3rd ATM-Challenge 2019 and Radioxenon Nuclear Explosion Signal Screening Inter-Comparison Exercise 2021

C. Maurer (ZAMG), J. Kusmierczyk-Michulec (CTBTO/IDC), P. Skomorowski (ZAMG), J. Baré (CTBTO/IDC), D. Arnold Arias (ZAMG), A. Hieden (ZAMG), B. Liu (CTBTO/IDC), B. T. Schrom (PNNL), J. M. Mendez (PNNL), R. S. Sarathi (PNNL), A. Tipka (CTBTO/IDC), A. Malo (CMC) , A. Crawford (NOAA-ARL) , J. Brioude (LACy), and M. Kalinowski (CTBTO/IDC)

ZAMG: Zentralanstalt fuer Meteorologie und Geodynamik, Vienna, Austria
CTBTO/IDC: Comprehensive Nuclear Test-Ban Treaty Organization/International Data Center, Vienna, Austria
PNNL: Pacific Northwest National Laboratory, Washington, Richland, USA
CMC: Canadian Meteorological Center, Québec, Dorval, Canada
NOAA-ARL: National Oceanic & Atmospheric Administration - Air Resources Laboratory, Maryland, College Park, USA
LACy: Atmosphere and Cyclone Lab, University de La Réunion, La Réunion, France

1. The 16 participants of the 3rd ATM-Challenge and their model set-ups

Main aims: Investigate the added of 1) stack emission data and 2) training an optimum ensemble

Name	Institution	ATM	Meteorology	Simulation length	Minor emitters included
Kihyun Park (Korea)	KAERI	LADAS	UM-GDAPS (KMA)	6 months	Yes
Arnaud Quérel (France)	IRSN	IdX-C3X (Eulerian)	ARPEGE (Météo France)	6 months	No (ZAMG‘s)
Akiko Furuno (Japan)	JAEA	WSPEEDI-II (WRF+GEARN)	GPV- Global (JMA)	3 months	No (ZAMG‘s)
Alain Malo (Canada)	CMC	MLDP	GDPS (CMC)	6 months	Yes
Donald Lucas (USA)	LLNL	LODI	NCEP-GFS/ADAPT	6 months	Yes
Paul Eslinger (USA)	PNNL	HYSPLIT	NCEP-GDAS	6 months	Yes
Yuichi Kijima (Japan)	JAEA	HYSPLIT	NCEP-GDAS	3 months	No (ZAMG‘s)
Rich Britton (UK)	UK-NDC/AWE	HYSPLIT	NCEP-GDAS	6 months	Yes
Blake Orr (Australia)	ARPANSA	HYSPLIT	ACCESS-G (BoM)	6 months	No (ZAMG‘s)
Alice Crawford (USA)	NOAA-ARL	HYSPLIT (-GEM)	NCEP-GDAS & ERA5	6 months	Yes
Anders Axelsson (Sweden)	FOI	HYSPLIT	NCEP-GDAS	6 months	Yes
Jolanta Kusmierczyk-Michulec (CTBTO)	CTBTO/IDC	FLEXPART 9.3.2	ECMWF-IFS	6 months	Yes
Christian Maurer (Austria)	ZAMG	FLEXPART 10.3	ECMWF-IFS	6 months	Yes
Michael Schoeppner (CTBTO)	CTBTO/OSI	FLEXPART 9.3.2	ECMWF-IFS	3 months	Yes
Petra Seibert (Austria)	BOKU	FLEXPART	ECMWF-IFS	6 months	Yes
Pham Kim Long (Vietnam)	VINATOM	FLEXPART	NCEP-GFS	6 months	Yes

2. A first glimpse on importantstatistics

What is the average benefit (over all four investigated stations CAX17, DEX33, SEX63, USX75 and all available samples for June to December 2014 and all submitted runs) of:

- using *actual historic daily stack emission*[1] versus *average literature emission* data for IRE and CNL facilities?
- including rough *estimates for NPPs' & and other facilities'* emissions?

$$\text{Rank} = R^2 + \left(1 - \frac{|\text{FB}|}{2}\right) + \text{F5} + \text{ACC}$$

1) „Rank" according to 2nd ATM-Challenge („R_2nd_Challenge"; 4 metrics combined)
2) „Rank" amended by distribution metric („R_KS"; 5 metrics combined)
3) „Seibert's Skill Score („SS"; 4 metrics combined)

Rank[2]	Actual daily stack emissions	Average literature emissions
NPP emissions included	**2.33 [1.45,2.70]**; 3.08 [2.01,3.56]; 0.46 [0.20,0.61]	**2.39 [1.47,2.73]**; 3.17 [2.04,3.62]; 0.45 [0.19,0.60]
NPP emissions not included	**1.92 [0.91,2.45]**; 2.56 [1.37,3.21]; 0.35 [0.07,0.58]	**2.09 [1.08,2.59]**; 2.78 [1.56, 3.39]; 0.39 [0.09,0.59]

[1]accessed via vDEC
[2]Reviewed data set accessed via vDEC + DEX33 altitude correction applied

Answers:

- No average benefit from daily stack data over all samples, independent of the score used
- Indication of a positive impact of roughly estimated emissions of NPPs and other facilities that adds up to ~20%

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

3. Different approaches for selecting samples

Hypothesis: Benefit of stack emission data depends on the samples selected

Selection methods applied:

- (1) $\frac{|measurement - MIPFs' \ contributions|}{measurement} \leq 50\% \ or \ 80\%$

  or (1a) $\frac{measurement - MIPFs' \ contributions}{measurement} \leq 50\% \ or \ 80\%$

Contributions are calculated based on (A1) FLEXPART V9 bwd runs or a (B1 & B1a) FLEXPART V9-CTBTO fwd run and (A1) 1° or (B1 & B1a) 0.5° meteorological input and output resolution (operational CTBTO/IDC set-up as of 2014 or set-up for 3rd ATM-Challenge 2019).

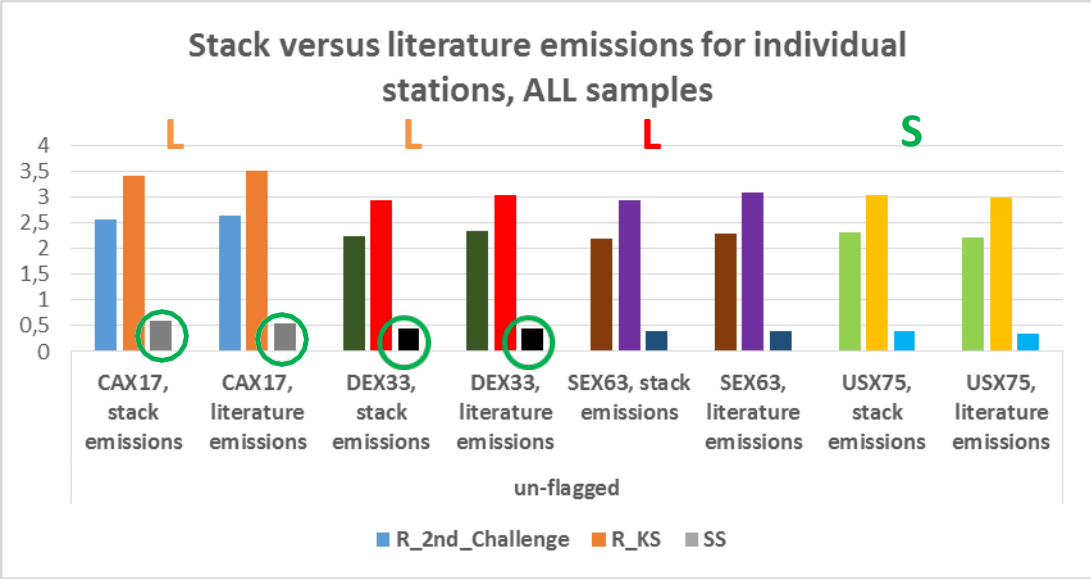
- (2) $\frac{NPPs' + NRRs' + other \ facilities' \ contributions}{simulated \ value} \leq 50\% \ or \ 80\%$

Contributions are calculated based on a FLEXPART V9-CTBTO fwd run and 0.5° meteorological input and output resolution (set-up for 3rd ATM-Challenge 2019).

- (3) Select subjectively a few outstanding daily stack emissions (outstanding with respect to the mean daily value as deduced from disaggregating the annual sum) and related samples predicted by the CTBTO fwd run.

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

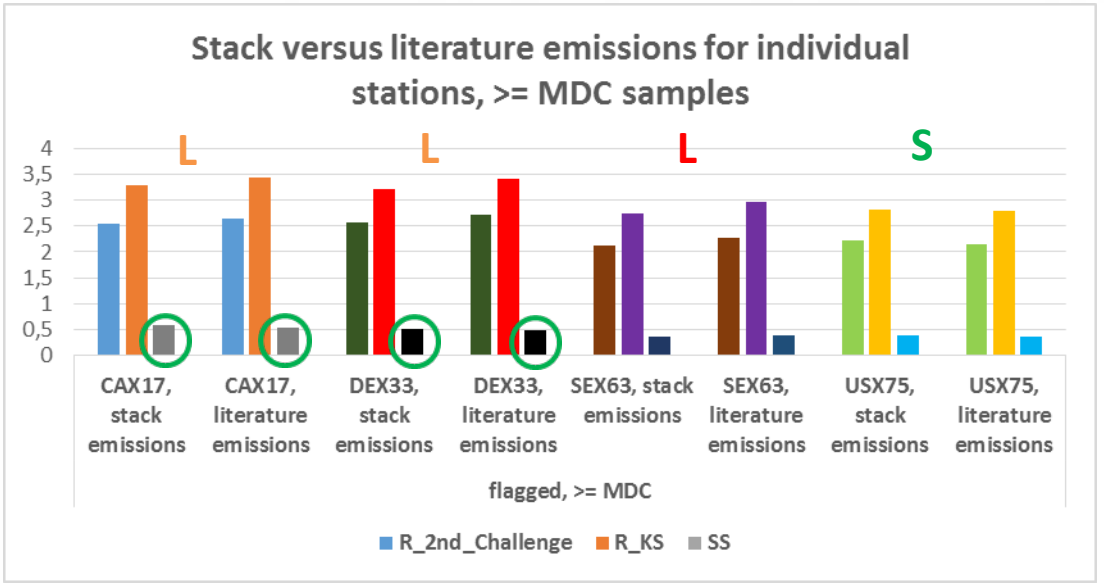
4. Statistics per station – flagging main emitters' influence – approach A1

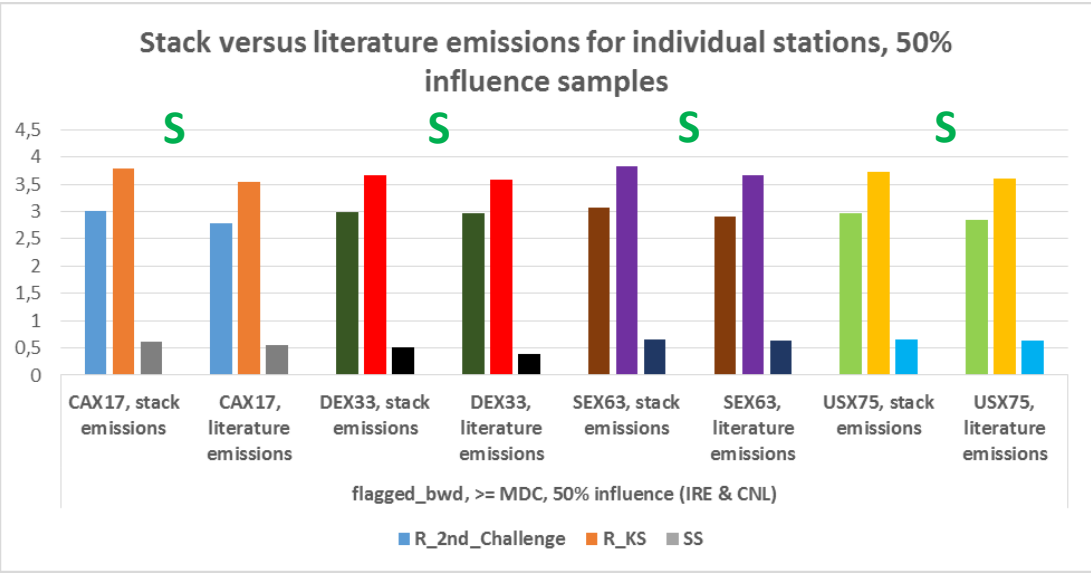


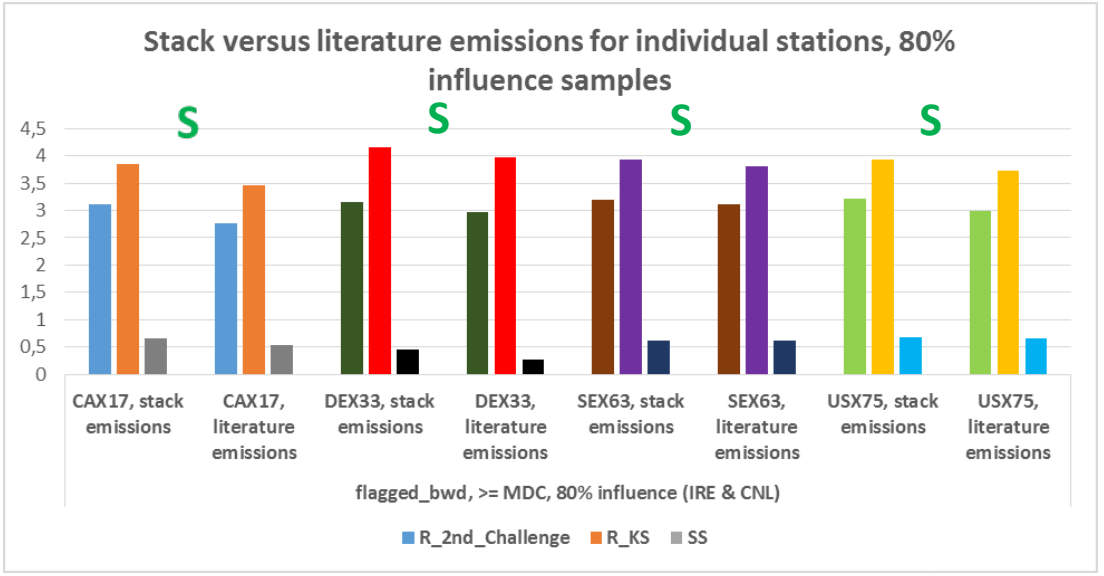
S: stack emissions outperform literature emissions based on all metrics

L: literature emissions outperform stack emissions based on Rank, but not on SS

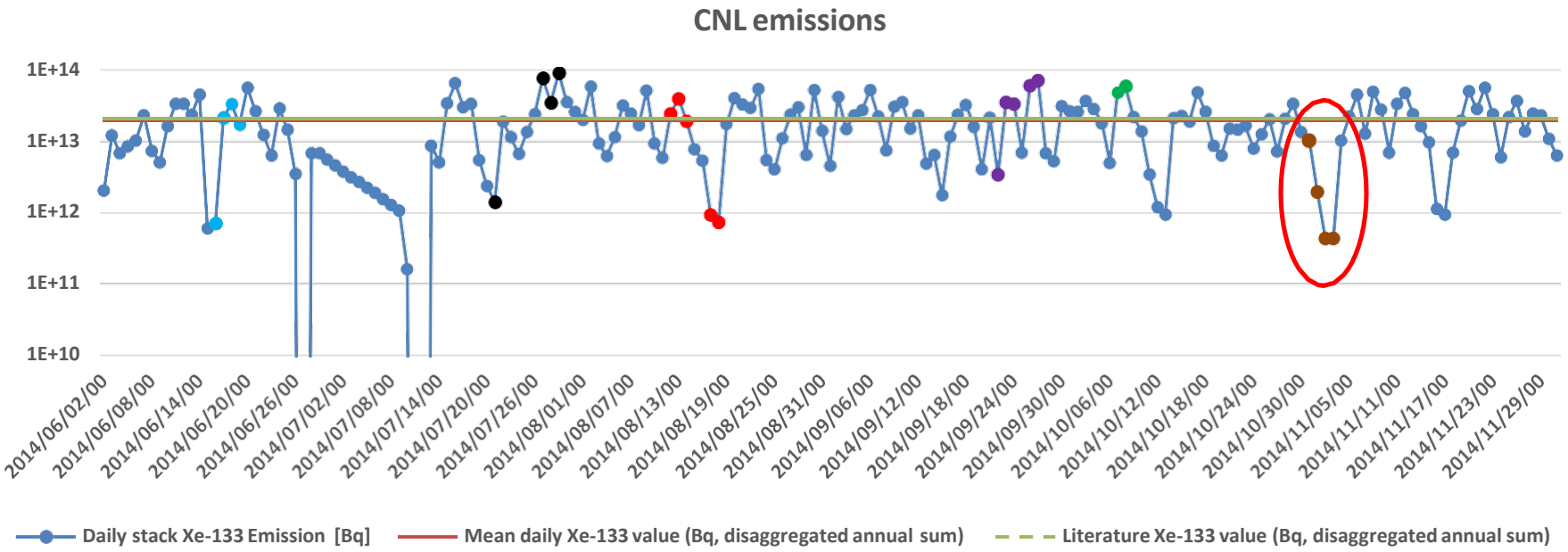
L: literature emissions outperform stack emissions based on all metrics







5.1 A different perspective – approach 3: Selecting samples based on the emission profile: CNL-USX75



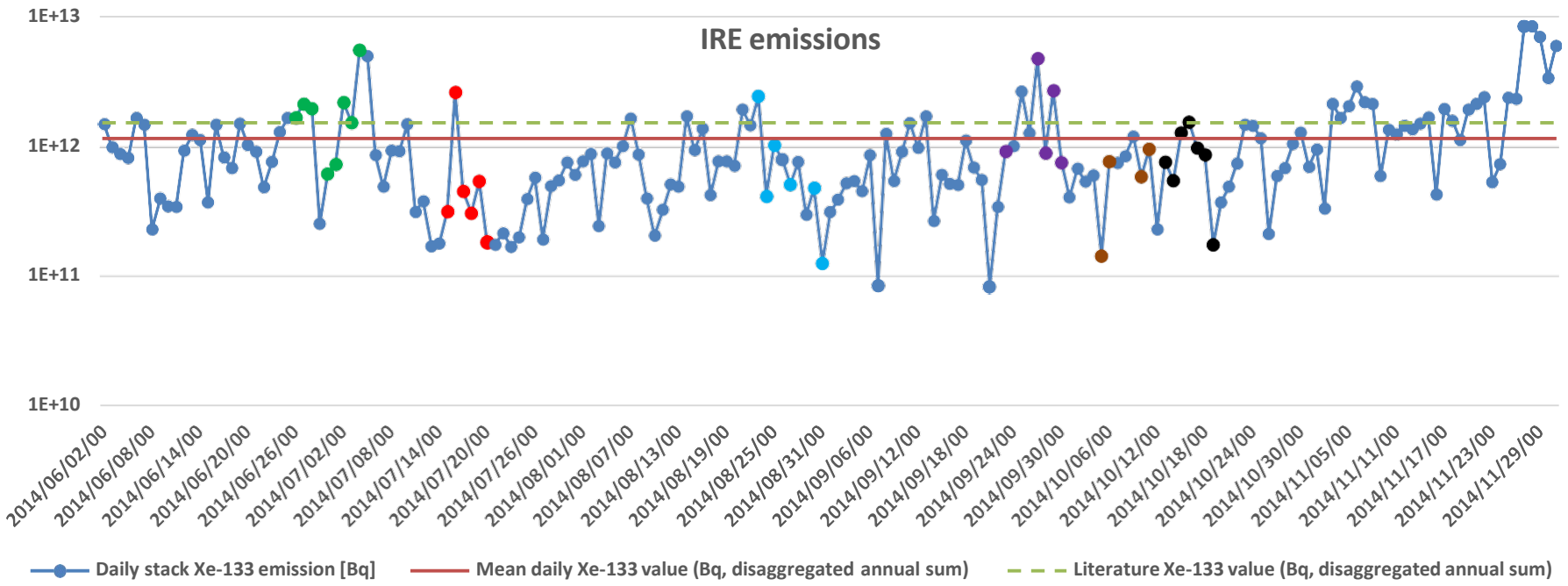
- In 2/3 of the cases the CNL contribution alone does not explain the signal -> *Something is missing!*
- NPPs+NRRs+other facilities` contributions are always and up to two orders of magnitude smaller than CNL stack emission based contributions -> *CNL is the driving force*
- *Stack data are benefitial in 2/3 of the cases*

USX75

Collection start [UTC]	Measured value [mBq/m3]	MIPFs contribution stack	MIPFs contribution literature	NPPs+NRRs+other facilites	Sum stack	Sum literature
20141102160000	0,53	0,88	20,70 S!	0,36 S!	1,24	21,06
20140802160000	*18,51*	3,92	1,07 S?	0,09 S	4,01	1,16
20140620160000	*5,63*	4,82	3,57 S?	0,13 S	4,95	3,71
20140820160000	*0,80*	0,17	0,23 L?	0,10 L	0,27	0,33
20141001160000	4,80	6,85	2,61 L	0,67 L	7,52	3,28
20141014160000	*0,95*	0,38	0,15 S?	0,05 S	0,43	0,20

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

5.2 A different perspective – approach 3: Selecting samples based on the emission profile: IRE-SEX63



- For all cases the IRE (+CNL) contribution alone does not explain the signal -> *Something is missing!*
- NPPs+NRRs+other facilities' contributions are in 2/3 of the cases and up to two orders of magnitude bigger than IRE stack emission based contributions -> *IRE is not the drivin gforc efor thesesamples*
- *Stack data are benefitial in 2/3 of the cases*

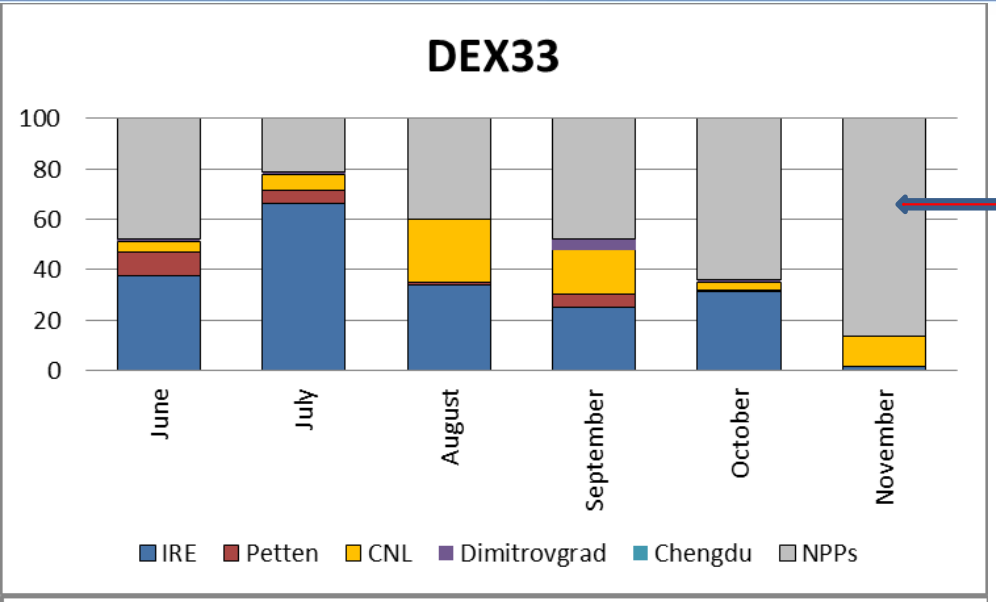
SEX63

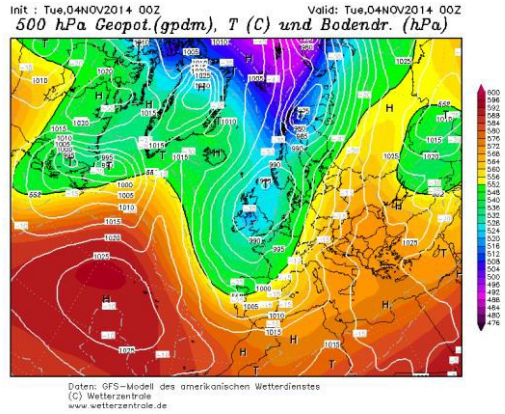
Collection start [UTC]	Measured value [mBq/m3]	MIPFs contribution stack	MIPFs contribution literature		NPPs+NRRs+other facilites		Sum stack	Sum literature
20140707000000	*1,71*	0,16	0,09	S?	0,11	S	0,27	0,20
20140727000000	*0,60*	0,05	0,08	L?	0,54	S	0,60	0,62
20140903000000	*0,41*	0,12	0,41	S?	0,29	S	0,42	0,70
20141024000000	*1,36*	0,02	0,03	L?	1,02	L	1,04	1,05
20141003000000	*1,01*	0,30	0,34	L?	1,52	S	1,82	1,86
20141013000000	*0,72*	0,19	0,48	L?	0,15	L	0,33	0,62

CNL influence!

ZAMG Zentralanstalt für Meteorologie und Geodynamik

6. To remember: Monthly Xe-background in 2014 (PTS pilot study)

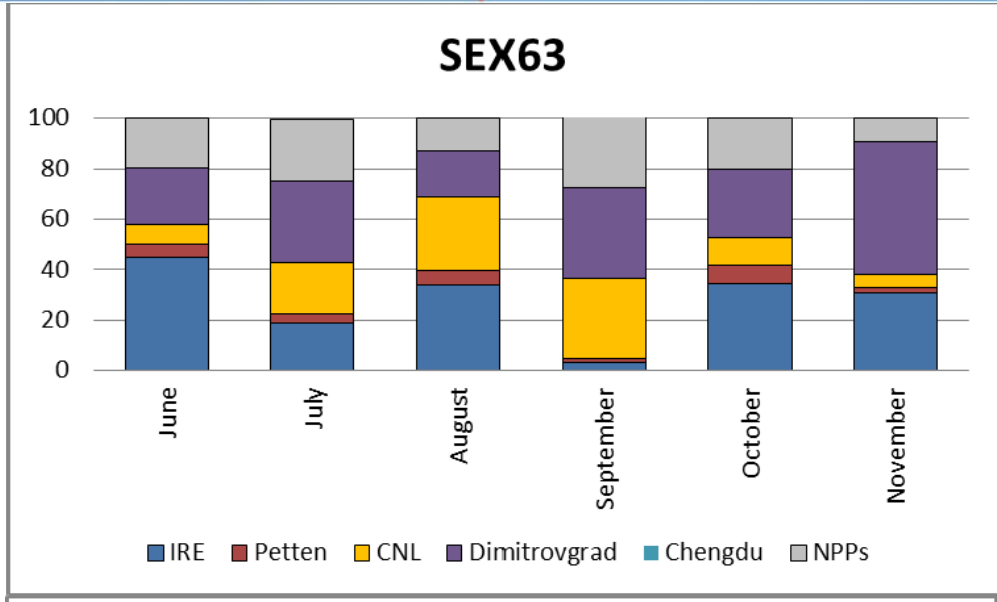


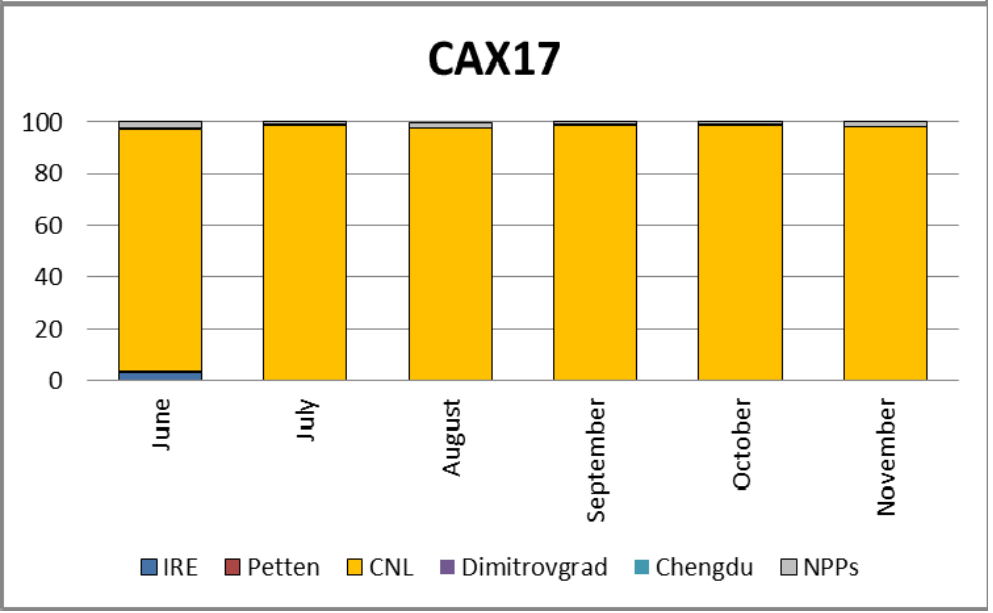


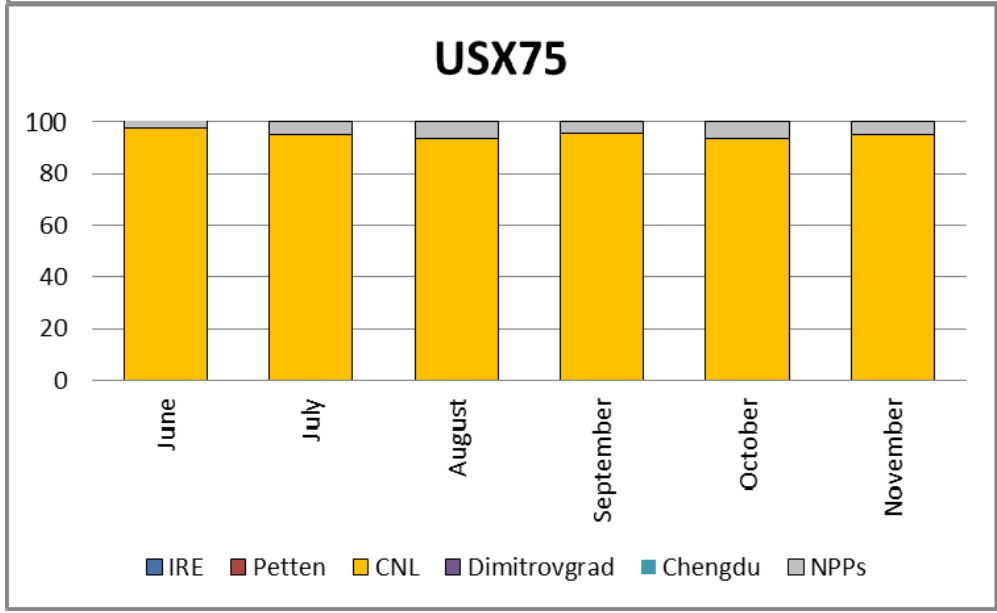
Gueibe et al. (2017):
IRE: 2E15 Bq/y
CNL: 1.5E16 Bq/y

- CNL had highest annual values!
- Knowing exact IRE emissions is clearly not enough for DEX33 and SEX63!

Percentage values are based on actual concentrations in Bq/m3







7.1 Conclusions I

- A huge data pool of modelling results has been created. Please request it from ZAMG/CTBTO (i.e., contact christian.maurer@zamg.ac.at and jolanta.kusmierczyk-michulec@ctbto.org). A more thorough statistical analysis (Phd?) would be desirable.
- It seems to be important to select samples appropriately to demonstrate an – on average – small added value of stack emissions. However, there is considerable benefit from stack data for individual samples (see approach 3).
- It is interesting to note that the mere selection of samples partly (at least to 50%) or predominately (at least to 80%) influenced by MIPFs pushes the scores up most. The relative increase in scores on average (data sets A1, B1 & B1a) adds up to ~15% when switching from all above MDC samples to those with 50% or 80% MIPF influence using literature emissions compared to 7% when additionally switching from literature to stack emissions for 50% or 80% influence samples. This demonstrates that 1) knowing a large emitter and its location as well as 2) a proper average emission is more important than knowing the exact emission profile. Implicitly suppressing samples with overprediction > 50% or 20% in the sample selection process (absolute difference in data sets A1 and B1) can further enhance the scores which demonstrates the effects of the transport error on scores.

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

7.2 Conclusions II

- Simulating the MIPF related radioxenon background at CAX17 without selecting samples according to MIPF influence on average seems to be especially promising since CAX17 is a remote station with (at the time of 2014) dominating CNL influence.
- It seems to be very important to gain more knowledge about non IRE and CNL-related emissions (for 2014). These emissions may be small individually (but can also be big, see MIPF Dimtrovgrad for SEX63), but in any case their sum (e.g., for DEX33) – depending on the predominant synoptic situation – is a decisive factor in accurately predicting the radioxenon background at IMS stations.

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

Radioxenon Nuclear Explosion Signal Screening Inter-Comparison Exercise 2021

- Problem: We need to be able to discriminate radioxenon nuclear explosion signals from industrial background.
- One way forward: CTBTO's contract awarded to ZAMG under EU Council Decision VIII funding *„Xenon Background Estimator: Development of an evaluation system and conducting a competition for the best method on a call-off basis"*. Project start: Jan, 1st, 2021; supposed to run for roughly one year.
- Approach and goals: Creation of a radioxenon test data set for 2014 to be used for subsequent radioxenon test data screening against hypothetical nuclear explosions[1].
  1. The screening procedure should be based on several criteria: Detection, location, magnitude determination, discrimination and timing capabilities (all criteria or additional ones to be tackled in a later phase of the project under the guidance of external experts)
  2. The screening procedure should be exemplarily performed in the frame of an inter-comparison exercise (autumn 2021) to find efficient available method(s) within the community dealing with CTBT verification.

[1]The study by *Axelsson et al. (2014)* was the first study on the complete verification capability on a network level. The design of the data set will partly follow the approach taken by FOI at that time, but will include additional aspects.

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

1. Elements of the test data set

- Explosion release scenarios: hypothetical radioxenon releases from pre-defined hypothetical underground and underwater nuclear explosions distributed over a global grid at different times of the day and year within 2014. No details yet…

Grid of *Axelsson et al. (2014):*

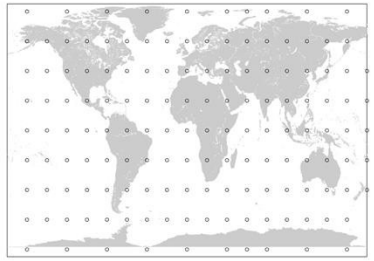
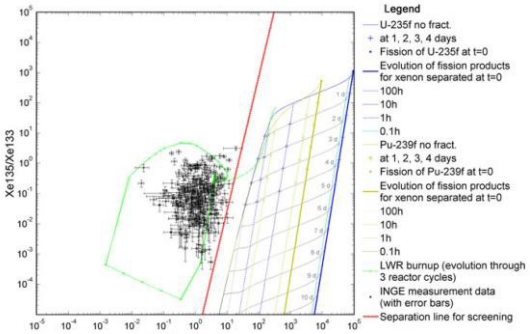


Figure 3.1: The distribution of the hypothetical nuclear explosions used in this study.

Kalinowski and Liao (2012):



- Xe inventories (Xe-133, Xe-133m, Xe-131m & Xe-135) of different reliability: for industrial emitters considering both radiopharmaceutical facilities (we have stack emissions from IRE, CNL and ANSTO!) and nuclear power plants as well as research reactors to model the radioxenon background if not measured – extension of the inventories as used for the 3rd ATM Challenge
- Xe background measurements where available: at ~30 IMS stations operating in 2014
- FLEXPART Source Receptor Sensitivities (SRSs): for all the existing or planned 79 IMS stations calculated in backward mode. Existing noble gas stations with available measurements will be prioritized.

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

2. Set-up of backward atmospheric transport calculations

Parallel (MPI) version of FLEXPART 10.4.1 employed to produce SRS fields. Important aspects:

- Hourly, global ECMWF meteorological re-analysis ERA5 fields (0.5°, 78 vertical levels up to ~100 hPa/~16 km a.s.l.) used as input. Currently probably the best available meteorology for historic times.
- Hourly releases at all 79 IMS stations of an inert tracer represented by 1.34E5 particles tracked back separately for 14 days to flexibly adapt to any (future) collection periods.
- No model time step adaption to Lagrangian time scales for runs at ZAMG's HPC and model time step adaption to Lagrangian time scales (computationally more demanding) at PNNL's HPC for the already existing 39 noble gas stations. The remaining 40 ones will be run at least without time step adaption on ZAMG's or PNNL's HPC.
- Using the mixing ratio option at the receptors. Via multiplication with surface standard density in a post-processing step we can account for radioxenon measurements being valid for a standard atmosphere.
- 0.5° x 0.5° x 100 m hourly output grid

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

3. Necessary post-processing steps

- Multiplication of the individual SRS backward fields with the actual (hourly or daily) time and space-dependent source terms and summing over 1) all source term contributions over time and space for each hourly release chunk and averaging over 2) hourly contributions falling within a given IMS station collection period. Radioactive decay will be applied for all four xenon isotopes. Ingrowth of Xe-133m to Xe-133 will be considered during transport and during sampling.
- Combining hypothetical nuclear explosion signals with real, measured background (for ~IMS 30 stations) or simulated background (for all 79 IMS stations based on the radioxenon inventories) including a proper, iosotpe and device dependent measurement uncertainty. An uncertainty-concentration relationship can be deduced empirically from measured data (see *Haas et al., 2017*) or an analytical relation (*Ringbom et al., 2015*).
- Calculating activity at acquisition start for istopic ratio formation and zero time estimates.
- All work is performed by the contractor and the output data set will be made available to participants.

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

4. Final data sets and inter-comparsion exercise

- 1) Data sets with only simulated or measured background for 79, ~30 stations, respectively. This aims at quantifying a robust false positive rate.
- 2) Data sets with simulated or measured background combined with hypothetical explosion signals. This means creating data subsets comprising all IMS station time series per explosion (no mixing of several explosions).
- Sub-periods of the data sets 1) & 2) consisting of simulated background or simulated background combined with explosion signals due to releases in different seasons and at different times of the day involving all IMS stations will form the basis for the inter-comparison exercise.
- Time line for the event screening exercise: September until December 2021. Please be aware that – unlike for the previous „ATM-Challenges"– you will need atmospheric transport as well as radionuclide expertise to perform the tasks of the inter-comparison exercise. Also, the time schedule is more tight.

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

5. References

- 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.
- P.W. Eslinger, T.W. Bowyer, P. Achim, T. Chai, B. Deconninck, K. Freeman, S. Generoso, P. Hayes, V. Heidmann, I. Hoffman, Y. Kijima, M. Krysta, A. Malo, C. Maurer, F. Ngan, P. Robins, J.O. Ross, O. Saunier, C. Schlosser, M. Schoeppner, B.T. Schrom, P. Seibert, A.F. Stein, K. Ungar, J. Yi (2016): International challenge to predict the impact of radioxenon releases from medical isotope production on a comprehensive nuclear test ban treaty sampling station. *Journal of Environmental Radioactivity* 157, 41-51, https://doi.org/10.1016/j.jenvrad.2016.03.001
- D. A. Haas, P. W. Eslinger, T. W. Bowyer, I. M. Cameron, J. C. Hayes, J. D. Lowrey and H. S. Miley (2017): Improved performance comparisons of radioxenon systems for low level releases in nuclear explosion monitoring. *Journal of Environmental Radioactivity* 178-179, 127-135.
- M. B. Kalinowski and Y-Y. Liao (2012): Isotopic Characterization of Radioiodine and Radioxenon in Releases from Underground Nuclear Explosions with Various Degrees of Fractionation. *Pure and Applied Geophysics*, https://doi.org/10.1007/s00024-012-0580-7
- C. Maurer, J. Baré, J. Kusmierczyk-Michulec, A. Crawford, P.W. Eslinger, P. Seibert, B. Orr, A. Philipp, O. Ross, S. Generoso, P. Achim, M. Schoeppner, A. Malo, A. Ringbom, O. Saunier, D. Quèlo, A. Mathieu, Y. Kijima, A. Stein, T. Chai, F. Ngan, S.J. Leadbetter, P. De Meutter, A. Delcloo, R. Britton, A. Davies, L.G. Glascoe, D.D. Lucas, M.D. Simpson, P. Vogt, M. Kalinowski, T.W. Bowyer (2018): International challenge to model the long-range transport of radioxenon released from medical isotope production to six Comprehensive Nuclear Test-Ban Treaty monitoring stations. *Journal of Environmental Radioactivity* 192, 667 – 686, https://doi:10.1016/j.jenvrad.2018.01.030
- A. Ringbom, A. Axelsson, M. Aldener, T. Fritioff, A. Mörtsell (2015): A novel approach to assess the verification capability of the IMS noble gas network. Oral presentation T4.1-O1 at the Science &Technology Conference 2015, Vienna, Austria.
- vDEC-Virtual Data Exploitation Centre. CTBTO, https://www.ctbto.org/specials/vdec/

THANK YOU FOR YOUR ATTENTION!

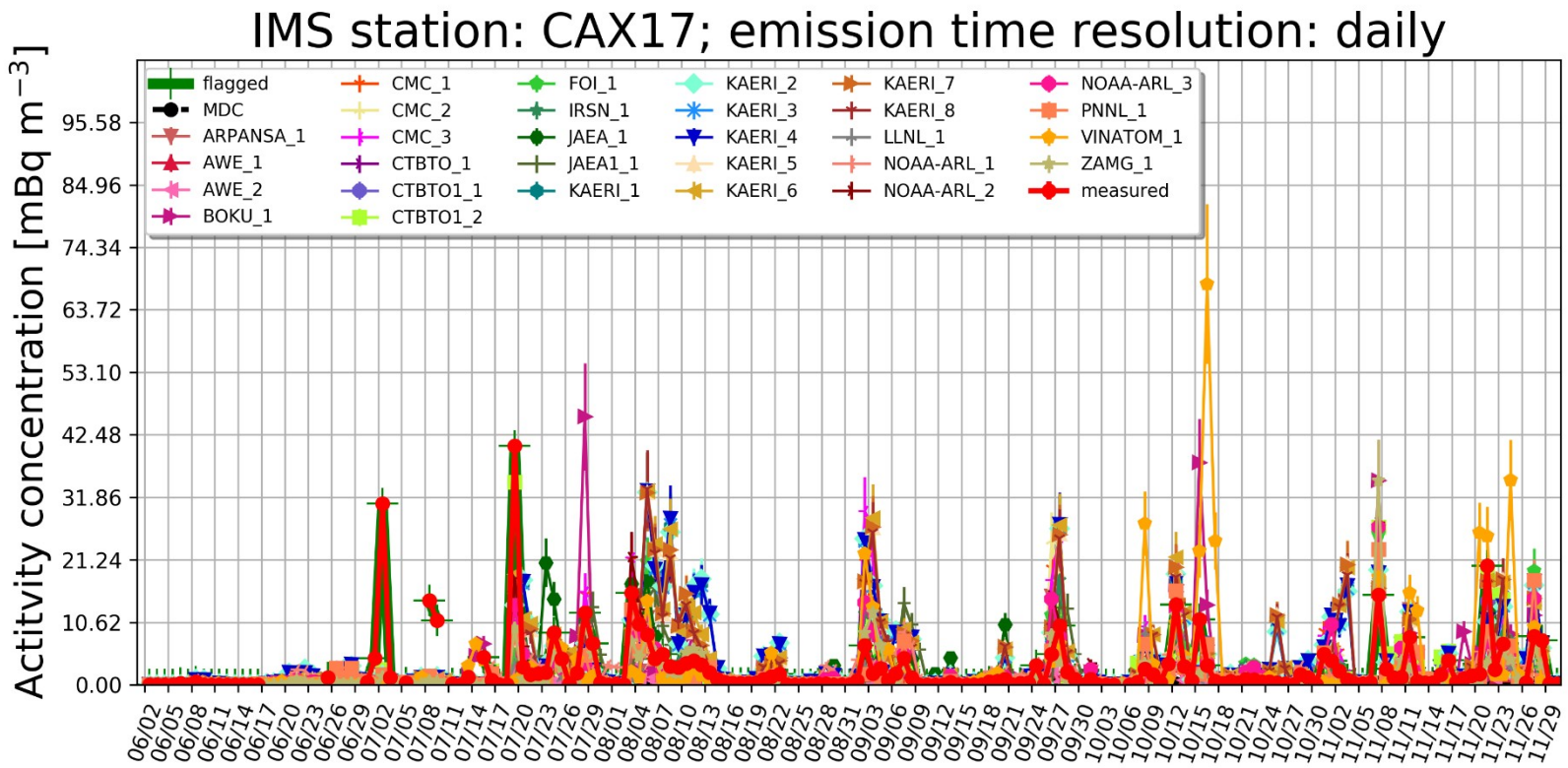
ZAMG
Zentralanstalt für Meteorologie und Geodynamik

Auxiliary material

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

An example: Time series for CAX17

- A lot of valuable data for half a year
- 28 to 31 runs per station (CAX17, DEX33, SEX63 & USX75)



- Ensemble approach has started: Appropriate files were sent to S. Galmarini (JRC, ensemble expert). Results will demonstrate how much independent and redundant information is inherent in the runs.

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

Correcting DEX33 results to STP conditions

- CTBTO-IMS Xe-measurements are valid with respect to STP (standard temperature and pressure, T =288.15 K, p =1013.25 hPa)
- All but BOKU and VINATOM submissions were referenced to ambient conditions ->simulations at DEX33 (~1200 ma.s.l.) are biased low due to reduced air density.
- Rough correction of activity concentrations via multiplying with the density quotient of STP density and average ambient density in the respective output layer.
- Average ambient density in the output layer calculated according to:

$$\rho = \rho_0 (T_0/(T_0 + \gamma h))^{(1+g_0 M/R\gamma)} \qquad (1)$$

$\rho_0 = 1.225\ \mathrm{kg/m^3}$, $T_0 = 288.15\ \mathrm{K}$, $\gamma =$ -0.0065 $\mathrm{K/m}$, $R = 8.314\ \mathrm{kg\,m^2/K\,mol\,s^2}$,
$M = 0.029\ \mathrm{kg/mol}$, $g_0 = 9.81\ \mathrm{m/s^2}$

- Correction on average improves scores just slightly (7% for one metric).
- Not unexpectedly, a positive effect is only pronounced for those runs, where upper output layers were sampled (e.g., ZAMG and CMC runs) and not just the first 100 or 200 mabove model topography.

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

A detailed look on the scores: all stations

Table 1: Average statistics per institution over all stations over all and for individual run-IDs ordered by Rank

Institution	R	FB	F5 [%]	RMSE	NMSE	KS [%]	ACC [%]	NAAD [%]	CRmax ($\hat{t}$)	Rank	Rank_KS	SS
AWE-1	0.18	-1.07	30	5.17	40	47	51	99	0.18 (-1)	**1.30**	**1.83**	**0.19**
AWE	0.16	-0.63	36	7.01	31	44	58	294	0.18 (-1)	**1.45**	**2.01**	**0.20**
AWE-2	0.14	-0.18	41	8.84	22	40	65	489	0.19 (-1)	**1.59**	**2.19**	**0.22**
LLNL	0.35	-0.51	63	4.52	11	33	66	92	0.35 (0)	**2.09**	**2.77**	**0.27**
FOI	0.36	-0.20	60	4.81	9	28	67	108	0.36 (-1)	**2.15**	**2.88**	**0.39**
VINATOM	0.27	0.48	56	11.80	18	21	70	181	0.27 (0)	**2.16**	**2.95**	**0.46**
PNNL	0.35	-0.20	60	4.61	8	34	69	100	0.35 (0)	**2.18**	**2.84**	**0.40**
JAEA	0.40	-0.15	61	4.37	6	25	65	101	0.40 (0)	**2.19**	**2.93**	**0.45**
CTBTO	0.47	-0.57	65	2.79	5	31	66	76	0.49 (0)	**2.25**	**2.94**	**0.45**
ARPANSA-1	0.41	-0.55	68	4.58	10	24	68	76	0.44 (-1)	**2.28**	**3.04**	**0.37**
ARPANSA	0.47	-0.57	70	3.90	9	25	67	74	0.49 (-1)	**2.33**	**3.09**	**0.35**
BOKU	0.32	-0.12	69	5.29	9	17	72	98	0.32 (0)	**2.38**	**3.21**	**0.58**
KAERI-6	0.39	0.34	67	5.71	7	24	75	132	0.41 (-1)	**2.40**	**3.17**	**0.62**
KAERI-5	0.37	0.30	68	5.76	7	22	75	126	0.40 (-1)	**2.42**	**3.21**	**0.62**
ZAMG	0.48	-0.44	70	4.19	8	25	71	70	0.48 (0)	**2.45**	**3.20**	**0.49**
KAERI-8	0.40	0.23	69	5.73	8	19	74	117	0.43 (-1)	**2.46**	**3.27**	**0.61**
KAERI-7	0.39	0.27	69	5.69	7	21	75	119	0.42 (-1)	**2.46**	**3.26**	**0.62**
JAEA1	0.45	0.05	71	4.10	4	18	73	108	0.54 (0)	**2.47**	**3.30**	**0.55**
NOAA-ARL-1	0.41	-0.00	69	4.76	6	28	76	91	0.42 (0)	**2.51**	**3.23**	**0.50**
NOAA-ARL	0.44	-0.23	71	4.56	8	24	75	79	0.46 (-1)	**2.52**	**3.28**	**0.49**
NOAA-ARL-3	0.47	-0.39	73	4.44	9	23	75	71	0.49 (-1)	**2.52**	**3.29**	**0.45**
KAERI	0.43	0.25	70	5.24	6	20	77	111	0.44 (-1)	**2.52**	**3.32**	**0.61**
NOAA-ARL-2	0.45	-0.30	72	4.49	8	22	75	74	0.49 (-2)	**2.53**	**3.31**	**0.52**
ARPANSA-2	0.70	-0.67	74	1.18	4	28	65	65	0.70 (0)	**2.56**	**3.28**	**0.26**
CTBTO1-1	0.48	0.07	72	5.20	11	16	78	100	0.49 (0)	**2.56**	**3.41**	**0.47**
CMC-3	0.43	0.03	76	4.47	6	21	78	92	0.43 (0)	**2.59**	**3.38**	**0.55**
KAERI-2	0.47	0.25	72	4.77	5	22	78	103	0.47 (0)	**2.60**	**3.38**	**0.60**
KAERI-4	0.48	0.19	72	4.75	5	19	79	98	0.48 (0)	**2.61**	**3.42**	**0.58**
KAERI-3	0.47	0.21	72	4.75	5	19	79	99	0.47 (0)	**2.61**	**3.42**	**0.60**
KAERI-1	0.47	0.19	73	4.73	5	18	79	96	0.47 (0)	**2.62**	**3.44**	**0.61**
CMC	0.48	-0.09	78	4.35	6	17	80	79	0.48 (0)	**2.68**	**3.51**	**0.56**
CTBTO1	0.51	0.03	74	4.34	8	15	79	89	0.52 (0)	**2.69**	**3.54**	**0.60**
IBSN	0.53	-0.36	77	4.00	7	14	78	66	0.53 (0)	**2.70**	**3.56**	**0.48**
CMC-1	0.50	-0.03	77	4.20	5	18	80	79	0.50 (0)	**2.72**	**3.54**	**0.60**
CMC-2	0.49	-0.28	80	4.38	7	12	81	67	0.49 (0)	**2.74**	**3.61**	**0.55**
CTBTO1-2	0.58	-0.05	78	2.63	3	13	82	69	0.58 (0)	**2.94**	**3.81**	**0.86**
Average over all institutions	**0.40**	**-0.20**	**66**	**4.99**	**10**	**24**	**71**	**108**	**0.42 (-1)**	**2.33**	**3.08**	**0.46**
Median over all institutions	**0.44**	**-0.20**	**69**	**4.44**	**8**	**24**	**70**	**95**	**0.45 (0)**	**2.35**	**3.14**	**0.47**
Maximum over all institutions	**0.53**	**0.48**	**78**	**11.80**	**31**	**44**	**80**	**294**	**0.54 (0)**	**2.70**	**3.56**	**0.61**
Minimum over all institutions	**0.16**	**-0.63**	**36**	**2.79**	**4**	**14**	**58**	**66**	**0.18 (-1)**	**1.45**	**2.01**	**0.20**

Ranking according to one metric does not necessarily do justice to the submissions!

Average Rank of 2nd ATM-Challenge was 2.06. However, the metrics of the two Challenges should not be compared because of different:

- participants, model set-ups (uniform vs. non-uniform output grid!), model versions
- coverage of seasons, hemispheres
- number of samples above MDC for the 2nd Challenge (very low)

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

A detailed look on the scores: CAX17

Table 2: Average statistics and individual statistics for station CAX17 per institution over all run-IDs and for individual run-IDs ordered by Rank

Institution	R	FB	F5 [%]	RMSE	NMSE	KS [%]	ACC [%]	NAAD [%]	CRmax ($\hat{t}$)	**Rank**	**Rank_KS**	**SS**
AWE-1	0.42	-1.46	23	5.25	25	46	54	86	0.42 (0)	**1.22**	**1.76**	**0.12**
AWE	0.43	-1.41	28	5.21	23	45	58	85	0.43 (0)	**1.33**	**1.89**	**0.13**
AWE-2	0.44	-1.37	32	5.17	20	43	61	83	0.44 (0)	**1.44**	**2.01**	**0.13**
JAEA1	0.12	-0.04	52	6.99	5	7	68	112	0.44 (-1)	**2.19**	**3.12**	**0.74**
LLNL	0.47	-0.89	74	4.87	9	29	78	69	0.47 (0)	**2.29**	**3.00**	**0.24**
VINATOM	0.23	0.03	54	8.04	9	21	74	111	0.23 (0)	**2.32**	**3.11**	**0.77**
CTBTO	0.48	-0.74	71	5.63	7	22	79	69	0.48 (0)	**2.36**	**3.14**	**0.25**
JAEA	0.42	-0.08	59	5.89	4	7	73	87	0.42 (0)	**2.45**	**3.38**	**0.79**
KAERI-4	0.43	0.46	69	6.38	4	14	88	125	0.43 (0)	**2.51**	**3.37**	**0.50**
KAERI-1	0.43	0.46	69	6.32	4	14	87	124	0.43 (0)	**2.52**	**3.38**	**0.51**
KAERI-3	0.43	0.46	70	6.32	4	14	87	124	0.43 (0)	**2.52**	**3.38**	**0.51**
KAERI-2	0.44	0.47	70	6.32	4	14	87	125	0.44 (0)	**2.52**	**3.38**	**0.50**
KAERI	0.45	0.42	71	6.22	4	13	87	118	0.45 (0)	**2.57**	**3.45**	**0.53**
KAERI-6	0.47	0.41	71	6.25	4	11	87	114	0.47 (0)	**2.60**	**3.49**	**0.54**
FOI	0.48	-0.34	75	4.79	5	16	80	69	0.48 (0)	**2.61**	**3.45**	**0.57**
PNNL	0.59	-0.53	70	4.31	5	18	83	63	0.59 (0)	**2.62**	**3.44**	**0.45**
KAERI-5	0.46	0.37	72	6.10	4	11	88	111	0.46 (0)	**2.62**	**3.51**	**0.57**
BOKU	0.37	-0.12	73	6.19	6	11	82	90	0.37 (0)	**2.63**	**3.52**	**0.77**
KAERI-7	0.47	0.37	72	6.09	4	11	88	111	0.47 (0)	**2.63**	**3.52**	**0.57**
ZAMG	0.53	-0.42	79	4.58	5	20	80	64	0.53 (0)	**2.66**	**3.46**	**0.52**
KAERI-8	0.46	0.34	73	5.98	4	11	89	109	0.46 (0)	**2.66**	**3.55**	**0.59**
ARPANSA	0.59	-0.39	78	4.28	4	13	83	65	0.59 (0)	**2.76**	**3.63**	**0.54**
NOAA-ARL-3	0.60	-0.31	79	4.24	4	12	84	63	0.60 (0)	**2.84**	**3.72**	**0.61**
NOAA-ARL-1	0.63	-0.08	70	4.11	3	26	79	72	0.63 (0)	**2.85**	**3.59**	**0.72**
NOAA-ARL	0.62	-0.20	76	4.16	3	17	83	66	0.62 (0)	**2.87**	**3.71**	**0.69**
NOAA-ARL-2	0.64	-0.21	79	4.12	3	12	84	63	0.64 (0)	**2.94**	**3.82**	**0.74**
CMC-2	0.56	-0.13	87	4.45	3	6	89	64	0.56 (0)	**3.01**	**3.95**	**0.80**
IRSN	0.62	-0.20	84	4.09	3	5	89	59	0.62 (0)	**3.02**	**3.97**	**0.71**
CMC-3	0.56	0.03	87	4.85	3	6	91	71	0.56 (0)	**3.07**	**4.01**	**0.89**
CMC	0.57	-0.03	88	4.59	3	5	91	68	0.57 (0)	**3.09**	**4.04**	**0.86**
CTBTO1-1	0.74	-0.15	79	3.53	2	11	87	58	0.74 (0)	**3.13**	**4.02**	**0.83**
CTBTO1	0.74	-0.12	79	3.53	2	10	87	58	0.74 (0)	**3.15**	**4.05**	**0.86**
CMC-1	0.60	0.01	90	4.47	3	3	93	68	0.60 (0)	**3.17**	**4.14**	**0.90**
CTBTO1-2	0.75	-0.09	79	3.53	2	9	88	58	0.75 (0)	**3.18**	**4.09**	**0.89**
Average over all institutions	**0.48**	**-0.32**	**69**	**5.21**	**6**	**16**	**80**	**78**	**0.50 (-1)**	**2.56**	**3.40**	**0.59**
Median over all institutions	**0.48**	**-0.20**	**74**	**4.83**	**5**	**15**	**81**	**69**	**0.48 (0)**	**2.61**	**3.45**	**0.63**
Maximum over all institutions	**0.74**	**0.42**	**88**	**8.04**	**23**	**45**	**91**	**118**	**0.74 (0)**	**3.15**	**4.05**	**0.86**
Minimum over all institutions	**0.12**	**-1.41**	**28**	**3.53**	**2**	**5**	**58**	**58**	**0.23 (-1)**	**1.33**	**1.89**	**0.13**

However, the highest Ranks tend to come with high „Seibert Scores"

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

A detailed look on the scores: DEX33

Table 3: Average statistics and individual statistics for station DEX33 per institution over all run-IDs and for individual run-IDs ordered by Rank

Institution	R	FB	F5 [%]	RMSE	NMSE	KS [%]	ACC [%]	NAAD [%]	CRmax ($\hat{t}$)	Rank	Rank-KS	SS
AWE-2	-0.17	1.79	10	16.44	22	90	63	1682	-0.07 (-3)	**0.87**	**0.97**	**0.04**
AWE	-0.13	0.99	34	9.03	13	61	64	903	-0.08 (-2)	**1.51**	**1.90**	**0.28**
VINATOM	-0.02	1.02	48	4.08	8	32	64	296	-0.02 (0)	**1.61**	**2.29**	**0.16**
FOI	0.16	0.72	45	3.01	6	29	57	213	0.16 (0)	**1.69**	**2.40**	**0.27**
PNNL	0.10	0.71	55	2.01	3	50	65	176	0.10 (0)	**1.85**	**2.35**	**0.36**
CMC-3	0.25	0.72	59	1.97	3	47	66	159	0.25 (0)	**1.95**	**2.48**	**0.39**
KAERI-6	0.45	0.98	58	2.78	4	48	66	223	0.45 (0)	**1.95**	**2.47**	**0.30**
LLNL	0.08	0.47	60	1.79	3	30	64	145	0.08 (0)	**2.01**	**2.71**	**0.41**
KAERI-5	0.39	0.87	62	2.98	5	36	69	203	0.39 (0)	**2.02**	**2.66**	**0.27**
JAEA	0.46	-0.71	65	1.60	4	39	59	70	0.46 (0)	**2.10**	**2.71**	**0.30**
ARPANSA-1	0.42	-0.55	66	1.31	4	19	58	81	0.42 (0)	**2.14**	**2.95**	**0.29**
AWE-1	-0.10	0.18	58	1.62	3	32	64	123	-0.10 (0)	**2.14**	**2.82**	**0.52**
CTBTO1-1	0.36	0.77	72	4.48	13	16	74	167	0.37 (3)	**2.20**	**3.04**	**0.17**
KAERI-7	0.47	0.78	68	2.70	5	33	69	174	0.47 (0)	**2.20**	**2.87**	**0.30**
KAERI-8	0.50	0.71	68	3.02	6	25	65	169	0.50 (0)	**2.22**	**2.97**	**0.27**
KAERI	0.55	0.70	63	2.10	3	38	67	153	0.55 (0)	**2.26**	**2.88**	**0.41**
BOKU	0.30	0.33	70	1.93	4	16	64	121	0.30 (0)	**2.26**	**3.10**	**0.55**
CMC-1	0.49	0.43	61	1.34	2	38	69	110	0.49 (0)	**2.32**	**2.94**	**0.51**
KAERI-2	0.63	0.66	60	1.43	2	50	66	129	0.63 (0)	**2.34**	**2.84**	**0.50**
ARPANSA	0.56	-0.61	70	1.24	4	24	62	73	0.56 (0)	**2.35**	**3.11**	**0.27**
CMC	0.44	0.34	67	1.48	2	31	72	111	0.44 (0)	**2.39**	**3.09**	**0.54**
KAERI-3	0.64	0.56	62	1.34	1	39	68	113	0.64 (0)	**2.43**	**3.04**	**0.53**
NOAA-ARL-1	0.54	0.37	65	1.23	2	49	68	104	0.54 (0)	**2.43**	**2.94**	**0.44**
CTBTO1	0.38	0.38	74	3.11	9	17	75	123	0.39 (1)	**2.45**	**3.28**	**0.50**
KAERI-4	0.65	0.53	62	1.31	1	37	67	109	0.65 (0)	**2.45**	**3.08**	**0.54**
KAERI-1	0.63	0.47	64	1.25	1	36	68	103	0.63 (0)	**2.49**	**3.13**	**0.56**
NOAA-ARL	0.55	0.15	64	1.19	2	36	68	89	0.55 (0)	**2.54**	**3.18**	**0.65**
ARPANSA-2	0.70	-0.67	74	1.18	4	28	65	65	0.70 (0)	**2.56**	**3.28**	**0.26**
NOAA-ARL-2	0.54	0.06	63	1.17	2	29	67	83	0.54 (0)	**2.57**	**3.28**	**0.75**
CTBTO	0.34	-0.11	78	1.64	2	21	73	84	0.34 (0)	**2.58**	**3.37**	**0.74**
ZAMG	0.50	-0.25	74	1.28	3	18	73	71	0.50 (0)	**2.59**	**3.41**	**0.65**
IRSN	0.51	-0.25	75	1.21	3	7	73	70	0.51 (0)	**2.61**	**3.54**	**0.53**
NOAA-ARL-3	0.55	0.03	65	1.16	2	31	69	80	0.55 (0)	**2.63**	**3.32**	**0.75**
CTBTO1-2	0.41	-0.01	77	1.73	4	17	76	79	0.41 (0)	**2.69**	**3.52**	**0.84**
JAEA1	0.61	-0.23	86	1.37	2	23	72	61	0.61 (0)	**2.83**	**3.60**	**0.71**
CMC-2	0.60	-0.14	81	1.12	2	7	81	64	0.60 (0)	**2.91**	**3.84**	**0.72**
Average over all institutions	**0.34**	**0.23**	**64**	**2.38**	**4**	**29**	**67**	**172**	**0.34 (-1)**	**2.23**	**2.93**	**0.46**
Median over all institutions	**0.41**	**0.33**	**66**	**1.71**	**3**	**30**	**66**	**116**	**0.42 (0)**	**2.31**	**3.09**	**0.46**
Maximum over all institutions	**0.61**	**1.02**	**86**	**9.03**	**13**	**61**	**75**	**903**	**0.61 (1)**	**2.83**	**3.60**	**0.74**
Minimum over all institutions	**-0.13**	**-0.71**	**34**	**1.19**	**2**	**7**	**57**	**61**	**-0.08 (-2)**	**1.51**	**1.90**	**0.16**

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

A detailed look on the scores: SEX63

Table 4: Average statistics and individual statistics for station SEX63 per institution over all run-IDs and for individual run-IDs ordered by Rank

Institution	R	FB	F5 [%]	RMSE	NMSE	KS [%]	ACC [%]	NAAD [%]	CRmax ($\hat{t}$)	**Rank**	**Rank KS**	**SS**
AWE-1	0.17	-1.63	19	2.22	63	61	38	91	0.17 (0)	**0.78**	**1.17**	**0.03**
AWE	0.12	-0.90	41	2.26	35	35	51	95	0.17 (0)	**1.48**	**2.14**	**0.31**
LLNL	0.19	-0.88	65	2.14	15	40	60	76	0.19 (0)	**1.84**	**2.44**	**0.15**
CTBTO	0.50	-0.89	61	0.93	4	42	51	73	0.60 (1)	**1.93**	**2.51**	**0.29**
FOI	0.17	-0.74	66	2.16	13	34	65	78	0.17 (0)	**1.97**	**2.63**	**0.21**
PNNL	0.17	-0.63	65	2.17	12	32	64	78	0.17 (0)	**2.00**	**2.68**	**0.25**
ARPANSA	0.19	-0.67	73	2.17	12	31	65	77	0.19 (0)	**2.09**	**2.78**	**0.26**
ZAMG	0.21	-0.57	72	2.17	11	30	66	75	0.21 (0)	**2.14**	**2.84**	**0.31**
JAEA	0.38	-0.44	63	0.99	3	27	59	80	0.38 (0)	**2.15**	**2.88**	**0.49**
AWE-2	0.08	-0.16	62	2.31	8	8	64	98	0.17 (1)	**2.18**	**3.10**	**0.59**
VINATOM	0.21	0.13	64	2.91	10	18	68	120	0.21 (0)	**2.30**	**3.12**	**0.72**
BOKU	0.17	-0.26	73	2.30	9	19	71	89	0.17 (2)	**2.34**	**3.15**	**0.54**
NOAA-ARL-1	0.22	-0.45	78	2.07	9	15	78	68	0.22 (0)	**2.38**	**3.23**	**0.22**
IRSN	0.22	-0.49	80	2.14	10	20	78	69	0.22 (0)	**2.38**	**3.18**	**0.32**
NOAA-ARL-3	0.23	-0.56	84	2.08	10	21	78	67	0.23 (0)	**2.40**	**3.19**	**0.21**
NOAA-ARL	0.21	-0.50	82	2.10	9	18	78	68	0.21 (-2)	**2.40**	**3.22**	**0.24**
NOAA-ARL-2	0.18	-0.48	83	2.14	10	18	79	70	0.19 (-4)	**2.41**	**3.23**	**0.29**
CMC-2	0.25	-0.44	81	2.06	9	16	77	66	0.25 (0)	**2.43**	**3.27**	**0.30**
CMC-3	0.24	-0.33	78	2.09	8	14	78	69	0.24 (0)	**2.46**	**3.32**	**0.41**
CMC	0.25	-0.35	80	2.08	8	14	78	68	0.25 (0)	**2.46**	**3.31**	**0.39**
CMC-1	0.25	-0.29	79	2.09	8	13	78	69	0.25 (0)	**2.49**	**3.36**	**0.44**
JAEA1	0.49	-0.45	79	0.89	2	19	69	65	0.49 (0)	**2.49**	**3.30**	**0.50**
CTBTO1	0.16	0.26	79	3.82	15	8	82	104	0.20 (-1)	**2.50**	**3.42**	**0.52**
KAERI-7	0.18	-0.01	74	2.42	8	11	75	93	0.18 (0)	**2.52**	**3.41**	**0.78**
KAERI-8	0.18	-0.02	75	2.42	8	13	76	93	0.18 (0)	**2.53**	**3.40**	**0.78**
KAERI-5	0.19	-0.01	75	2.36	7	12	76	92	0.19 (0)	**2.54**	**3.42**	**0.77**
KAERI-6	0.18	0.01	75	2.42	8	10	77	93	0.18 (0)	**2.55**	**3.45**	**0.78**
KAERI	0.19	0.01	77	2.33	7	10	79	89	0.19 (0)	**2.59**	**3.49**	**0.76**
KAERI-1	0.20	0.03	78	2.25	6	7	82	86	0.20 (0)	**2.63**	**3.56**	**0.74**
KAERI-3	0.19	0.02	79	2.25	6	7	82	85	0.19 (0)	**2.64**	**3.57**	**0.74**
KAERI-4	0.19	0.03	79	2.26	6	8	82	86	0.19 (0)	**2.64**	**3.56**	**0.74**
KAERI-2	0.20	0.03	80	2.24	6	8	83	85	0.20 (0)	**2.65**	**3.57**	**0.74**
Average over all institutions	**0.24**	**-0.46**	**70**	**2.10**	**11**	**25**	**68**	**82**	**0.25 (0)**	**2.19**	**2.94**	**0.39**
Median over all institutions	**0.20**	**-0.49**	**73**	**2.16**	**10**	**24**	**67**	**77**	**0.21 (0)**	**2.23**	**3.00**	**0.32**
Maximum over all institutions	**0.50**	**0.26**	**82**	**3.82**	**35**	**42**	**82**	**120**	**0.60 (2)**	**2.59**	**3.49**	**0.76**
Minimum over all institutions	**0.12**	**-0.90**	**41**	**0.89**	**2**	**8**	**51**	**65**	**0.17 (-2)**	**1.48**	**2.14**	**0.15**

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

A detailed look on thescores: USX75

Table 5: Average statistics and individual statistics for station USX75 per institution over all run-IDs and for individual run-IDs ordered by Rank

Institution	R	FB	F5 [%]	RMSE	NMSE	KS [%]	ACC [%]	NAAD [%]	CRmax ($\hat{t}$)	Rank	Rank_KS	SS
AWE-1	0.21	-1.39	21	11.60	71	50	48	95	0.21 (-1)	**1.04**	**1.54**	**0.09**
AWE	0.22	-1.19	40	11.53	54	34	61	94	0.22 (-1)	**1.46**	**2.12**	**0.10**
AWE-2	0.23	-0.99	58	11.45	37	18	73	93	0.23 (0)	**1.87**	**2.69**	**0.12**
JAEA	0.36	0.64	55	8.98	13	28	69	167	0.36 (0)	**2.05**	**2.77**	**0.22**
ARPANSA	0.45	-0.58	56	10.54	19	32	65	83	0.55 (-4)	**2.12**	**2.80**	**0.39**
CTBTO	0.54	-0.52	49	2.96	5	37	60	78	0.54 (0)	**2.13**	**2.76**	**0.51**
NOAA-ARL-2	0.43	-0.58	62	10.55	19	28	69	80	0.59 (-4)	**2.21**	**2.93**	**0.32**
NOAA-ARL-3	0.51	-0.71	63	10.26	21	28	69	74	0.57 (-4)	**2.22**	**2.94**	**0.22**
LLNL	0.66	-0.72	52	9.28	17	32	64	78	0.66 (0)	**2.24**	**2.92**	**0.29**
PNNL	0.52	-0.34	52	9.96	13	35	64	84	0.52 (0)	**2.26**	**2.91**	**0.53**
NOAA-ARL	0.40	-0.38	62	10.81	17	25	72	91	0.48 (-2)	**2.27**	**3.02**	**0.39**
BOKU	0.42	-0.43	61	10.75	17	22	71	94	0.42 (0)	**2.28**	**3.06**	**0.46**
FOI	0.60	-0.43	54	9.28	13	31	65	72	0.63 (-1)	**2.35**	**3.04**	**0.49**
JAEA1	0.56	0.91	67	7.16	6	22	84	193	0.61 (1)	**2.37**	**3.15**	**0.25**
NOAA-ARL-1	0.26	0.15	62	11.62	11	20	77	120	0.28 (2)	**2.38**	**3.18**	**0.62**
VINATOM	0.67	0.75	59	32.17	45	12	73	195	0.67 (0)	**2.39**	**3.27**	**0.20**
KAERI-8	0.44	-0.11	59	11.52	14	27	67	96	0.57 (-1)	**2.41**	**3.14**	**0.80**
ZAMG	0.67	-0.49	56	8.72	12	32	66	70	0.67 (0)	**2.42**	**3.10**	**0.47**
CTBTO1	0.67	-0.58	57	8.96	14	28	70	69	0.67 (0)	**2.42**	**3.14**	**0.38**
KAERI-7	0.45	-0.05	62	11.56	13	27	69	98	0.55 (-2)	**2.49**	**3.22**	**0.84**
KAERI-5	0.45	-0.04	62	11.62	13	27	69	99	0.55 (-2)	**2.50**	**3.23**	**0.85**
KAERI-6	0.45	-0.04	62	11.40	13	25	71	98	0.53 (-1)	**2.52**	**3.27**	**0.85**
CMC-2	0.56	-0.40	72	9.88	14	20	77	74	0.56 (0)	**2.60**	**3.40**	**0.36**
KAERI	0.54	-0.13	70	10.30	12	21	74	86	0.59 (-1)	**2.67**	**3.46**	**0.73**
IRSN	0.76	-0.48	71	8.56	11	24	73	64	0.76 (0)	**2.78**	**3.54**	**0.37**
CMC	0.64	-0.32	76	9.26	11	19	78	71	0.64 (0)	**2.79**	**3.61**	**0.46**
KAERI-4	0.65	-0.27	78	9.04	10	16	78	72	0.65 (0)	**2.84**	**3.68**	**0.54**
KAERI-3	0.63	-0.20	79	9.08	10	16	78	74	0.63 (0)	**2.86**	**3.70**	**0.63**
KAERI-1	0.63	-0.20	79	9.09	10	16	78	74	0.63 (0)	**2.86**	**3.70**	**0.63**
KAERI-2	0.63	-0.16	79	9.09	9	15	78	74	0.63 (0)	**2.88**	**3.73**	**0.68**
CMC-3	0.67	-0.29	78	8.98	10	18	79	69	0.67 (0)	**2.88**	**3.70**	**0.50**
CMC-1	0.68	-0.26	77	8.92	10	18	79	70	0.68 (0)	**2.90**	**3.72**	**0.53**
Average over all institutions	**0.54**	**-0.27**	**58**	**10.58**	**17**	**27**	**69**	**99**	**0.56 (-1)**	**2.31**	**3.04**	**0.39**
Median over all institutions	**0.55**	**-0.43**	**57**	**9.28**	**13**	**28**	**69**	**84**	**0.60 (0)**	**2.31**	**3.05**	**0.39**
Maximum over all institutions	**0.76**	**0.91**	**76**	**32.17**	**54**	**37**	**84**	**195**	**0.76 (1)**	**2.79**	**3.61**	**0.73**
Minimum over all institutions	**0.22**	**-1.19**	**40**	**2.96**	**5**	**12**	**60**	**64**	**0.22 (-4)**	**1.46**	**2.12**	**0.10**

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

Details on statistical metrics

Given N predictions P_j and measurements M_i at times t_j and t_i with mean values $\overline{P}$ and $\overline{M}$ as well as minimum detectable concentrations MDC_j and MDC_i the statistical scores in subsection 2.8 are formally defined as:

$$\mathrm{FB} = 2\frac{(\overline{P} - \overline{M})}{(\overline{P} + \overline{M})} \tag{A.1}$$

$$R = \frac{\sum (M_i - \overline{M})(P_i - \overline{P})}{\sqrt{(M_i - \overline{M})^2 (P_i - \overline{P})^2}} \tag{A.2}$$

Let T denote the number of fractions satisfying:

$$0.2 \leq \frac{P_i}{M_{i|M_i > 0.0}} \leq 5.0 \tag{A.7}$$

and Q the number of pairs with $P_{i|P_i \geq MDC_i}$ and/or $M_{i|M_i \geq MDC_i}$ then F5 is defined as:

$$\mathrm{F5} = \frac{T}{Q}100 \tag{A.8}$$

Given the number of correctly forecasted above MDC values A (true positives) and below MDC values B (true negatives) as well as the number of not correctly forecasted above MDC values (false positives) C and below MDC values D (false negatives), the Accuracy is given as:

$$\mathrm{ACC} = \frac{A + B}{A + B + C + D}100 \tag{9}$$

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

Details on statistical metrics

The *Kolmogorov-Smirnov parameter* (KSP) is defined as:

$$\mathrm{KSP} = Max|D(M_k) - D(O_k)| * 100\% \tag{A.7}$$

where D is the cumulative distribution of the predicted and measured (or other predicted) concentrations over the range of k values such that D is the probability that the concentration will not exceed M_k or O_k. The score measures the ability of the model to reproduce the measured (or another predicted) concentration distribution regardless of space and time. The maximum difference between any two distributions cannot be more than 100%.

Further visual evaluations in Kioutsioukis and Galmarini [19] are based on comparing *Cumulative Density Functions (cdf)* of the models and the observations or on showing *Taylor diagrams* (i.e, a combination of correlation coefficient, root mean square error and standard deviation, Figure 2). The distance between the reference (black point in Figure 2) and model points in a Taylor diagram is then the BC_RMSE. According to Taylor [42] model standard deviation, BC_RMSE and R can be combined into a single skill score S_r.

$$S_r = 2(1+R)(\frac{\sigma_m}{\sigma_o} + \frac{\sigma_o}{\sigma_m})^{-2} \tag{20}$$

with σ_m and σ_o being the standard deviations of predictions and observations.

The correlation contribution becomes important for large values of BC_RMSE. If one wants to include also the relative bias FB into the skill score, Seibert [34] suggests:

$$S_b = \frac{1}{1 + bFB^2} \tag{21}$$

ZAMG
Zentralanstalt für Meteorologie und Geodynamik

Details on statistical metrics

A value of $b = 10$ appears to give a relationship fulfilling Seibert's [34] subjective idea about such a skill score. Finally, both skill scores can be combined into a total skill score S:

$$S = \alpha S_r + (1 - \alpha) S_b \tag{22}$$

The value of α is rather arbitrary and would depend on the application. $\alpha = 0.5$ might be acceptable. An additive and not multiplicative combination of the two scores is suggested because a model that has skill either in reproducing the mean or in reproducing the patterns should be attributed some total skill; the product of the two scores would be zero with one of the factors being zero. In any case, data sets with strongly non-normal distributions might better be transformed before applying any of the measures.