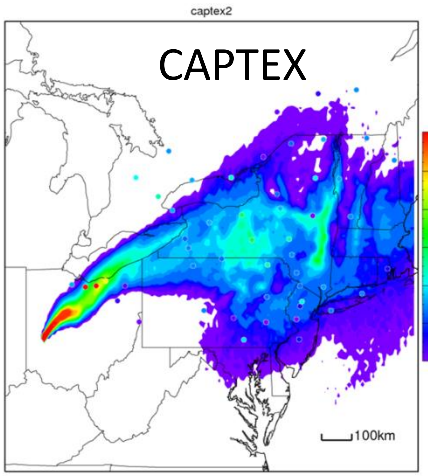
HYSPLIT Inverse Modeling

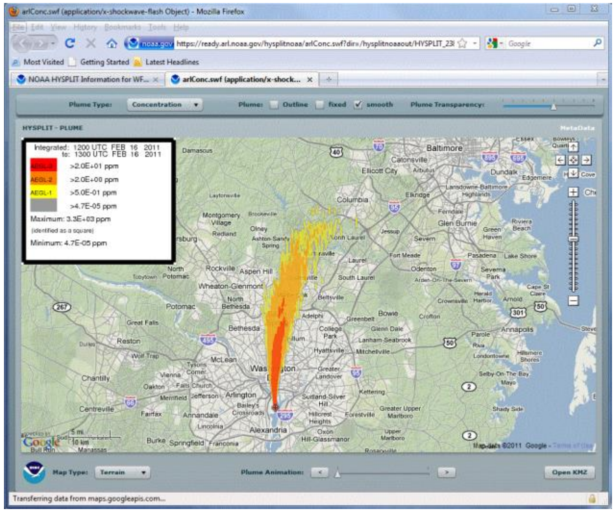
Tianfeng Chai[1,2], Ariel Stein[2], and Fong Ngan[1,2]
1:Cooperative Institute for Climate & Satellites – Maryland, University of Maryland
2: NOAA Air Resources Laboratory

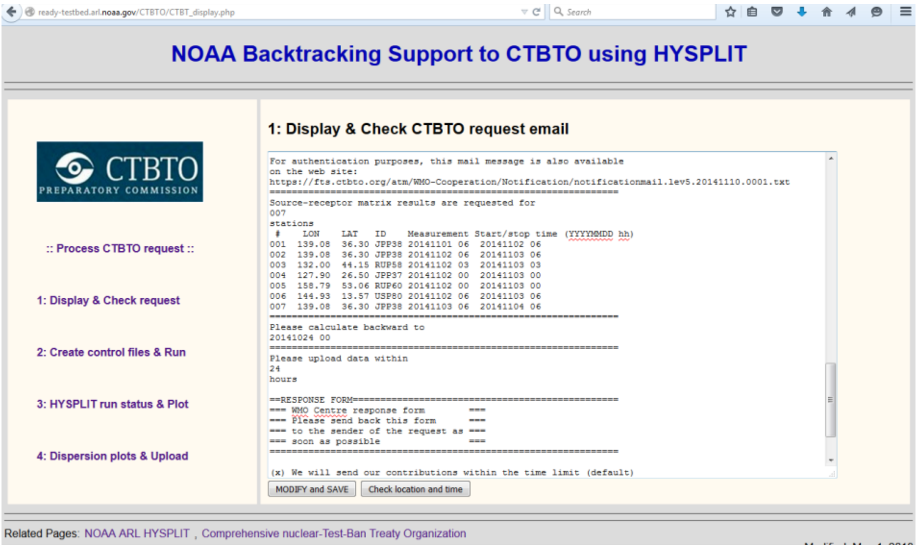
HYSPLIT

- HYSPLIT model allows Lagrangian representations of the transported air masses w/ 3D particles, puffs, or a hybrid
- Applications include the simulation of atmospheric tracer release experiments, radionuclides, smoke originated from wild fires, volcanic ash, mercury, and wind-blown dust, etc.



Emergency Response Chemical Releases





HYSPLIT Inverse Modeling

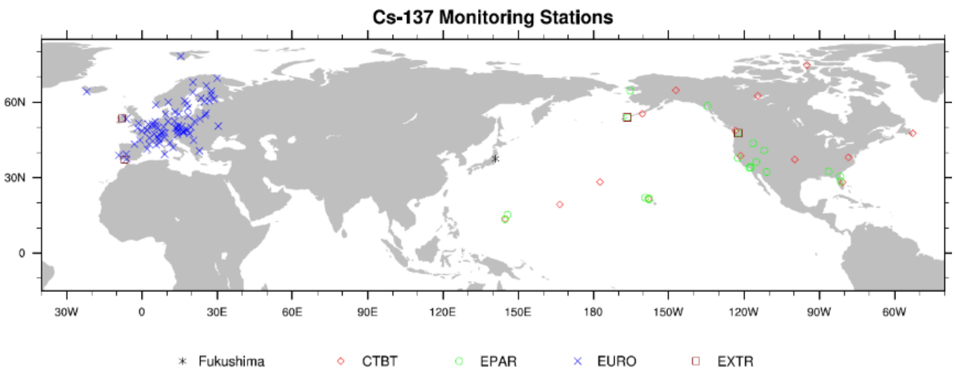
- In a top-down approach, the unknown emission terms are obtained via searching the emissions that would provide the best model predictions closely matching the observations by minimizing a cost function F.

$$\mathcal{F} = \frac{1}{2}\sum_{t=1}^{T}\sum_{k=1}^{K}\sum_{i=1}^{I}\frac{(q_{ikt} - q_{ikt}^{b})^2}{\sigma_{ikt}^2} + \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{M}\frac{(c_{nm}^{h} - c_{nm}^{o})^2}{\epsilon_{nm}^2} + \mathcal{F}_{other}$$

- Unknowns: emission rate q_{ikt} at location *i*, at height *k* and period *t*. c^{o}_{nm} is the *m-th* observed concentration or mass loading at time period *n* and c^{h}_{nm} is the HYSPLIT counterpart;
- A background term is included to measure the deviation of the emission estimation from its first guess q^{b}_{ikt}. The background terms ensure that the problem is well-posed even when there are not enough observations available.
- The background error variances σ^{2}_{ikt} measure the uncertainties of q^{b}_{ikt}. The observational error variances ε^{2}_{nm} represent the uncertainties from both the model and observations as well as the representative errors.
- F_{other} refers to the other regularization terms that can be included in the cost function.
- The optimization problem can be solved using many minimization tools, such as L-BFGS-B package, to get the final optimal emission estimates.

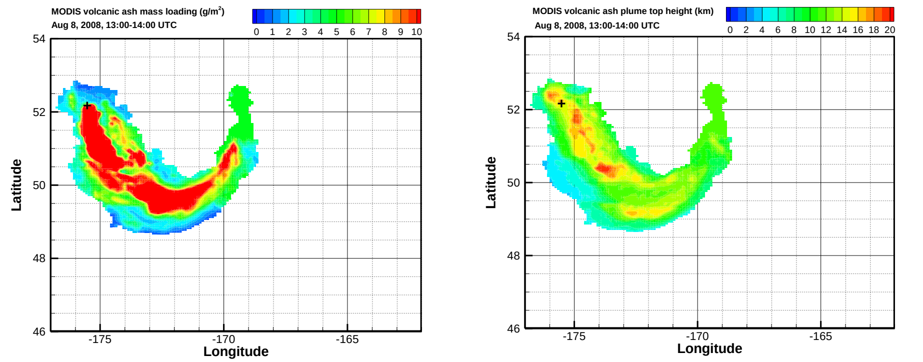
HYSPLIT inverse modeling

1. Fukushima source term estimation



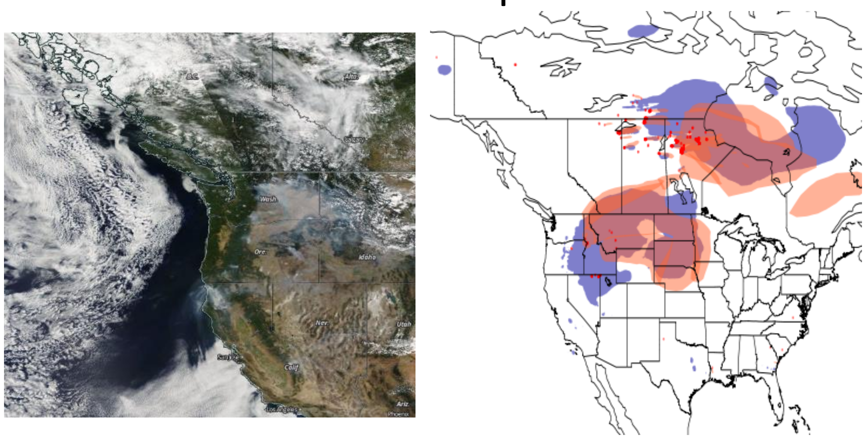
Source term estimation using air concentration measurements and a Lagrangian dispersion model–Experiments with pseudo and real cesium-137, T Chai, R Draxler, A Stein – Atmos. Environ., 2015

2. Volcanic ash application-Kasatochi eruption



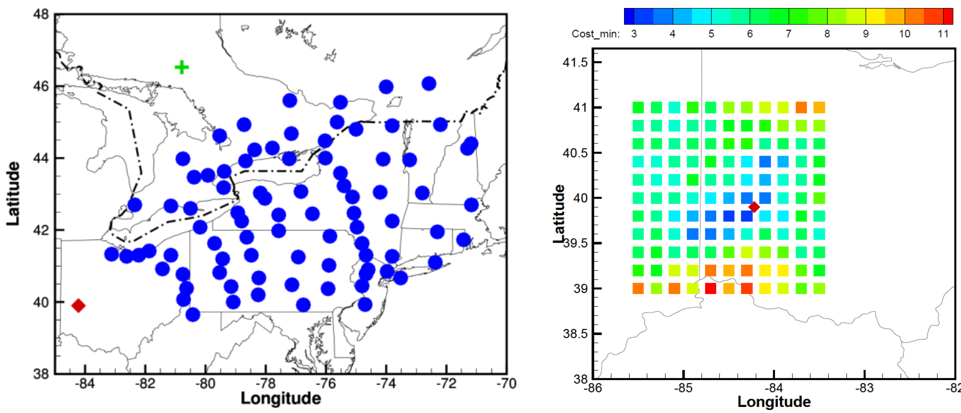
Improving volcanic ash predictions with the HYSPLIT dispersion model by assimilating MODIS satellite retrievals, Chai, T., et al. Atmos. Chem. Phys. 17, 2017

3. Smoke forecast with top-down emissions



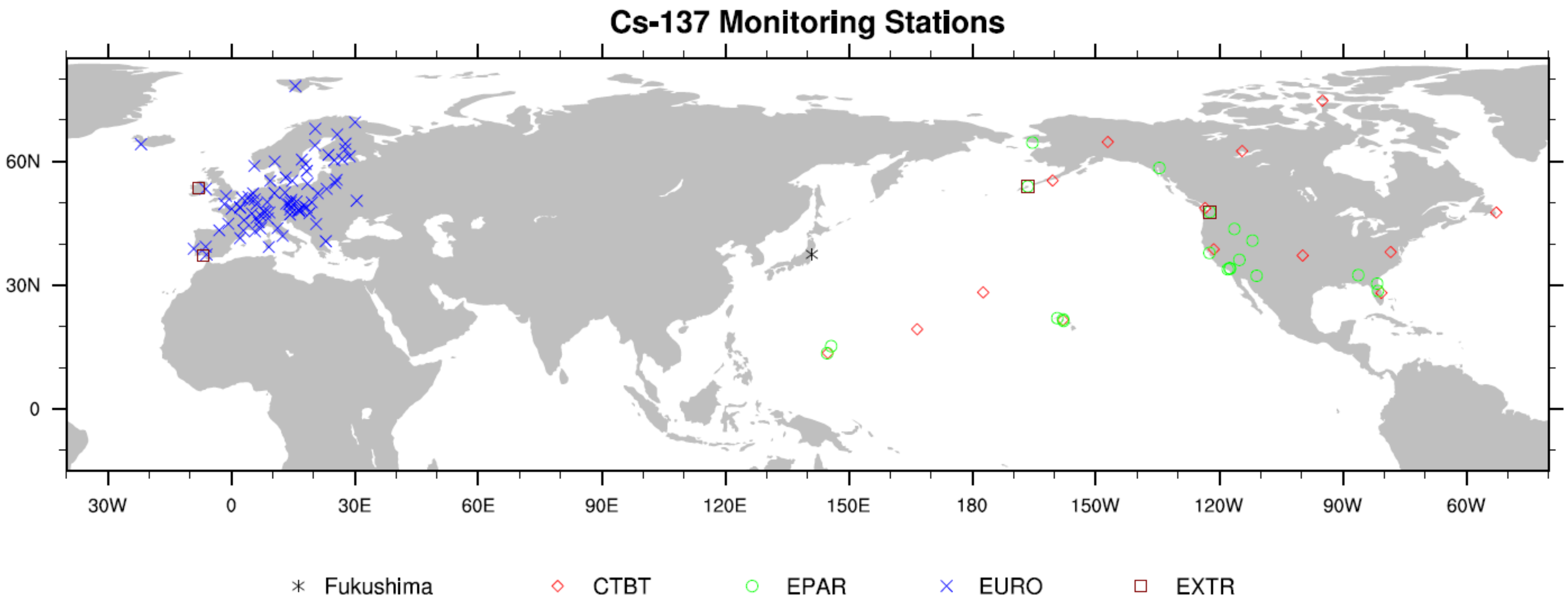
NOAA Award NA16OAR4590121: "Top-down estimation of wildfire smoke emission based on HYSPLIT model and NOAA NESDIS GOES Aerosol/Smoke products (GASP) to improve smoke forecasts in the US"

4. Model uncertainty study using CAPTEX data



Weak-constraint inverse modeling using HYSPLIT Lagrangian dispersion model and Cross Appalachian Tracer Experiment (CAPTEX) observations – Effect of including model uncertainties on source term estimation, Chai, T., A. Stein, and F. Ngan, *Geosci. Model Dev.*, https://doi.org/10.5194/gmd-2018-159, 2018

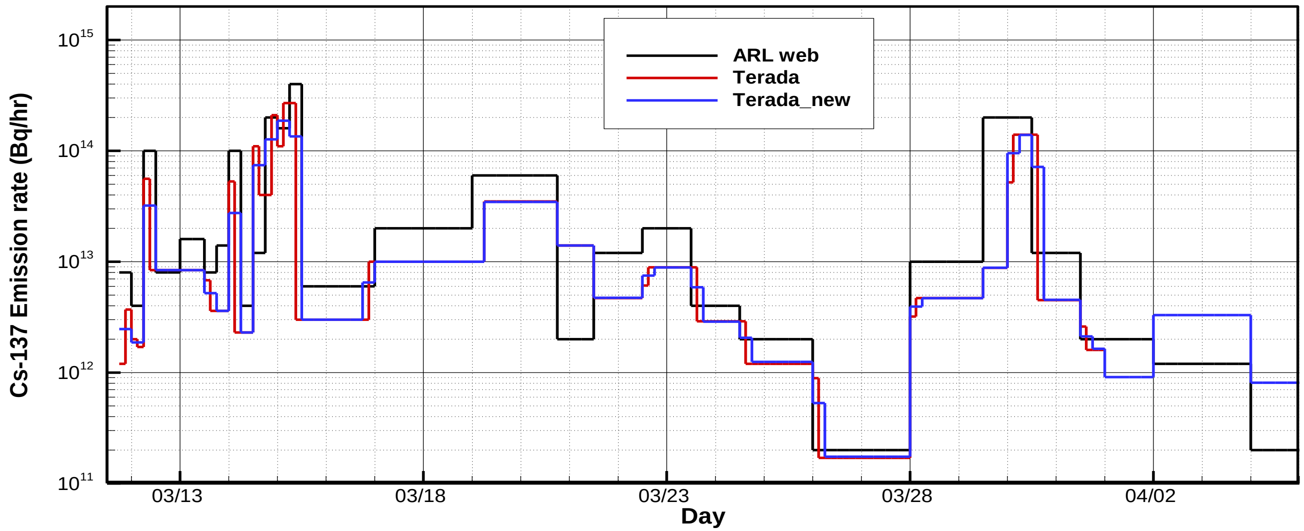
Fukushima air concentration measurements



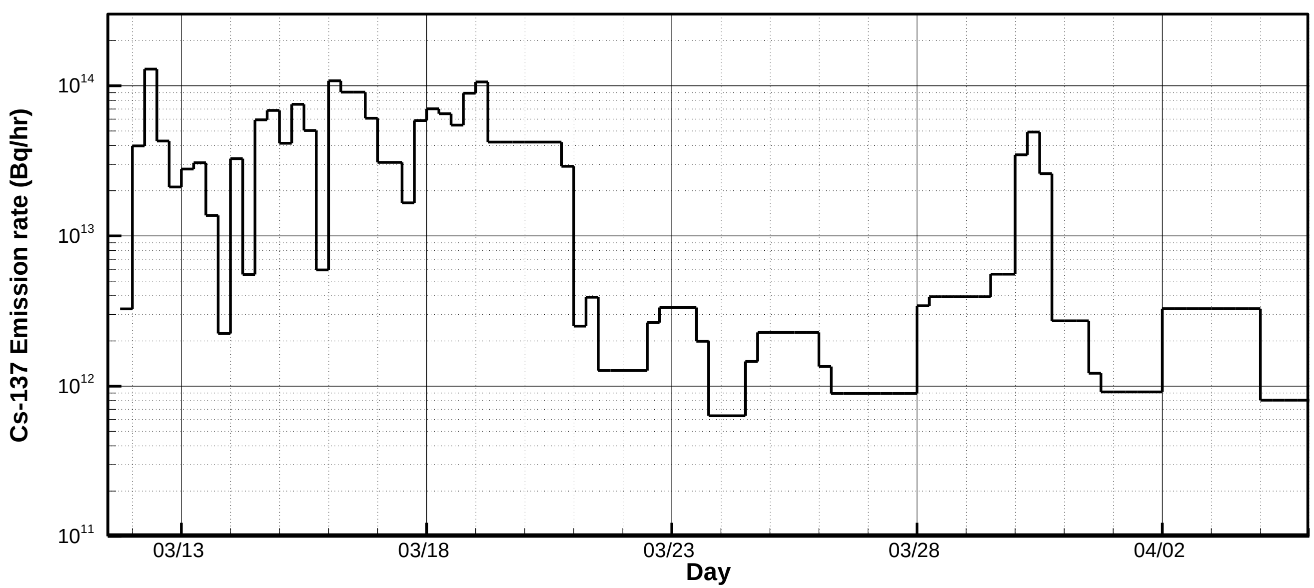
Data source	Number of monitoring stations	Count of total samples
CTBT	14	417 (421)
EPAR	19 (20)	35 (39)
EURO	78 (80)	785 (797)
EXTR	4	59 (61)
Total	115 (118)	1296 (1318)

Cs-137 emission rates

Chino et al. (2011);
Katata et al. (2012);
Terada et al. (2012)

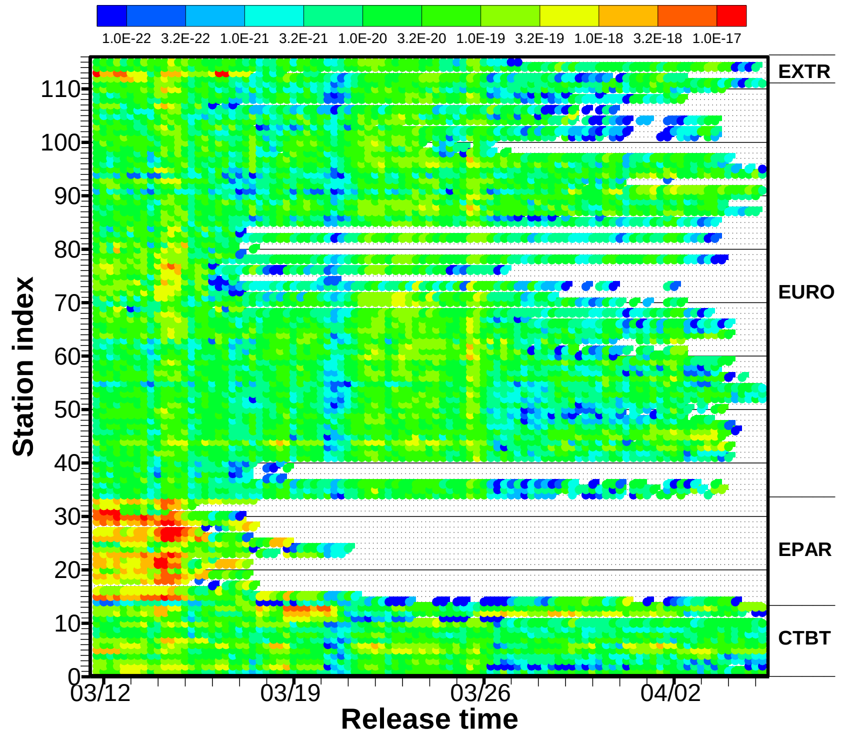


Katata et al. (2015)

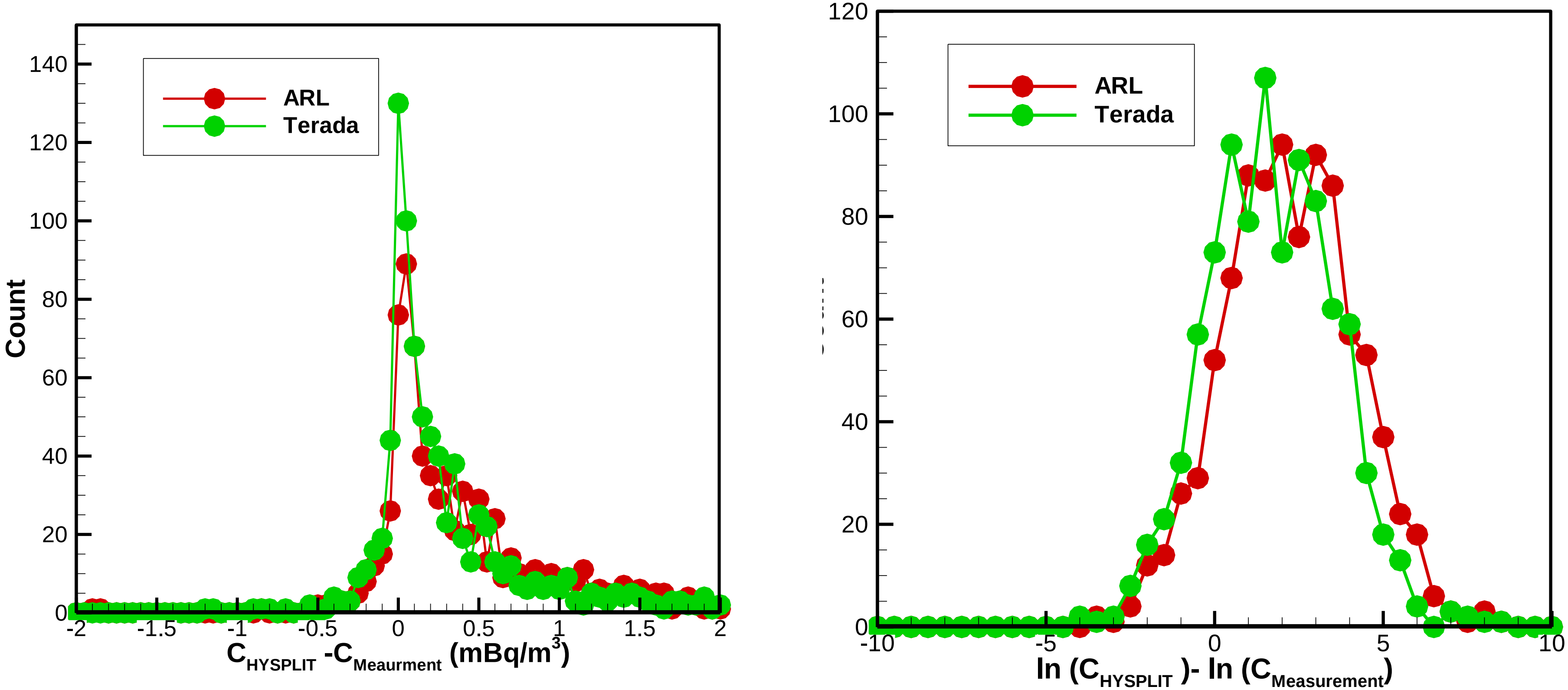


Transfer coefficient matrix (TCM)

$$\begin{pmatrix} c_1^h \\ c_2^h \\ \vdots \\ c_M^h \end{pmatrix} = \begin{pmatrix} H_{1,1} & H_{1,2} & \cdots & H_{1,N} \\ H_{2,1} & H_{2,2} & \cdots & H_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ H_{M,1} & H_{M,2} & \cdots & H_{M,N} \end{pmatrix} \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_N \end{pmatrix}$$



Metric variable – C or ln(C) ?



Twin experiments

- In twin experiments, pseudo-observations are generated with known sources

$$\mathbf{c^p} = \mathbf{H} \cdot \mathbf{q} + \varepsilon$$

Where $$\varepsilon_m^p = (f_m^p \times c_m^p + a_m^p) \times r_m, \quad m = 1, \cdots, M$$

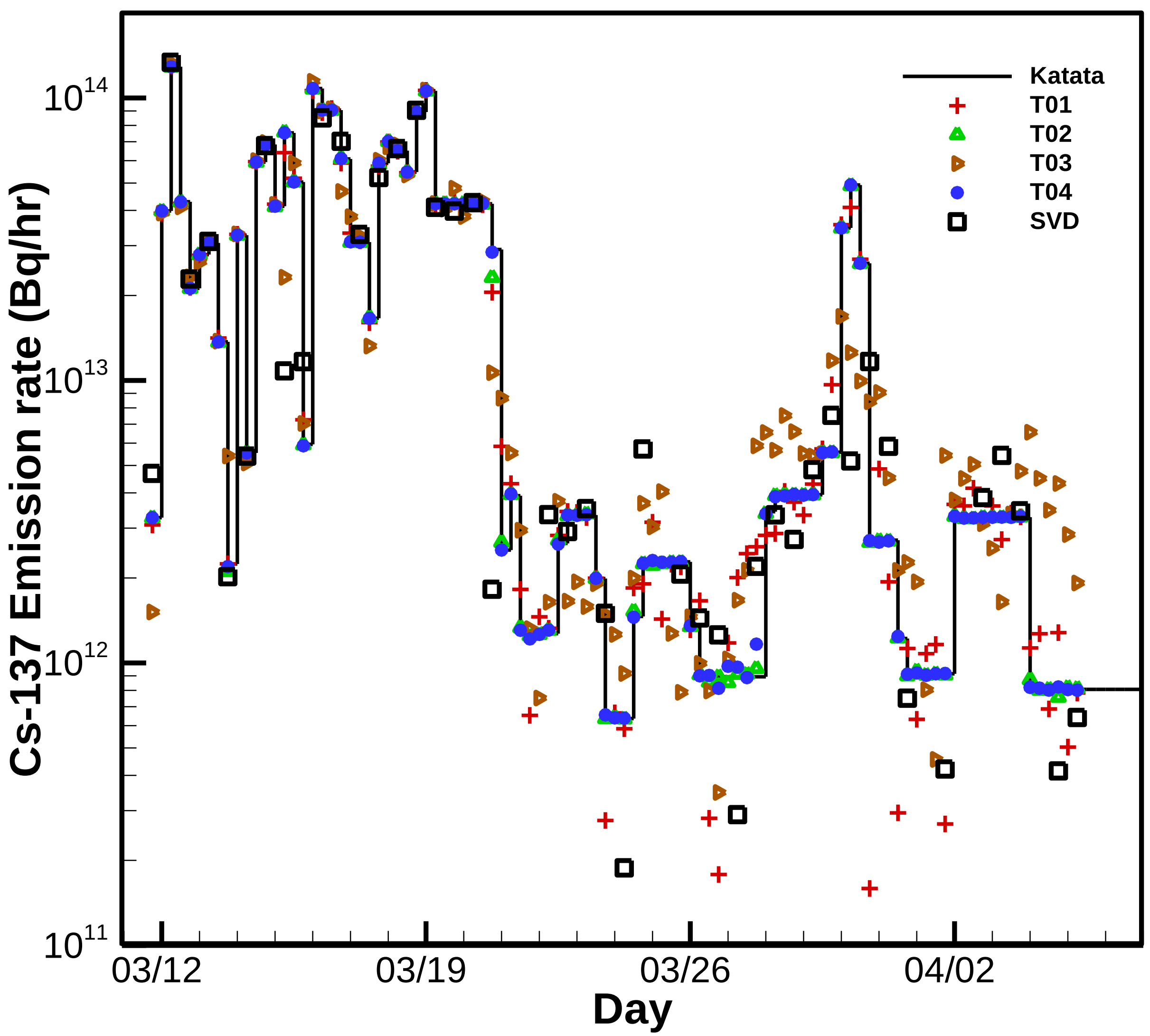
- They are used to address the following issues:
  1. How to measure the concentration/release differences?
  2. How does the cost functional approach compare to SVD?
  3. Will the results be sensitive to measurement errors?
  4. Will the results be sensitive to the first guess?

Control and metric variables

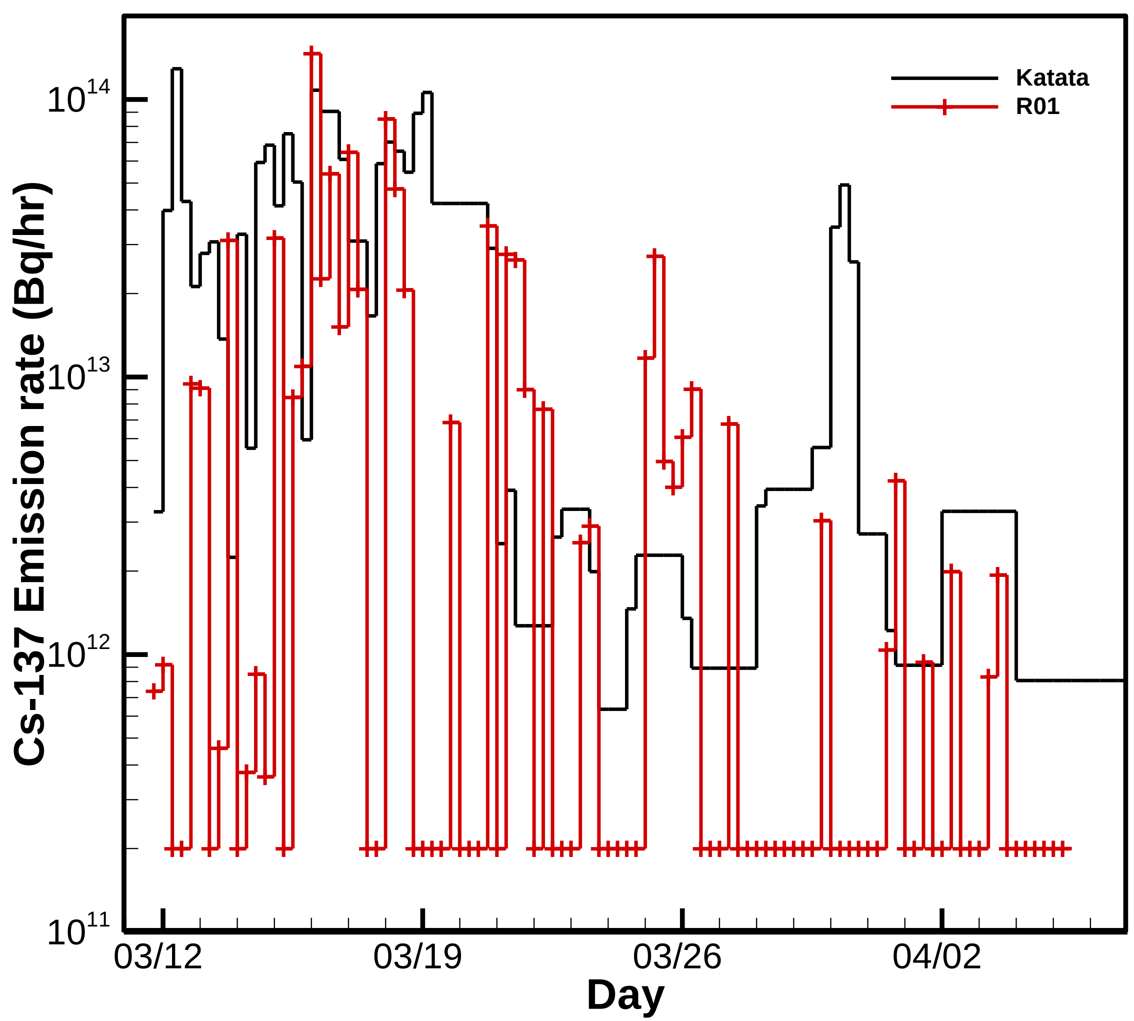
Case	x	y	f_m^p	a_m^p	f_m^c	a_m^c	f_i^q	a_i^q	q^b
T01	q	c	0	0	0.001	10^{-5}	10^5	10^9	10^{13}
T02	q	ln(c)	0	0	0.001	10^{-5}	10^5	10^9	10^{13}
T03	ln(q)	c	0	0	0.001	10^{-5}	10^5	10^9	10^{13}
T04	ln(q)	ln(c)	0	0	0.001	10^{-5}	10^5	10^9	10^{13}

Case	Cs-137 release			
	MAE	RMSE	MRE	R
T01	0.90×10^{12}	1.88×10^{12}	0.229	0.998
T02	0.11×10^{12}	0.61×10^{12}	0.012	1.000
T03	3.32×10^{12}	7.66×10^{12}	0.740	0.965
T04	0.03×10^{12}	0.08×10^{12}	0.011	1.000
SVD	3.57×10^{12}	9.06×10^{12}	0.588	0.952

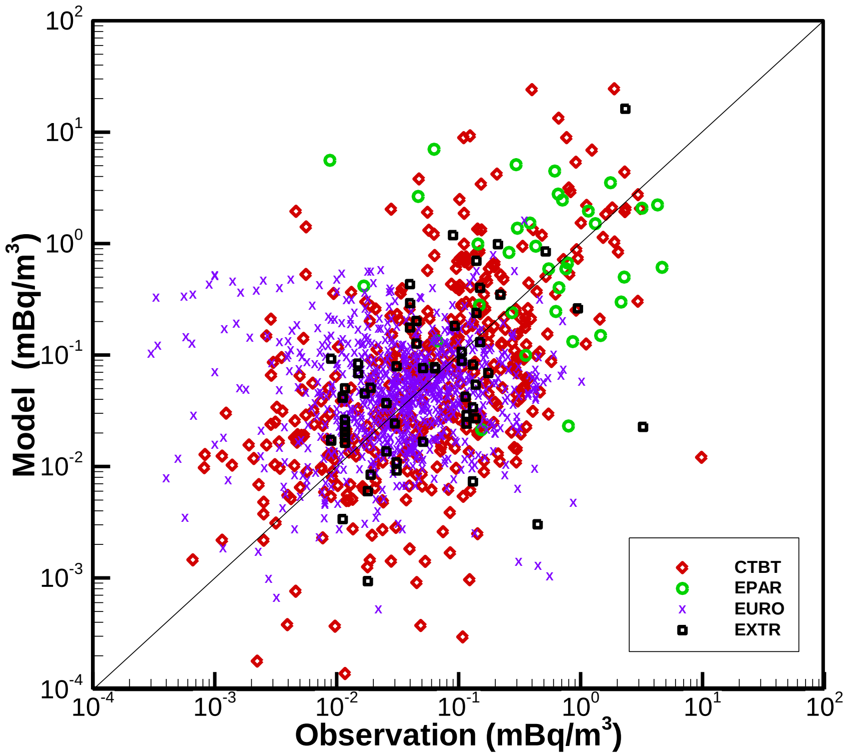
MAE: mean absolute error; RMSE: root-mean-square error; MRE: mean relative error; R: linear correlation coefficient. Units of MAE and RMSE: mBq/m3 for concentrations, Bq/h for releases.

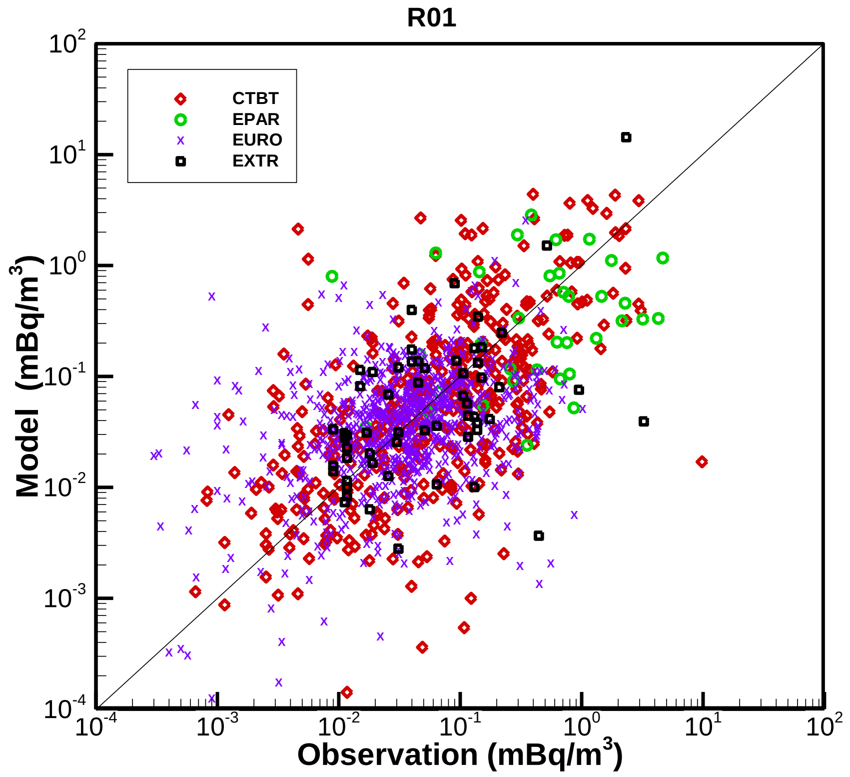

Cs-137 Emission rate (Bq/hr)
Day
10^{14}
10^{13}
10^{12}
10^{11}
03/12
03/19
03/26
04/02
Katata
T01
T02
T03
T04
SVD

R01: without smoothness



Katata et al. (2015)

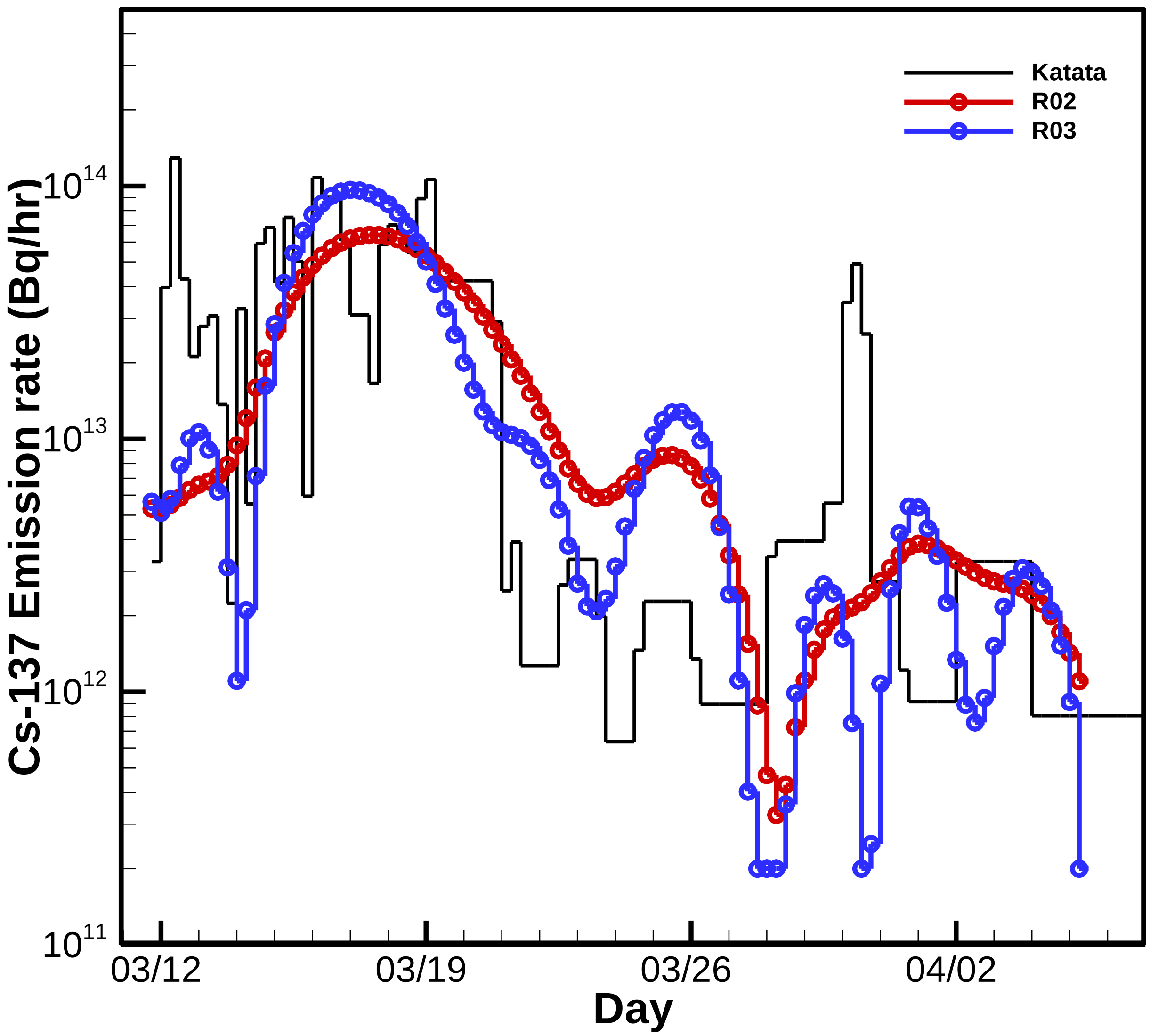




Rc=0.255, $R_{ln(c)}$ = 0.376

Rc=0.327, $R_{ln(c)}$ = 0.518

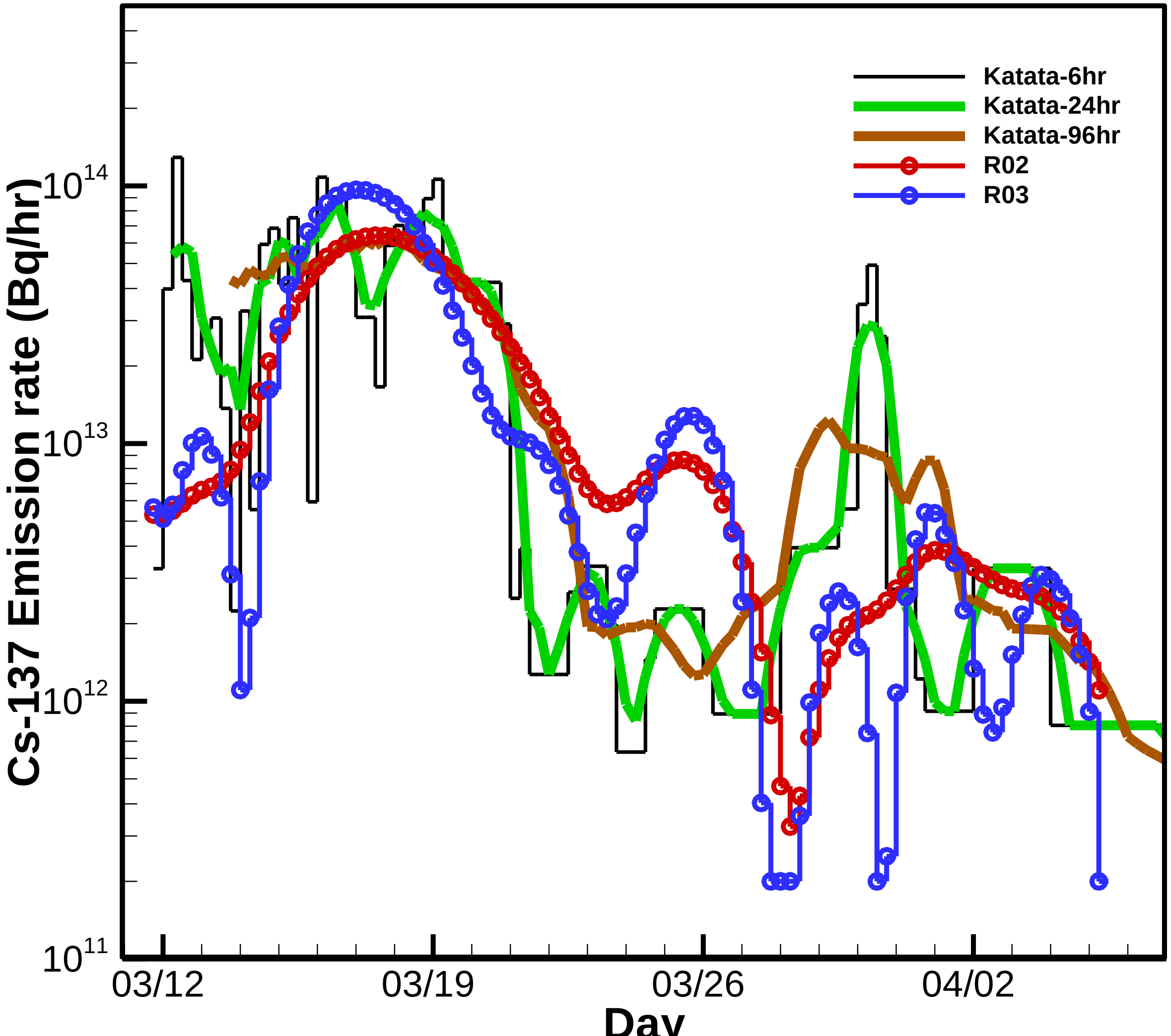
Smoothness penalty term



R02: Csm = 1.0

R03: Csm = 0.1

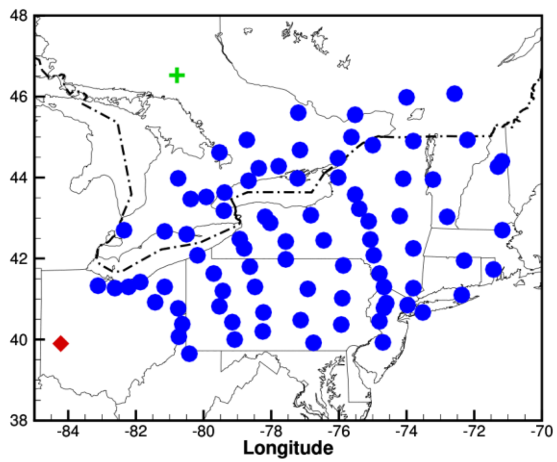
Compared to running-average



Cross Appalachian Tracer Experiment (CAPTEX)

- It consisted of a near-surface release of the inert tracer perfluro-monomethylcyclohexane (PMCH) , from Dayton, Ohio, USA and Sudbury, Ontario, Canada;
- Samples were collected at 84 different measurement sites distributed from 300 to 800 km downwind of the emission source, as either 3- or 6-hour averages up to 60 hours after each release.

#	Site (latitude, longitude)	Release time	Amount	M_{obs}
1	Dayton (39.80°, -84.05°)	1700-2000Z, Sep. 18, 1983	208 kg	395
2	Dayton (39.90°, -84.22°)	1705-2005Z, Sep. 25, 1983	201 kg	400
3	Dayton (39.90°, -84.22°)	1900-2200Z, Oct. 02, 1983	201 kg	404
4	Dayton (39.90°, -84.22°)	1600-1900Z, Oct. 14, 1983	199 kg	367
5	Sudbury (46.62°, -80.78°)	0345-0645Z, Oct. 26, 1983	180 kg	357
6	Dayton (39.90°, -84.22°)	1530-1600Z, Oct. 28, 1983	32 kg	-
7	Sudbury (46.62°, -80.78°)	0600-0900Z, Oct. 29, 1983	183 kg	358



Inverse results with $\varepsilon = f^o \times c^o + a^o$

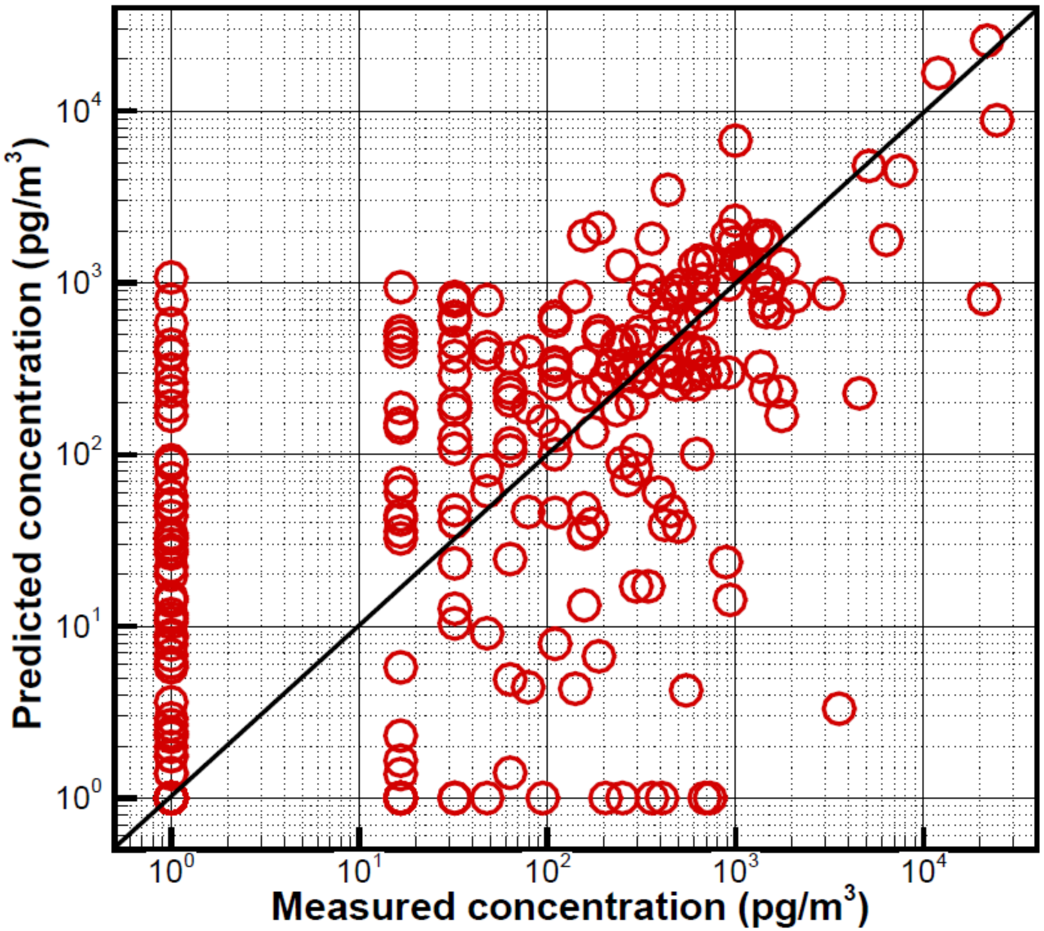
Release strength estimates of R2. Concentration is used as the metric variable.

Emission (kg/hr)	a^o =10 pg/m^3	a^o =20 pg/m^3	a^o =50 pg/m^3
$f^o = 10\%$	7.1	11.1	17.4
$f^o = 20\%$	4.1	7.1	12.6
$f^o = 30\%$	2.9	5.2	10.0
$f^o = 50\%$	1.8	3.4	7.1

Release strength estimates of R2. Logarithmic concentration is used as the metric variable.

Emission (kg/hr)	a^o =10 pg/m^3	a^o =20 pg/m^3	a^o =50 pg/m^3
$f^o = 10\%$	115.2	119.8	124.7
$f^o = 20\%$	106.3	112.9	119.8
$f^o = 30\%$	101.2	108.5	116.3
$f^o = 50\%$	94.4	101.2	109.6

HYSPLIT predictions with exact source terms



$$\mathcal{F} = \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{(q_{ij}-q_{ij}^{b})^2}{\sigma_{ij}^2}$$

$$+\frac{1}{2}\sum_{m=1}^{M}\frac{(c_m^h-c_m^o)^2}{\epsilon_m^2}$$

Comparison between the predicted and measured concentrations for Release 2 during the CAPTEX experiment. In the HYSPLIT simulation, at the exact release location, an emission rate of 67 kg/hr was applied from 17Z to 20Z on September 25, 1983.

With "dynamic" $\varepsilon^2 = (f^o \times c^o + a^o)^2 + (f^h \times c^h + a^h)^2$

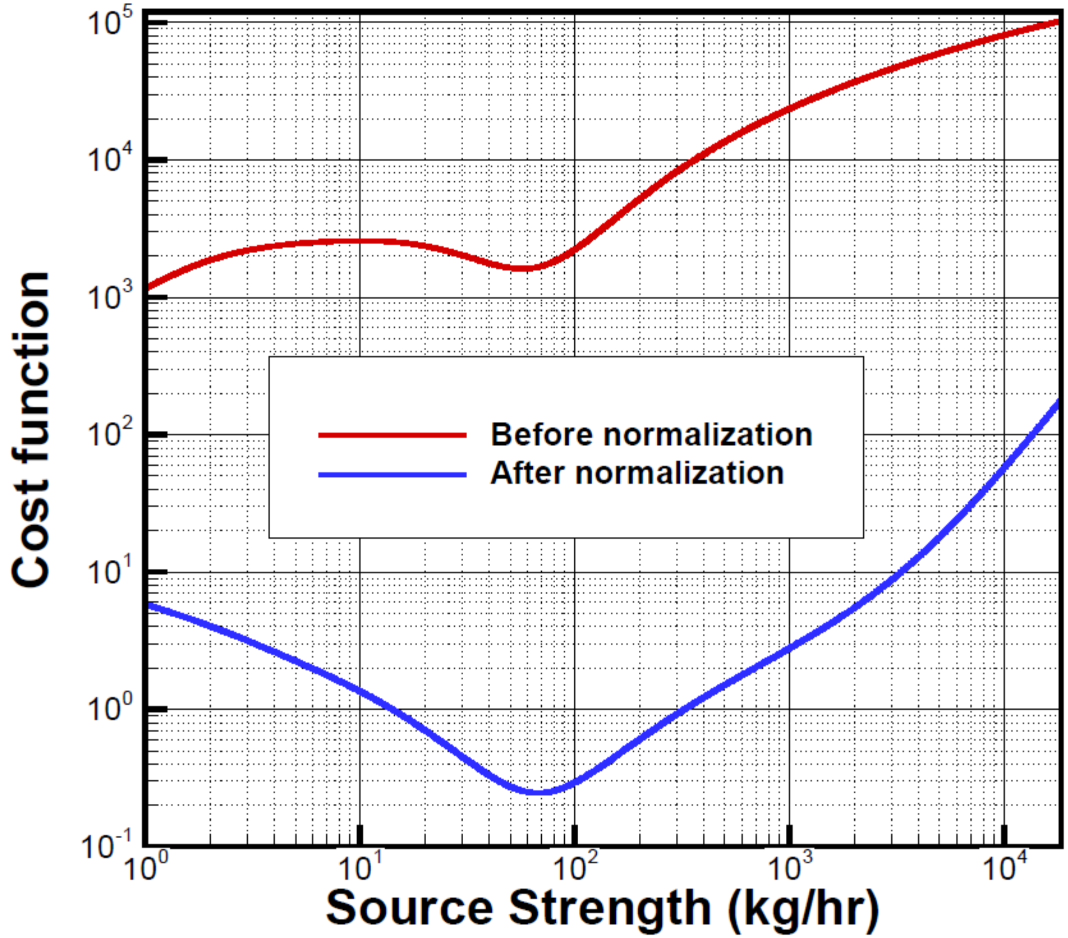
Release estimates of R2. Concentration is used as the metric variable. $f^o = 20\%$, $a^o = 20\ pg/m^3$.

Emission (kg/hr)	$a^h = 10\ pg/m^3$	$a^h = 20\ pg/m^3$	$a^h = 50\ pg/m^3$
$f^h = 0$	7.7	9.1	13.6
$f^h = 10\%$	15.9	22.1	32.9
$f^h = 20\%$	48.5	50.4	53.5
$f^h = 50\%$	114.0	111.8	104.3

Estimates of R2. Logarithmic concentration is used as the metric variable. $f^o = 20\%$, $a^o = 20\ pg/m^3$

Emission (kg/hr)	$a^h = 10\ pg/m^3$	$a^h = 20\ pg/m^3$	$a^h = 50\ pg/m^3$
$f^h = 0$	64.7	58.5	53.5
$f^h = 10\%$	61.5	55.7	49.4
$f^h = 20\%$	58.5	53.0	46.6
$f^h = 50\%$	55.1	49.4	42.6

Cost function normalization



$$\mathcal{F} = \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{(q_{ij}-q_{ij}^{b})^2}{\sigma_{ij}^2} +$$

$$\frac{1}{2}\sum_{m=1}^{M}\frac{(c_m^h - c_m^o)^2}{\epsilon_m^2} \times \frac{\sum_{m=1}^{M}\frac{1}{\epsilon_m^{b\ 2}}}{\sum_{m=1}^{M}\frac{1}{\epsilon_m^2}}$$

Cost function as a function of source strength before and after cost function normalization, with f_h = 0, a_h =50 pg/m3, f_o = 10%, and a_o =20 pg/m^3.

Results with cost function normalization

Release estimates of R2. Concentration is used as the metric variable. f^o = 20%, a^o =20 pg/m^3.

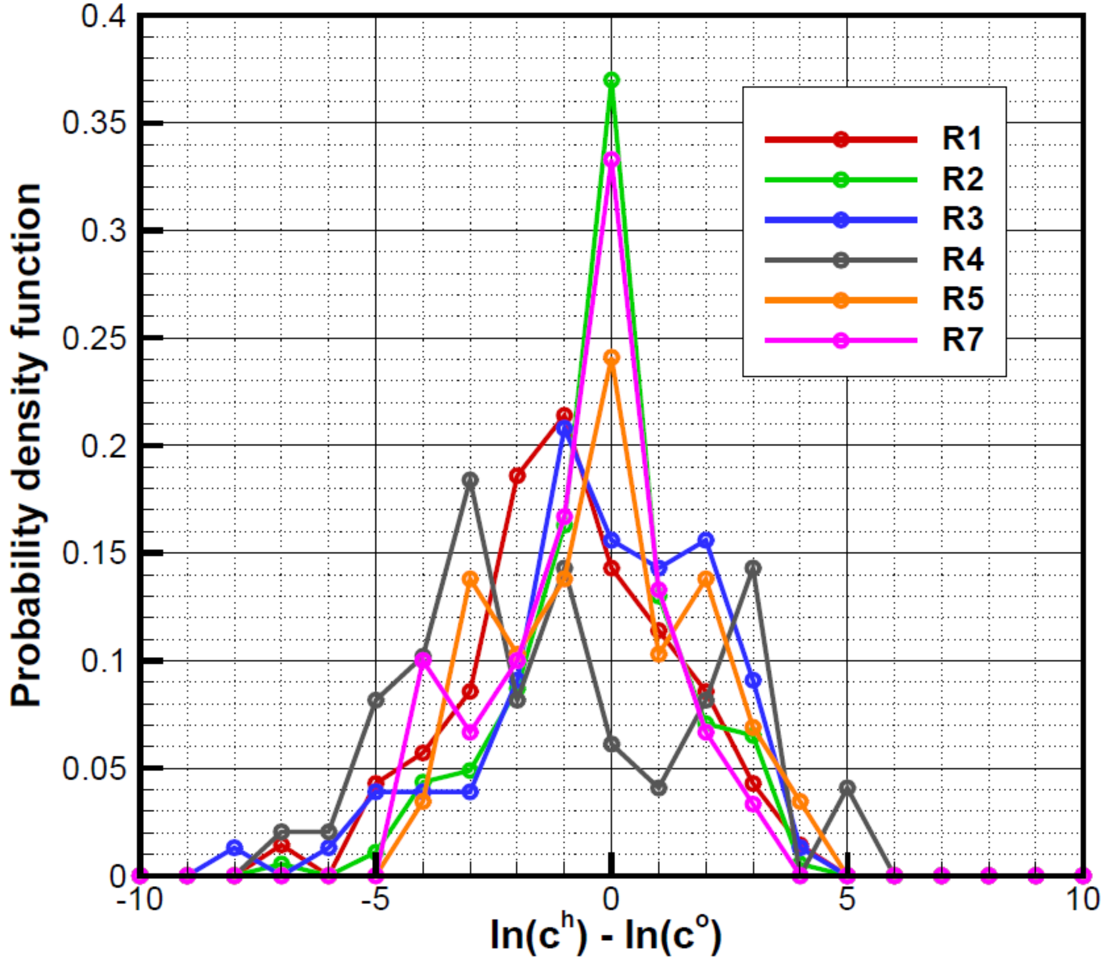
Emission (kg/hr)	a^h =10 pg/m^3	a^h =20 pg/m^3	a^h =50 pg/m^3
$f^h = 0$	7.7	9.1	13.6
$f^h = 10\%$	10.9	15.1	26.4
$f^h = 20\%$	32.9	35.6	41.3
$f^h = 50\%$	64.7	64.7	65.3

Estimates of R2. Logrithmic oncentration is used as the metric variable. f^o = 20%, a^o =20 pg/m^3.

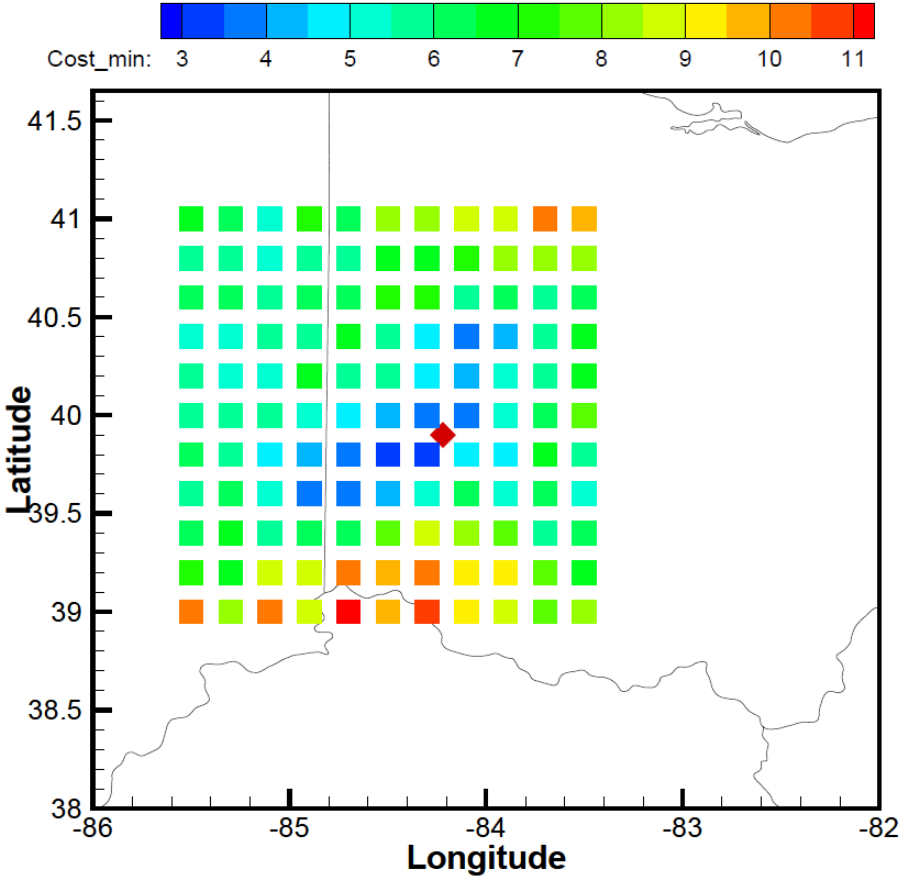
Emission (kg/hr)	a^h =10 pg/m^3	a^h =20 pg/m^3	a^h =50 pg/m^3
$f^h = 0$	69.3	64.0	62.1
$f^h = 10\%$	67.3	63.4	60.9
$f^h = 20\%$	65.3	61.5	59.1
$f^h = 50\%$	61.5	58.0	55.1

Why is it better to choose ln(c) as metric variable ?

Probability density function (pdf) of $\ln(c_h)-\ln(c_o)$ for the six CAPTEX releases



Source location identification



Distribution of 121 candidate source locations for release 2. The minimal cost function at each location associated with an optimal release strength is indicated by color. The cost function defined in Equation 5 is calculated with f^o = 20%, a^o =20 pg/m^3, f^h = 20%, and a^h =20 pg/m^3. The actual source location , Dayton, Ohio, U.S., is shown as a red diamond.

Inverse results for all releases

The source location (latitude, longitude) and release rate q_{min} identified by the minimal normalized cost function F_{min} for each CAPTEX releaseΔ is the distance between the point with F_{min} and the actual release site. q′ is the estimated release rate by assuming that the actual release location is known. Logarithm concentration is taken as the metric variable.

	Source location (latitude, longitude)		Δ(km)	Release rate (kg/hr)			
#	Actual	Estimated		Actual	q_{min}	q'	$\epsilon_{q'}$
1	39.80°, -84.05°	41.0°,-83.9°	134.2	69.3	23.9	106.3	6.2
2	39.90°, -84.22°	39.8°,-84.5°	26.4	67.0	48.5	61.5	1.8
3	39.90°, -84.22°	40.8°,-85.3°	135.8	67.0	63.4	41.7	2.6
4	39.90°, -84.22°	40.2°,-85.5°	114.1	66.3	185.7	75.1	4.6
5	46.62°, -80.78°	46.2°,-81.0°	49.7	60.0	72.9	42.6	3.0
7	46.62°, -80.78°	47.4°,-81.2°	92.5	61.0	201.0	66.0	3.9

Summary

- HYSPLIT inverse modeling was applied to the Fukushima nuclear accident and the release estimation agrees well with others;
- It is found that computing ln(c) differences between model and observations is better than using the concentration differences;
- Using CAPTEX data, concentration metric variable results in severe underestimation while ln(C) metric results in overestimation before introducing model uncertainty terms;
- Adding “dynamic” model uncertainty terms improves results for both choice of the metric variables in the cost function while ln(C) metric variable gives better results;
- A cost function normalization scheme is introduced to avoid spurious minimal source term solutions and proves to be effective;
- At last, the system is tested for its capability to find a single source location as well as its source strength.

Thank you!