A computational EXFOR database

Georg Schnabel

Uppsala University

ND2019, Beijing, 2019/5/22
### EXFOR Database

- About **18 million data points** and growing!
- No covariance matrices for many measurements
- Plenty of dependencies

#### Database as of: 2019-05-07

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of ENTRY</td>
<td>22615</td>
</tr>
<tr>
<td>Number of SUBENT</td>
<td>156824</td>
</tr>
<tr>
<td>Number of Datasets</td>
<td>174048</td>
</tr>
<tr>
<td>Number of Datapoints</td>
<td>17880333</td>
</tr>
<tr>
<td>Experimental works</td>
<td></td>
</tr>
<tr>
<td>Data tables</td>
<td></td>
</tr>
<tr>
<td>Data tables of reactions</td>
<td></td>
</tr>
<tr>
<td>Total number of data points</td>
<td></td>
</tr>
</tbody>
</table>

- **About 18 million data points**: The EXFOR database contains a vast collection of experimental nuclear reaction data, with over 18 million data points as of 2019. This number continues to grow as new experiments are conducted and data is added.
- **No covariance matrices for many measurements**: Many of the measurements in the database do not have covariance matrices associated with them, which can affect the interpretation of the data.
- **Plenty of dependencies**: The database contains a multitude of dependencies that are crucial for understanding the behavior of nuclear reactions.
Statistics & Machine Learning

- Explosion of research output, e.g., papers published on May 10 tagged as ML: **376**, about **100,000 per year**

- Stunning performances, e.g., search for "Microsoft creates AI that can read a document and answer questions about it as well as a person"

- Is there something useful for nuclear data, especially nuclear databases?
Some work done

From: J.A. Hirdt and D.A. Brown, Identifying Understudied Nuclear Reactions by Text-mining the EXFOR Experimental Nuclear Reaction Library
What is the problem?

EXFOR format

<table>
<thead>
<tr>
<th>ENTRY</th>
<th>SUBENT</th>
<th>BIB</th>
<th>INSTITUTE</th>
<th>REFERENCE</th>
<th>AUTHOR</th>
<th>TITLE</th>
<th>FACILITY</th>
<th>INC-SOURCE</th>
<th>METHOD</th>
<th>DETECTOR</th>
<th>MONITOR</th>
<th>ERR-ANALYS</th>
<th>STATUS</th>
<th>HISTORY</th>
</tr>
</thead>
</table>

ENDBIB | 25 | NOCOMMON | 0 | NOVISION | 0 | ENDSUBENT | 28 | 37 |

High-level languages

Python and R

Many community contributed packages for ML

Improvements over raw EXFOR

Original EXFOR

<table>
<thead>
<tr>
<th>SUBENT</th>
<th>23171003</th>
<th>20170913</th>
<th>20180129</th>
<th>20180126</th>
<th>2265</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIB</td>
<td>11</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>REACTION</td>
<td>(26-FE-56(N,2N)26-FE-55,,SIG)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DECAY-DATA</td>
<td>(26-FE-55,2.73YR)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FACILITY</td>
<td>1(VDG,22ZZGEL) Van de Graaff accelerator at IRMM.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2(ACCEL,2AUSINK) For AMS (accelerator mass spectrometry)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>at VERA lab. for radionuclides 10Be, 14C, 26Al, 55Fe,</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(CW,2GERDRE)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>METHOD</td>
<td>1(ACTIV) Activation technique combined with</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2(AMS) mass spectrometric technique</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAMPLE</td>
<td>Natural iron samples were irradiated at TU Dresden</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>and IRMM.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INC-SOURCE</td>
<td>T(d,n)He-4 .</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INC-SPECT</td>
<td>1 Quasi-monoenergetic neutrons with energies between</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13.4 and 14.8 MeV;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 from 13 to 20 MeV.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FLAG</td>
<td>(1.) Experiment in 2007 yr, TUD/VERA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.) Experiment in 2010 yr, TUD/VERA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.) Experiment in 2010 yr, IRMM/VERA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MRC</td>
<td>EN-ERR+EN-RSL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERR-ANA</td>
<td>(EN-ERR) finite neutron energy distribution and the</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>absolute uncertainty in the neutron energy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(namely it is EN-ERR+EN-RSL).</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STATUS</td>
<td>(PRELM) Preliminary results (declared by A.Wallner,</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2017-08-30) under STATUS).</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HISTORY</td>
<td>(TABLE) Data received from the author (2017-08-30)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SD: Data received from the author were added. FLAG was added. BIB update according to comments from author.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENDBIB</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOCOMMON</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DATA</td>
<td>5</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EN</td>
<td>EN-ERR</td>
<td>DATA</td>
<td>DATA-ERR</td>
<td>FLAG</td>
<td></td>
</tr>
<tr>
<td>MEV</td>
<td>MEV</td>
<td>MB</td>
<td>MB</td>
<td>NO-DIM</td>
<td></td>
</tr>
<tr>
<td>13.35</td>
<td>0.15</td>
<td>300.</td>
<td>100.</td>
<td>3.</td>
<td></td>
</tr>
<tr>
<td>13.49</td>
<td>0.04</td>
<td>322.4</td>
<td>16.1</td>
<td>1.</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENDDATA</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENDSUBENT</td>
<td>48</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

X4 format

```
##---EXFOR Request #33113
##---EXFOR Standard Output---
#DATASET 23171003
#NOW 2019/05/16:10:47:16
#SUBENT 23171003 20170913
#ENTRY 23171 20170913
#AUTHOR1 A.Wallner+
#YEAR 2011
#X4REF1 J,KPS,59,(2),1378,2011
#REFERENCE1 Jour: Journal of the ...
#REACTION 26-FE-56(N,2N)26-FE-55,,SIG
#D4REAC R0#
#C4Reaction (N,2N),SIG
#ReactionType CS
#Quantity Cross section
#IndVarFamCode [0 2]
#ExpectedUnits [B]
#xVariables 1
#+ Y = Y(X1)
#COLUMNS 4
!DATA DATA-ERR EN EN-ERR
!MB MEV MEV
!Y.Value Y.Err+- X1.Value X1.Err+-
!0.1 0.911 1.1 1.911
!variable variable variable variable
!0.001 0.001 1000000.0 1000000.0
!B MB EV EV
!#DATA 15 4 12
!300. 100. 13.35 0.15
!322.4 16.1 13.49 0.04
!...
!#ENDDATA 15 4

Suited for physics codes (e.g., in Fortran)
```

Textual information can be valuable for ML
Wish list

- Convenient access to any information in the original EXFOR database
- Flexible and powerful search capabilities accessible from a high-level programming language
- Rapid conversion of EXFOR data to any (e.g. X4-like) format with the information a user needs by the user
- Don’t reinvent the wheel, use proven solutions

How can we do it?
Long live EXFOR!

Hierarchical format

SUBENT 23171003 20170913 20180129 20180126 2265
BIB 11 26
REACTION (26-FE-56(N,2N)26-FE-55,,SIG)
DECAY-DATA (26-FE-55,2.73YR)
FACILITY 1(VDG,2ZZZGEL) Van de Graaff accelerator at IRMM.
2(ACCEL,2AUSIRK) For AMS (accelerator mass spectrometry)
at VERA lab. for radionuclides 10Be, 14C, 26Al, 55Fe,
(CCW,2GERDRE)
METHOD 1(ACTIV) Activation technique combined with
2(AMS) mass spectrometric technique
SAMPLE Natural iron samples were irradiated at TU Dresden
and IRMM.
INC-SOURCE1(D-T) T(d,n)He-4.
INC-SPECT 1 Quasi-monoenergetic neutrons with energies between
13.4 and 14.8 MeV;
2 from 13 to 20 MeV.
FLAG (1.) Experiment in 2007 yr, TUD/VERA
(2.) Experiment in 2010 yr, TUD/VERA
(3.) Experiment in 2010 yr, IRMM/VERA
ERR-ANALYS (EN-ERR) finite neutron energy distribution and the
absolute uncertainty in the neutron energy
(namely it is EN-ERR+EN-RSL).
STATUS (PRELM) Preliminary results (declared by A.Wallner,
2017-08-30) under STATUS).
HISTORY (20170913R) SD: Data received from the author (2017-08-30)
SD: Data received from the author were added. FLAG was added. BIB update according to
comments from author.
ENDBIB 26
NOCOMMON 0 0
DATA 5 15
EN MEV 13.35 13.49
EN-ERR MEV 0.15 0.04
DATA MB 300. 322.4
DATA-ERR MB 100. 16.1
FLAG NO-DIM 3. 1.
ENDDATA 17
ENDSUBENT 48

<table>
<thead>
<tr>
<th>EN(MEV)</th>
<th>EN-ERR(MEV)</th>
<th>DATA(MB)</th>
<th>DATA-ERR(MB)</th>
<th>FLAG(NO-DIM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.35</td>
<td>0.15</td>
<td>300.</td>
<td>100.</td>
<td>3.</td>
</tr>
<tr>
<td>13.49</td>
<td>0.04</td>
<td>322.4</td>
<td>16.1</td>
<td>1.</td>
</tr>
</tbody>
</table>

...
JSON format

A light-weight data-interchange format. It is easy for humans to read and write. It is easy for machines to parse and generate.

from https://www.json.org

- Supports strings, numbers, and **hierarchical organization**
- Convert original EXFOR (sub)entries to JSON
- Rely on wide support for the JSON format in many programming languages to read and process information
```
{
  "ID": "23171003",
  "DATEMOD": "20170913",
  "BIB":{
    "REACTION": "(26-FE-56(N,2N)26-FE-55,,SIG)",
    "DECAY-DATA": "(26-FE-55,2.73YR)",
    "FACILITY": [
      "(VDG,2ZZGEL) Van de Graaff accelerator ...",
      "(ACCEL,2AUSIRK) For AMS ... at VERA lab ..."
    ],
    "ERR-ANALYS": "(EN-ERR) finite neutron energy distribution and the absolute uncertainty in the neutron energy (namely it is EN-ERR+EN-RSL).",
    ...
  },
  ...
}
```

R Code for creation of JSON

```
library(exforParser)

filepath = "<PATH TO EXFOR FILE>"

exforText = readChar(filepath, file.info(filepath)$size)

entry = parseEntry(exforText)

convToJSON(entry$SUBENT[[3]])
```

https://github.com/gschnabel/exforParser
Convenient access

From R

```r
...
entry = parseEntry(exforText)

entry$SUBENT[[3]]$BIB$REACTION

entry$SUBENT[[3]]$DATA$UNIT

entry$SUBENT[[3]]$DATA$TABLE
```

From Python

```python
import json

with open('exforEntryFile') as json_file:
    entry = json.load(json_file)

entry['SUBENT'][3]['BIB']['REACTION']

...
What about searching?

- SQL databases not ideal for hierarchically organized data
- But document-oriented databases (DOB) are!
- Choose one of the publicly available DOB* and insert the JSON files created from EXFOR entries
- Use the powerful search capabilities provided by the DB implementation

Example: MongoDB

```r
library(MongoEXFOR)
db <- connectExfor("entries","exfor","mongodb://localhost")

queryStr <- makeQueryStr(and("BIB.REACTION: { $regex: "26-FE-56.*SIG", $options: "" }",
'BIB.REACTION: { $not: { $regex: "\) *\[+-*/\) *\(", $options: "" }}',
'DATA.TABLE.DATA: { $exists: true }',
'DATA.TABLE.EN: { $exists: true }
))

resDt <- db$find(queryStr, {
  list(SUBENT = ID,
       REAC = BIB$REACTION)
})
```

```
<table>
<thead>
<tr>
<th>SUBENT</th>
<th>REAC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: 10022010</td>
<td>(26-FE-56(N,P)25-MN-56,,SIG)</td>
</tr>
<tr>
<td>2: 10031005</td>
<td>(26-FE-56(N,P)25-MN-56,,SIG)</td>
</tr>
<tr>
<td>3: 10037004</td>
<td>(26-FE-56(N,EL)26-FE-56,,SIG)</td>
</tr>
<tr>
<td>4: 10037005</td>
<td>(26-FE-56(N,TOT),,SIG)</td>
</tr>
<tr>
<td>5: 10037015</td>
<td>(26-FE-56(N,INL)26-FE-56,PAR,SIG)</td>
</tr>
</tbody>
</table>

https://github.com/gschnabel/MongoEXFOR
```
How can it be used?

Construction and evaluation of experimental covariance matrices (at scale)

- exp.point #1
  - method
    - time-of-flight
  - exp.point #2
  - exp.point #3
- exp.point #4
- exp.point #5
- exp.point #6
- exp.point #7
  - method
    - activation
Correlation due to method

Search specification

```
queryStr <- makeQueryStr(and(
  'BIB.REACTION: { $not: { $regex: "\) *\[+-*/\] *\("", $options: '' }
  'BIB.METHOD: { $exists: true }
  'DATA.TABLE.EN: { $exists: true }
  'DATA.TABLE.DATA: { $exists: true }
))
```

Data retrieval

```
expDt <- db$find(queryStr, {
  if (length(BIB$REACTION) > 1) NULL
  else if (length(BIB$METHOD) > 1) NULL
  else
    list(EXPID = ID,
         DIDX = seq(nrow(DATA$TABLE)),
         METHOD = BIB$METHOD,
         DATA = DATA$TABLE$DATA)
})
```

Additional preparation

```
pat <- "^\((([A-Z]+)\).*$"
expDt <- expDt[!is.na(DATA),]
expDt <- expDt[DATA > 1,]
expDt <- expDt[grepl(pat, METHOD)]
expDt <- expDt[, METHOD := sub(pat, "\\1", METHOD)]
expDt[, UNC := 0.1 * DATA]
expDt[, IDX := seq_len(.N)]
```
## Correlation due to method

<table>
<thead>
<tr>
<th>EXPID</th>
<th>DIDX</th>
<th>METHOD</th>
<th>DATA</th>
<th>UNC</th>
<th>IDX</th>
</tr>
</thead>
<tbody>
<tr>
<td>10001002</td>
<td>1</td>
<td>TOF</td>
<td>25.00</td>
<td>2.500</td>
<td>1</td>
</tr>
<tr>
<td>10001002</td>
<td>2</td>
<td>TOF</td>
<td>30.00</td>
<td>3.000</td>
<td>2</td>
</tr>
<tr>
<td>10001002</td>
<td>3</td>
<td>TOF</td>
<td>26.00</td>
<td>2.600</td>
<td>3</td>
</tr>
<tr>
<td>10001002</td>
<td>4</td>
<td>TOF</td>
<td>28.00</td>
<td>2.800</td>
<td>4</td>
</tr>
<tr>
<td>10001002</td>
<td>5</td>
<td>TOF</td>
<td>29.00</td>
<td>2.900</td>
<td>5</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>METHOD</th>
<th>numPoints</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: TOF</td>
<td>1281743</td>
</tr>
<tr>
<td>2: ACTIV</td>
<td>27312</td>
</tr>
<tr>
<td>3: FNB</td>
<td>10489</td>
</tr>
<tr>
<td>4: EDE</td>
<td>8208</td>
</tr>
<tr>
<td>5: COINC</td>
<td>4882</td>
</tr>
<tr>
<td>6: GSPEC</td>
<td>4816</td>
</tr>
<tr>
<td>7: TRN</td>
<td>4662</td>
</tr>
<tr>
<td>8: BCINT</td>
<td>4375</td>
</tr>
<tr>
<td>9: SLODT</td>
<td>1539</td>
</tr>
<tr>
<td>10: PHD</td>
<td>1293</td>
</tr>
<tr>
<td>11: ASEP</td>
<td>883</td>
</tr>
<tr>
<td>12: CHSEP</td>
<td>856</td>
</tr>
<tr>
<td>13: STTA</td>
<td>783</td>
</tr>
<tr>
<td>14: TTM</td>
<td>661</td>
</tr>
<tr>
<td>15: AMS</td>
<td>639</td>
</tr>
<tr>
<td>16: MAGFR</td>
<td>374</td>
</tr>
<tr>
<td>17: SITA</td>
<td>242</td>
</tr>
<tr>
<td>18: EDEG</td>
<td>226</td>
</tr>
<tr>
<td>19: BSPEC</td>
<td>203</td>
</tr>
<tr>
<td>20: REC</td>
<td>194</td>
</tr>
</tbody>
</table>

49 different methods (TOF, ACTIV, ...)

~1.3 million correlated TOF points

~1.4 million data points
Generalized $\chi^2$ calculation

library(nucdataBaynet)
normHandler <- createSysCompNormHandler("DATA")
normHandler$addSysUnc("METHOD",
  levels = unique(expDt$METHOD),
  vals = 0,
  uncs = 0.1, rel = TRUE)
sysDt <- normHandler$createSysDt()
sysDt[, IDX := seq_len(.N)]

library(Matrix)
x <- runif(nrow(expDt))
S <- normHandler$map(expDt, sysDt, ret.mat = TRUE)
P <- Diagonal(x = sysDt$UNC^2)
D <- Diagonal(x = expDt$UNC^2)

invD <- solve(D)
invP <- solve(P)
Z <- t(S) %*% invD %*% S + invP
invD_x <- t(S) %*% invD %*% x
xt_invCov_x <- t(x) %*% invD %*% x - t(invD_x) %*% solve(Z, invD_x)

Calculation of $\exp(-0.5*\chi^2)$ involving 1.4 mio correlated data points in 0.5 secs!
Summary & Conclusion

- EXFOR is a hierarchial format and powerful free and open-source options exist to interact with data in such format.
- The conversion of EXFOR into a mainstream document-oriented database facilitates data handling in high-level languages such as Python and R.
- Data analysis becomes easier, formats similar to, e.g., X4 can be created by the user suitting their specific needs.
- EXFOR becomes amenable to a treatment with statistics & ML solutions provided as packages by statisticians and ML researchers for high-level languages.
- Code packages available at https://github.com/gschnabel
  Mostly under the GPL license: Do whatever you want with the codes but contribute back to the community.