduce the quality.

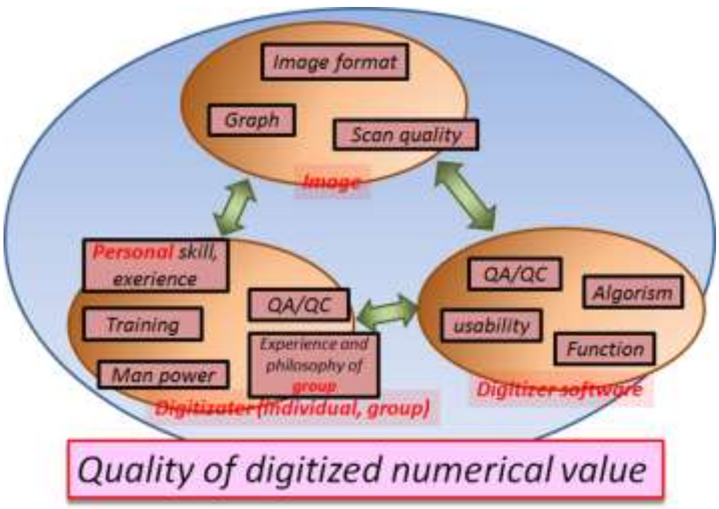


FIG. 11 (colour online). Factors and sources which influences the quality of digitized numerical value.

The quality of digitization is not only determined by digitization software, but individual skill and experience are very important. Training, man power, QA and QC system for each group are also important.

The quality of digitization is also determined by quality of the figure image. GSYS can now treat BMP, JPEG, and PNG format (bitmap or Raster scan for). However vector format is good for quality, so it is useful for treating PDF format, since recent original figure is distributed by PDF (vector format). If the file is scanned, scan quality also influence the quality of data and quality of original figure also should be considered, since sometimes old experimental data is written manually.

6. Summary

GSYS has been developed over 7 years. Originally, GSYS is targeted to NRDF compilation. Nowadays, GSYS is widely used for EXFOR compilation by other NRDC members [12], and also used even for the purpose of non-nuclear activities.

At GSYS2.X, the author aimed to improve usability and simplicity and include some interesting features, for example, feedback function, magnifying glass function and automatically point recognition function. The author believes that these features or functions improve the quality of digitized data.

As discussed in this paper, many kind of factors and sources, which influences the quality of digitized numerical value, exist. Digitizing software will contribute to improve the quality of data further, but it is difficult to take care of the factor especially come from human factor. In the initial work using the feedback function [7] and benchmark test [13, 14] shows that the mistake of axes setting causes the big uncertainties. This kind of mistake happens easily, so each group establish their own digitizing protocol to reduce this kind of error.

It is evident that there is a growing international awareness of the importance of the quality of numerical data [13, 14]. The author hopes that the effort to provide and improve good numerical data continues in the future.

7. Future plan

The next major updated version of GSYS (GSYS2.6) is underway to mainly implement the “undo” and “redo” functions and refactoring the system. At GSYS2.6, it is also planned to implement the interface to change the magnification factor and size of magnify glass function. It is often plan to improve the auto point recognition function.

Acknowledge

I would to acknowledge several individuals in the JCPRG members (present or past). Dr. Koji Arai for his developing the first GSYS and providing source code. Dr. Shinya Ito for his releasing GSYS2.2 and developing GSYS2.4 (at 2006). Dr. Hirokazu Ohmi for his giving advice to me. Dr. Ayumi Minoguchi, who named the ’feedback function’, and also cooperated on translating GSYS2 manual into English and for her encouraging. Dr. Kenichi Naito for teaching me profundity of programing language. Regarding to GSYS2.4, I would like to thank Hideki Murakami (for management of the development) and Vidya Devi (for reviewing GSYS2.4 manual). I wish to thank the users of GSYS, including Ms. Takako Ashizawa for giving us valuable comments. I wish to thank Ms. Hitomi Yoshida and Dr. Naohiko Otuka for strong support from the development and release of GSYS2.X.

I also acknowledge Dr. Naoki Miyamoto (medical physics at Hokkaido University) for his explaining me the template matching technique. At last, I would like to express deepest heartfelt gratitude to my beloved family.

References

[1] H. Tanaka and H. Kazama, NRDF Annual Report 1985, 1-2, 1985. (Japanese)
[2] H. Kazama, NRDF Annual Report 1985, 3-14, 1985. (Japanese)
[3] H. Kazama, NRDF Annual Report 2, 2-16, 1988. (Japanese)
[4] S. Okabe, NRDF Annual Report 4, 2-8, 1990. (Japanese)
[5] H. Ohmi, Development and User’s Manual of Graph Reading System with Customized Image Analysis Software, NRDF Annual Report 12, 2-20, 1998 (Japanese); Development, installation and user’s manual of SyGRD (System of Graph Reading and numerical data Displaying with image analysis software), NRDF Annual Report 15, 50-85, 2001. (http://www.jcprg.org/gsys/sygrd/)
[6] GSYS: Development and usage of a software to read-in and digitize the graphical data, NRDF Annual Report 18, 78-89, 2004. (Japanese); A. Minoguchi, K. Arai, N. Otuka, K. Naito, GSYS: Development and usage of a software to read-in and digitize the graphical data, INDC(JPN)-194/U, 19-22, 2005. (http://www.jcprg.org/gsys/ver1/gsys-e.html).
[7] R. Suzuki, GSYS2 Manual, NRDF Annual Report 19, 10-44, 2005. (http://www.jcprg.org/gsys/ver2/gsys-e.html).
[8] S. Ito and R. Suzuki, GSYS2.2 Manual, NRDF Annual Report 20, 39-76, 2006. (http://www.jcprg.org/gsys/ver22/gsys-e.html).

[9] R. Suzuki, GSYS2.4 Manual, NRDF Annual Report 20, 3-25, 2010. (Japanese); GSYS2.4 Manual, http://jcprg. hokudai.ac.jp/gsys/2.4/gsys24-e.pdf. (English). (http://jcprg. hokudai.ac.jp/gsys/2.4/).
[10] GSYS website, http://jcprg.hucc.hokudai.ac.jp/gsys/.
[11] InpGraph, G. Pikulina et al., Sarov, Russia, 2012, Web: http://www-nds.iaea.org/nrdc/nrdc_sft/inpgraph-201211.zip
[12] H. Shirato, et al., Four-dimensional treatment planning and fluoroscopic real-time tumor tracking radiotherapy for moving tumor. Int. J. Radiat. Oncol. Biol. Phys. 48, 435-442, 2000.
[13] V. Semkova, Benchmarking of Digitization Software (Memo CP-D/761): Results and Discussions, Summary, Appendix D in this report.
[14] N. Otuka and S. Dunaeva, Quality and consistency of digitized data, Appendix E in this report.

Appendix: Short manual

1. Download application from the web site of GSYS [10].
2. Start GSYS by double clicking the application.
3. Drag and drop the figure image on to the GSYS main window, or by selecting “Open Image File” from “File” menu and load the image file.
4. Enlarge the image file sufficiently by clicking <u>Magnify</u> button or using loupe function (press <u>Loupe</u> button and drag the region to be magnified).
5. Select the axis using following a) or b):
    a) Press <u>Auto</u> button, and enclose X-axis dragging on the image. Select the start point and end point of X-axis from detected positions. Set the Y-axis same way.
    b) Press <u>Xa</u> button, and select start and end point of X-Axis. Similarly, press <u>Ya</u> button and select start and end point of Y-Axis. If the start point of Y-axis is same as X-axis, press <u>*Ya</u> instead, and select end point of Y-Axis only.
6. Input the values of start and end point of X-axis and Y-axis to “Start” and “End” menus in axis dialog. Also select the scale type (Linear or Log).
7. Press <u>Ad</u> button, and click the data points.
8. After finishing adding the data points, read error information of X value by using <u>Xerr(sy)</u> or <u>Xerr(Asym)</u> for symmetric or asymmetric error, respectively. After pressing the button, select the point to be added error, and click one endpoint. If you add the asymmetric error, you should click both endpoints.
9. Same operation can be done for errors for Y value of data by using <u>Yerr(sy)</u> or <u>Yerr(Asym)</u>.
10. Output by using the Output Data window which can be displayed by selecting “Output Numerical data” from “File” menu.

Tips:

- For modifying the position of data points, error bars, and start or end point of axis, dragging by mouse is easy way. However, cursor key is useful for fine adjustment. Magnifying glass function is effective for this kind of correction.

- Before submit the digitized data, it is recommended to check the data point, error bars, start and end point of axis. Feedback function is useful for this kind of check.

Introduction of the Digitization Software GDgraph

Chen Guochang, Jin Yongli, Wang Jimin
China Nuclear Data Center, China Institute of Atomic Energy, Beijing, 102413, China

The evaluators and experimenters always desire to have full and latest experimental data sets. However, the data are often published on figures without numerical values for some publications or journals. Furthermore, the quality of figures is not always good enough, especially for some figures scanned from the hardcopy of old publications. Furthermore the researchers would like to retrieve the data directly from EXFOR database. So digitization of figures is only one method to obtain the data and correlative uncertainty from old publications. For the requirements from evaluation, measurement and EXFOR compilation, we need to develop software for digitization at China Nuclear Data Center (CNDC).

Before 2000, there was no common software to digitize experimental and evaluated data. And the quality of digitization results couldn't fit the requirements of evaluation and measurement using the traditional coordinate paper or rule. The end of twenty century, the personal computer is developing so quickly that to develop software for digitization purpose become possible. Since 1997, China Nuclear Data Center devotes to develop software for digitization. 4 years later, the first version of digitization software GDgraph was released and developed using VC++ computer language. Although, the functions of the 1st version of GDgraph is fit the basic requirements of digitization only, in which can digitize one group data excluding data uncertainty, BMP image format only, and it couldn't randomly delete digitizing points. However, the mold of GDgraph software can obtain higher quality digitizing results and efficiency than the traditional way.

After 5 years to use GDgraph, we collected much feedback information on update and some bugs on this software. The 2nd version of GDgraph software was released at 2006, in which the whole software was re-written using Perl computer language to obtain more comfortable condition for programming and update requirements. Some new functions are listed as following:

1) The graphics can be opened including mostly image format such as jpg, png, bmp etc.
2) The graphics could be zoom-in and zoom-out.
3) Using new button to realize the graphics rotation, so the coordinate of image could fit well with the software one.
4) The image size could automatically fit to the GUI window.
5) Revert image size after the original one is been zoom-in or zoom-out by the digitizer.
6) Allow to add and delete data groups. The maximum number of data group is three and use blue, red and green colour to represent each data group, respectively.
7) Allow to randomly add and move the digitizing point by mouse.
8) Output data could be saved as a data file or clipboard.
9) Allow to digitize y error with mouse.
10) Allow to set fix value for y uncertainty using relative (%, in per cent) or absolute value.

The 3rd version of GDgraph software was released at 2011. Some new functions are list as following:

1) Allow to copy image directly from clipboard.
2) Allow to select thin or thick line for axis and error bar line of digitization.
3) Except for the rotation figure function, there are allow user to set rotated angle to realize slightly degree rotation.
4) Partially zoom-in of the selected point is available for checking the position of digitizing point.
5) The digitizing axis and image could zoom-in and zoom-out together.
6) Allow to move the digitizing point by keyboard for slightly adjusting the position.

The latest version of GDgraph is 4.4 and released at 2012. Some new features are list as following:

1) Allow to digitize x error and use keyboard to move xy error bars.
2) Add Project and Remarks function. The Project function allows to save image, digitizing results, coordinate setting and other setting together for next time to check and modify. The Remarks function allows to keep some marks and memo text for checking, modification and memory.
3) Add Setting function, so to change the colour, size and symbol of point is available. The size of partially zoom-in window is allowed to be changed. Allow to change the magnification of partially zoom-in from 200% to 400% for obtaining more clear point image. Allow to change the length of error bar cross line to fit the error bar in image. The grid and the setting of the space is available.
4) Allow to move partially zoom-in window by mouse.

Since 1997, the digitization software GDgraph is developed to fit the requirements of evaluation, measurement and EXFOR compilation also. From the mold software to present version 4.4, GDgraph is mainly fit the requirements, although there are some aspects need to modify and also add some new functions also.

Proposals on the Development of Software on Data Digitizing for Data Entering into the EXFOR Library

G. Pikulina
Russian Federal Nuclear Center – VNIIEF, Sarov 607188, Russia

At present we actively use two digitizers. The EXFOR-Digitizer of our center is intended to digitize low quality graphic data presented in old literature sources. The cross-platform digitizer GSYS of Hokkaido University is very convenient for digitizing of high quality graphic data. Both of these digitizers provide a manual mode of data input.

In our opinion many difficulties are connected with digitization of curves with a lot of points or solid lines. Implementation of automatic or semi-automatic digitization could simplify the compilation of graphic data for EXFOR.

We propose two ways for such implementation:

- Use of suitable freeware or shareware soft products after careful testing of trial versions;
- Development of a special algorithm of automatic or semi-automatic digitization and including a plug-in into the existing Digitizers.

As for the first way there are a lot of different freeware or shareware soft products for conversion of scanned bit images into a vector format. We have tested several programs and come to the following conclusion: automatic digitizing mode of such programs requires colour bit images of high quality.

All these programs have their own advantages and disadvantages. They solve the common problem of automatic digitization but, possibly, don't meet particular requirements of EXFOR data. For example there may be problems with digitization errors.

At the first step the image pre-processing ought to be available. This processing could include the following operations: image filtration (in case of noises), image rotation (if figure was scanned with a bent); erasing coordinate lines and captions

There are a lot of algorithms for automatic digitizing mode. For example, the image binarization could be used. A binarization threshold is selected for every image. It is determined by histogram analysis of background local intensity. Then the most intensive points would be found.

The skeletonization algorithm for curves digitization could also be applied.

To our mind the problem of automatic vectorization may be solved if an image contains one curve. It also could be implemented for noncrossing curves.

If curves cross or overlay the semiautomatic technique could be applied. In this case areas of curves are selected and then automatically digitized.

We are greatly interested in fruitful discussion on these questions.

GetData Digitizing Program Code: Description, Testing, Training

S. Taova
Russian Federal Nuclear Center – VNIIEF, Sarov 607188, Russia

90 percents of compilation in our center is obtained by data digitizing. So we are rather interested in the development of different techniques of data digitizing.

Plots containing a great amount of points and solid lines are most complicated for digitizing. From our point of view including to the Exfor-Digitizer procedures of automatic or semi-automatic digitizing will allow to simplify significantly this process.

We managed to test some free available program codes. Program GETDATA Graph Digitizer (www.getdata-graph-digitizer.com) looks more suitable for our purposes.

GetData Graph Digitizer is a program for digitizing graphs, plots and maps. Main features of GetData Graph Digitizer are:

- supported graphics formats are TIFF, JPEG, BMP and PCX;
- two algorithms for automatic digitizing;
- convenient manual digitizing;
- reorder tool for easy points reordering;
- save/open workspace, which allows to save the work and return to it later;
- obtained data can be exported to the clipboard;
- export to the formats: TXT (text file), XLS (MS Excel), XML, DXF (AutoCAD) and EPS (PostScript).

GetData Graph Digitizer includes two algorithms for automatic digitizing.

Auto trace lines:
This method is designed to digitize solid lines. Choose the starting point, and the program will trace the line, stopping at it's end. To trace the line use Operations=>Auto trace lines menu or context menu ("Auto trace lines" item). To choose starting point click left mouse button, or click right mouse button to additionally choose direction for line tracing.

Digitize area:
The second way is to set digitizing area.
This method works for any type of lines, including dashed lines. Data points are set at the intersection of grid with the line. You can choose the type of grid (X grid or Y grid), and set the distance between grid lines.
You can also make the grid be shifted in such a way, that it will pass through a specific X (or Y) value. To digitize area use *Operations →Digitize area menu*.

You may use Clipboard to output data or a special procedure “Export to text file”. Some additional options can be set.

You also have an opportunity to save the Workspace. This allows you to save your work, and return to it later. Workspace files are saved in XML format.

So, we think this program code may be useful when you deal with lines and big arrays of points.

There are only two important conditions for such automatic digitizing:

1) There should be curves on a figure well separated from each other;
2) The quality of plot should be high enough: no spots, no figures.

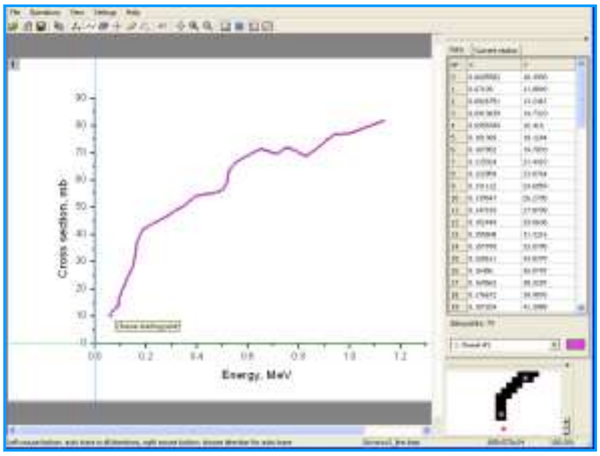


FIG. 1 (colour online). A window of GetData Graph Digitizer.

An Overview of Indian EXFOR Compilation (Feedback and Suggestions)

Lalremruata Bawitlung
Department of Physics, Mizoram University, Aizawl-796004, India

To cater to various needs of the department, the Nuclear Data Physics Centre of India (NDPCI) was formed which has been successful in pursuing all aspects of nuclear data, viz., measurements, analysis, compilation and evaluation involving national laboratories and universities in India. EXFOR compilation on a regular basis has been going on for the last few years, through funds given by NDPCI to Universities. In this Presentation, a brief introduction on the Indian EXFOR Compilation procedure, and the peoples involved have been presented. Feedbacks and Suggestions of the three Indian Compilers (Ms. R.Gosh, Ms. S. Badwar, Dr. R. Mandal) have also been discussed. The error messages in the Russian EXFOR Digitizer version 2.1, such as I/O error 32, I/O error 103 and the sources as well as the reproduction procedures of the error messages have also been demonstrated. The error encountered with GSYS2.4.2, error message 'java.lang.Array IndexOutOfBounds Exception' has also been reported. The old version of EXFOR dictionary used by the Russian EXFOR Digitizer version 2.1 has also been reported with examples of Sb-129 missing. The problem faced by Indian compilers such as, "the file does not exist" message they get from running the graph_new.exe file is discussed. Some of the suggestions for improvement of the Russian EXFOR software such as, having control over the number of digits in the Digitizer output, choice of fixed or floating decimal number for the linear and logarithmic scales, to be able to digitize the asymmetric error bars, etc have also been reported. The need for the Japanese GSYS digitizer to evaluate the digitizer uncertainty based on the suggestions by the Indian compiler is also discussed.

Digitizing Process: Comments, Real Error Estimation, Recommendations

M. Mikhaylyukova
Institute for Physics and Power Engineering, 249020 Obninsk, Russia

At present time there are lot of articles, where data are given on figures only and authors' data are not available by a request to the authors. In such cases the digitization is only one possible way to get the data to be inserted in EXFOR data base. This is the reason, why quality of digitization software is considered to be very important.

In CJD the digitizing code InpGraph (Sarov) is used for digitization during last 10 years. Recently also the GSYS code started to be used for digitizing.

A. Questions, Comments, Proposals for Improvement of Sarov's Digitizer

The main advantage of Sarov's digitizer (InpGraph.exe of 26.08.2010, graf_new.exe of 29.02.2004) is that it gives errors due to scales digitization. Comments are listed below.

At 2009 workshop

1) Number of meaning digits in DATA and DATA-ERR at digitizing - now is usually 5. Could be defined at digitizing process by pushing some new button?

   - Number of meaning digits in DATA-ERR has to be 3 (maximally 4, it's enough, but not 5, as now).
   - Number of meaning digits in DATA has to correspond to number of meaning digits in DATA-ERR.

   As example:

```
DATA = 0.01354   DATA-ERR = 0.00256
       5.3527               0.0425
       1.234E+2             0.456E+2
```

2) There is button "Inserting at the cursor place". Could be defined - "before" or "after" cursor place (line, where cursor is)?
3) Problems with definition of axes scales.
   - If I forget to add scale and try to add after digitizing – I receive a message "err 103" already after digitizing. Could be scales checked before digitization of data?
   - Also scale non-linearity is defined by program after the digitization of data. Could it be done before the process of data digitization? So, all errors concerning scale definition could be output just after this definition?
4) Could be outputs about errors defined more detail? Not as only error head?
5) ERR-DIG given by digitizer is really too low…
6) When one has defined data heading for X axis (printed as example EN), and then tried to define data unit and find it from list, it would be good, if units in list would correspond to data heading (in case of data heading EN – list of possible units for energy only, for ANG – angle units and so on). Now it's full list of all units and it's not convenient to find.

At 2011 year Workshop

1) Sarov's digitizer uses only BMP- files as input for digitizing.
   Could be PDF-format also introduced? And may be also others formats of figures?
   Because articles are sent usually in PDF- format. And it would be good to use as BMP as PDF (and may be others?) format files to avoid additional efforts of conversion into BMP-format.
2) Now Sarov's digitizer consists of two codes – first for digitizing process InpGraph.exe, second for producing a file in EXFOR format graf_new.exe. Could these two codes be combined in one?
   As example, add in first code a button "produce EXFOR file" to run second code.

3) When start to run second code to produce EXFOR file, one needs to print file name with exact way for directory, where this file is. It's not convenient and could be avoided, if 2) will be introduced.

This CM 2012

1) Buttons ErrX+, ErrX-, ErrY+, ErrY- could be introduced to digitize non-symmetrical errors?
2) Not all possible units are available in list (For example 1/MEV is absent)
3) Proposal - to make points for data, err-X, Err-Y by different colours. For example: data – green, X-errors – blue, Y-errors – yellow – during digitizing process.
4) If I mark point and then ERR-Y and only then ERR-X, I obtain wrong result…
   Could be the order "ERR-X and only then ERR-Y " be implemented by "ERR-Y and then ERR-X" – so, to introduce errors in any order?
5) Modes "Check" and "View" look like of the same possibilities. What is difference?
6) Could be "Scale" changed into "Zoom". (Scale is understandable as scales X and Y, but not as size of figure)?
7) Number of digits in errors could be reduced 5 -> 4 (or even 3) (see proposals of 2009 year).
8) Fist position is useless in data presentation (data < 1, E format.) Example:

such data	give as
0.1898	1.898E-01
0.4750E-01	4.750E-02
0.1152E-01	1.152E-02
0.8008E-02	8.008E-03

9) Text in ERR-ANALYS for digitizing errors could be improved by addition of more detailed digitizing error explanations.
10) Data symbols of red colour are not seen in additional window, which is open for higher size for selected point.
11) Button "END" could be deleted? If new button is pushed (e.g. LNY, this means "END" of previous option (e.g.,SCX)

B. Comments for GSYS Digitizer

The advantages of GSYS are that it uses modern, progressive and innovative method of digitizing and also is more convenient than Sarovs' digitizer for compilers usage. Comments are listed below.

1) Digitizer can define value of non-perpendicularity of X and Y axes?
2) Could error due to non-perpendicularity be estimated?
3) Could digitizer be improved for digitizing of figures produced by scanner?

C. Comparison of Digitization Results

The data of an experiment (made in IPPE) were published in different articles on figures (**Fig. 1**):

- Fig.3 of S,ISINN-18,153,2011
- Fig.3 of J,EPJ/CS,21,03005,2012
- Fig.3 of C,2010KRAKOW,,287,2010

These three figures contain the same data, reported at different conferences.

- ISINN article in pdf-format was made by scanner from ISINN-18 proceedings hard copy.
- EPJ/CS and 2010KRAKOW articles in pdf-format were made by computer conversion from doc-files.

Two figures (ISINN and EPJCS) were copied from the articles with the same zoom 300% in the article pdf-file, converted in BMP-files by the same manner and then were digitized by Sarov's digitizer. Third figure 2010Krakow: content copying is not allowed due to security password in the article – by Adobe Reader, figure was copied by PrintScreen button from pdf-file with the same zoom 300% and then saved as BMP-file by the same manner as previous two figures; additional fields (top, bottom, left, right, due to PrintScreen) were not deleted, and then this figure was also digitized by Sarov's digitizer.

- Figures in EPJCS and 2010KRAKOW look like of better quality than Figure in ISINN.

- Figures in EPJCS and 2010KRAKOW look like identical.

These three figures were digitized by Sarov's digitizer. Digitized data are compared with author's data (received from Vitaly Khryachkov, IPPE) on **Fig. 2**.

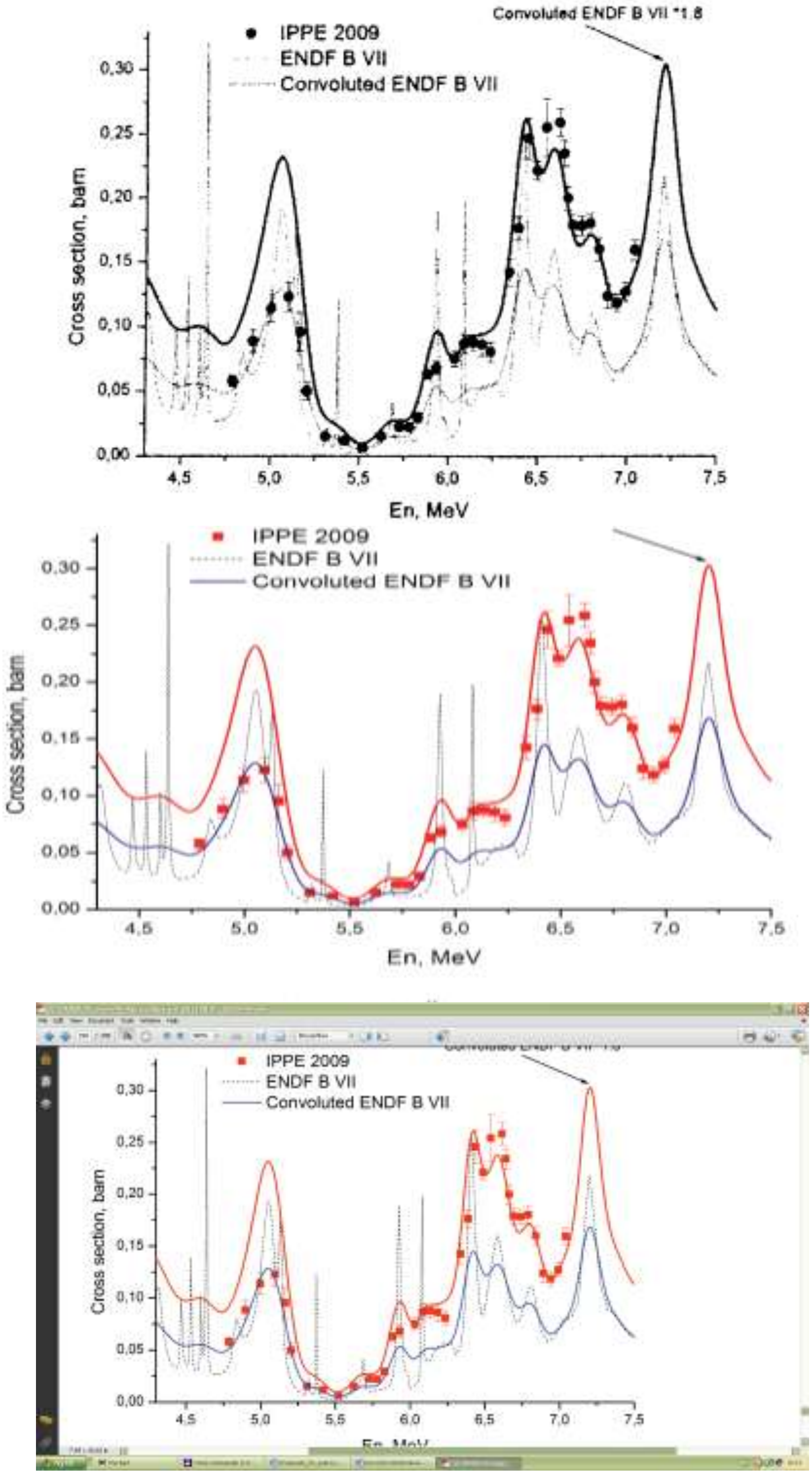


FIG. 1 (colour online). Figure images used for comparison of digitization results. (Top) Fig.3 of S,ISINN-18,153,2011. (Middle) Fig.3 of J,EPJ/CS,21,03005,2012. (Bottom) Fig.3 of C,2011KRAKOW,,287,2010.

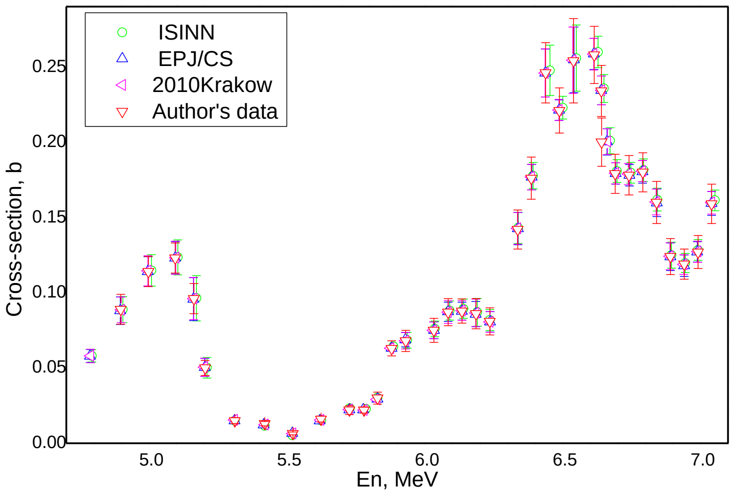


FIG. 2 (colour online). Comparison of absolute values of digitized data with author’s ones.

First point is absent in author’s data. Total error is given for author’s data.

C.1 Deviation of digitized data from author’s data was analyzed

Deviation as “author’s data minus digitized data” were calculated and analyzed.

Deviation in EN (X scale) See**. Fig. 3**:

In all cases – error given by digitizer is lower than real error of digitized data

For digitized data of ISINN figure (which was scanned by scanner using the hard-copy of article) the absolute deviation of digitized data from author’s data is essentially higher than for figures, produced by doc-files – more than 3.3 times.

TABLE I. Mean deviations in EN digitized by Sarov’s digitizer.

	ISINN	**EPJ/CS**	**2010Krakow**
Mean value of EN deviation (Fig.2), MeV	-0.00626	-5.71429E-6	3.91429E-4
Absolute mean value of EN deviation (Fig.3) , MeV	0.00637	0.00178	0.00193
EN-ERR-DIG given by digitizer, MeV	0.0054906	0.0012531	0.0015172
Ratio of abs.mean value to EN-ERR-DIG,no-dim.	1.16	1.42	1.27

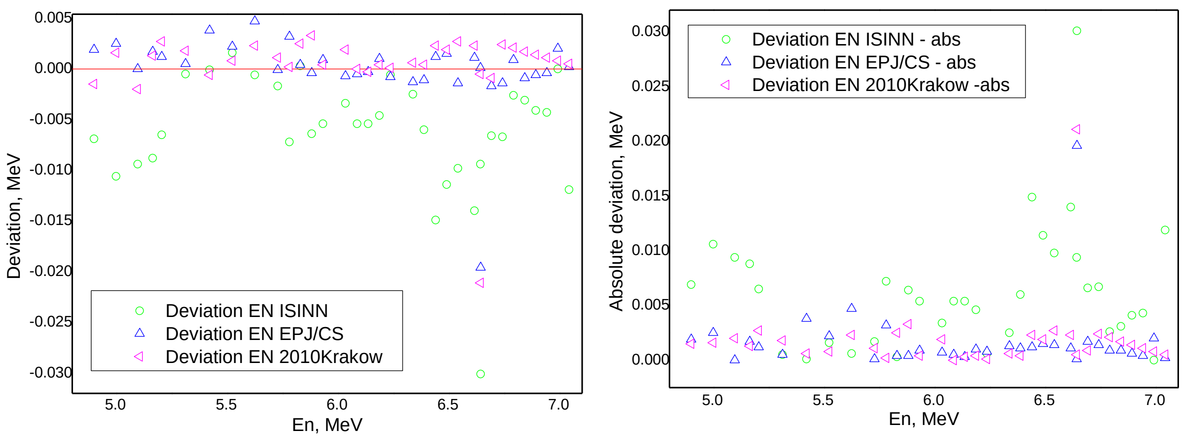


FIG. 3 (colour online). Deviation (left) and absolute deviation (right) of digitized EN values from author’s data.

Deviation in DATA (Y scale). See **Fig. 4**:

In all cases – error given by digitizer is lower than real error of digitized data.

For digitized data of ISINN figure (which was scanned by scanner using the hard-copy of article on paper) the absolute average deviation of digitized data from author's data is essentially higher than for figures, produced by doc-files – more than 2.85 times.

TABLE II. Mean deviations in DATA digitized by Sarov's digitizer

	ISINN	**EPJ/CS**	**2010Krakow**
Mean value of DATA deviation (Fig.4), b	-8.96403E-4	-1.22943E-5	-1.96017E-4
Absolute mean value of DATA deviation (Fig.5),b	1.12E-3	3.55666E-4	3.91926E-4
ERR-DIG given by digitizer, b	0.70250E-03	2.0735E-4	2.2050E-4
Ratio of abs.mean value to ERR-DIG, no-dim.	1.594	1.715	1.777

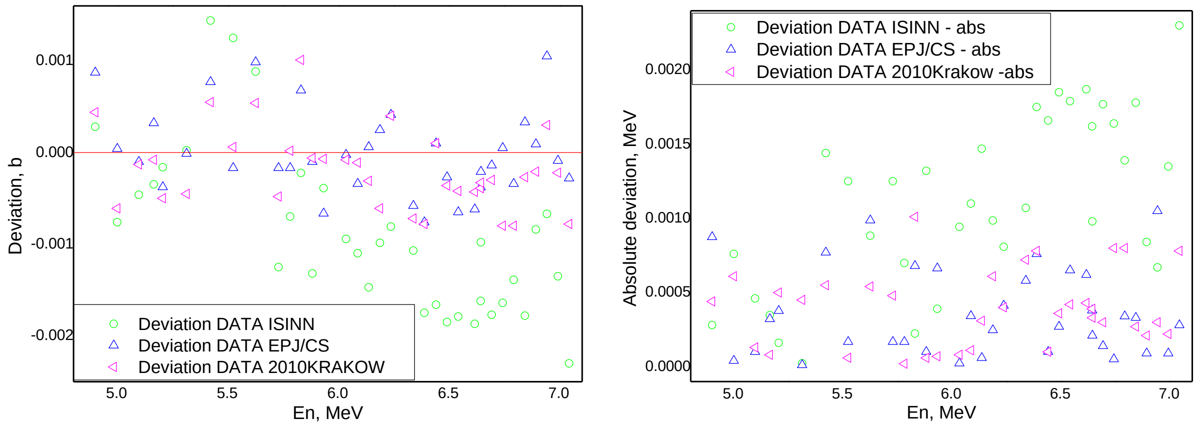


FIG. 4 (colour online). Deviation (left) and absolute deviation (right) of digitized DATA values from author's data.

C.2. Comparison results from Sarov's digitizer and GSYS for ISINN figure

ISINN figure was digitized by Sarov's digitizer and GSYS digitizer. Results of the absolute deviation from author's data are compared on **Fig. 5**.

Both digitizers could be improved for digitizing of "bad" quality plots..

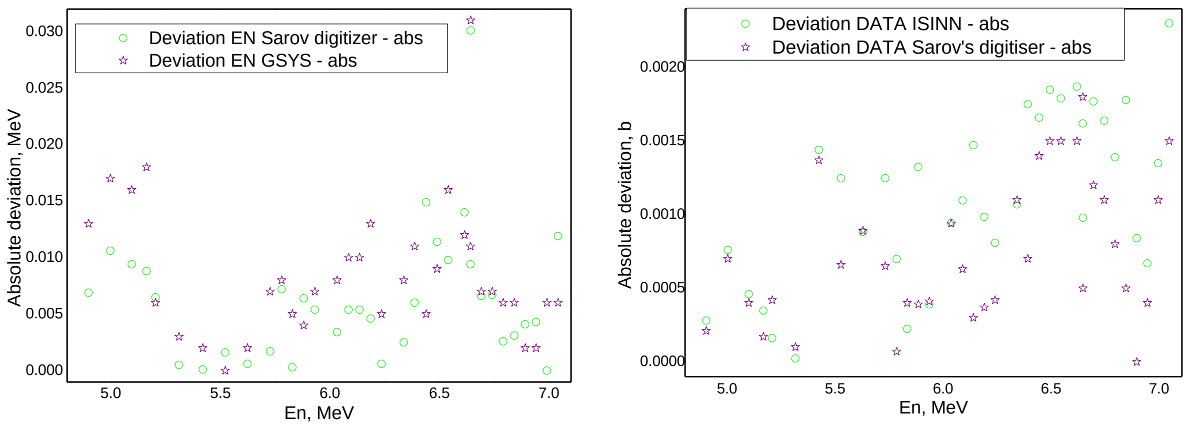


FIG. 5 (colour online). Absolute and deviation of ISINN data for EN (left) and DATA (right) digitized by two digitizers.

Table III. Mean deviations in EN/DATA digitized by Sarov's and GSYS digitizers.

Absolute mean value of	**Sarov's**	**GSYS**	**GSYS/Sarov's**	**Sarov's/GSYS**
EN deviation,MeV	0.00637	0.00854	1.340	0.746
DATA deviation, b	1.12E-3	7.45686E-4	0.666	1.501

C.3. Relative Deviations (Sarov's digitizer)

For very low value of data the relative deviation could be very high.

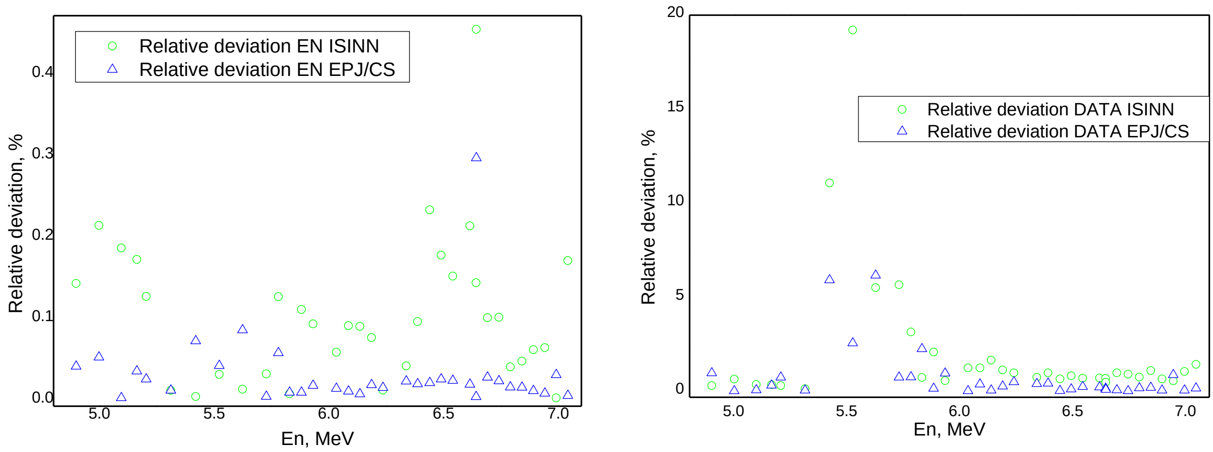


FIG. 6 (colour online). Relative deviations for EN (left) and DATA (right).

D. Proposal to produce new modern digitizer on the base of "symbol recognition"

Digitizers could be improved by "symbol recognition" method: after scales definitions, compiler selects one or more symbols (from proposed list, with or without error bars of different types) and then he/she pushes a button like "digitize points for defined symbols" and as result obtains:

- DATA block of digitized data for defined symbol (with and/or without errors) and
- plot with initial and produced digitized data – to check and be sure, that all points were included correctly.

Manual deletion of selected points and/or addition of new points has to be available for compiler as it's done now.Such digitizer could essentially decrease time spent by compiler for digitizing, also help to avoid mistakes, duplicative points, wrong errors and so on.

E. Comments/recommendations for digitizing process (as now)

Process of digitizing of data for EXFOR consists of several steps:

- producing figure for digitizing in format available in digitizing codes,
- digitizing by software,
- correction of file produced by digitizing software and producing corresponding Subent in EXFOR format.

Respectively, quality of digitized data in EXFOR depends on:

- quality of figure (quality of printing, quality of scanning, size of figure, size of symbols, overlapping of symbols for different data and so on)
- quality of software for digitization,
- art of compilation of compiler.

.

Proposals of recommendations for compilers in digitization process:

- *Always try to receive data from author.* Explain to author, that authors' data are more preferable than digitized from figures.

- Read the article attentively – some data could be found in text. As example: energy bins or energy points – for scale X – are sometimes given in text of article.
- Analyze figures attentively. As for example, on Fig.1a (received from NDS and proposed for digitization) it’s clear enough, that for scale X (Eout) the energy step is 1 MeV, and data are given for energy intervals 1-2, 2-3, 3-4,…, 9-10 MeV. Also sometimes data are multiplied by a coefficients like 10., 100. and so on, what has to be taken into account for respective data.
- At conversion from PDF into BMP-format, use as large size of figure (zoom) as possible for you, and also the same at digitizing process (this helps to do not loose accuracy)
- Special case – digitizing of smooth curves. Digitize reasonable number of points on curve.
- Compare data of initial plot with digitized data.

F. Other proposal

To communicate with INDC and ask to support requests (sent to authors by compilers) of experimental data for EXFOR data base, discuss, in which form such support could be done – mention in INDC/IAEA report, in report at Conf. “Nuclear data for science and technology”, copy of e-mail, some other way.

Expression of Digitized Values in EXFOR Entries Based on Digitization Accuracy

N. Otuka
IAEA Nuclear Data Section, Vienna International Centre, A-1400 Vienna, Austria

1. Introduction

Digitization is not avoidable for compilation of both old and new articles. In addition to accurate digitization of values and uncertainties from the data points and error bars, the compiler should process the outputs from the digitization software appropriately when the outputs are included in EXFOR entries. Usually experimentalists report values and uncertainties by the same number of digits (*e.g.*, 1.00 ± 0.03 barn, 1.000 ± 0.031 barn). The notation of uncertainties by parenthesized values (*e.g.*, 1.00(3) barn) supports this rule, and EXFOR compilers are also expected to follow the same convention. Note that the usage of the parenthesized uncertainty expression may depend on authors (*e.g.*, 1.00(3.1) or 1.00(31) for 1.00 ± 0.031 barn) and the authors are expected to describe its usage explicitly in their articles (e.g., Section 7 of [1]). This paper discusses rounding of digitized values and evaluation of digitization accuracy.

2. Rounding of digitized values

Compilers often receive digitized results with many digits (*e.g.*, 1.001238 … ± 0.031453…) from digitization tools, and they need a criterion to determine the number of digits to be kept in the EXFOR entry. Suppose a compiler digitized a data point, and received 1.001238… ± 0.031453… with "digitization accuracy" δ= 0.001012… from a digitization tool. The compiler may keep (i) 1.001 ± 0.031 barn, (ii) 1.0012 ± 0.0315 barn, (iii) 1.00124 ± 0.03145 barn *etc*, by considering δ. The first keeps the minimum number of digits. By the same option, an existing entry F0001.004 in **Fig. 1** (left) may be processed to Fig. 1 (right). Sometimes rounding may help the compiler to find out the value given by the authors. In Fig. 1 (right), we observe digitized ^{3}He incident energy values (EN) are close to integers. Actually the authors explicitly describes that "^{6}Li(^{3}He,*p*)2α reaction at bombarding energies of 3, 4, 5, and 6 MeV" [2].

```
SUBENT        F0001004   20110808
BIB                  6         10
REACTION   (3-LI-6(HE3,P)4-BE-8,PAR,SIG)
…
ERR-ANALYS (EN-ERR-DIG) digitizing error
           (ERR-DIG)    digitizing error
           (DATA-ERR)   estimating accuracy
….
ENDBIB               9
COMMON               3          3
EN-ERR-DIG DATA-ERR   ERR-DIG
MEV        PER-CENT   MB
 0.04       15.        0.3
ENDCOMMON            3
DATA                 3          8
E-LVL      EN         DATA
MEV        MEV        MB
 0.         2.943      11.63
 0.         3.985      10.22
 0.         4.967       9.395
 0.         5.979      10.63
 2.9        2.96       53.73
 2.9        3.972      53.73
 2.9        4.955      50.48
 2.9        5.97       41.03
ENDDATA             10
ENDSUBENT           28
```

```
SUBENT        F0001004   20110808
BIB                  6         10
REACTION   (3-LI-6(HE3,P)4-BE-8,PAR,SIG)
…
ERR-ANALYS (EN-ERR-DIG) digitizing error
           (ERR-DIG)    digitizing error
           (DATA-ERR)   estimating accuracy
….
ENDBIB               9
COMMON               3          3
EN-ERR-DIG DATA-ERR   ERR-DIG
MEV        PER-CENT   MB
 0.04       15.        0.3
ENDCOMMON            3
DATA                 3          8
E-LVL      EN         DATA
MEV        MEV        MB
 0.         2.94       11.6
 0.         3.99       10.2
 0.         4.97        9.4
 0.         5.98       10.6
 2.9        2.96       53.7
 2.9        3.97       53.7
 2.9        4.96       50.5
 2.9        5.97       41.0
ENDDATA             10
ENDSUBENT           28
```

FIG. 1. Example of rounding based on digitization error. Left: Before rounding. Right: After rounding.

It is known that the rounding option (i) sometimes leads to repetition of the same number in the independent variable even if the two data points are distinguishable on the figure image. A typical example is seen in excitation functions of charged-particle induced reactions in resonance regions measured by Van de Graaff accelerators. When the compiler applies rounding to values digitized from such an image, the compiler should check whether repetition of the same value of an independent variable (*e.g.*, incident energy) is not seen after rounding. It could be a good practice to keep one or two more digits (the option (ii) and (iii) above) when the compiler is aware such a trouble may happen. It is known that such a repetition is also made by *digitization of the same symbol twice*, and this mistake can be detected if we apply such checking procedure after rounding.

3. Evaluation of digitization accuracy for rounding

In order to perform rounding discussed above, the compiler need to determine the digitization accuracy δ introduced above. There could be various sources of errors, and it is not practical to estimate δ by proper error propagation from these sources. Suppose the digitization accuracy is expressed by a fixed-size rectangle on the figure image $\delta x \times \delta y$ at a point (x, y), and the x- and y-scales are linear and logarithmic respectively. If x-values $x_{1,d}$, $x_{2,d}$, ... are digitized from original x-values x_1 , x_2, ... on the linear scale, one may expect the digitization errors $|x_{1,d} - x_1| \sim |x_{2,d} - x_2| \sim \ldots \sim \delta x$, namely the error is approximated to a constant in the absolute unit (*e.g.*, MeV). If y-values $y_{1,d}$, $y_{2,d}$, ... were digitized from original y-values y_1 , y_2, ... on the logarithmic scale, another relation $|(y_{1,d} - y_1) / y_1| \sim |(y_{2,d} - y_2) / y_2| \sim \ldots \sim \delta y / y$ is expected, namely the error is approximated to a constant in relative unit (*e.g.*, %). This supports the expression of digitized values (fixed and floating decimal point numbers for values digitized from the linear and logarithmic scale, respectively).

Some data centres (*e.g.*, CAJaD, CNPD) have utilized values printed on the tics for evaluation of δ. In this method n tics on the scale x_i (i = 1,n) are digitized, and a value δ evaluated by

$$\delta = \sqrt{\sum_{i=1}^{n} \Delta^2 /(n-1)} \quad \text{with} \begin{cases} \Delta = x_{i,d} - x_i & \text{for linear scale} \\ \Delta = (x_{i,d} - x_i)/x_i & \text{for log scale} \end{cases}$$

is compiled as the digitization accuracy (digitization error) under the heading `EN-ERR-DIG`, `ERR-DIG` etc. The digitized results submitted by the meeting participants show that δ obtained by above prescription does not always explain the real deviation of the digitized value from the original value. Nevertheless it could be still utilized as a currently available criterion for rounding of digitized values.

It would be noteworthy to propagate δ (*i.e.*, the accuracy of the value digitized from a symbol) to the accuracy of the error bar size determined by taking difference between two digitized points (ε) assuming that the error propagation raw for standard deviations is valid for propagation from δ to ε. Suppose we want to digitize a symbol with an error bar for which the original value is σ ± Δσ (mb). In the linear scale, both the location of the symbol y=σ (mb) and the ends of upper / lower error bars y=σ ± Δσ (mb) are determined with the same accuracy δ (mb), and therefore the absolute accuracy of the error bar size ε (mb) is determined by

$$\varepsilon = \sqrt{\delta^2 + \delta^2} = \sqrt{2}\delta ,$$

namely ε (mb) is expressed by a constant of the data set in the absolute unit (mb), and it is slightly larger than δ. For the logarithmic scale case, both the location of the symbol y = σ (mb) and the end position of upper / lower error bars y = σ ± Δσ (mb) are determined with the same accuracy δ (%). In the absolute unit, they are σ · δ/100 (mb) and (σ ± Δσ)·δ/100 (mb), respectively. Therefore the relative digitization accuracy of the error bars ε (%) is

$$\varepsilon = \sqrt{(\sigma\delta)^2 + (\sigma \pm \Delta\sigma)^2 \delta^2} \,/\, \Delta\sigma = \delta \sqrt{1 \pm 2\left(\frac{\sigma}{\Delta\sigma}\right) + 2\left(\frac{\sigma}{\Delta\sigma}\right)^2} ,$$

namely the ε(%) is no longer a constant of the data set, and it depends on the experimental relative uncertainty of each data point (Δσ/σ). When digitization accuracy δ ~ 1% and experimental uncertainty Δσ/σ ~ 10%, the accuracy of the digitized error bar ε ~ 15%. This result also implies that the number of digits kept for the uncertainties digitized from logarithmic scale cannot be determined by the constant digitization accuracy δ%. It should be determined so that it is consistent with the number of digits of digitized values. Example: Suppose a digitization tool outputted 100.1234…. ± 1.2345… for a data point with error bar on the logarithmic scale with the digitization accuracy δ ~ 1%. If the compiler decided to keep 3 digits after the decimal point in the floating decimal point number expression, the digitized value and uncertainty must be recorded in the EXFOR entry as `1.001E+02±0.012E+02` rather than `1.001E+02±1.235E+00`. This is a natural selection if we recall that the value and uncertainty is expressed as $(1.001 \pm 0.012) \times 10^2$ in scientific literature.

References

[1] A.L. Nichols, Radioactive decay data: powerful aids in medical diagnosis and therapy, analytical science and other applications, Radiochim. Acta **100** (2012) 615.

[2] C.R. Gould *et al.*, “Cross section measurements for charged particle fusion reactors: the ^{6}Li(^{3}He,p)2α reaction”, Nucl. Sci. Eng. **60** (1976) 477.

Benchmarking of Digitization Software (Memo CP-D/761): Results and Discussions

V. Semkova

IAEA Nuclear Data Section, Vienna International Centre, A-1400 Vienna, Austria

1. Background

A considerable part of the numerical data included in the EXFOR database has been obtained by digitization. Compilers use different software programs for digitizing numerical data from graphs. There have been developments in digitization software since the last benchmark in 2005. A benchmark exercise was proposed by Memo CP-D/761 distributed on 18 of September 2012. The aim of the benchmark is to check quality and consistency of the digitized data and propose recommendations to compilers as well as to software developers in order to improve accuracy of the digitization.

2. Method

The Benchmark includes:

1. Distribution of graphs for digitization. **Figure I** consists of 6 graphs in linear-logarithmic scale from Fig. 7 of Phys. Rev. C 83, 034604 (2011) [1], EXFOR subentry 14290002. **Figure II** and **III**, both in linear-linear scales, correspond to Fig 3 and Fig. 4 from Nucl. Instrum. Methods. Phys. Res. B 269 (2011) 2032 [2], EXFOR subentry D0655005 (angles 150 and 115 degrees respectively)
2. Digitization of data on graphs by the participants
3. Submission of digitized data to NDS in the form of EXFOR entry
4. Comparison of digitized data and uncertainties with authors' data (tables and plots)
5. Evaluation of the "accuracy" of digitization
6. Distribution of recommendations to compilers and software developers.

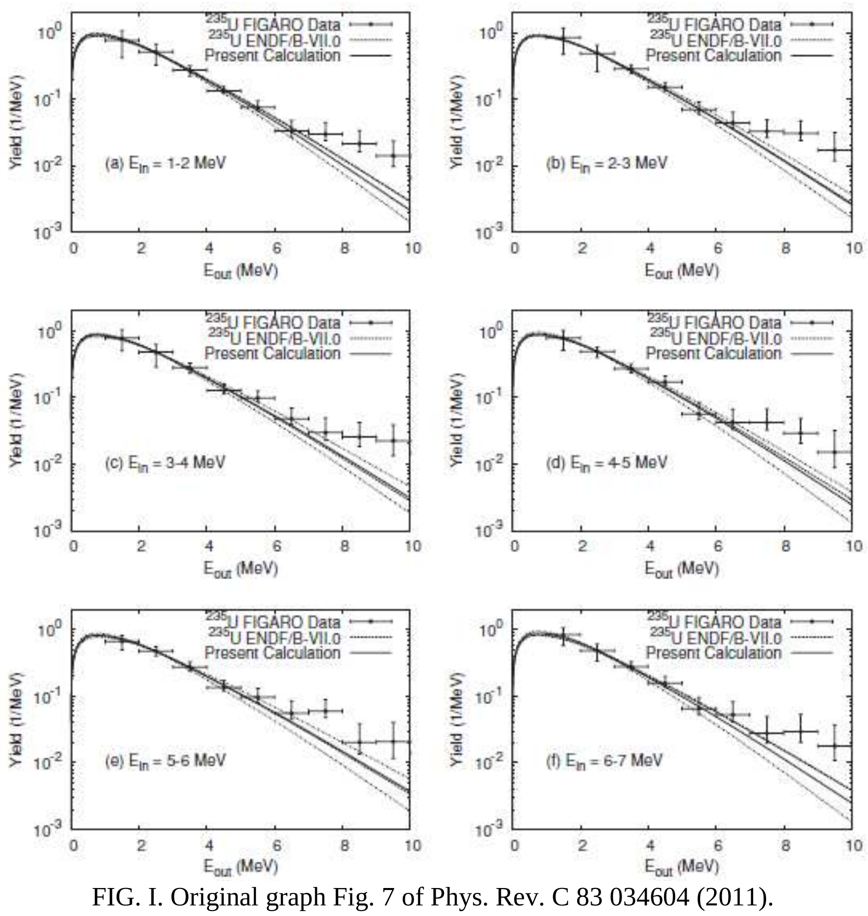


FIG. I. Original graph Fig. 7 of Phys. Rev. C 83 034604 (2011).

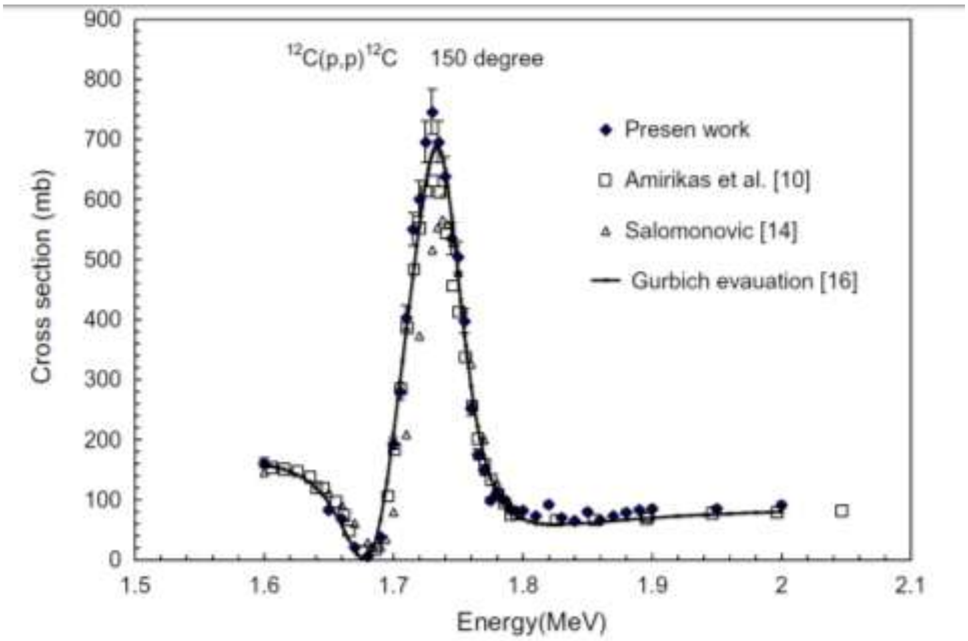


FIG. II. Original graph Fig 3 of Nucl. Instrum. Methods Phys. Res. B 269 (2011) 2032.

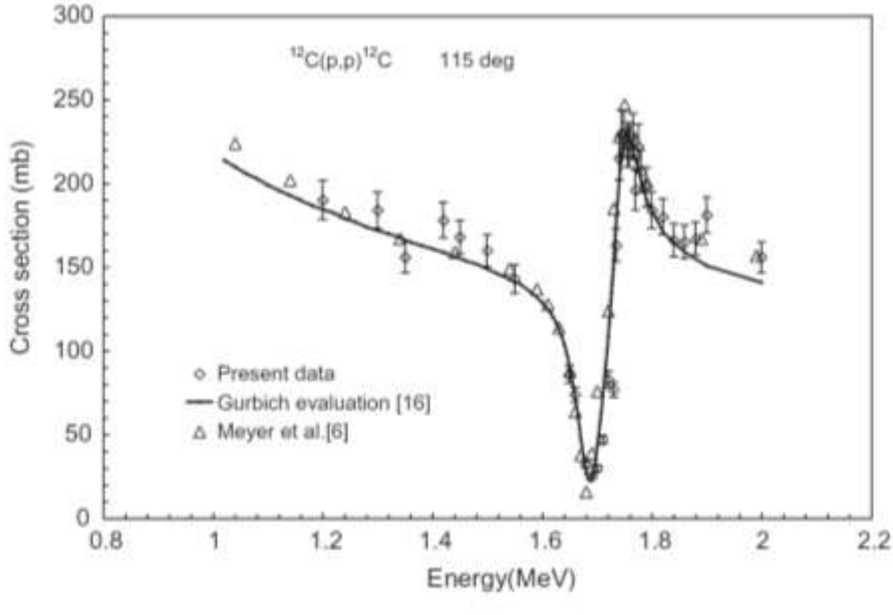


FIG. III. Original graph Fig. 4 of Nucl. Instrum. Methods Phys. Res. B 269 (2011) 2032.

3. Results and discussion

Digitized data in a form of EXFOR Entries were received from 14 participants.

Summary of centres/compilers and software used for digitization are included in **Tables I** and **II** as well as the formats chosen by compilers to present numerical data in the submitted files.

The precision of the Y and ΔY values for Fig. I in all cases exceeds the accuracy of the measurement (for most of the data points above 10%).

TABLE I. Summary of centres/compilers, software used for digitization, and format of the digitized data for Fig.I.

Centre/Compiler	Digitizer	Format of the digitized data*			
		X **(MeV)**	**Δ*X*** **(MeV)**	***Y*** **(PT/FIS/MEV)**	**Δ*Y*** **(PT/FIS/MEV)**
Author subent.14290002		1 decimal	1 decimal	4 digits	4 digits
CNDC	GDgraph v.5.0	1 decimal	1 decimal	3 digits	3 digits
T. Ashizawa	GSYS v.2.3.21	4 digits	4 digits	4 digits	4 digits
L. Bawitlung	GSYS v.2.4	1 decimal	1 decimal	3 digits	3 digits
A. Makinaga	GSYS v.2.4.0	4 digits	4 digits	4 digits	4 digits
M.Mikhaylyukova	InpGraph	1 decimal	1 decimal	5 digits	5 digits
N. Otsuka	GSYS v 2.4.2.	1 decimal	1 decimal	4 digits	4 digits
R. Mandal	GSYS	3 decimal	3 decimal	4 digits	4 digits
R. Ghosh	GSYS v.2.4.2	3 decimal	3 decimal	3 digits	3 digits
V. Semkova	GSYS v.2.4.0	2 decimal	2 decimal	3 digits	3 digits
R. Suzuki	GSYS 2.5.5	3 decimal	3 decimal	4 digits	4 digits
S. Badwar	GSYS v.2.4.1	2 decimal	2 decimal	3 digits	3 digits
S. Takacs	GSYS v.2.4.0	1 decimal	1 decimal	3 digits	3 digits
S. Taova	InpGraph	4 decimal	5 decimal	6 digits	6 digits

* Example of data format: "1.23E+45" 3 digits, "12.345": 3 decimal.

Regarding Fig. II and Fig. III there are cases were X and ΔX are presented with less precision than the author's data. This causes repetition of same values of the independent variable despite the fact that the points are not overlapping, but rather the X values of some points are close due to the steep change of the value of the variable.

Table II. Summary of centres/compilers, software used for digitization, and format of the digitized data for Fig. II and Fig. III.

Centre	Digitizer	Format of the digitized data*		
		***X*(MeV)**	***Y*(mb)**	**Δ*Y*(mb)**
Author 5 subentry D0655005		3 decimal	0 decimal	6 %
CNDC	GDgraph v.5.0	3 decimal	1 decimal	1 decimal
T. Ashizawa	GSYS v.2.3.21	4 digits	4 digits	4 digits
L. Bawitlung	GSYS v.2.4	3 decimal	2 decimal	2 decimal
L. Vrapcenjak	GSYS v.2.4.2	5 decimal	5 decimal	5 decimal
A. Makinaga	GSYS v.2.4.0	4 digits	4 digits	4 digits
M.Mikhaylyukova	InpGraph	4 decimal	3 decimal	3 decimal
N. Otsuka	GSYS v.2.4.2.	3 decimal	1 decimal	1 decimal
R. Mandal	GSYS	3 decimal	3 decimal	3 decimal
R. Ghosh	GSYS v.2.4.2	2 decimal	3 digits	3 digits
V. Semkova	GSYS v.2.4.0	3 decimal	1 decimal	1 digits
R. Suzuki	GSYS v.2.5.5	3 decimal	3 decimal	3 decimal
S. Badwar	GSYS v.2.4.1	2 decimal	3 digits	3 digits
S. Takacs	GSYS2 v.4.0	2 decimal	1 decimal	1 decimal
S. Taova	InpGraph	4 decimal	2/3 decimal	3/4 decimal

* Example of data format: "1.23E+45" 3 digits, "12.345": 3 decimal.

The precision of the digitized data should reflect the resolution of the measurement (uncertainty of the independent variable X) and uncertainty of the measured quantity Y.

The digitised data received from the participants were plotted together with the authors' data on **Fig 1**, **5** and **6** for Figure I, Figure II and Figure III respectively.

The accuracy of the digitized data were also evaluated by the parameter (C/T)-1, where C is the digitized value, T is the value from authors.

3.1 Evaluation of digitization accuracy for Fig. I.

The results for the Y as well as for ΔY values for Fig. I are shown on **Figs. 2-4**.

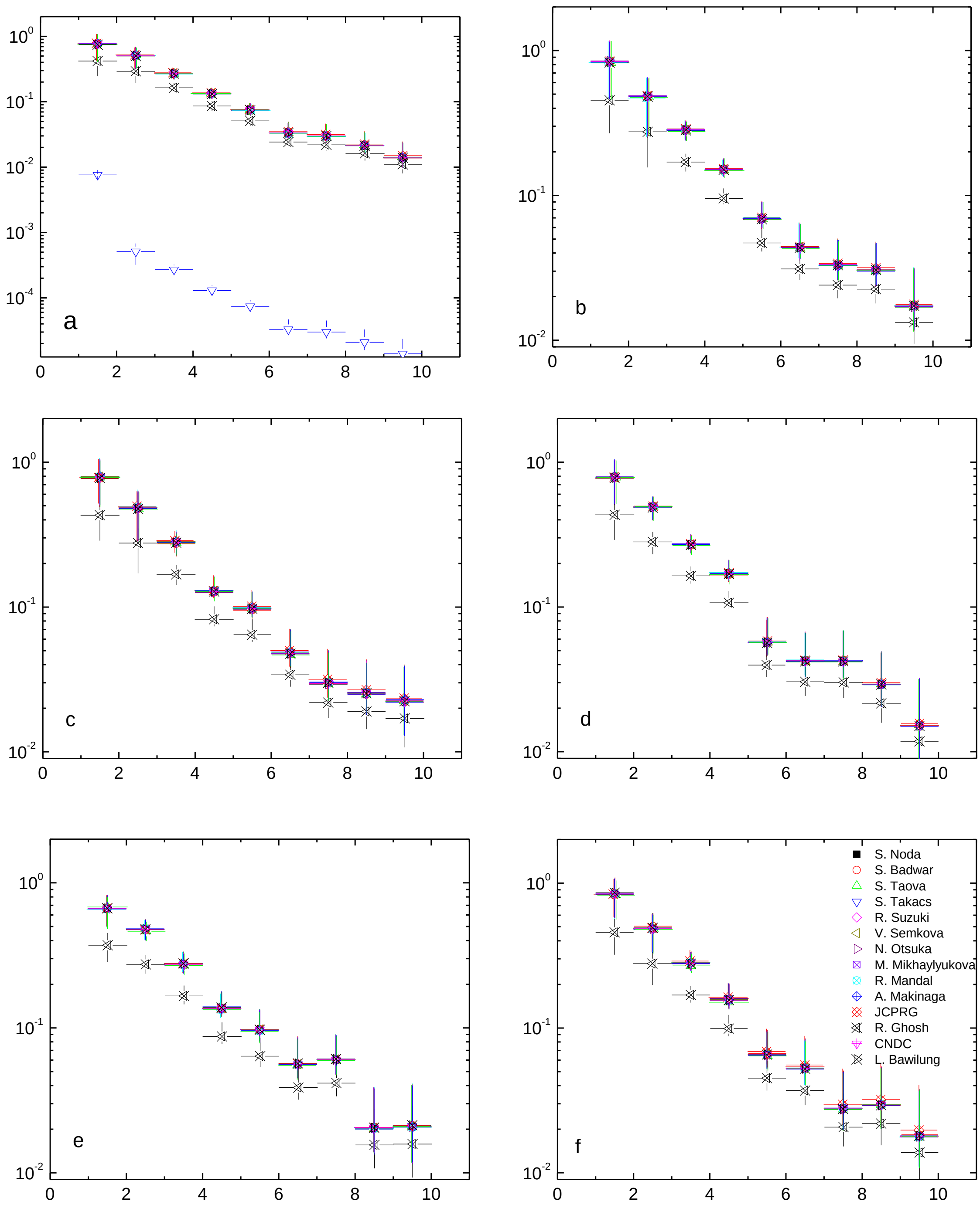


FIG. 1 (colour online). Comparison between digitized and authors' (S. Noda) data for Fig. I.

Very good agreement 1% and less for the digitized Y values have been achieved by using different digitizers, namely: CNDC with GDgraph v.5.0, R. Suzuki with GSYS v.2.5.5, and M. Mikhaylyukova with InpGraph. In other cases deviations of 2-3 % and in some cases up to 5% exist. It was confirmed after the meeting that the discrepancies between digitized data received from R. Ghosh and authors' data are due to mistakes in settings of y-axes. Good knowledge of the software, accumulation of experience, and attentiveness are recommended in order to ensure precise digitization results.

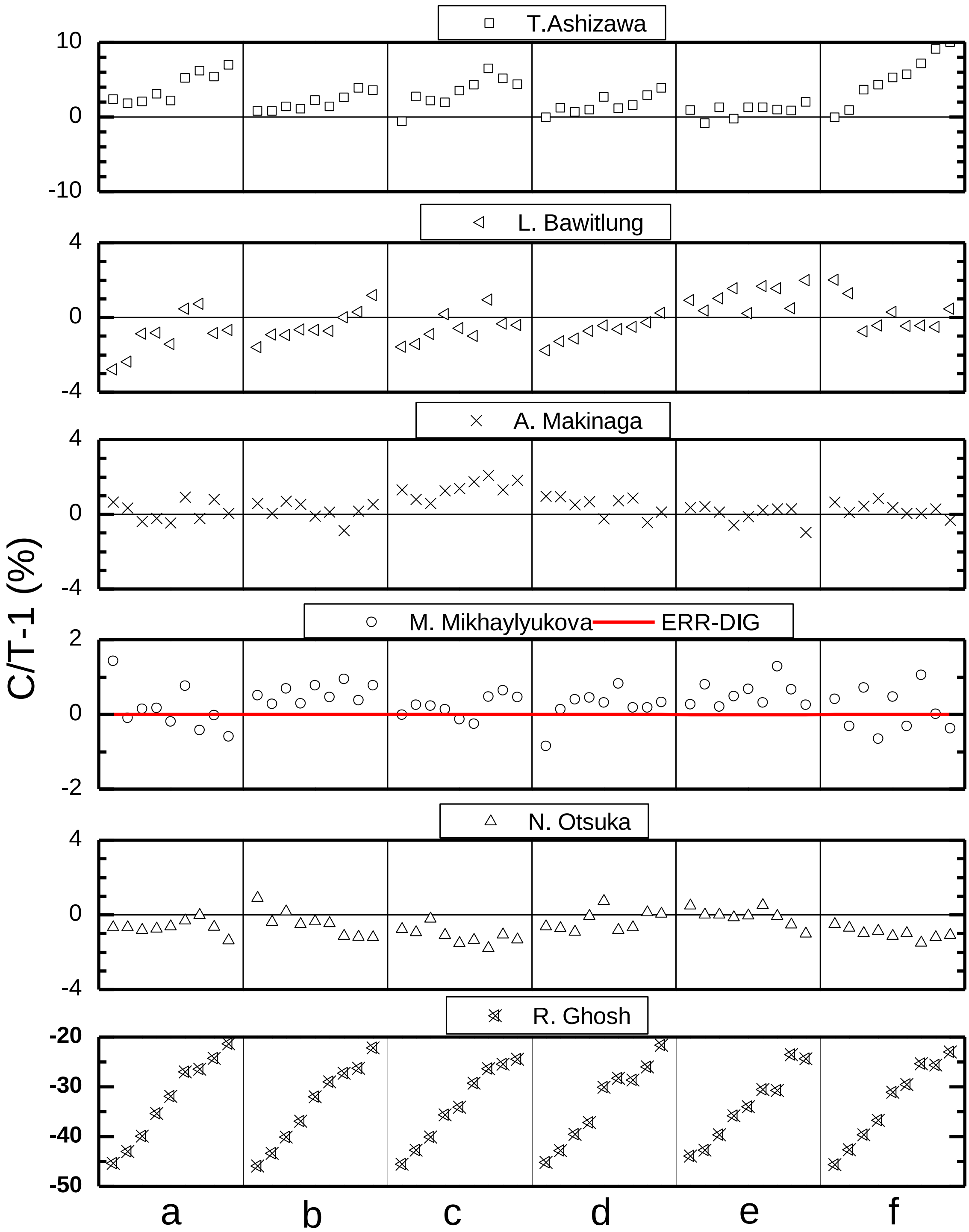


FIG. 2 part I (colour online). Accuracy of digitization of Y values for Fig. I a), b), c), d), e), and f).

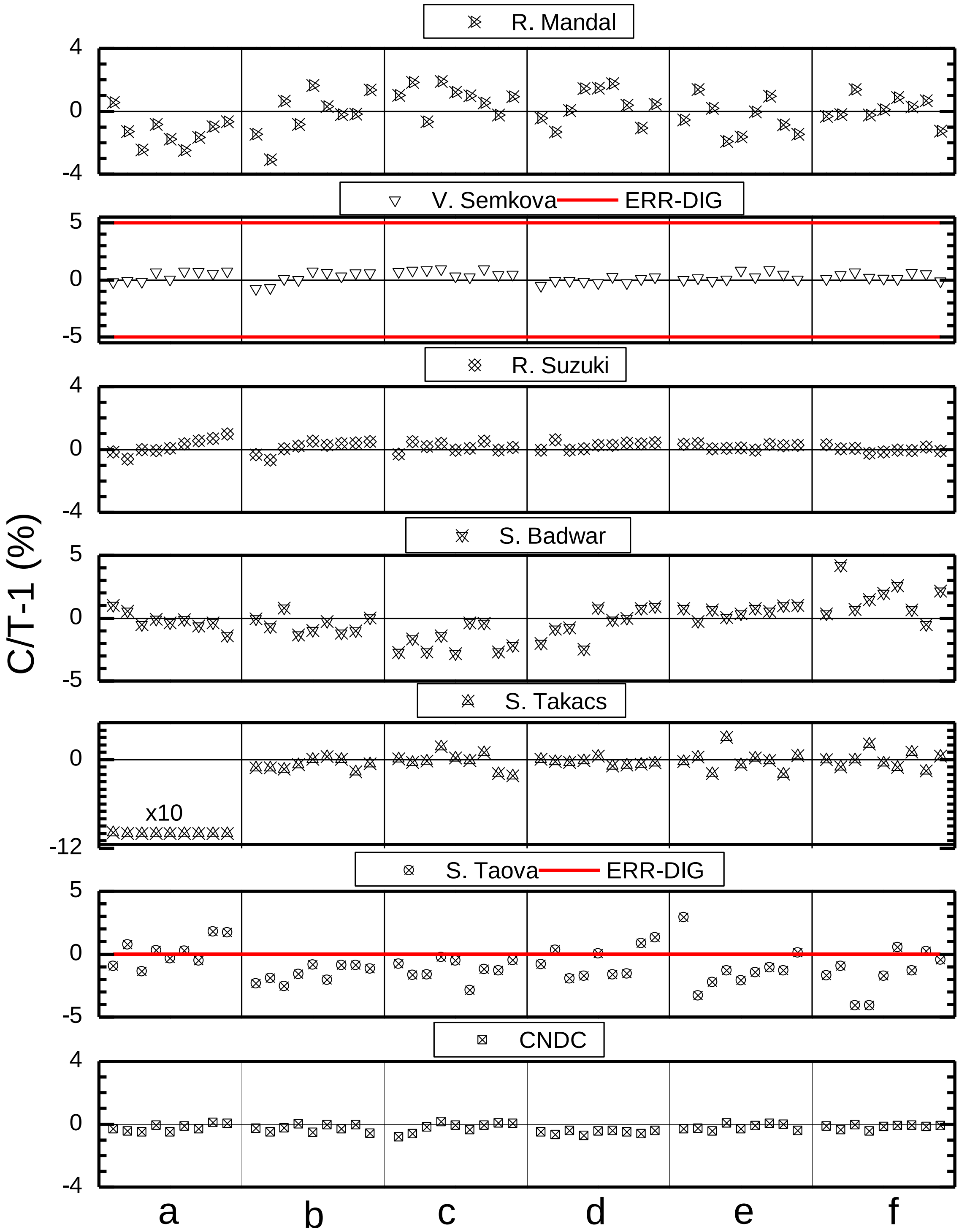


FIG. 2 part II (colour online). Accuracy of digitization of Y values for Fig. I a), b), c), d), e), and f).

Deviations of the digitized values both for positive and negative error bars from the authors' uncertainties are more prominent. Referring to the comparative analysis of digitized uncertainties of the data values and corresponding error bars presented in the contribution of N. Otuka [3] we may consider that deviation of 10 to 15 % for the error bars is reasonable.

Very few of the submitted files contain estimation of the digitization uncertainties. Only InpGraph among the used digitizing codes provides such information shown in red curve, however there are estimates unreasonably small in Figs. 3 and 4.

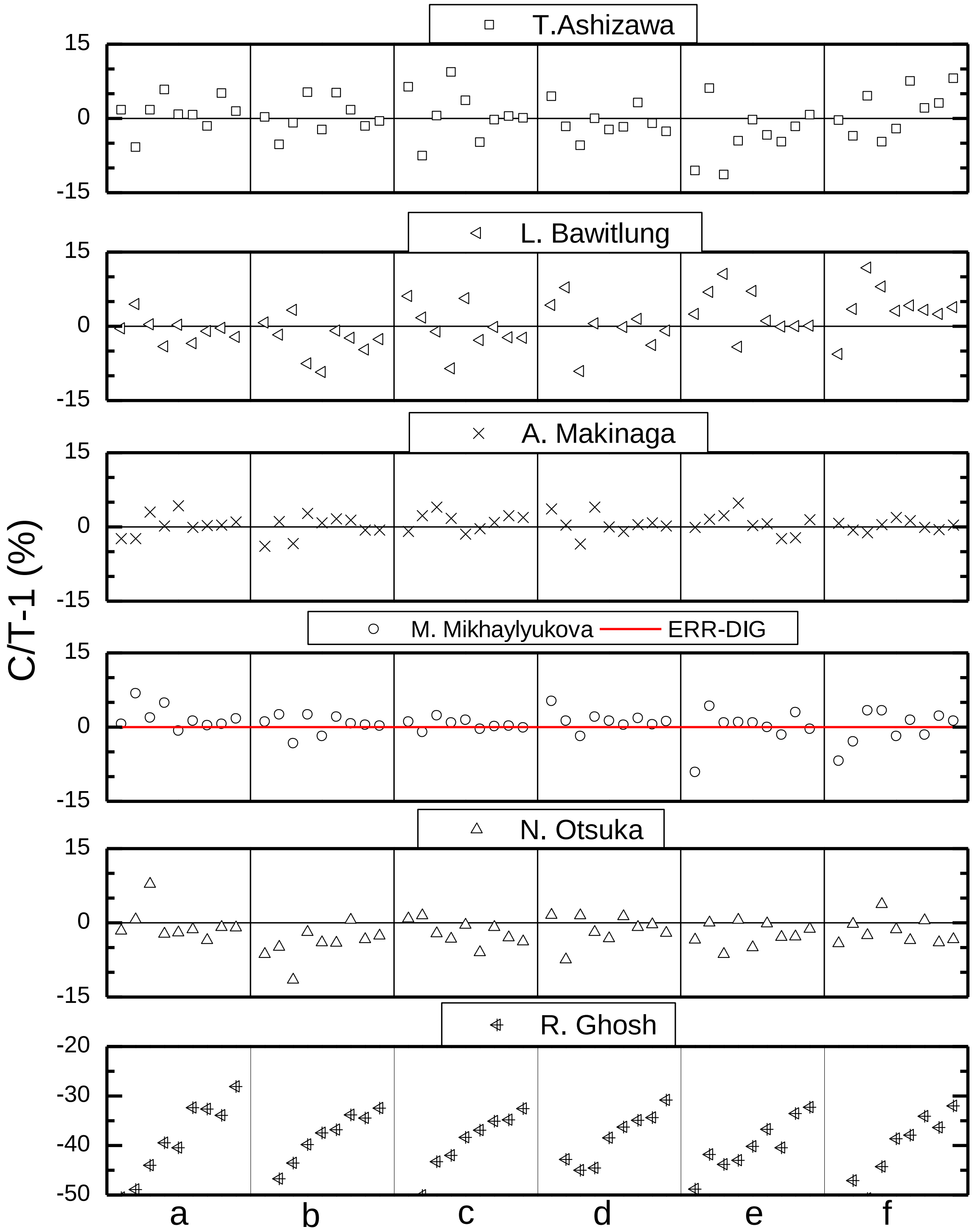


FIG. 3 part I (colour online). Accuracy of digitization of positive ΔY values for Fig. I a), b), c), d), e) and f).

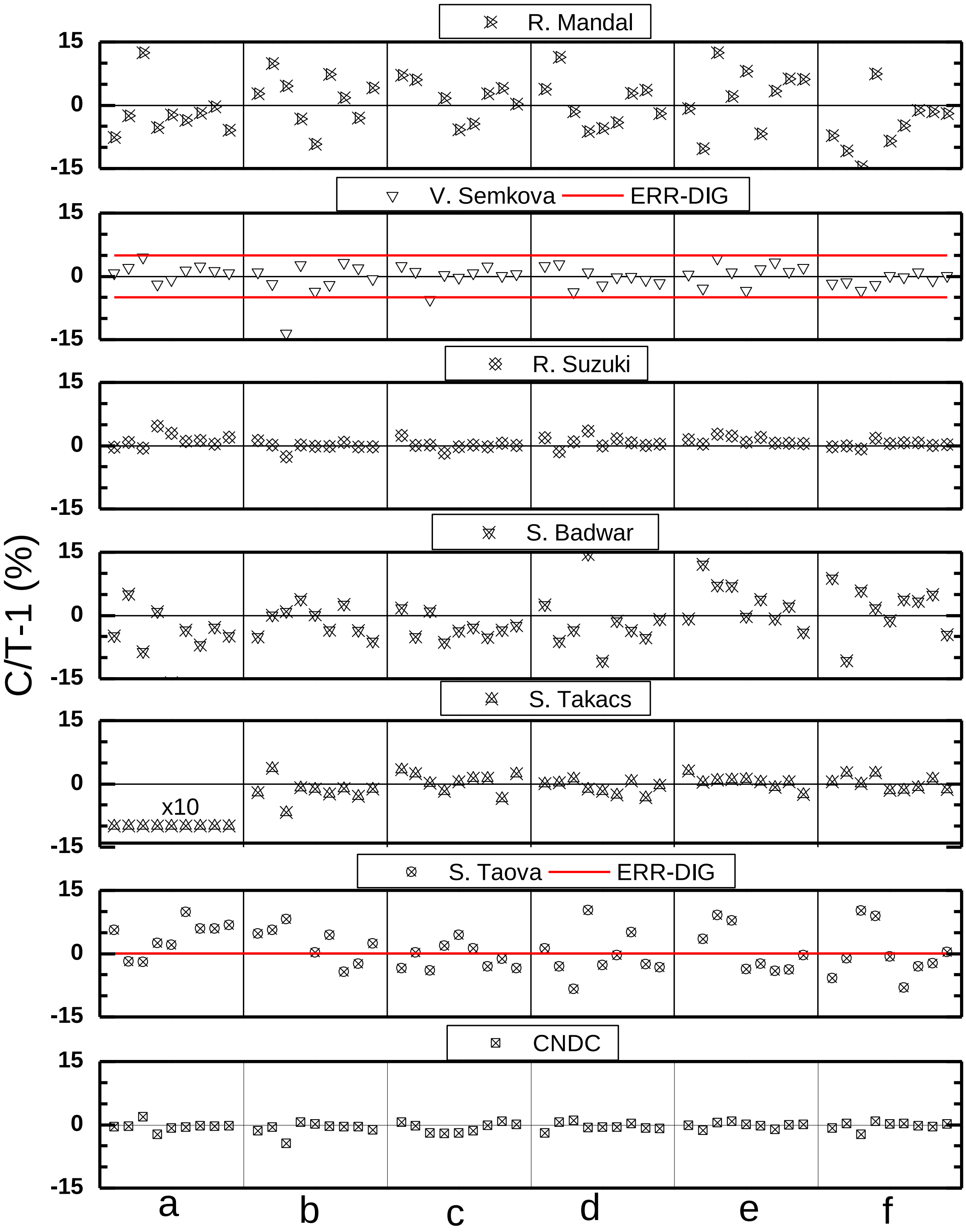


FIG. 3 part II (colour online). Accuracy of digitization of positive ΔY values for Figure I a), b), c), d), e), and f).

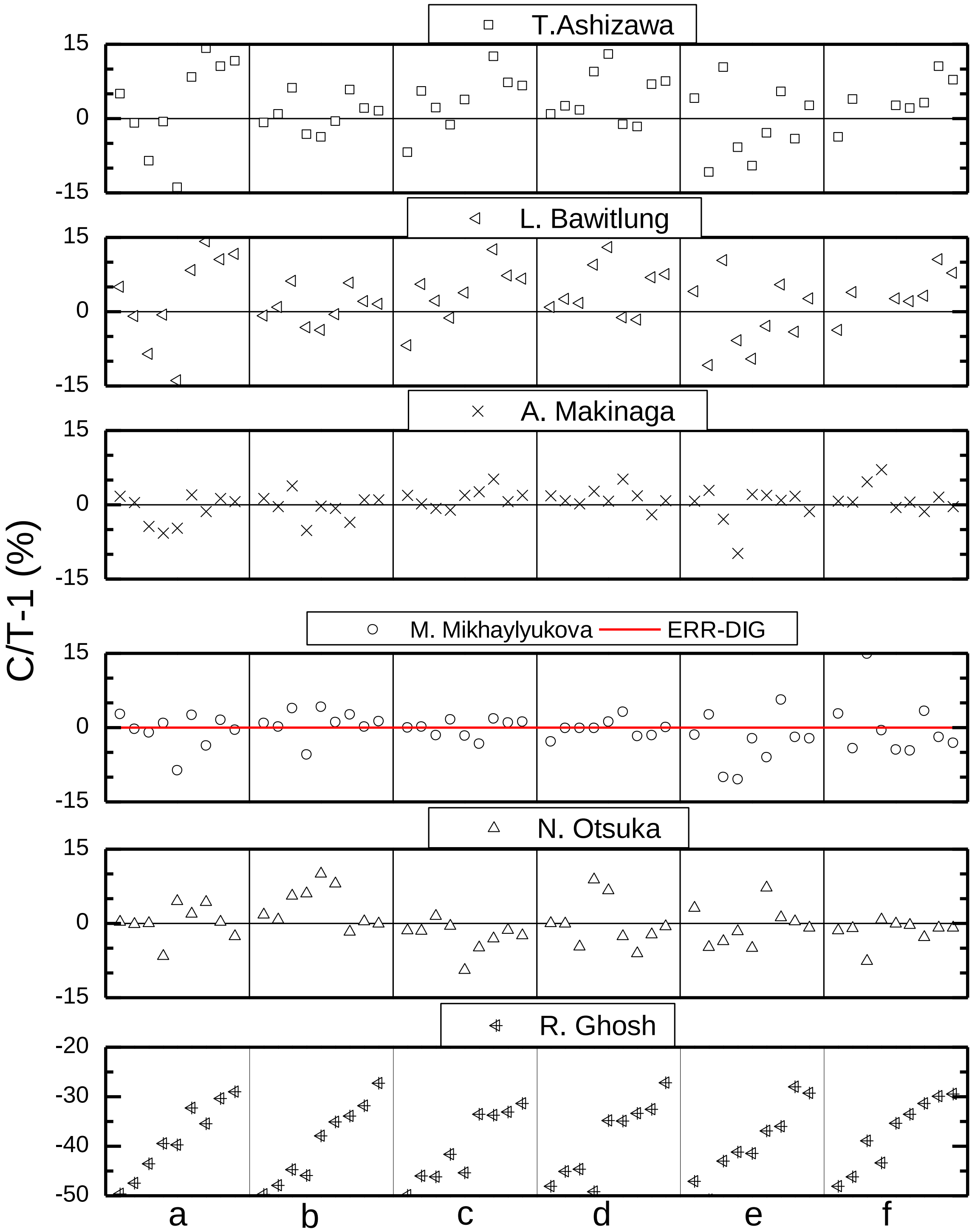


FIG. 4 part I (colour online). Accuracy of digitization of negative ΔY values for Figure I a), b), c), d), e), and f).

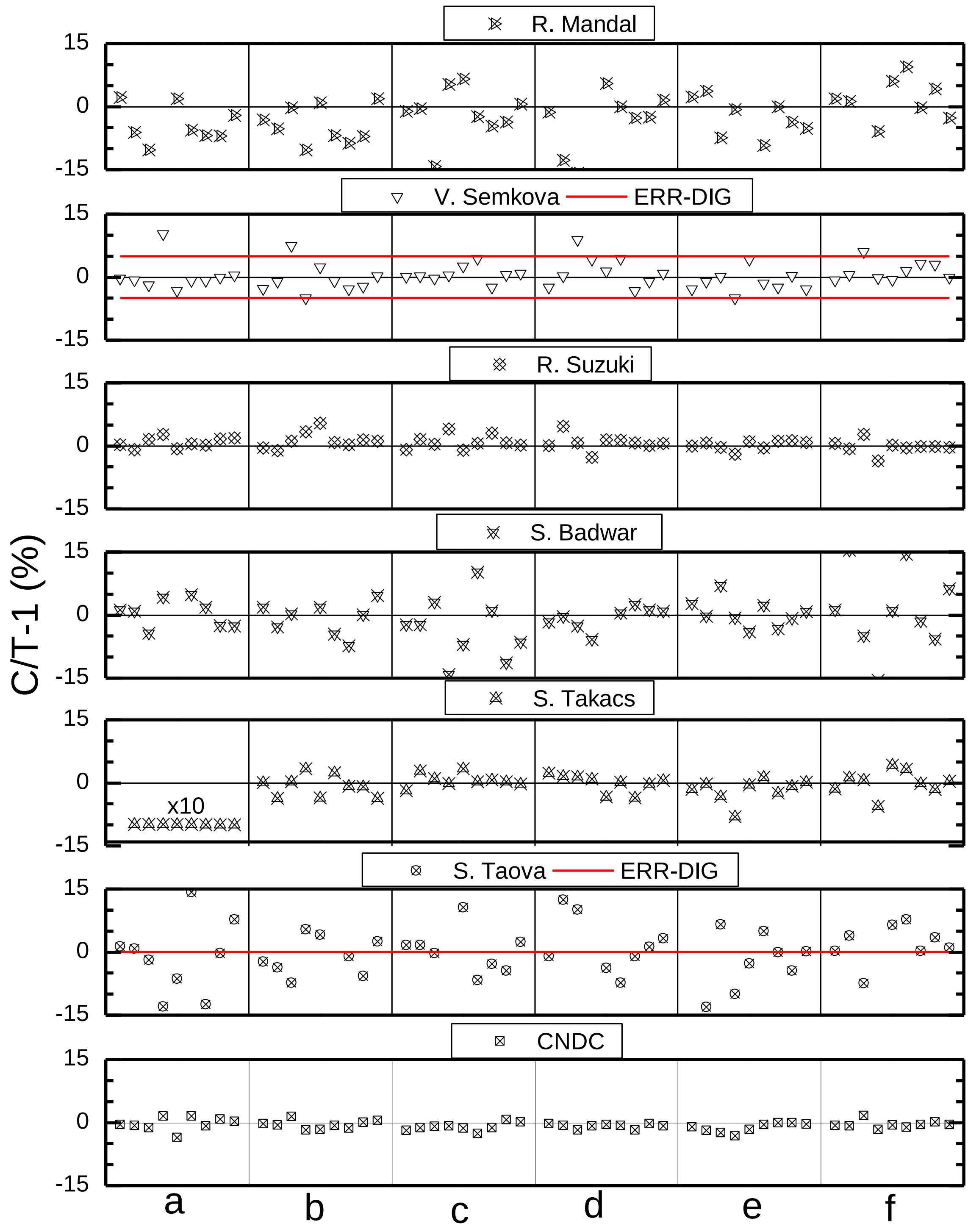


FIG. 4 part II (colour online). Accuracy of digitization of negative ΔY values for Figure I a), b), c), d), e), and f).

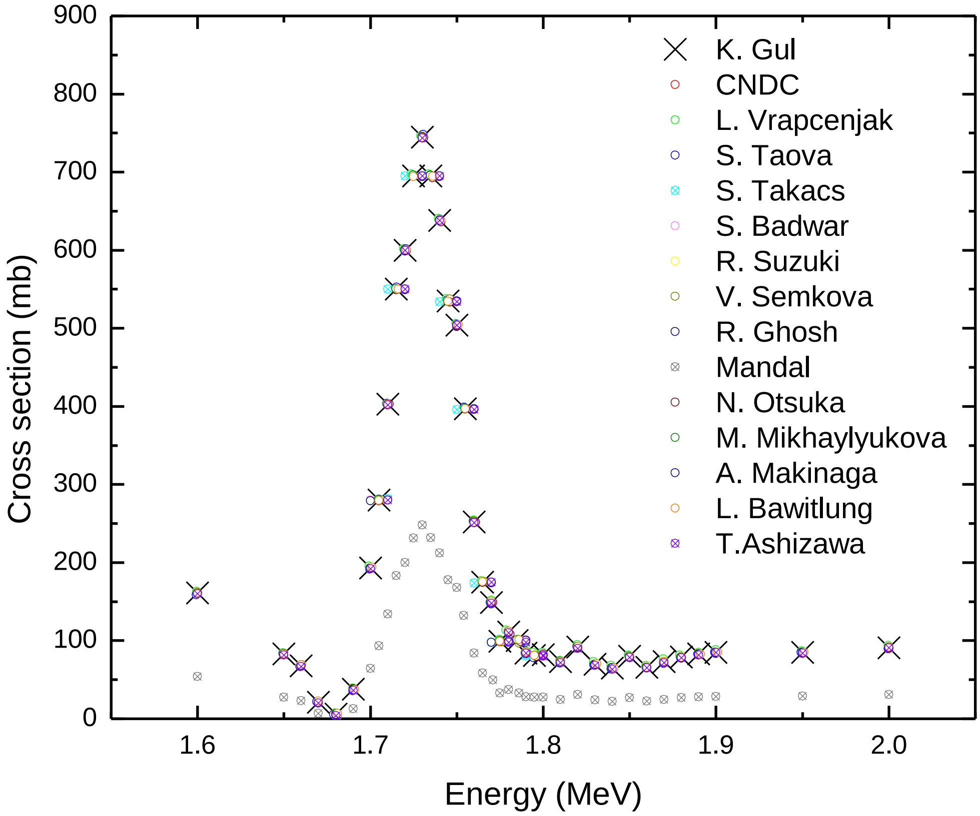


FIG. 5 (colour online). Comparison between digitized and authors' (K. Gul) data for Fig. II.

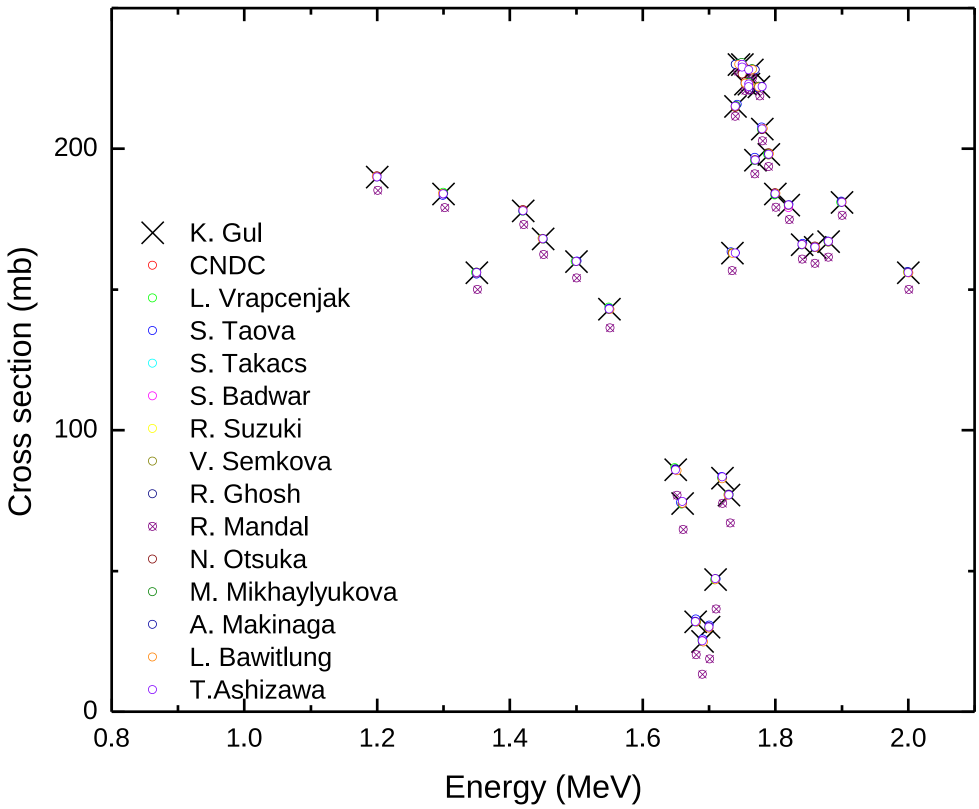


FIG. 6 (colour online). Comparison between digitized and authors' (Gul) data for Fig. III.

3.2 Evaluation of digitization accuracy for Figs. II and III.

The accuracy of the digitization of the Y values for Fig. II and Fig. III (linear-linear scale) is shown on **Fig. 7**. For most of the data points the parameter C/T-1 is less that 1% except for the 2-3 points in the 1.67-1.69 energy range were the cross sections are relatively small and consequently relative uncertainties of digitization become larger as we expect in the digitization of linear-linear scale.

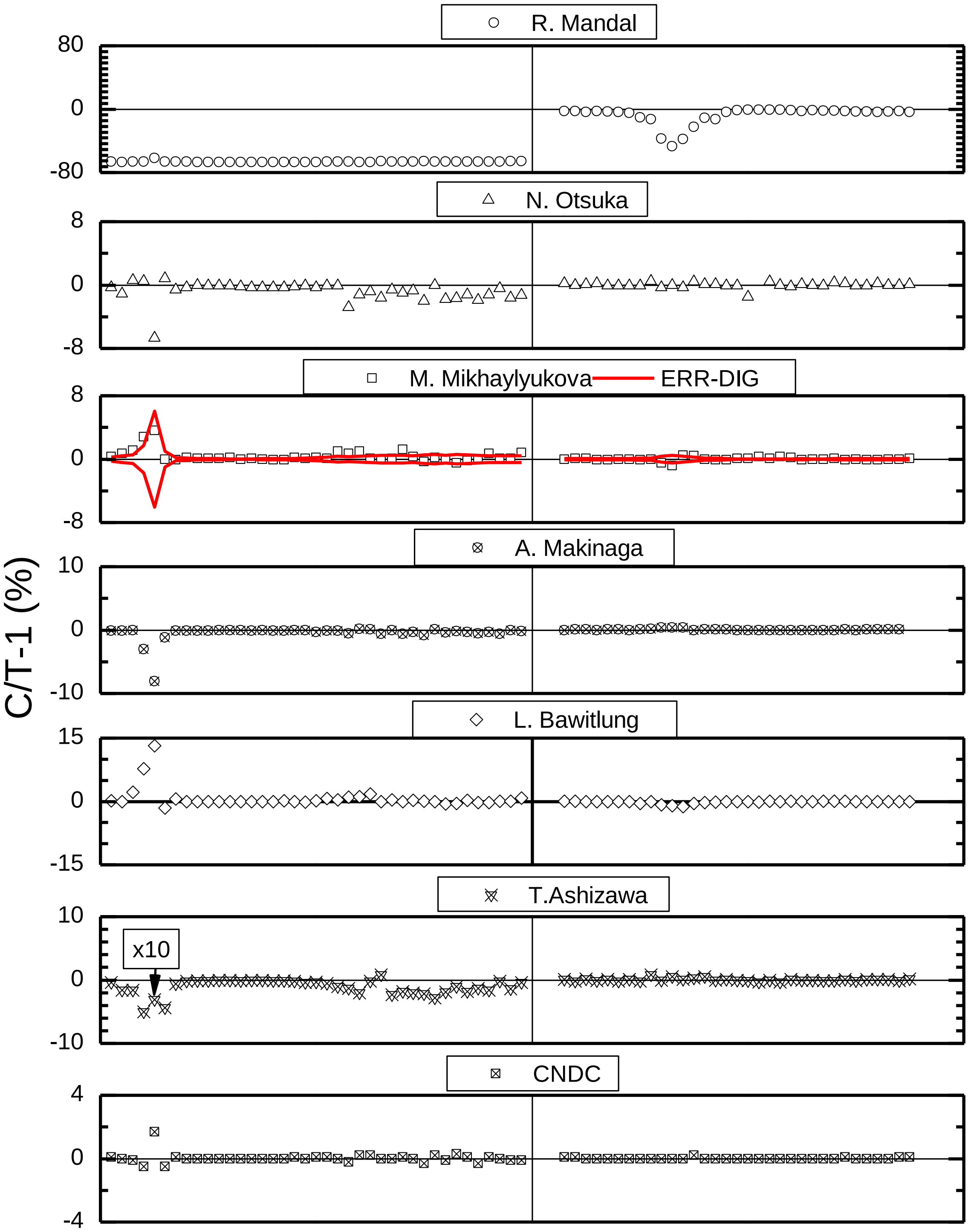


FIG. 7 part I (colour online). Accuracy of digitization of Y values linear-linear scale. Fig. II left part and Fig. III right part.

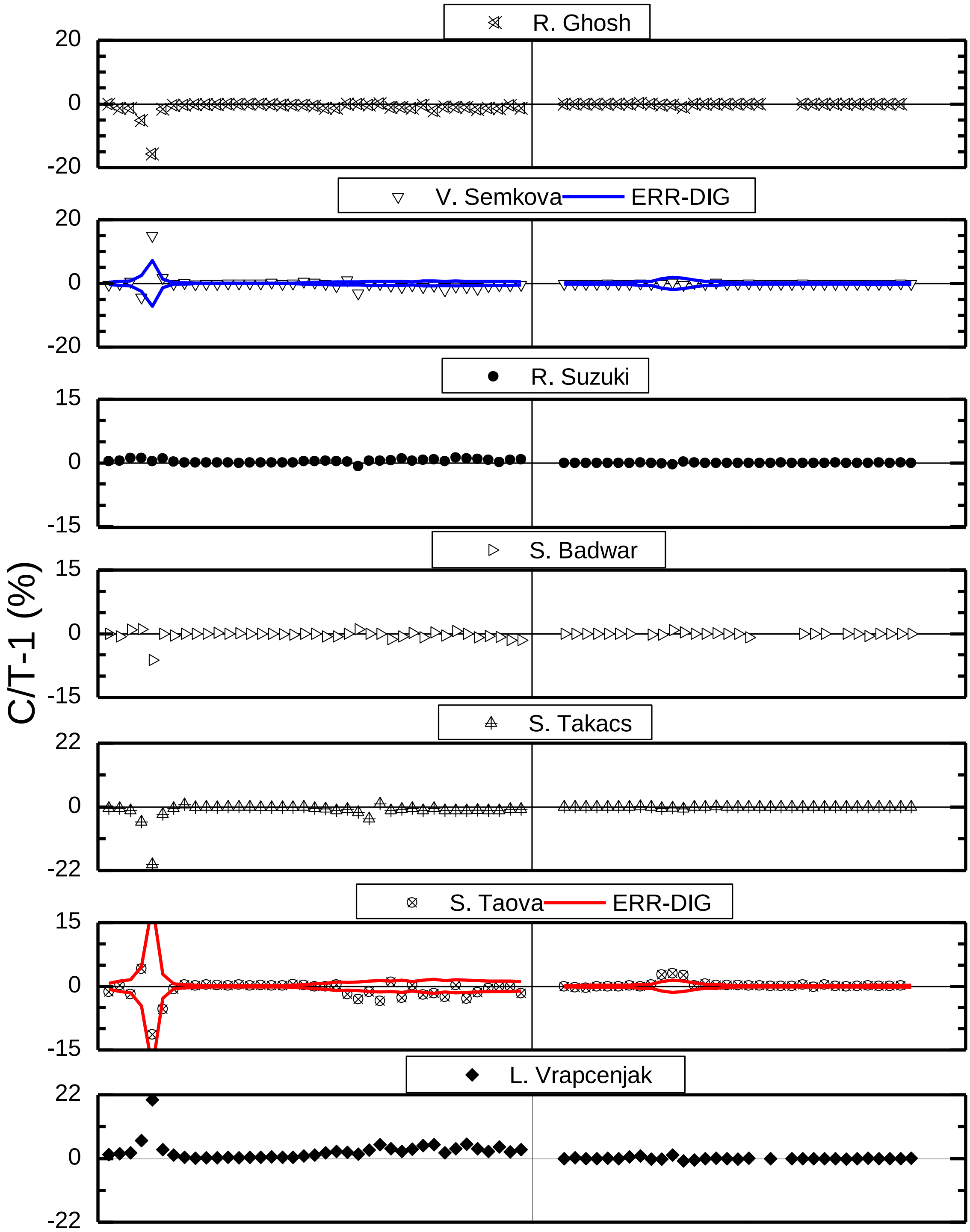


FIG. 7 part II (colour online). Accuracy of digitization of Y values linear-linear scale. Fig. II left part and Fig. III right part.

Uncertainties in the digitized ΔY values for Fig. III in many cases are quite satisfactory (less that 5%). However, for Fig. II the accuracy is in order of 20 % or higher. For some of the data points on Fig II the error bars are comparable with the size of the symbols. In some cases attempts to provide an estimation of the data uncertainties were made, however the differences with the authors' data have considerably increased. At alternative solution by FLAG and free text comment in the BIB is given in the entry submitted by M. Mikhaylyukova.

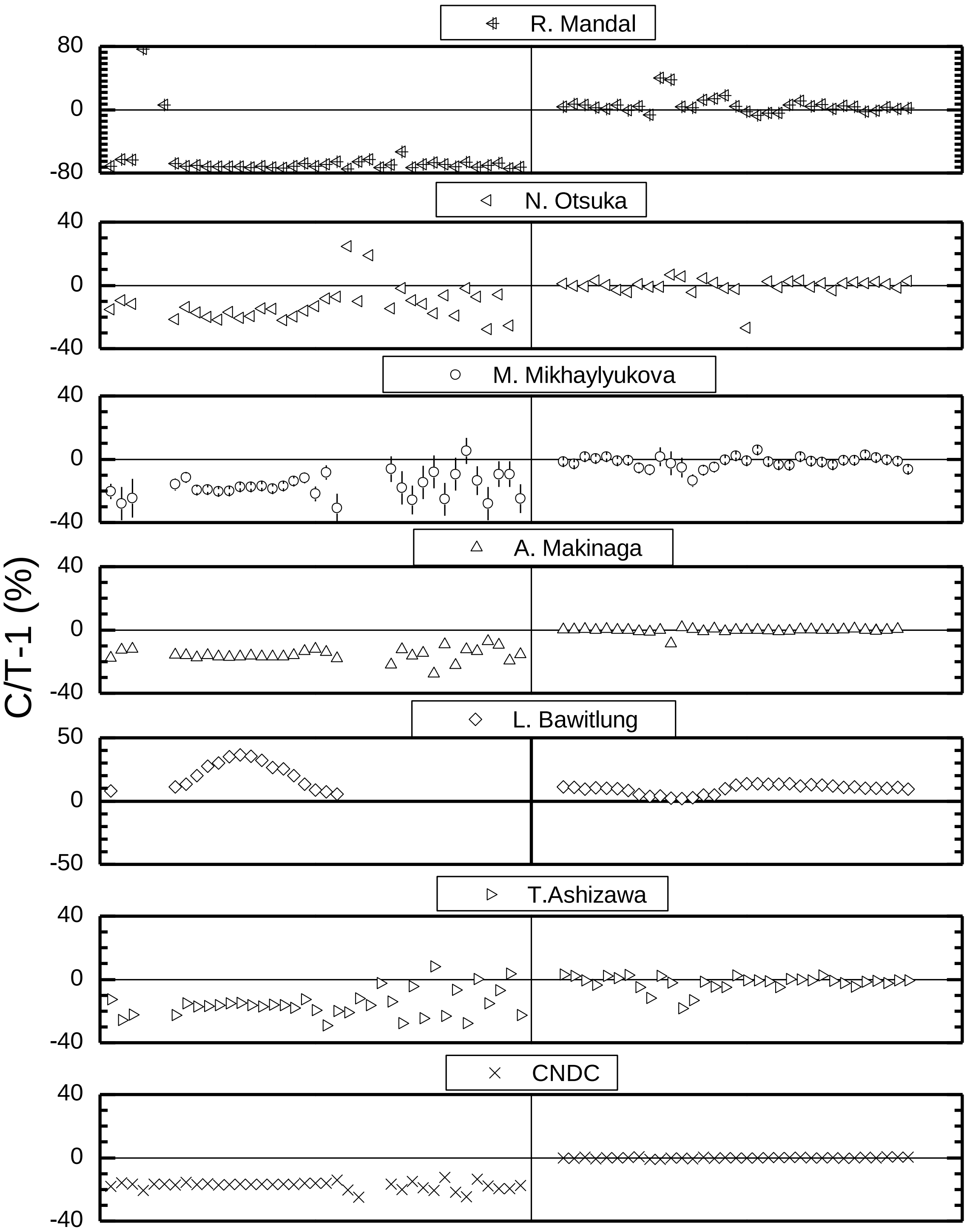


FIG. 8 part I. Accuracy of digitization of ΔY values linear-linear scale. Fig. II left part and Fig. III right part.

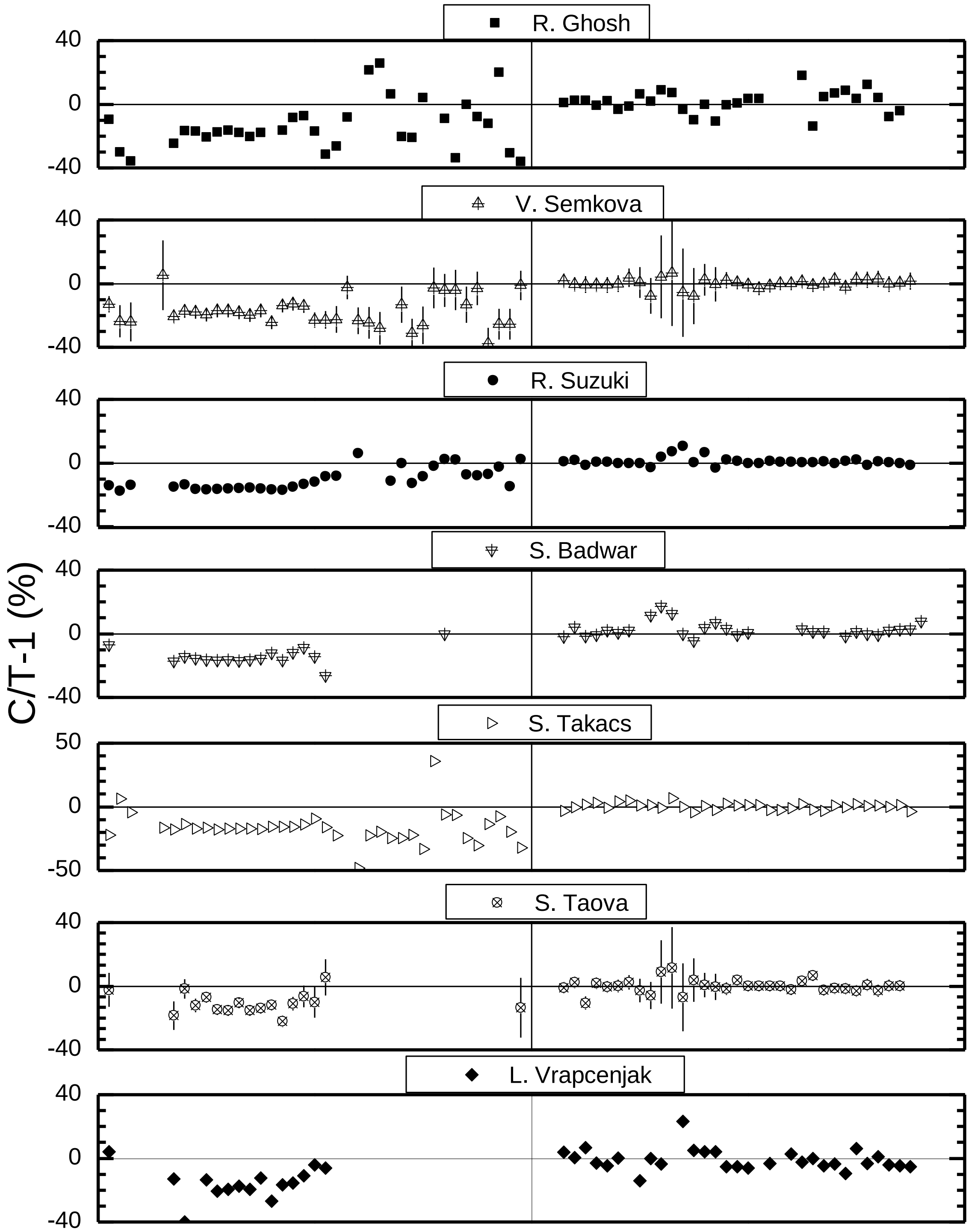


FIG. 8 part II. Accuracy of digitization of ΔY values linear-linear scale. Fig. II left part and Fig. III right part.

4. Conclusions

The Benchmark of Digitizing software provided a basis for testing quality and consistency of the digitization procedure for different types of graphs. The results show good overall agreement between digitized and authors' data. The programs used in the Benchmark work with comparable accuracy since for each type of software there are sets of data where accuracy below 1% for the digitized X and Y values have been achieved. The spread of the C/T-1 values for different datasets digitized with the same software shows the contribution of the "human factor" to the final results. In the case of GSYS, chosen by most of the compilers, uncertainties in order of 1%, 2-3%, 4%, and in an extreme case up to 50% for the digitized Y values can be observed. Good knowledge of the software, accumulation of experience, and attentiveness are recommended in order to ensure precise digitization results. Further software development and implementation of the users' feedback will provide better performance. Test examples and list with recommendations for compilers' training will be provided by IAEA-NDS.

References

[1] S. Noda et al., "Prompt fission neutron spectra from fission induced by 1 to 8 MeV neutrons on ^{235}U and ^{239}Pu using the double time-of-flight technique", Phys. Rev. C **83**, 034604, 2011.

[2] K. Gul et al., "Experimental study of proton scattering on carbon", Nucl. Instrum. Methods. Phys. Res. B **269** (2011) 2032.

[3] N. Otuka, "Expression of Digitized Values in EXFOR Entries Based on Digitization Accuracy", Appendix D in this report.

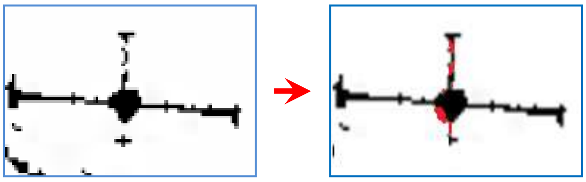
Some Requirements for Digitizing Software and Using Advanced Plotting for Checking Results

V. Zerkin

IAEA Nuclear Data Section, Vienna International Centre, A-1400 Vienna, Austria

All three digitizers [1-3] presented on the Workshop presented show their essential progress. The most promising direction of the programs' development is further automation of operations. In particular, using automatic adjustment (focusing), i.e. search of the optimal location of the elements (images of symbol and scales) demonstrated by GSYS can allow reducing influence of human factor and finally to reach more stable results of digitization. There are still a lot of possibilities for improvements:

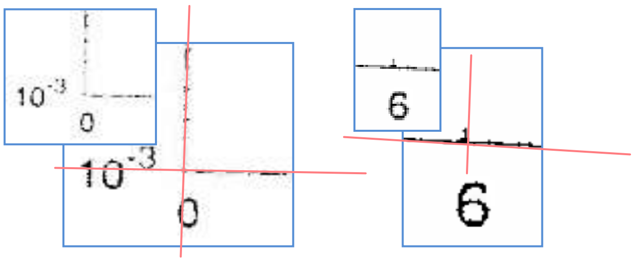
1) Auto-focusing of symbols: to implement in digitizer InpGraph, to extend in GSYS for cases with overlapping symbols (operator input number of symbols)
2) Editing defects on symbols and error-bars (?) and filling symbols (to make autofocusing more stable). Example: improving image for automatic operations by editing (add/remove pixels)



3) Precise calibration of scales using analytical approximation if necessary, e.g. as it is usually done for treatment of gamma spectra (calibration energy, FWHM and efficiency):

Energy(nChannel): $E(n)= a_0 + a_1 \cdot n + a_2 \cdot n^2$
FWHM(E)=$\sqrt{a_0 + a_1 \cdot E}$
Efficiency(E): $\ln(Eff(E)) = a_0 + a_1 \cdot \ln(E) + a_2 \cdot \ln^2 (E) + a_3 \cdot \ln^3 (E) \ldots + a_7 \cdot \ln^7(E)$

4) Auto-focusing of scales' crossings and strokes (needs recognition of lines, or "insert vectors and calculate their crossing")



5) Independent plotting (e.g. EXFOR-Uploading + Web-ZVView) can be used as objective checking of digitization result, e.g. it can allow finding mistakes in the scale definition and correct mistakes (labelling of axis) which cannot be found by "feedback" function of digitizers. Fig.1 presents direct comparison of the plotted digitized data with original plot (picture on the top was produced using existing tools without any software development).

Using ZVView plotting system with scale calibration for independent checking results of digitization.

Recently plotting package ZVView [4] was extended by 2-D scales calibration and output transformation functions. **Fig.2** shows how manual 2-D scales calibration looks. **Fig.3** was produced by ZVView distorting output plot using results of 2-D scales calibration from Fig.2. **Fig.4** and **5** show direct comparison of the original plot with plot of the data - result of digitizing, distorted by ZVView (we can see too small upper error bars of the experimental data points between 5 and 10 MeV). **Fig.6** shows an example of another way of output plot transformation implemented in ZVView – using analytical functions. The options of transformation of the plots can be also added to Web-ZVView package if users are interested.

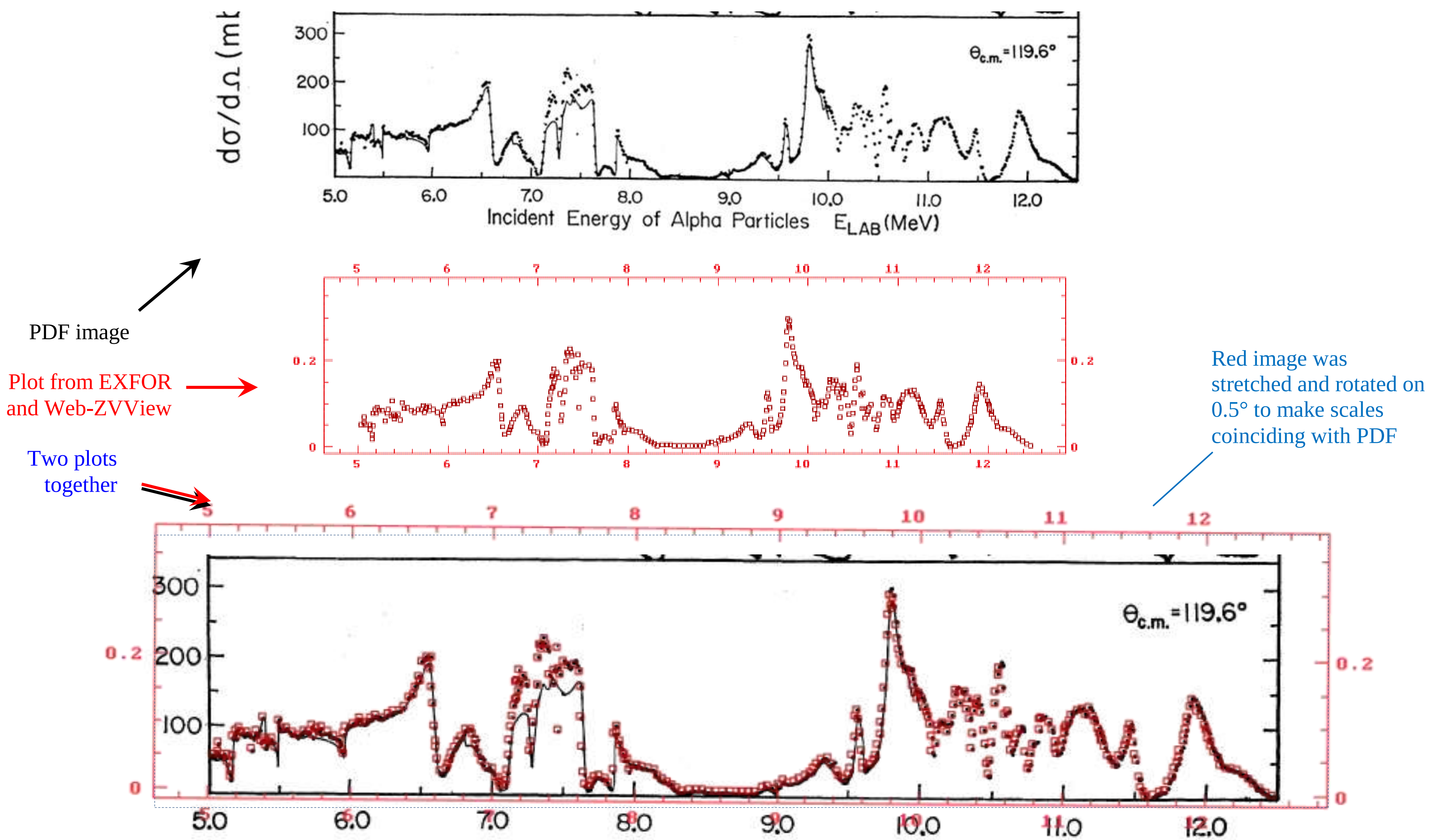


FIG. 1 (colour online). Independently produced plot (red) on top of PDF image (black) with precise coincidence of the scales.

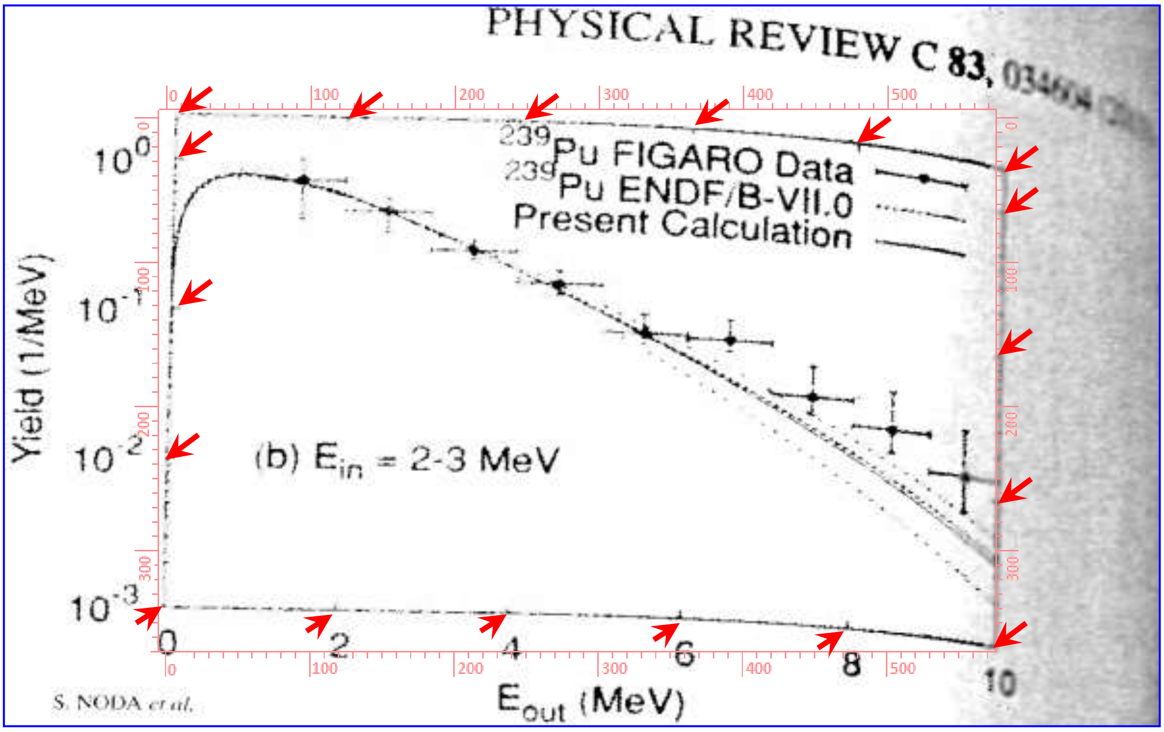


FIG. 2 (colour online). Difficult case: needs calibration of the scales and advanced algorithms to restore rectangular grid correctly.

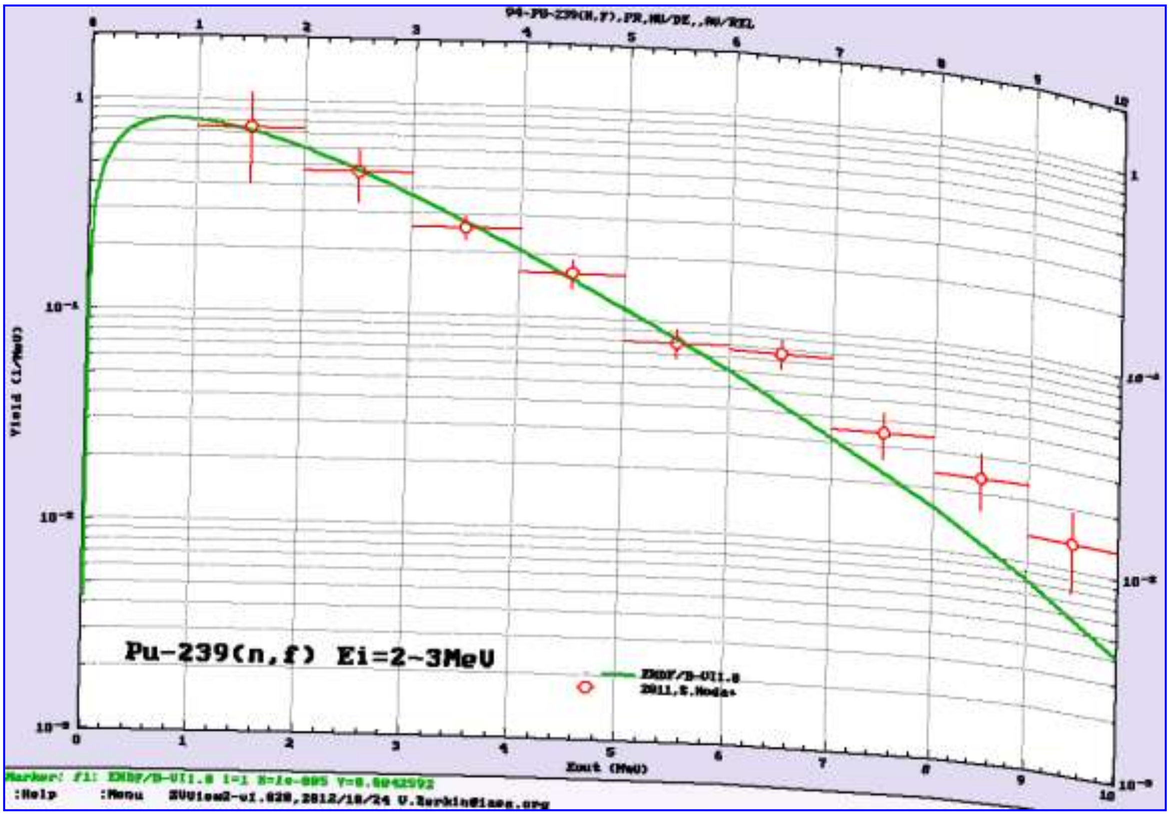


FIG. 3 (colour online). Plot of digitizing results distorted by extended version of ZVView. Transformation was done using 2-D scales calibration procedures (Fig. 2).

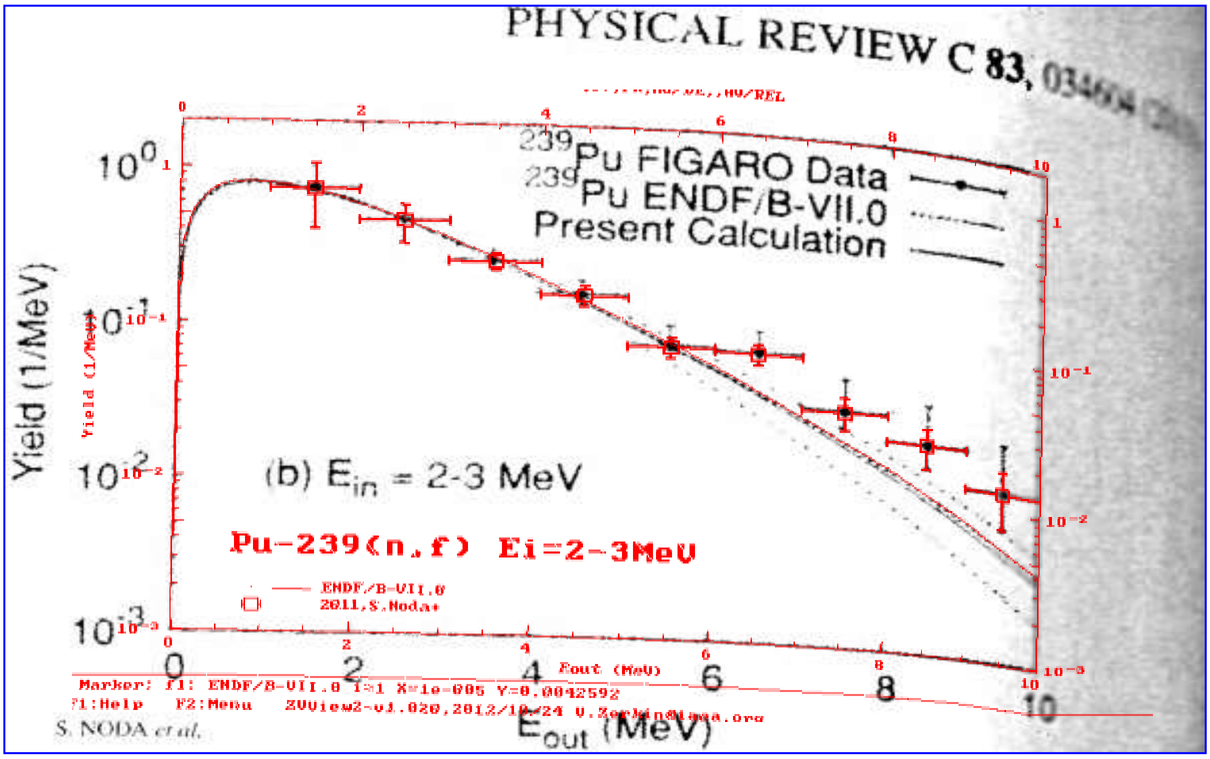


FIG. 4 (colour online). Direct comparison of original plot and the plot of the results of digitizing distorted by ZVView using calibration of the scales (Fig. 2).

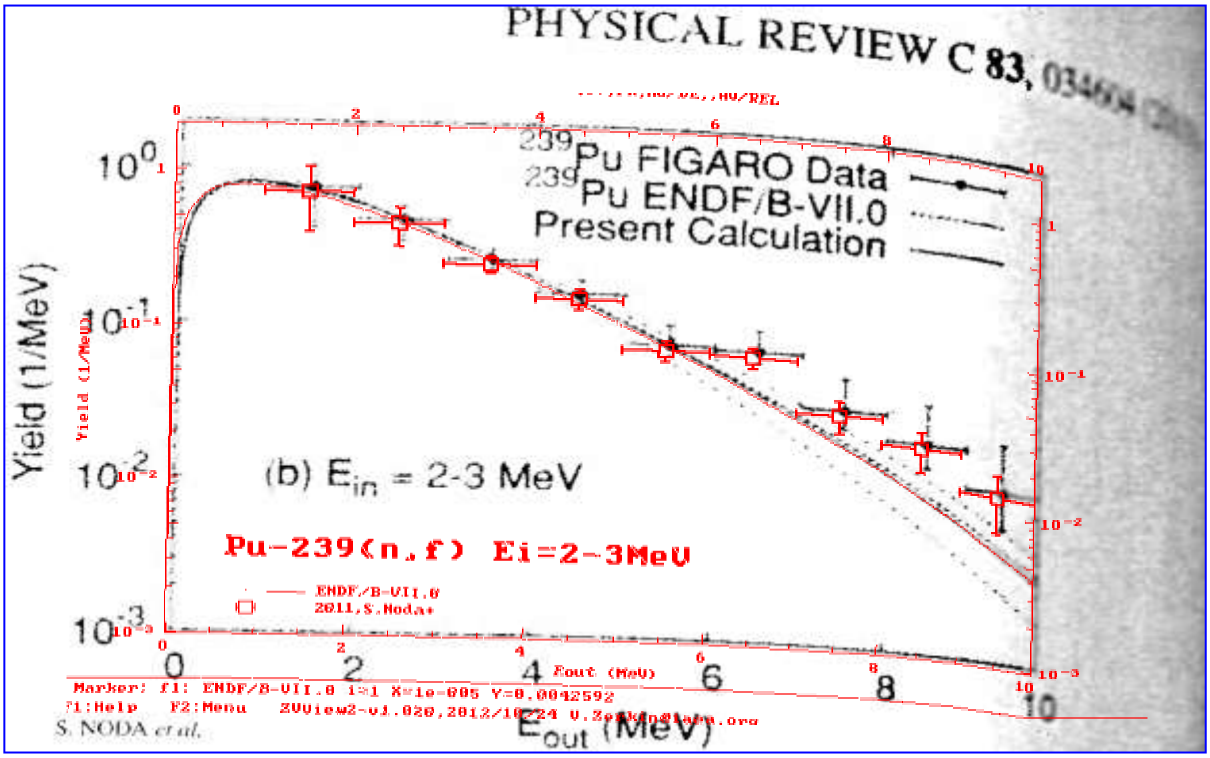


FIG. 5 (colour online). Direct comparison of the original plot and distorted plot of digitized data: the same as Fig.4, but with shifted beginning of coordinates for better visibility of the differences in error bars.

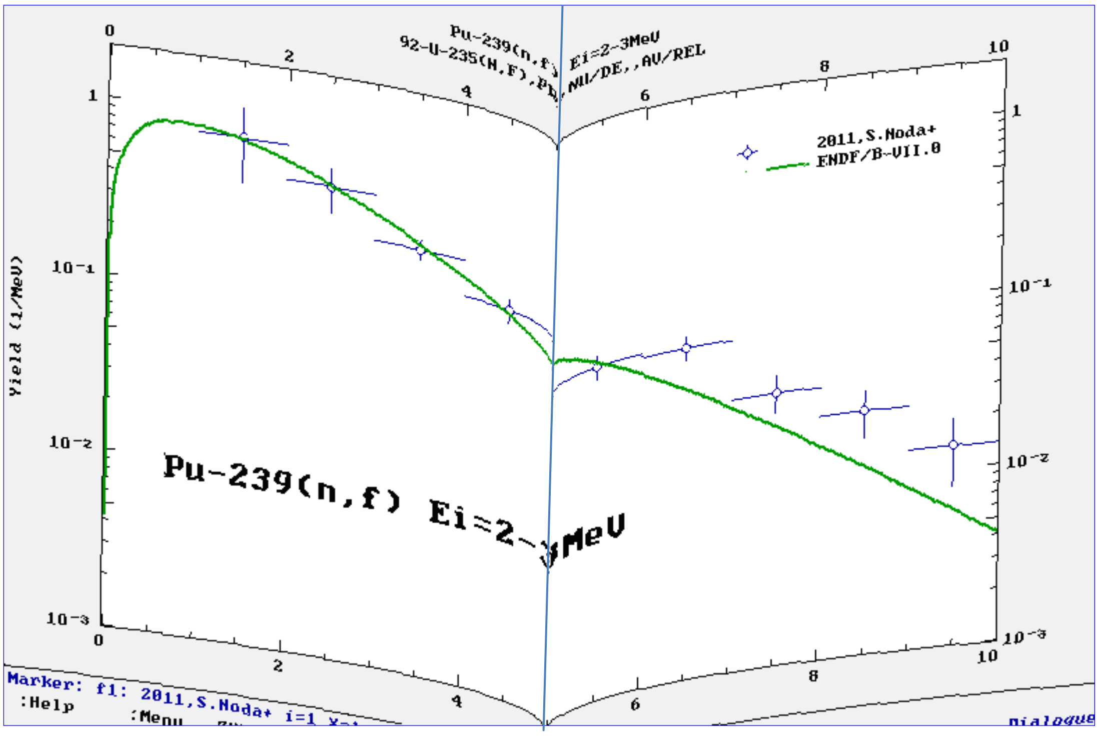


FIG. 6 (colour online). Transformed plot imitating photocopy of scanned book pages (produced by ZVView using analytical functions for transformation).

References

1. EXFOR Digitizer InpGraph, G.Pikulina et al., Sarov, Russia, 2012
   Web: http://www-nds.iaea.org/nrdc/nrdc_sft/inpgraph-201211.zip
2. Graph Suchi Yomitori System (GSYS), R.Suzuki et al., Sapporo, Japan, 2005-2012
   Web: http://www.jcprg.org/gsys/
3. Introduction of the digitizing software GDgraph, G.Chen, 2012
   http://www-nds.iaea.org/digitization/docs/chen.pdf
4. Interactive nuclear data plotting package ZVView, V.Zerkin, Kiev-IAEA, 1996-2012
   Web: http://www-nds.iaea.org/ndspub/zvview/

Appendix E

QUALITY AND CONSISTENCY OF DIGITIZED DATA*

N. Otuka†

Nuclear Data Center, Japan Atomic Energy Research Institute, Tokai-mura, 319-1195 Ibaraki, Japan
Department of Physics, Hokkaido University, 060-0810 Sapporo, Japan

S. Dunaeva‡

IAEA Nuclear Data Section, Wagramerstraße 5, A-1400 Vienna, Austria

At present, EXFOR 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 recognize 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 on 1 April, 2005 (Memo CP-D/415). EXFOR compilers were invited to digitize a set of linear-log plots (angular distribution, **Fig. 1**). The source of the figure (Fig.2 of [1]) was unknown for EXFOR compilers before their submissions.

Five EXFOR compilers from NNDC, CAJaD, CDFE, JCPRG and CNPD digitized data, and submit them to NDS (**Table I**). The deviation of the digitized value from the original value (*i.e.*, value provided from the author) was analyzed by JCPRG and NDS. 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. Example of data format: "1.23E+45" is 3 digits, "12.345": 3 decimal.

Centre	Digitizer	Developer	Format of digitized data		
			x (deg)	y (mb/sr)	Δy (mb/sr)
NNDC	GSYS	JCPRG	3 digits	3 digits	3 digits
CAJaD	(unknown)	CAJaD	2 decimal	3 digits	(in %)
CDFE	(unknown)	CDFE?	2 decimal	3-5 decimal	3-5 decimal
JCPRG	GSYS	JCPRG	3 digits	3 digits	3 digits
CNPD	Graf_new	CNPD	3 decimal	5 digits	(in %)

For brief comparison between digitized data and original data, digitized data and original data are shown together in **Fig.2**. The original data are well reproduced by digitized data in general. Some large discrepancy is seen in the y values of the following data sets:

- NNDC: Underestimate authors' data at all angles at Ex = 993 keV.
- CAJaD: Overestimate authors' data at all angles at Ex = 240 keV.
- CNPD: Over- and underestimate authors' data at small and large angles at Ex = 814 keV.

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

For more detailed comparison, we 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. 1). Additional curves given by CNPD and CAJaD show digitization error coded under

* Reproduced from the working paper WP2005-24 distributed in the NRDC 2005 Meeting (with minor corrections).

† Present address: IAEA Nuclear Data Section, Vienna International Centre, A-1400 Vienna, Austria

‡ Retired

ANG-ERR-D (for x-values) and ERR-DIG (for y-values). These would be regarded as “allowable deviation” in the present checking. Short comments are given for each figure below:

Fig. 2 (accuracy of 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. 3 (accuracy of 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 situations are also seen in Fig. 1. 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,).

Fig.4 (accuracy of y-error values):

Error bars in the original figure are very small, and JCPRG, CNPD, NNDC and CAJaD skipped digitization of error bars for many data points. It is very difficult to digitize such a small error bar depicted in logarithmic scale with good accuracy. CAJaD data of 2405 keV are involved in the trouble as mentioned in the discussion of Fig. 1.

This exercise 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.

Reference

[1] D. Bucurescu *et al.*, Nucl. Phys. **A674** (2000) 11.

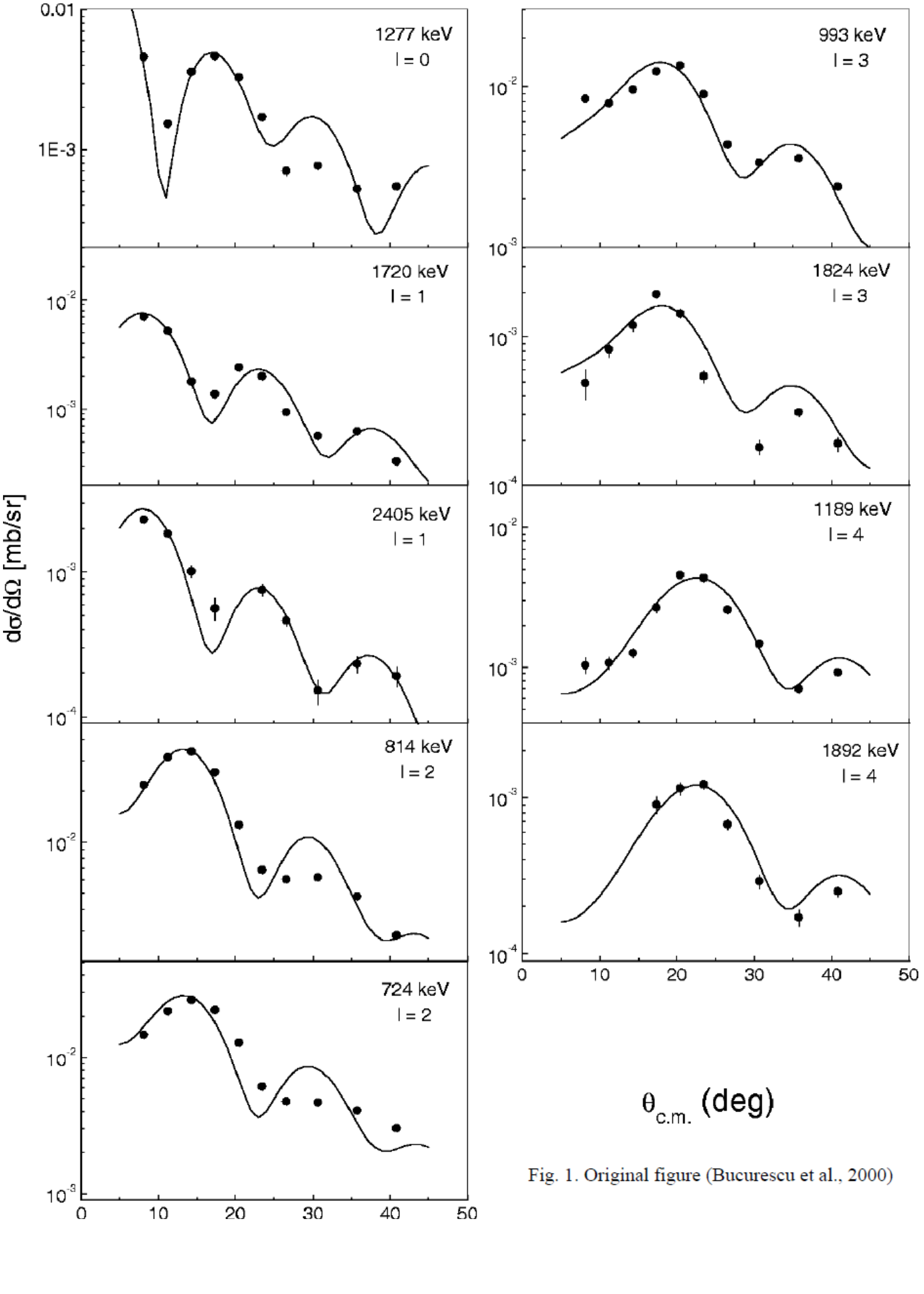


Fig. 1. Original figure (Bucurescu et al., 2000)

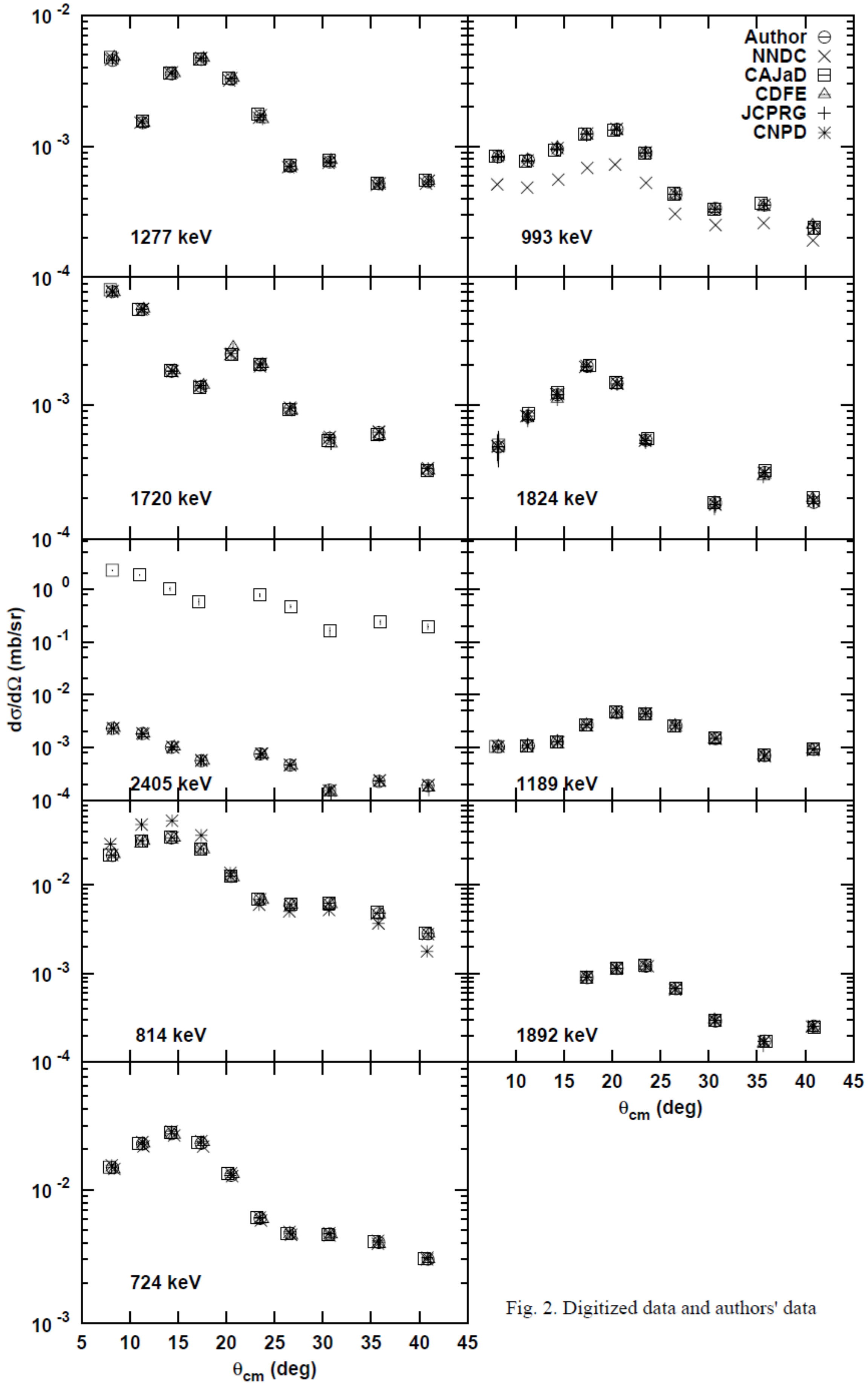


Fig. 2. Digitized data and authors' data

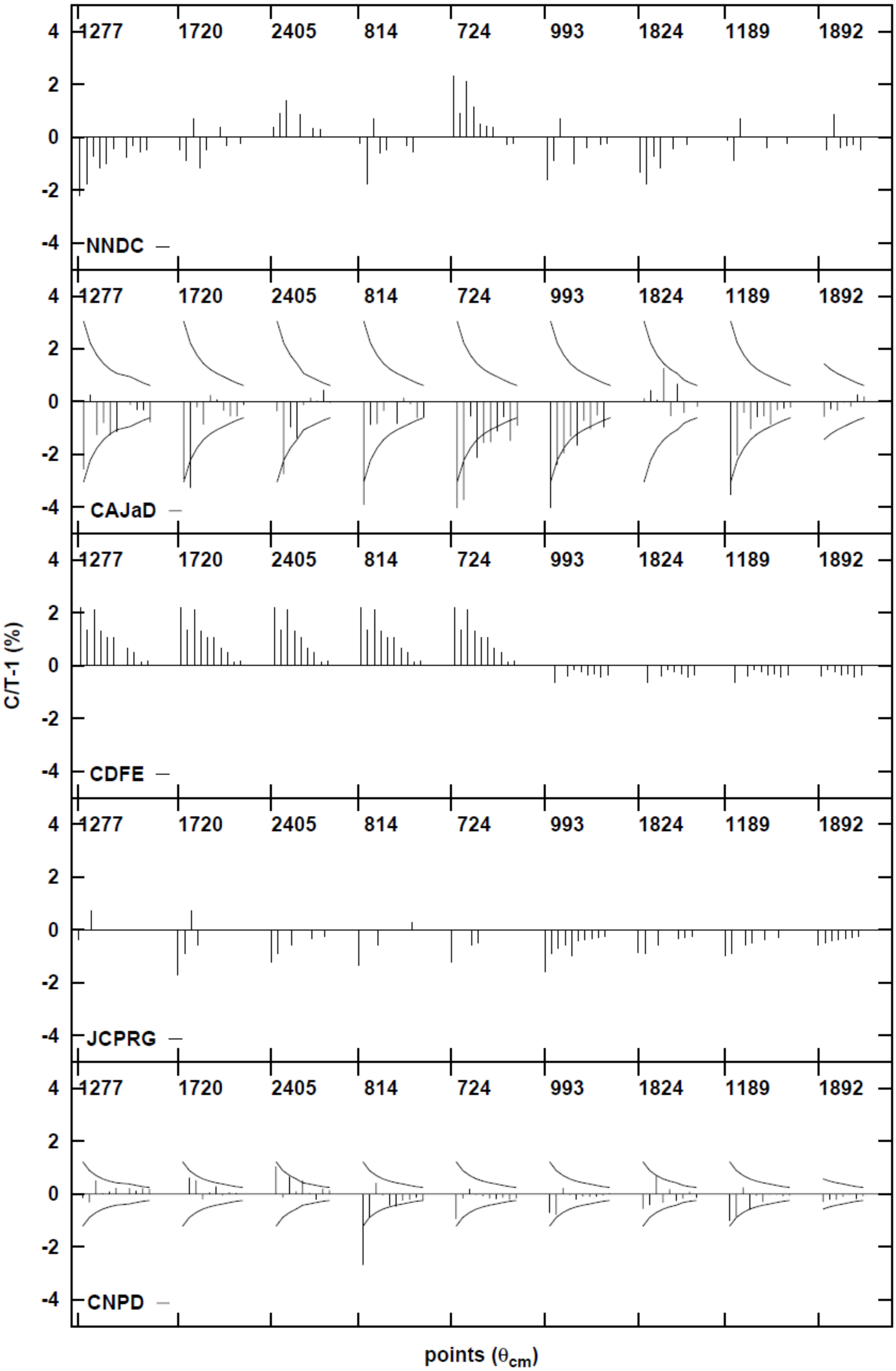


Fig. 3. Accuracy of digitization for x value at each data point.

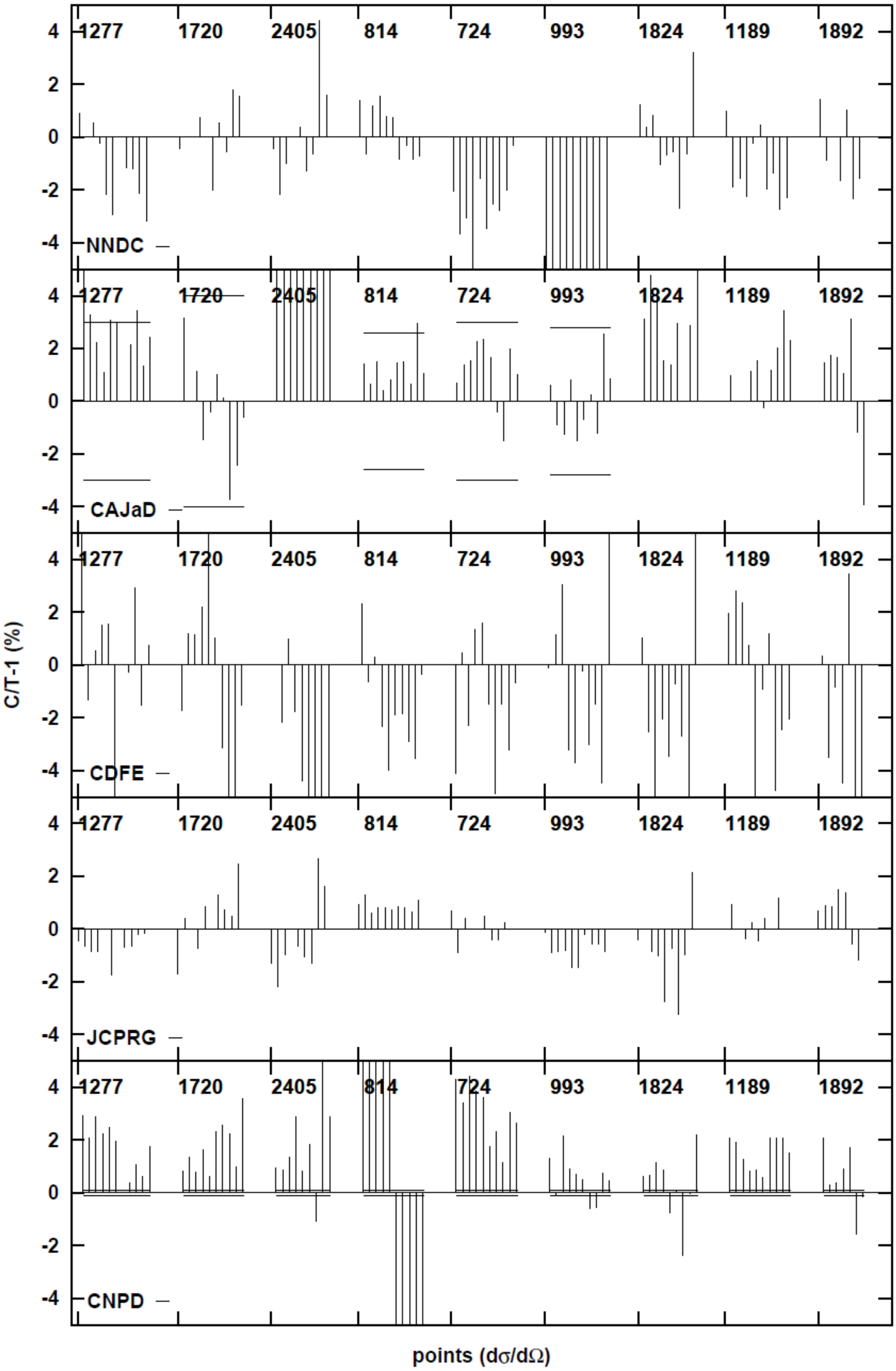


Fig. 4. Accuracy of digitization for y value at each point.

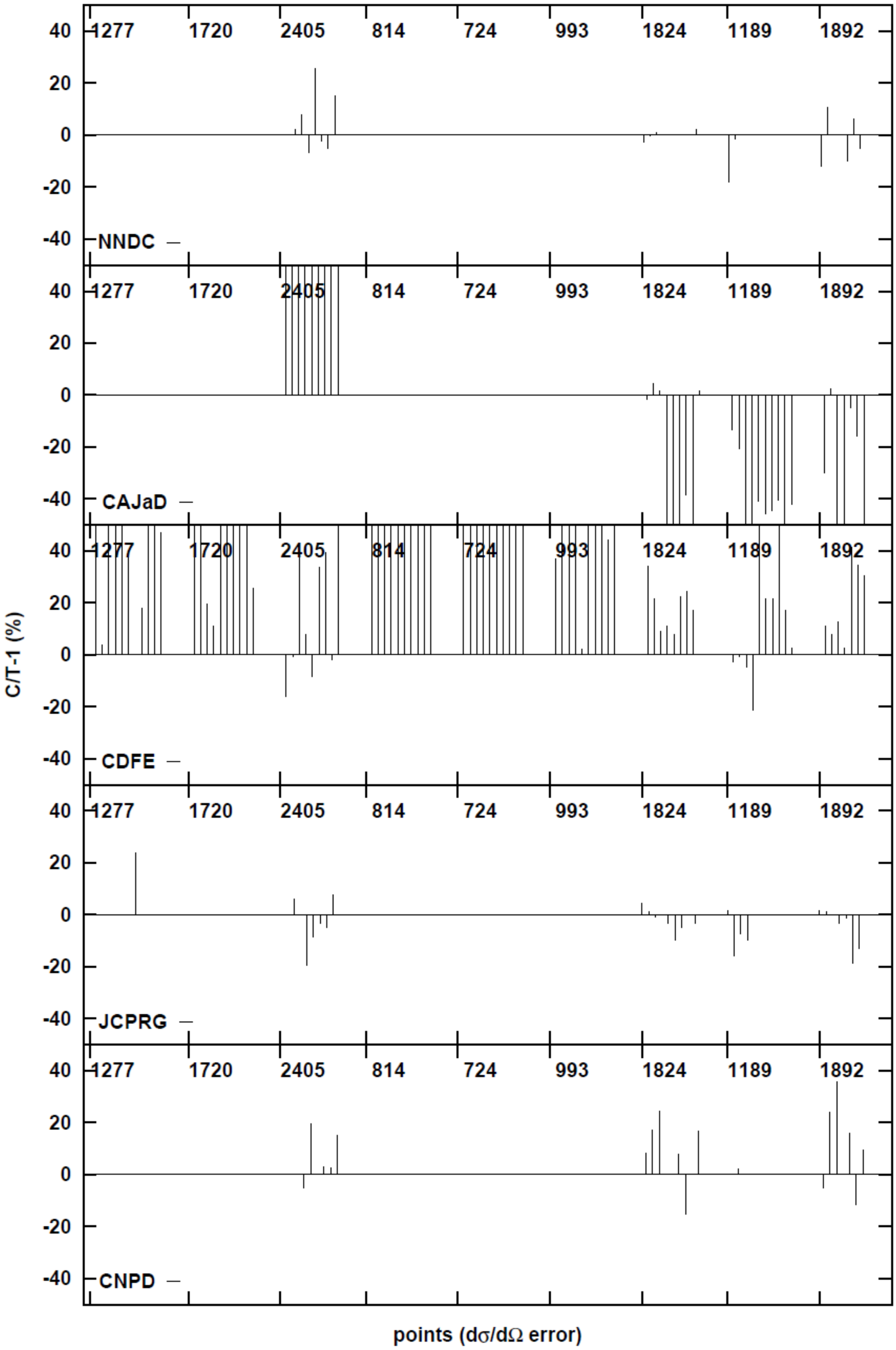


Fig. 5. Accuracy of digitization for y-error value at each data point.

Nuclear Data Section
International Atomic Energy Agency
Vienna International Centre, P.O. Box 100
A-1400 Vienna
Austria

e-mail: nds.contact-point@iaea.org
fax: (43-1) 26007
telephone: (43-1) 2600-21710
Web: http://www-nds.iaea.org/