IAEA 60 Years
Atoms for Peace and Development

EXFORTABLES: An analysis of EXFOR
+
Determination of isotopes, energy range and observables

Arjan Koning

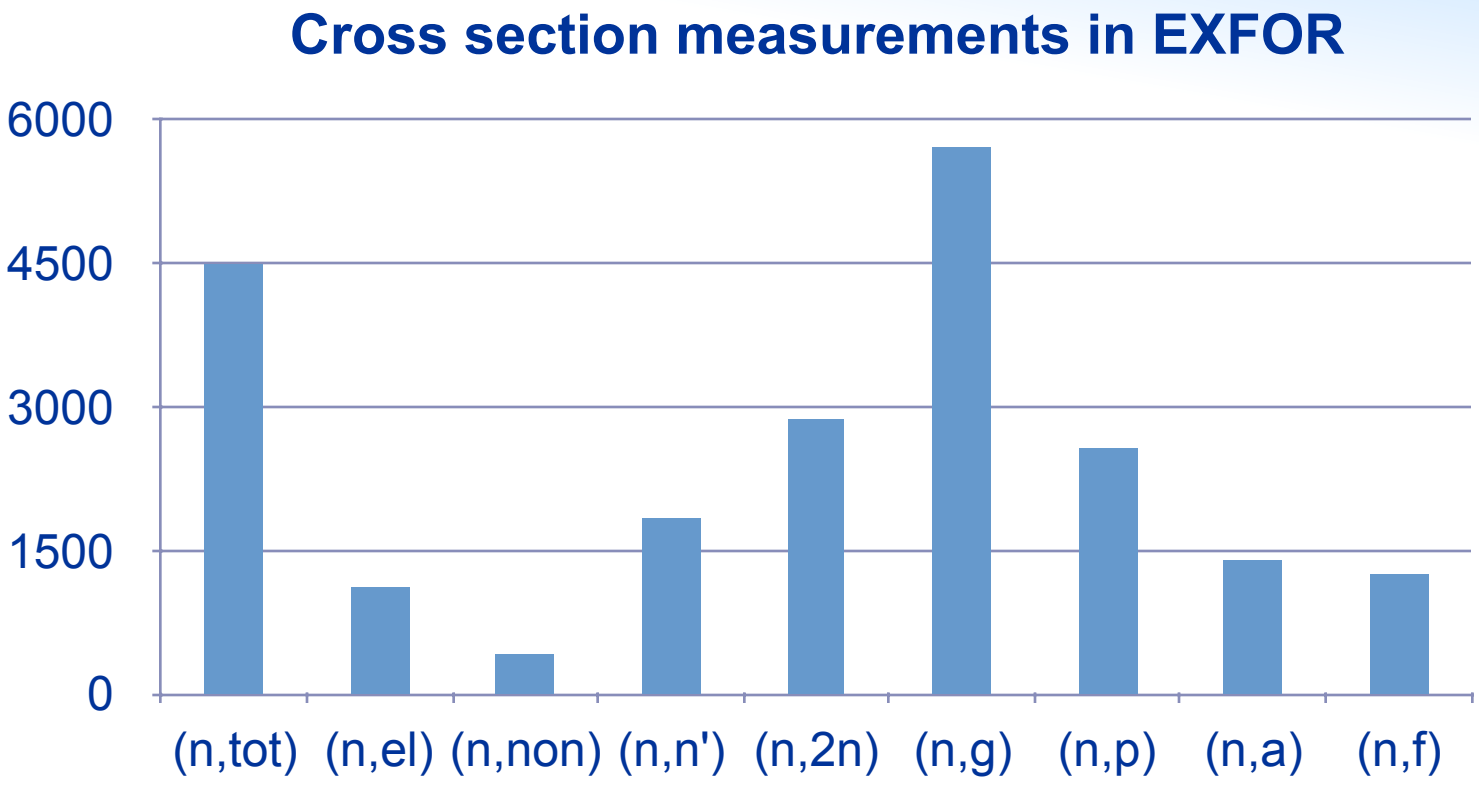
Head of Nuclear Data Section
International Atomic Energy Agency, Vienna

WPEC-SG50, September 14-15 2020, NEA, Boulogne-Billancourt
(Video meeting)

Contents

- EXFOR as a data library
- Experimental nuclear data evaluation
- Global testing of EXFOR
- Machine learning?
- Scope: Isotopes, energy range and observables
- Summary

EXFOR database
(Nuclear Reaction Data Center Network: IAEA, NNDC, NEA, JAEA, etc)



Total estimated cost of EXFOR (private comm.): between 20 – 60 Billion USD
Total estimated value of EXFOR : priceless

The issues

- EXFOR is not a machine-readable data library
  - Flexible and general command-line API needed!
- EXFOR is not an evaluation but is (and should remain) a numerical compilation of published data and metadata:
  - Many historical experimental data **evaluations** are hidden in E(“Encrypted”)NDF files
- EXFOR contains compilation errors and could be improved regarding consistent formatting

- EXFOR is such an important nuclear database that all above issues need to be addressed, some inside and some outside EXFOR.

- The eventual curated experimental database will be outside EXFOR

Nuclear data evaluation and validation in ~~2030~~ 2040

- No more GUI's in the data flow
  - getlib n U 233 -lib endfb9.0 -MT 18 -format GNDS
  - getexp n U 233 -normalization Zolotarev -format json
  - Super nuclear model code < input > output (including data library)
  - NJOY21 < data library
  - geticsbep U 233 -sort energy (+ getirphe, getsinbad)
  - Monte Carlo or determ. code < all ICSBEP, etc. input files
  - Final analysis (including GUI) and loop back to 1 if necessary

If we would have a high-quality, machine-readable, evaluated experimental reaction data library

- Efficient use in nuclear data evaluation
  - Neutron standards
  - Data projects could do more than a few isotopes per year
  - Much better complement to model-based data evaluation
- Machine learning:
  - Cross section trend analysis without using model codes
  - Evaluation tables from sentiment analysis, etc. etc.
- Flexible automated plotting
- Flexible automated optimization, Bayesian Monte Carlo etc.
- Global assessment of predictive power of nuclear model codes
- Efficient nuclear model and systematics development, e.g.
  - New optical models
  - New Kalbach systematics
  - New Fission Yield systematics, etc etc
  - (often > 50% of the time is wasted by collecting and massaging data)

Quality scoring for EXFOR

- Quality scores for 28455 EXFOR subentries
  - Natalia Dzysiuk for activation c.s. + (all c.s.) Ni: 2336 subentries
  - Erwin Alhassan for proton induced reactions: 166 subentries
  - Natalie Gaughan for proton induced reactions: 103 subentries
  - Arjan Koning for neutron activation cross sections: NEA report NEA/DB/DOC(2017)1: ***Statistical verification and validation of the EXFOR database: (n,gamma), (n,n'), (n,2n), (n,p), (n,alpha) and other neutron-induced reaction cross sections***: 25850 subentries

Quality scoring table in YAML 28455 subentries

subentry: 30336037
evaluator : Natalia Dzysiuk (2018)
quality : 0
comment :
data were not used for evaluation
an overestimated value of monitor cross section

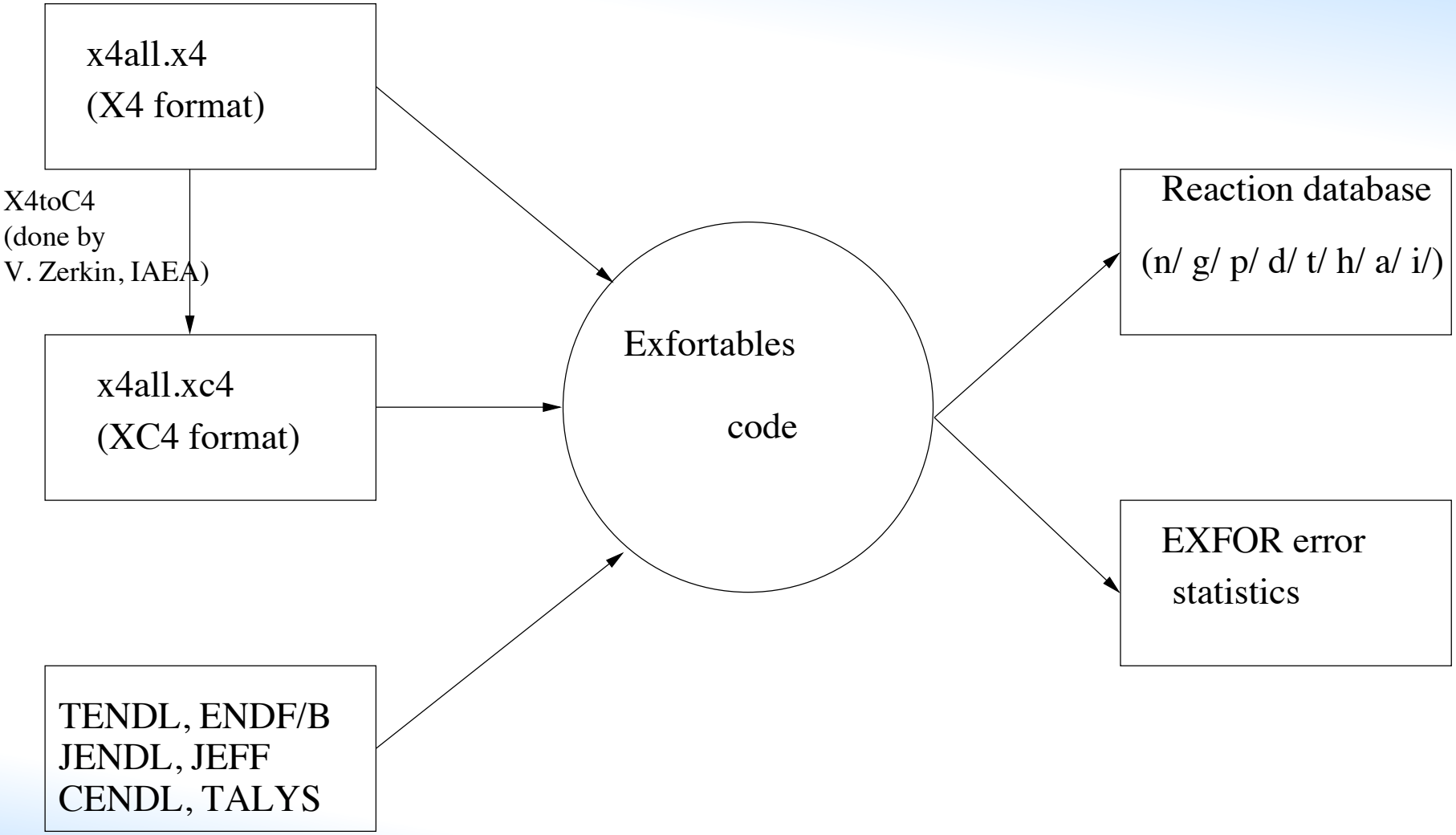
subentry: P0019006
evaluator : Erwin Alhassan (2019)
quality : 0
comment :
Erwin Alhassan (PSI, 2018): 0
(1 -> accept and 0 -> reject)
Reasons for inclusion/exclusion:
----------#
1) Experimental data set not consistent with other experiments such as Takacs (2005) and Tarkanyi (1994)

subentry: B0048005
evaluator : Natalie Gaughan (2019)
quality : 1
comment :
IAEA-TECDOC-1211 - Data selected.

subentry: O1062003
evaluator : Natalie Gaughan (2019)
quality : 0
comment :
Cross section data from this publication was selected by IAEA-TECDOC-1211. However, O1062003 is yield data, and cross section data for this publication and reaction was found in EXFOR.

Will be available to SG50 and also in IAEA EXFOR correction system

EXFORTABLES code and database



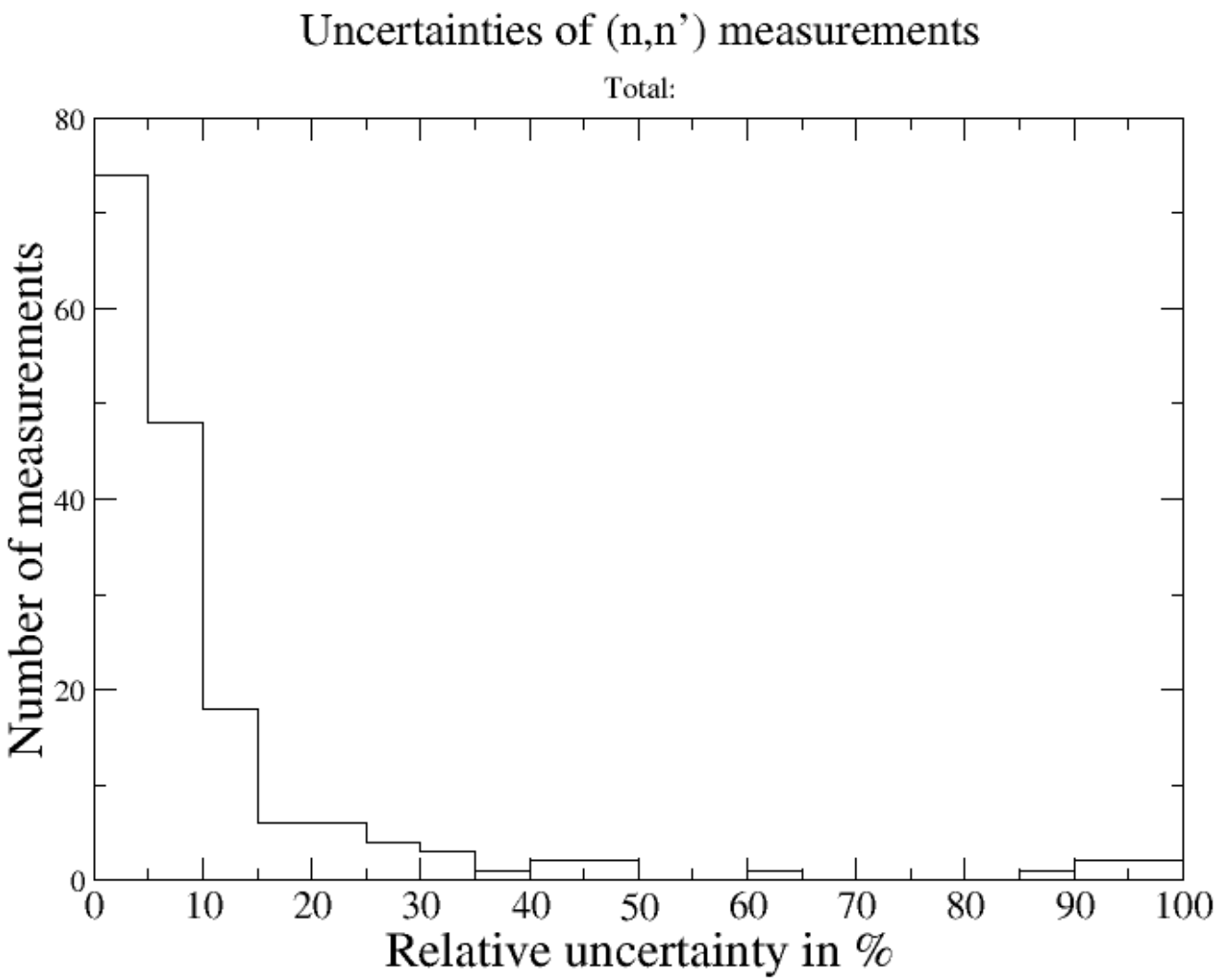
Very complete output of statistics of EXFOR entries (quality vs data libraries)

F (~ C/E) value per EXFOR entry

- Suggests a ‘quality’ per paper
- These F values are sorted over the entire EXFOR, from ‘best’ paper to ‘worst’ paper
- Can lead to a ‘quality’ per author
- …but we should use that with care!!!!

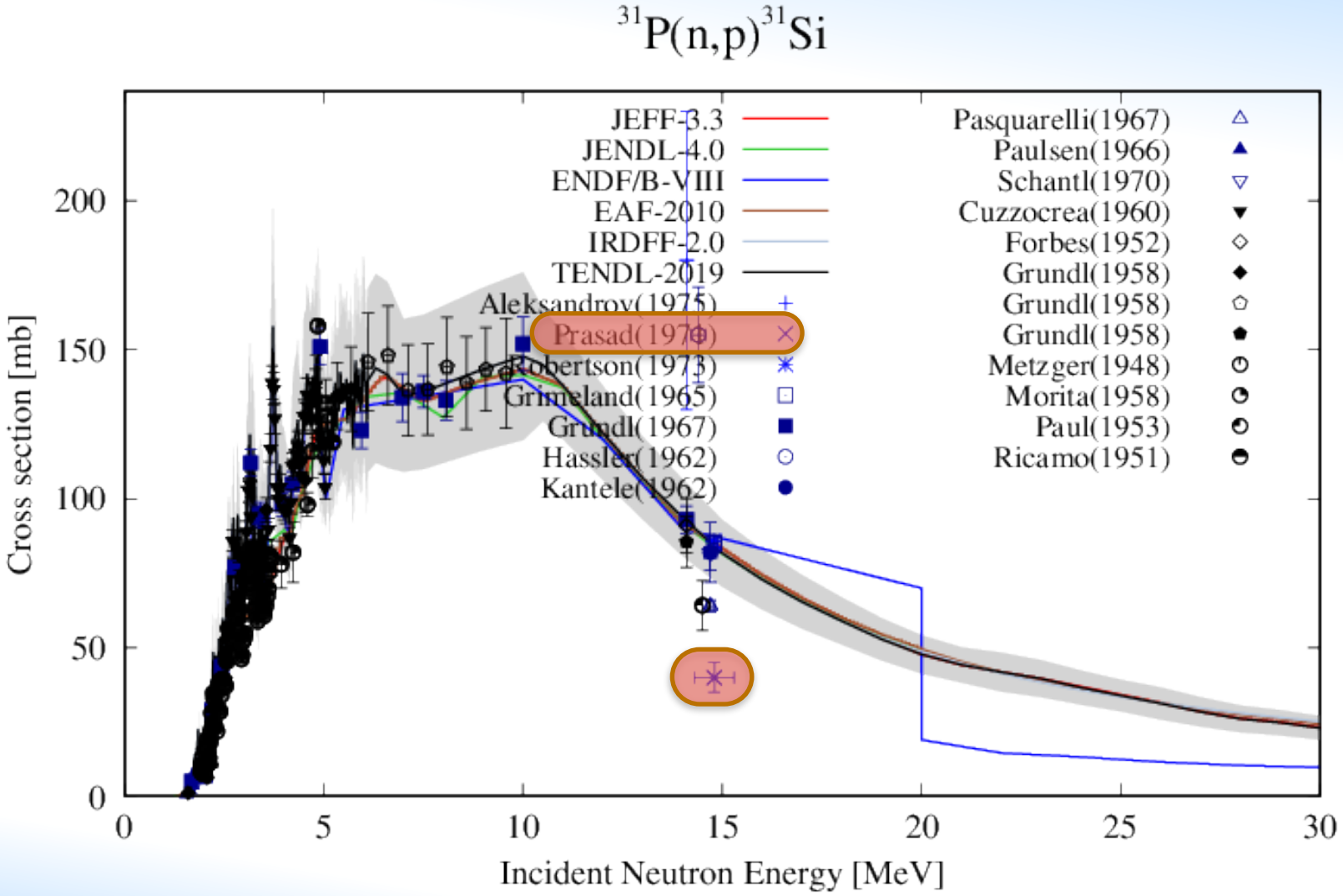
#ENTRY AUTHOR YEAR Evaluation Quality
22820 V.Semkova+ 2004 F= 1.19 T1: 3 T2: 0 T3: 0 N1: 0 N2: 0 N3: 0 R1: 7 R2: 0 R3: 0

#SUBENT	N	Reaction	F	Quality	endfb8.0	tendl.2017
22820002	12	28-NI-58(N,P)27-CO-58,,SIG	1.22	R1	1.20	1.16
22820004	12	28-NI-58(N,X)27-CO-57,,SIG	1.05	T1	1.04	1.04
22820006	7	28-NI-58(N,P)27-CO-58-M,,SIG	1.35	R1	0.00	1.31
22820007	9	28-NI-58(N,2N)28-NI-57,,SIG	1.07	R1	1.06	1.06
22820008	4	28-NI-60(N,P)27-CO-60-M,,SIG	1.11	R1	0.00	1.12
22820009	1	28-NI-61(N,X)27-CO-60-M,,SIG	1.18	T1	0.00	1.12
22820010	5	28-NI-61(N,P)27-CO-61,,SIG	1.15	R1	1.06	1.11
22820011	4	28-NI-62(N,X)27-CO-61,,SIG	1.54	T1	1.29	1.75
22820012	13	27-CO-59(N,2N)27-CO-58,,SIG	1.06	R1	1.07	1.04
22820013	5	27-CO-59(N,2N)27-CO-58-M,,SIG	1.10	R1	0.00	1.14
#SUBENT	N	Reaction	F	Quality	endfb8.0	tendl.2017
		Average:	1.19		1.13	1.19

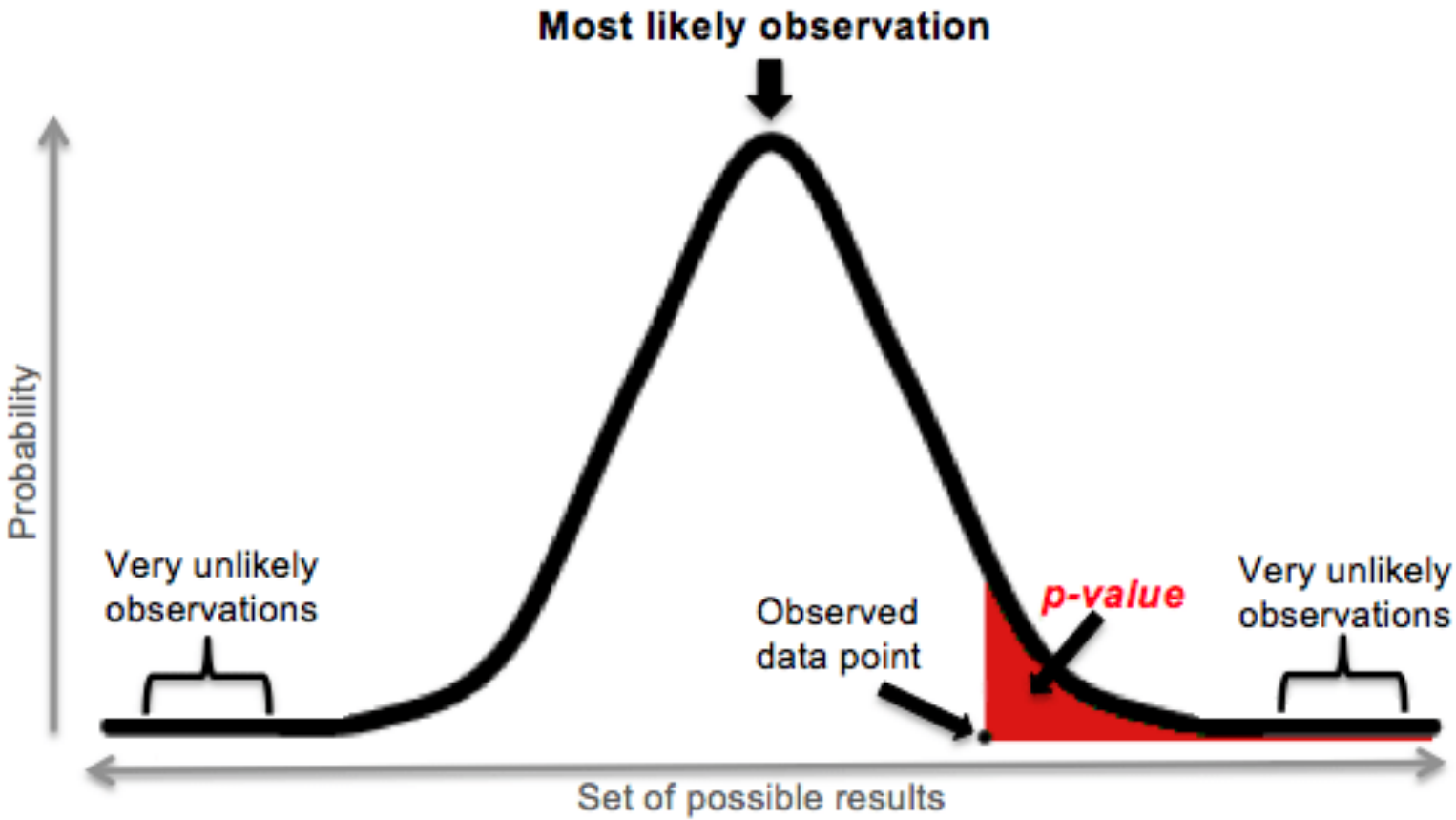


If a measurement has no uncertainty, but we still want to use it,
should we assign uncertainties on the basis of all other measurements?
(of course, OUTSIDE EXFOR)

Exforcism: Identifying and expelling evil data

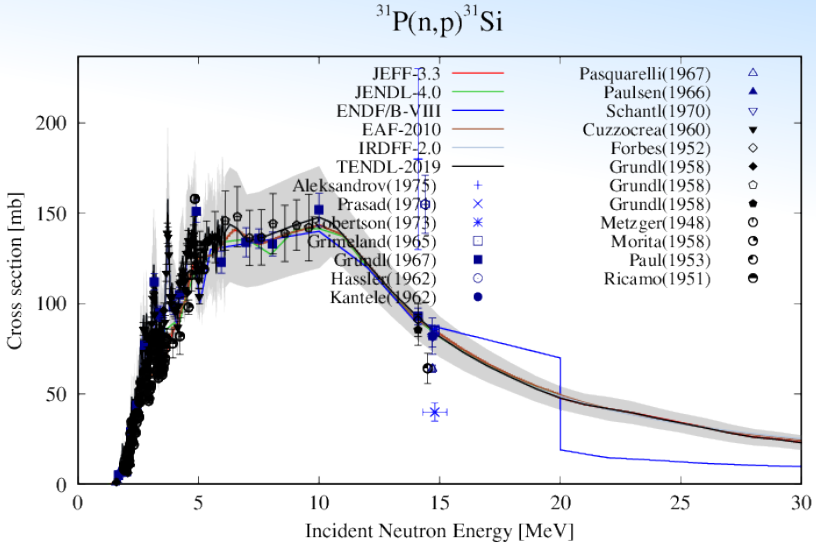


Towards a unified measure for outliers: p values



A *p-value* (shaded red area) is the probability of an observed (or more extreme) result arising by chance

Loose translation for EXFOR: the p-value is the probability that the true cross section value is the observed data point or more extreme than that



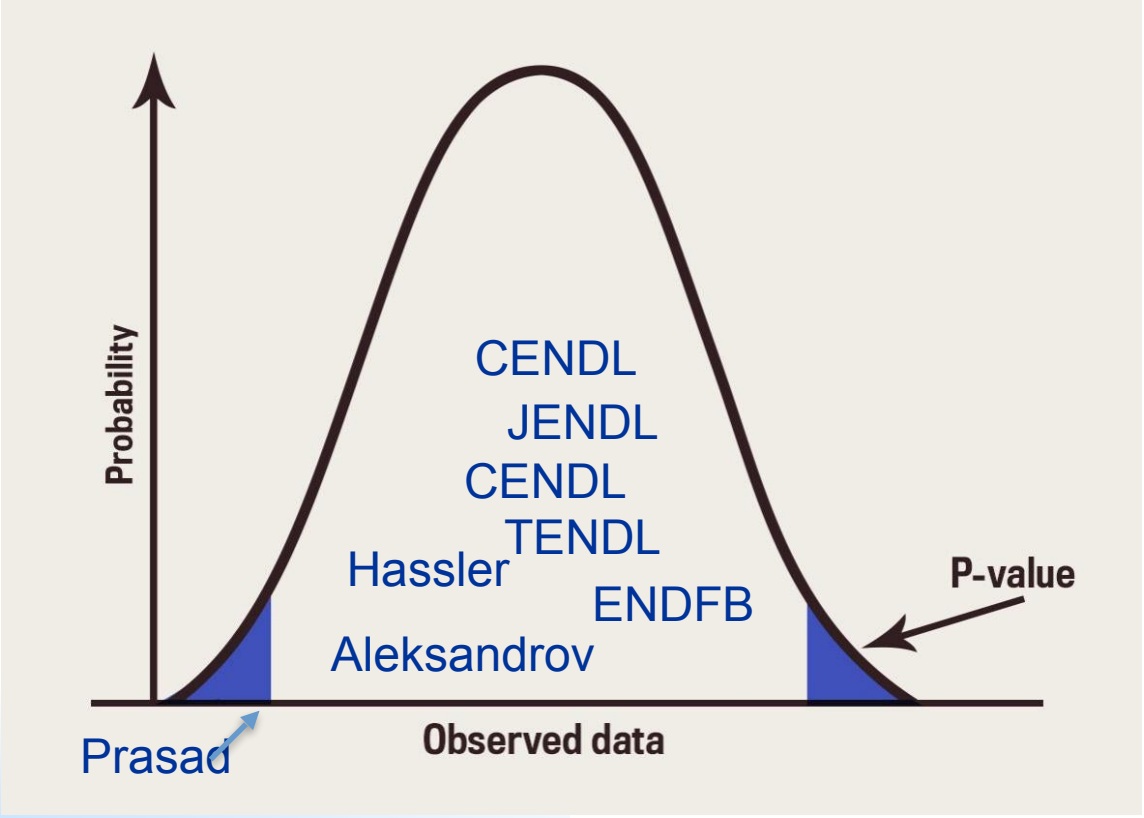
For every single EXFOR data point a probability distribution can be constructed from all alternative values, from libraries and other measurements

Cut at 14.8 MeV P31(n,p)
probability distribution for Prasad measurement

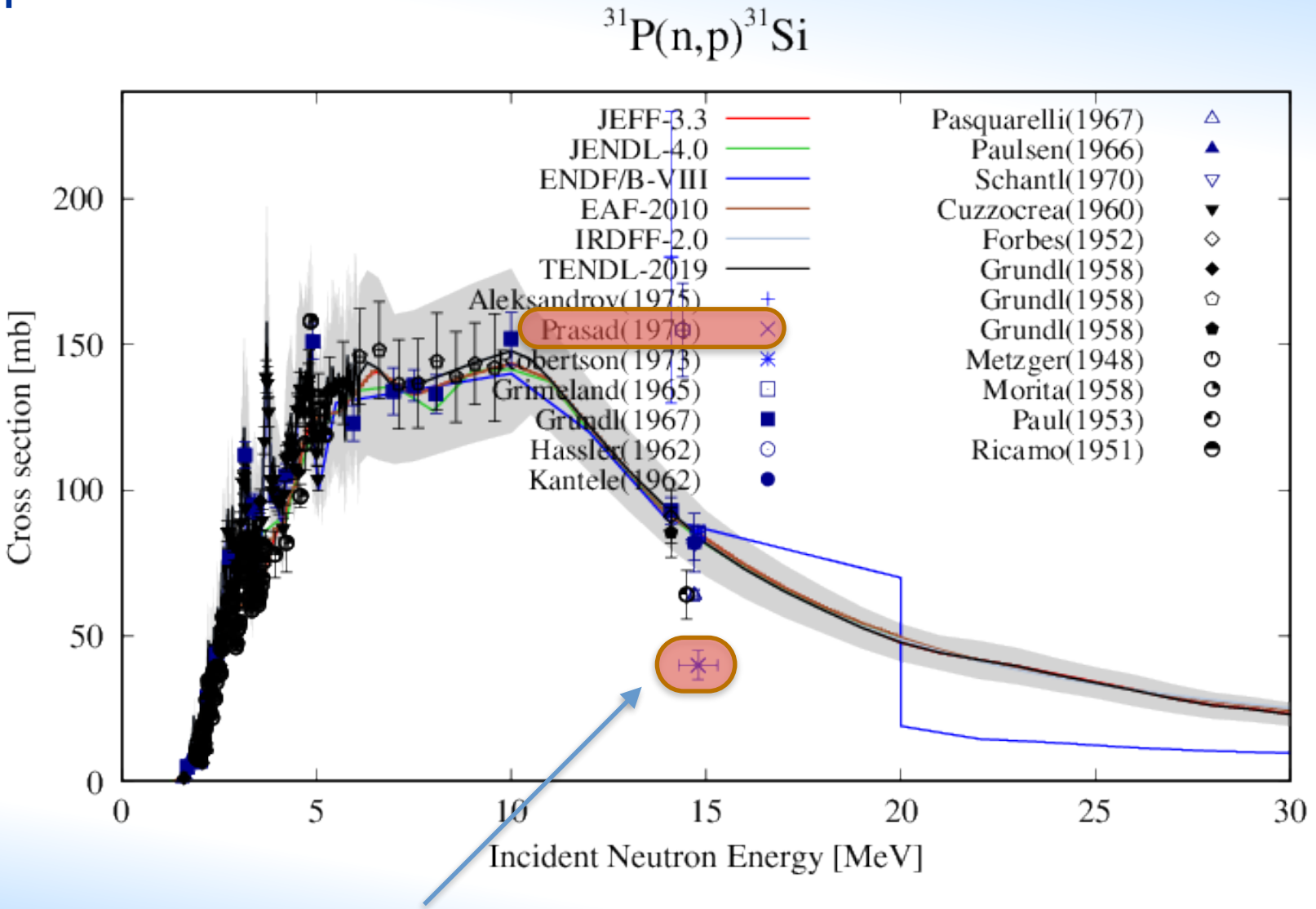
The more independent information exists around the average, the worse it is for the outlier

For (semi-)automatic nuclear data evaluation: discard all measurements for e.g p < 0.05

Prasad(1970) p value: 0.016



Exforcism: Can we click on graph and declare data points as outlier?



Mouse click and direct coupling to subentry number

Machine learning

- Front-end (EXFOR compilation)
  - Identification of EXFORable articles: web crawling + probabilistic methods
  - Semi-automatic production of EXFOR entries from the pdf files
- Back-end (data evaluation)
  - Sentiment analysis, e.g.

(from a publication or report)
…"As usual, Koning et al did not take multiple scattering corrections into account so we discarded their measurement. We have more confidence in the data of Plompen et al, since…."

EXFOR nr.	Author	Score
123456.789	Koning	0.1
987654.123	Plompen	0.95

Scope: Isotopes, energy range and observables
A. Global

- Unify metadata as much as possible (for machine-readability and machine-interpretation)
- Can we have numerical control over all data?
  - Cross sections: (basically) yes, though still inconsistencies with e.g. exclusive vs activation cross sections
  - Angular distributions, mixture of degrees and Legendre coefficients
  - Spectra, non-ambiguous storage of data, difficult to get data out sorted by incident energy and emission energy and angle
  - Fission Yields, danger of hiding important data under non-suggestive reaction identifiers (recent FY CRP)
  - etc.
- (?) Produce alias tables, when equivalent data for a user are stored under different EXFOR identifiers

Scope: Isotopes, energy range and observables
B. High-priority (which is very subjective)

- CIELO isotopes: ^{1}H, ^{16}O, ^{56}Fe, ^{235,238}U, ^{239}Pu
- n-induced: 0 - 20 MeV
- Observables:
  - Cross sections
  - nubar
  - Prompt fission neutron spectra
  - Fission yields
  - Angular distributions (?)
- Consistency of metadata in general, systematic uncertainty components
- Reason 1: may get more interested people on board
- Reason 2: high-quality (low uncertainty) constraints, will have positive impact on other isotopes, energy ranges as well

Summary: Essential steps

1. Make EXFOR machine-readable via XC4, JSON, etc, for all data
2. Make experimental data retrieval totally flexible and accessible by command-line API's ("getexfor" or "getexp")
3. Statistical comparison of all exp. data with other data and nuclear data libraries, model codes: global quality assessments and feedback (compilation errors, etc.)
4. Establish a curated experimental reaction data library by performing automating operations on EXFOR (IAEA system?)
5. Develop advanced templates for consistent and complete metadata (classification of systematic uncertainties etc.)
6. Develop Machine Learning techniques for both front-end (compilation) and back-end (data trends, sentiment analysis, etc.)

Finally

- EXFORtables will soon be made available
- YAML files with quality scores are available

IAEA 60 Years
Atoms for Peace and Development

Thank you!