wosmip
Workshop on Signatures of Man-made Isotope Production

1st Nuclear Explosion Signal Screening Open Inter-Comparison Exercise 2021

Christian Maurer[1], Paul Skomorowski[1], Ramesh S. Sarathi[2], Alexander Hieden[1], Boxue Liu[3], Jonathan Baré[3], Jerome Brioude[4], Delia Arnold Arias[1], Yuichi Kijima[3], Brian T. Schrom[2], Jennifer M. Mendez[2], Anne Tipka[3], Jolanta Kusmierczyk-Michulec[3], Martin Kalinowski[3], and Robin Schoemaker[3]

[1]*Zentralanstalt fuer Meteorologie und Geodynamik (ZAMG)*
[2]*Pacific Northwest National Laboratory (PNNL)*
[3]*Comprehensive Nuclear-Test-Ban Treaty Organization/International Data Center (CTBTO/IDC)*
[4]Laboratoire de l'Atmosphère et des Cyclones *(LACy)*

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

1. Test data set structure

48 date-times * 53 grid points * 23 IMS time series = 58512 data files

Four per month

Uniformly distributed over the globe with denser sampling of latitudes compared to longitudes

Samples for the whole year 2014 according to actual data availability

Each data file potentially impacted by one explosion

2544 explosion scenarios:

Immediate release, 0.92% venting	24 hours containment, 10% release within one hour

Burnett et al. (2019) underwater & IDC underground source term:

- **No mixing of explosions.**
- **At maximum 14 days are influenced by a hypothetical explosion.**
- Sample **meta data included** (MDC, LC ect.).
- The **full data set cannot be processed/handled during the *1st Nuclear Explosion Signal Screening Open Inter-Comparison Exercise 2021*** -> **reduction to 424 scenarios** (8 date-times, four target periods for ATM) **- <u>not necessarily explosions</u>**

ZAMG Zentralanstalt für Meteorologie und Geodynamik

2. Participants

Name	Institution	Country	Confidential emission data (IRE & ANSTO) requested	ATM + meteorology combination	Level 1 results (ATM only) submitted	Level 2 & 3 results (screening of test data set with own methods) to be submitted
P. de Meutter, A. Delcloo & C. Gueibe	SCKCENRMI	Belgium	Yes	FLEXPART V10.4 + ECMWF	Yes	Yes
S. J. Leadbetter	MetOffice	UK	Yes	NAME 8.3 + MetOffice Unified Model	Yes	No
J. Kusmierczyk-Michulec	CTBTO („XeBet“)	Austria	-	FLEXPART V9 + ECMWF	Yes (Xe-133 only)	No
M. Schoeppner	IAEA	Austria	Yes	FLEXPART V9 + ECMWF	Yes	Not likely
P. Tayyebi	NSTRI, AEOI	Iran	Yes			
J. Roberts, J. Lucas	US NDC	US	Yes			
S. Wang, Q. Li, Y. Zhao	BRL	China	Yes			
U. A. Kadiri, H. A. Muhammed, I. Dodo	CGG	Nigeria	Yes			
A. Quérel, D. Quélo, O. Saunier	IRSN	France	Yes			
M. Goodwin, D. Chester	AWE	UK	Yes			
R.S. Sarathi	PNNL	US				

So far 4 participants, international interest

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

3a. Evaluation: *Detection Power*

- **Question**: *"Is a measurement an anomaly (regardless of what has caused it)?"*
- **Approach based on ATM of civil sources** (use of Level 1 results - tricky):

1. Calculate residuals between the test data set values and a participant's civil background estimates per IMS station and separately for all radioxenon isotopes.
2. Filter the test data set according to LC in order to prevent accounting for samples below LC that could be solely due to the detector background.
3. Claim a detection if a certain percentile value of all the residuals is exceeded for a sample.
4. Calculate the true positive and false positive rates for any of the four xenon isotopes.
5. Optionally: Apply a moving average [t-1,t+1] to both time series before residual calculation to prevent relying on single sample values.

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

3b. Evaluation: *Screening Power*

- **Question:** *„Has an underground or underwater nuclear explosion to be assumed based on isotopic ratios?"*
- **Approach:**

  Based on all claimed (true and false) positives according to detection power evaluation and on multi-isotope detections (2 to 4 isotopes) evaluate true positive and false positive rates for:

  I. Three and four xeonon isotope discrimination relations (*Kalinowski et al., 2010*):

  $$R = \frac{AC_x}{AC_y};\ u^2(R) = R^2\left(\frac{u^2(AC_y)}{AC_y^2} + \frac{u^2(AC_x)}{AC_x^2}\right);\ u^2(AC_x) = ERR_{AC}{}^2_{x,testset} + \frac{AC^2_{x,modelled}}{S_x}$$; S_x: detector sensitivity

  $$R_{a,b} < K_{a,b,c,d}\, R_{c,d}{}^{m_{a,b,c,d}}$$

  II. Comparison to xenon flags for xenon isotope pairs: Xe-133m/Xe-131m > 2, Xe-135/Xe-133 > 5, Xe-133m/Xe-133 > 0.3 and Xe-133/Xe-131m > 1000

  a) Bayesian limits (*Zaehringer and Kirchner, 2008*):

  $$AC_x^- = AC_x + u(AC_x)NORMSINV(1 - 0.975NORMDIST(AC_x/u(AC_x))$$
  $$AC_x^+ = AC_x + u(AC_x)NORMSINV(1 - 0.025NORMDIST(AC_x/u(AC_x))$$
  $$R^-_{x,y} = AC^-_x/AC^+_y;\ R^+_{x,y} = AC^+_x/AC^-_y$$

  b) Fieller's theorem (*Axelsson et al., 2014*):

  $$R^{\pm}_{x,y} = \frac{1}{(AC_x^2 - 4u_x^2)}\left\{(AC_x\,AC_y - \rho 4u_x u_y) \mp \sqrt{(AC_x\,AC_y - \rho 4u_x u_y)^2 - (AC_x^2 - 4u_x^2)(AC_y^2 - 4u_y^2)}\right\}$$

ZAMG Zentralanstalt für Meteorologie und Geodynamik

3c. Evaluation: *Timing Power*

- **Question**: *„Can we determine time zero +/- uncertainty within a predefined time window?“*
- **Approach**:

1. For Xe-133 and Xe-133m: $R_{133m/133}(t) = \frac{e^{-\lambda_{133m}t}}{e^{-\lambda_{133m}t}\frac{\lambda_{133}}{\lambda_{133}-\lambda_{133m}}\left(1-e^{-(\lambda_{133}-\lambda_{133m})t}\right)+\frac{1}{R_{133m/133(0)}}e^{-\lambda_{133}t}}$
2. <u>If Xe-133m is not present:</u> E.g., $R_{135/133}(t) = R_{135/133}(0)e^{-(\lambda_{135}-\lambda_{133})t}$

   Analogous, simple relations for Xe-133m/Xe-131m and Xe-133/Xe-131m (no parent-daughter decay).
3. <u>If Xe-133m is present</u>: E.g., $R_{135/133}(t) = \frac{e^{-\lambda_{135}t}}{\frac{1}{R_{135/133m(0)}}e^{-\lambda_{133m}t}\frac{\lambda_{133}}{\lambda_{133}-\lambda_{133m}}\left(1-e^{-(\lambda_{133}-\lambda_{133m})t}\right)+\frac{1}{R_{135/133(0)}}e^{-\lambda_{133}t}}$

   Analogous relation for Xe-133/Xe-131m (Parent-daughter decay to be considered if Xe-133 is involved).
4. Evaluate timing success rates based on single samples which where found to be true positives after detection and screening power evaluation and on a 10% tolerance criterion.

	Uncertainty	Tolerance (10% of the total uncertainty)
Xe-135/Xe-133	57 h	6 h
Xe-133/Xe-131m	45 d	108 h
Xe-133m/Xe-131m	24 d	58 h
Xe-133m/Xe-133	16 d	38 h

For the purpose of estimating the uncertainty the release scenarios include one case at hour zero (immediate release) and another at 24 hours as well as U-235 and Pu-239 fission materials.

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

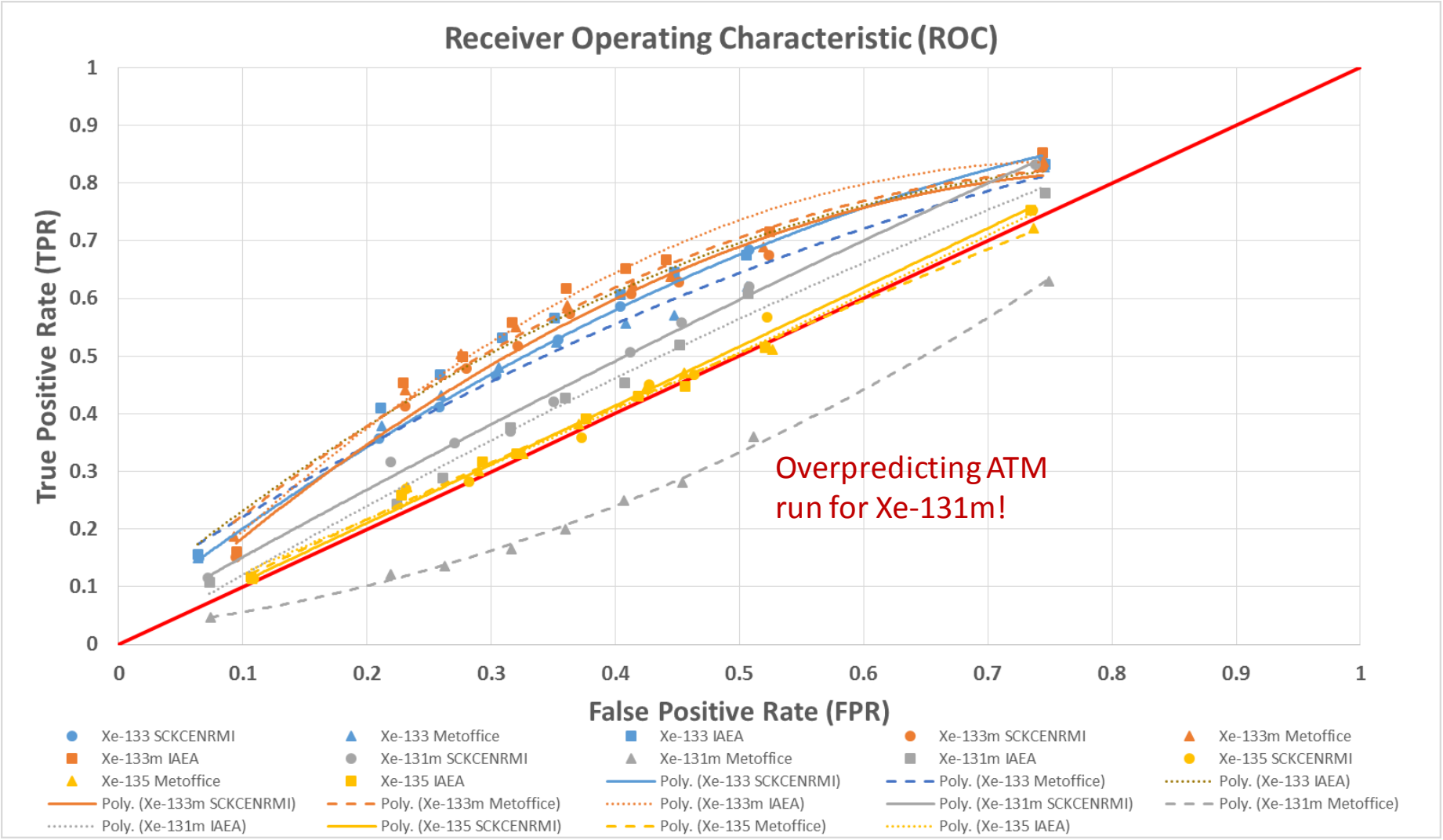
3d. Evaluation: *Location and Magnitude estimation Power*

- **Approach: Very limited evaluation**
  1. Location Power: Calculate the percentage of cases for which there are 1) two, 2) three or 3) more than three detections related to a nuclear explosion regardless of the isotope. (PSR fields can be calculated blending different isotopes as well as detections and non-detections. Minimum is one detection and two non-detections.)
  2. Magnitude estimation power: If there are two detections related to a nuclear explosion, location and releases for two isotopes could be estimated. If there are three detections related to a nuclear explosion, location and releases for three isotopes could be estimated. If there are four detections, location and releases for two or up to four isotopes could be estimated (depending on whether there are different two- or three-isotope ratios involved in case two two- or three-isotope ratios are present). Count the number of different detected isotopes for each of the different above settings (i.e., 1) two, 2) three or 3) more than three detections regardless of the isotope).

  Include only samples in the statistics which where found to be true positives after detection and screening power evaluation.

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

4a. Detection power based on ATM for civil sources



- **Models** tend to **produce similar output** (-> see ensemble analysis of 3rd ATM Challenge).
- There is **some skill for Xe-133 and Xe-133m.**
- There is **hardly/no skill for Xe-131m and Xe-135.** But **ATM runs** (especially source terms) **need to be checked!**
- The **optimum residual threshold** can be **empirically determined**, **ranging** approximately **from the 55th to 70th percentile** (depending on the isotope and the specific ATM run).

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

4b. Detection power for different data sets

Jouden index = Sensitivity (= TPR) + Specificity (= 1-FPR) -1; [-1,1]

Table 1: Jouden-indices (J_{70th}) for *detection power*. [1]: only 19 instead of 23 stations modeled, 4 stations omitted due to inappropriate collection times, FLEXPART runs as of 2014, Xe-133 only

Run	**All tests (321)**	All tests moving average	Underground (100)	Underwater (221)	*Pretended (103)*	Tropics (149)	Extratropics (172)	January (106)	July (106)
Xe-133									
SCKCENRMI-1Mio	**0.16**	0.18	0.33	-0.01	*0.69*	0.21	0.15	0.11	0.19
SCKCENRMI-5Mio	**0.16**	0.18	0.32	-0.01	*0.69*	0.21	0.14	0.11	0.19
Metoffice	**0.17**	0.17	0.34	0.01	*0.69*	0.22	0.16	0.00	0.37
IAEA	**0.22**	0.22	0.38	0.05	*0.69*	0.23	0.22	0.16	0.32
CTBTO[1]	**0.04**	0.06	0.09	-0.01	*0.69*	0.05	0.03	-0.07	0.18
CTBTO[1] - whole year of 2014	**0.07**	0.09	0.14	-0.01	*0.70*	0.09	0.06	-0.05	0.21
Xe-133m									
SCKCENRMI-1Mio	**0.20**	0.18	0.26	0.13	*0.68*	0.26	0.17	0.34	0.04
SCKCENRMI-5Mio	**0.20**	0.18	0.26	0.13	*0.68*	0.26	0.17	0.34	0.04
Metoffice	**0.23**	0.19	0.29	0.17	*0.68*	0.28	0.21	0.33	0.05
IAEA	**0.24**	0.21	0.28	0.20	*0.68*	0.30	0.23	0.41	0.08

- **Slightly higher overall detection power for Xe-133m than for Xe-133** (-> source term + civil Xe background)
- **Higher detection power for underground compared to underwater tests** (-> source term)
- **No detection power for Xe-133, but small one for Xe-133m for underwater tests** (-> source term)
- **Higher detection power for tropics compared to extratropics** (-> lower civil Xe background in the tropics?)
- **Considerable differences between seasons** (-> ?)
- Longer period with civil background predictions -> better results

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

4c. Screening & timing power with/without ATM support

Table 2: Jouden-indices (J_{70th}) for *screening power*. [1]: Samples are selected according to samples predicted by ATM for the four periods of interest.

Approach	2-isotope ratios (# TP detection results)	3-isotope ratios (# TP detection results)	4-isotope ratio (# TP detection results)
With ATM (SCKCENRMI-1Mio)	0.15 (272)	0.10 (52)	0.71 (7)
Without ATM[1]	0.08 (701)	0.08 (194)	0.58 (19)

Table 3: *Timing success rates* for true positive screening results. [1]: Samples are selected according to samples predicted by ATM for the four periods of interest.

Approach	2-isotope ratios (# TP screening results)	3-isotope ratios (# TP screening results)	4-isotope ratio (# TP screening results)
With ATM (SCKCENRMI-1Mio)	0.29 (414)	0.27 (5)	0.55 (5)
Without ATM[1]	0.16 (611)	0.2 (15)	0.57 (11)

- **Use of ATM enhances screening and timing power results** to different extents. Largest improvements are seen for 2-isotope screening and subsequent timing.
- Only **combination of ATM + 4 isotope ratio screening enables a more save claim of a nuclear test** ($J > 0.7$)

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

4d. Location and magnitude power counting statistics

Table 4: *Location* and *detection magnitude estimation power*. Percentages of tests with a specific amount of true positive detections based on ATM.

Run	2 True Positives	3 True Positives	> 3 True Positives - based on 4-isotope screening	> 3 True Positives - based on 2 or 3-isotope screening combinations
SCKCENRMI-1Mio	24%	2%	1%	21%
Metoffice	24%	2%	1%	18%
IAEA	21%	2%	1%	17%

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

5. The problem of false positives

Table 5: Percentage of tests with at least one TP detection and with Jouden-index (J_{70th}) above a specific threshold both based on ATM.

Run	# tests with at least one TP detection	# tests with Jouden-Index ≥ 0.6
Xe-133		
SCKCENRMI-1Mio	53%	11%
Metoffice	52%	11%
IAEA	50%	8%
Xe-133m		
SCKCENRMI-1Mio	34%	6%
Metoffice	38%	10%
IAEA	33%	9%
Xe-131m		
SCKCENRMI-1Mio	37%	11%
Metoffice	17%	4%
IAEA	35%	12%
Xe-135		
SCKCENRMI-1Mio	31%	7%
Metoffice	30%	8%
IAEA	29%	8%

We **could detect half of the tests based on Xe-133, 23 IMS stations and radioxenon systems as of 2014**. But **accompanied by a very high average false positive rate** per test!

J above 0.7 is only reached for one test, for Xe-133 and two participants!

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

6. Preliminary conclusions

- **Overall detection power based on different ATM runs is similar.**
- **Detection power per isotope** based on ATM **depends on** the combined **effects of explosion source term magnitude, decay and magnitude of average civil background** (as well background representation by ATM). **ATM results need to be checked.**
- There is a **slight overall positive impact on detection power for Xe-133** (J ranges from 0.16 to 0.22) **and for Xe-133m** (J ranges from 0.20 to 0.24). This is likely related to high fission yields in combination with long half-lifes of these radioxenon isotopes.
- **There is a measurable positive impact on screening and timing power results from detection power analysis** based on ATM.
- **Civil background calculated via ATM needs to be clearly improved. Approach of nudging ATM simulations towards (IMS) observations** as outlined in *Zwaaftink et al. (2018,* https://gmd.copernicus.org/articles/11/4469/2018/gmd-11-4469-2018-assets.html) **to overcome effects of source term and transport errors.**

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

7. Remarks and references

- **Please mind the exercise deadline of June, 30th, as well as templates for submitting results (Level 1 and Level 2+3) !**
- Publication *„Third international challenge to model the medium- to long-range transport of radioxenon to four Comprehensive Nuclear-Test-Ban Treaty monitoring stations"* has just been accepted by the *Journal of Environmental Radioactivity.*

- A. Axelsson, A. Ringbom, M. Aldener, T. Fritioff, and A. Mörtsell (2014): The Impact of System Characteristics on Noble Gas Network Verification Capability for CTBT. Report No. FOI-R-3856-SE, ISSN-1650-1942, Stockholm, Sweden.
- M. B. Kalinowski, A. Axelsson, M. Bean, X. Blanchard, T. W. Bowyer, G. Brachet, S. Hebel, J. I. McIntyre, J. Peters, C. Pistner, M. Raith, A. Ringbom, P. R. J. Saey, C. Schlosser, T. J. Stocki, T. Taffary, and R. K. Ungar (2010): Discrimination of Nuclear Explosions against Civilian Sources Based on Atmospheric Xenon Isotopic Activity Ratios. *Pure and Applied Geophysics* **167**, 517–539.
- vDEC-Virtual Data Exploitation Centre. CTBTO, https://www.ctbto.org/specials/vdec/
- M. Zähringer and G. Kirchner (2008): Nuclide ratios and source identification from high-resolution gamma-ray spectra with Bayesian decision methods. *Nuclear Instruments and Methods in Physics Research A*.

THANK YOU FOR YOUR ATTENTION!

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

Auxiliary material I

Differences regarding the metrics as used in this study compared to the FOI study:

- *Detection Power*: Percentile is used as threshold instead of the MDC. The use of the MDC for the purpose of detecting a nuclear explosion is challenged by the project team in general. The use of ATM to model the civil background probably makes the use of a threshold that depends on the individual modeled time series at a specific IMS station more appropriate.
- *Location Power*: Sample counting approach only – very limited evaluation
- *Rejection Power* in the FOI study vs. *Screening Power* in the current evaluation: No generation of false scenarios, model trajectories, respectively.
- *Timing Power*: Xe-135/Xe-133 is not the only ratio considered, the current evaluation also covers Xe-133/Xe-131m, Xe-133m/Xe-131m and Xe-133m/Xe-133. But no least-square fitting for multiple ratios is applied.

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

Auxiliary material II

- 2 isotope ratios calculated directly for test data set values (no residual approach): Ratios Xe-133/Xe-131m and Xe-133m/Xe-131m are never evaluated likely because of the simultaneous occurence of Xe-133 and Xe-133m with >= LC values. Thus, Xe-133m/Xe-133 vs. Xe-133m/Xe-131m can be evaluated.
- 3 isotope ratios calculated directly for test data set values (no residual approach): Ratio: Xe-135/Xe-133 vs. Xe-133m/Xe-133 is never evaluated likely because of the simultaneous occurence of Xe-135 and Xe-131m with >= LC values. Thus, the 4-isotope relation can be evaluated.
- Test data set (excluding ACs < 0 and ACs impacted by explosions) versus related background values averaged over all stations and tests: Xe-133m and Xe-135 source terms too low?
  1. SCKCENRMI-1Mio: Xe-133: 0.428 vs. 0.248, ***Xe-133m: 0.141 vs. 0.002 (factor 70)***, Xe-131m: 0.052 vs. 0.003, ***Xe-135: 0.212 vs. 0.005 (factor 40)***
  2. IAEA: Xe-133: 0.453 vs. 0.446, ***Xe-133m: 0.150 vs. 0.006 (factor 25)***, Xe-131m: 0.056 vs. 0.050, ***Xe-135: 0.210 vs. 0.007 (factor 30)***
  3. MetOffice: **Xe-133: 0.438 vs. 0.739**, *Xe-133m: 0.143 vs. 0.010*, **Xe-131m: 0.055 vs. 0.262**, *Xe-135: 0.212 vs. 0.022* **(Run needs to be checked - OVERPREDICTING!)**
- Spurious differences in overall level of Xe-131m predicted by participants (SCKCENRMI and IAEA):
  1. SCKCENRMI-1Mio: Xe-133: 0.193, Xe-133m: 0.002, **Xe-131m: 0.003**, Xe-135: 0.004
  2. IAEA: Xe-133: 0.344, Xe-133m: 0.004, **Xe-131m: 0.051**, Xe-135: 0.005
  3. MetOffice: **Xe-133: 0.745**, Xe-133m: 0.007, **Xe-131m: 0.243**, **Xe-135: 0.018** **(Run needs to be checked - OVERPREDICTING!)**

ZAMG
Zentralanstalt für Meteorologie und Geodynamik