Inverse modeling with HYSPLIT Lagrangian dispersion model

Tianfeng Chai

NOAA Air Resources Laboratory,
Atmospheric Sciences and Modeling Division (ASMD)
Cooperative Institute for Satellites Earth System Studies (CISESS),
University of Maryland

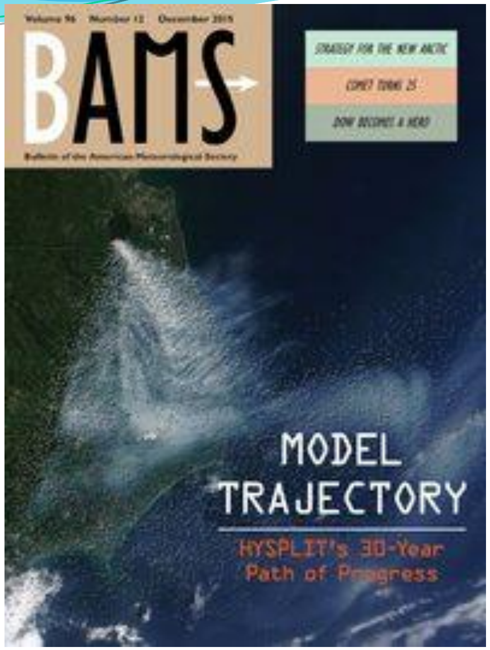
Atmospheric Constituents Data Assimilation and Inverse Modeling
Banff International Research Station
for Mathematical Innovation and Discovery, March 19-24, 2023

Outline

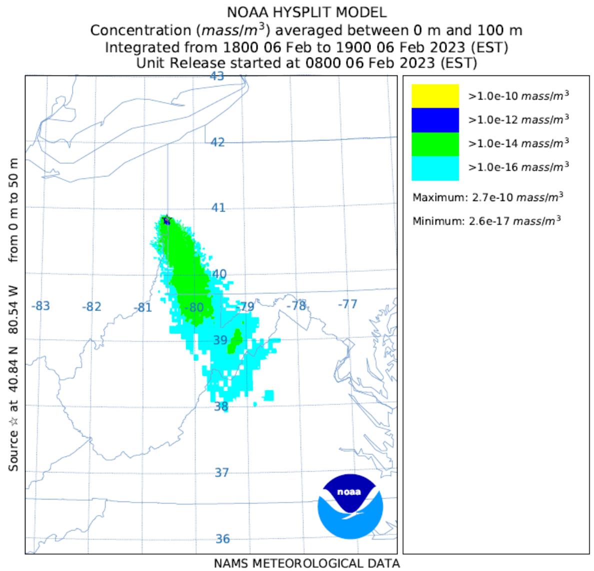
- HYSPLIT dispersion model
- Fukushima source term estimation
- Volcanic Ash application
- Wildfire application
- CAPTEX field experiment
- Power plant SO_2 emission estimation
- Summary

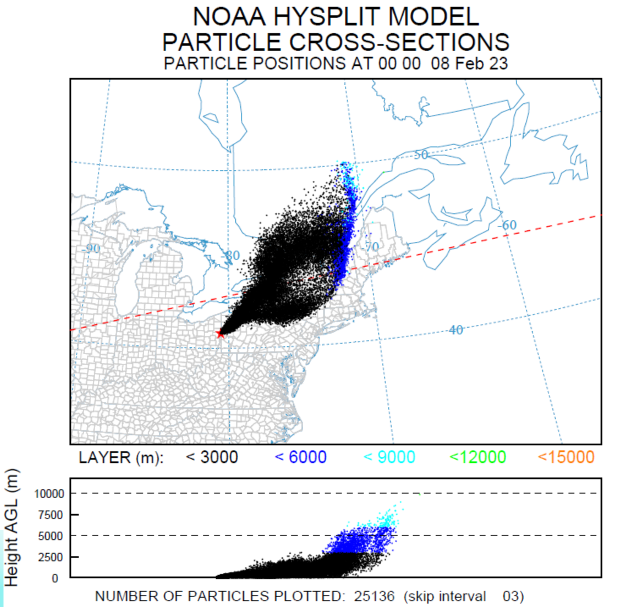
HYSPLIT model

- HYSPLIT (Hybrid Single Particle Lagrangian Integrated Trajectory Model) is more than a “trajectory model”
- HYSPLIT 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

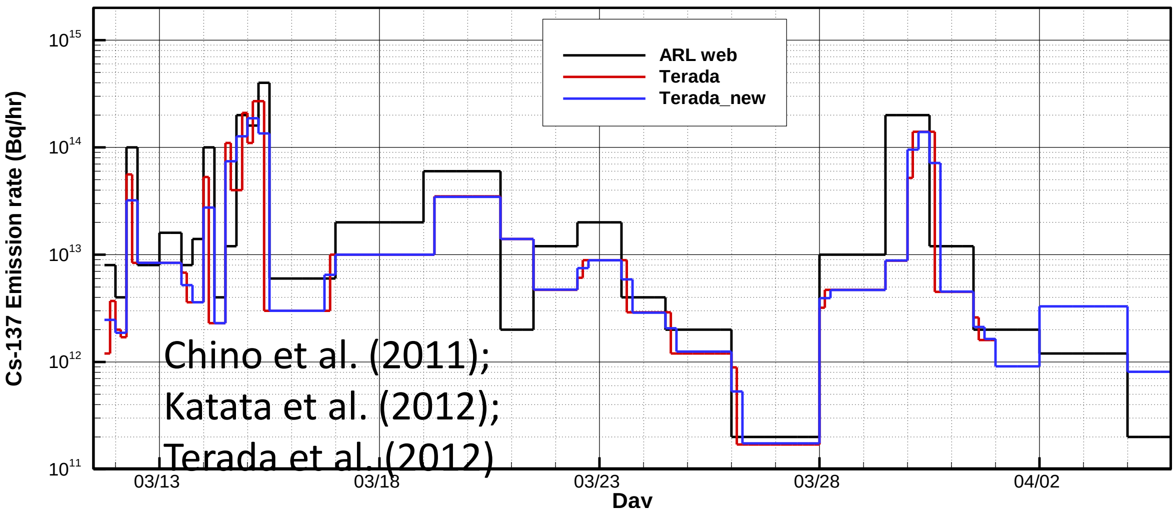


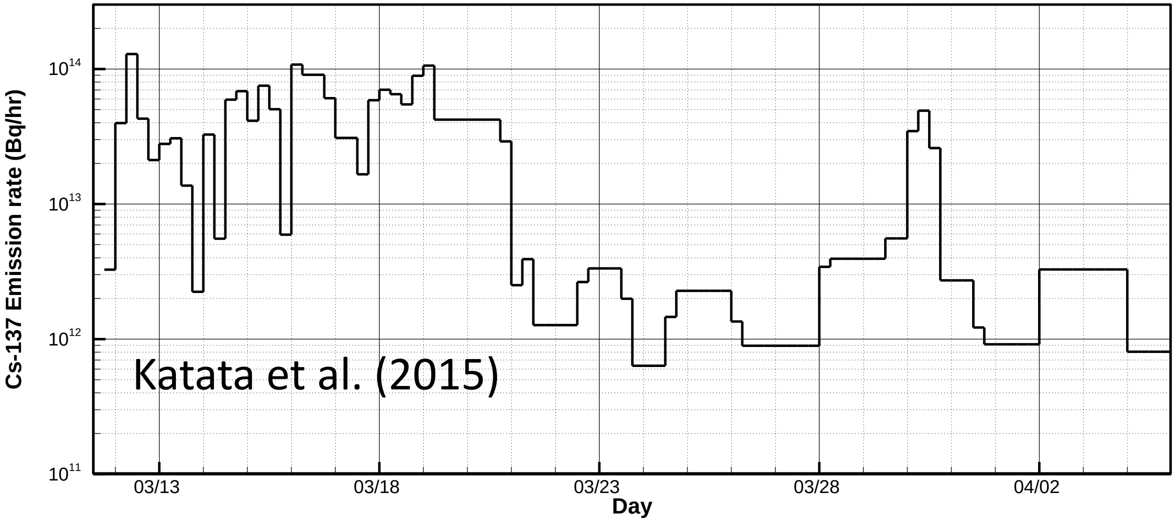


Train derailment in East Palestine, Ohio, on Feb. 3, 2023

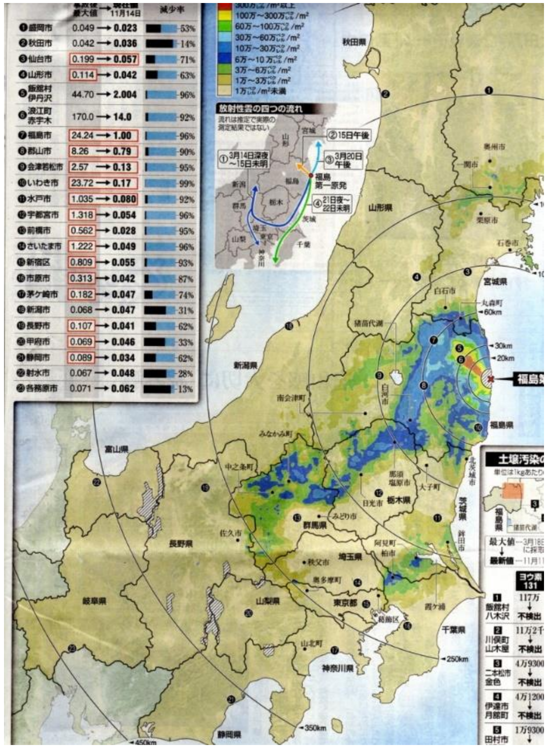
Fukushima Accident

Cs-137 emission estimations using local measurements

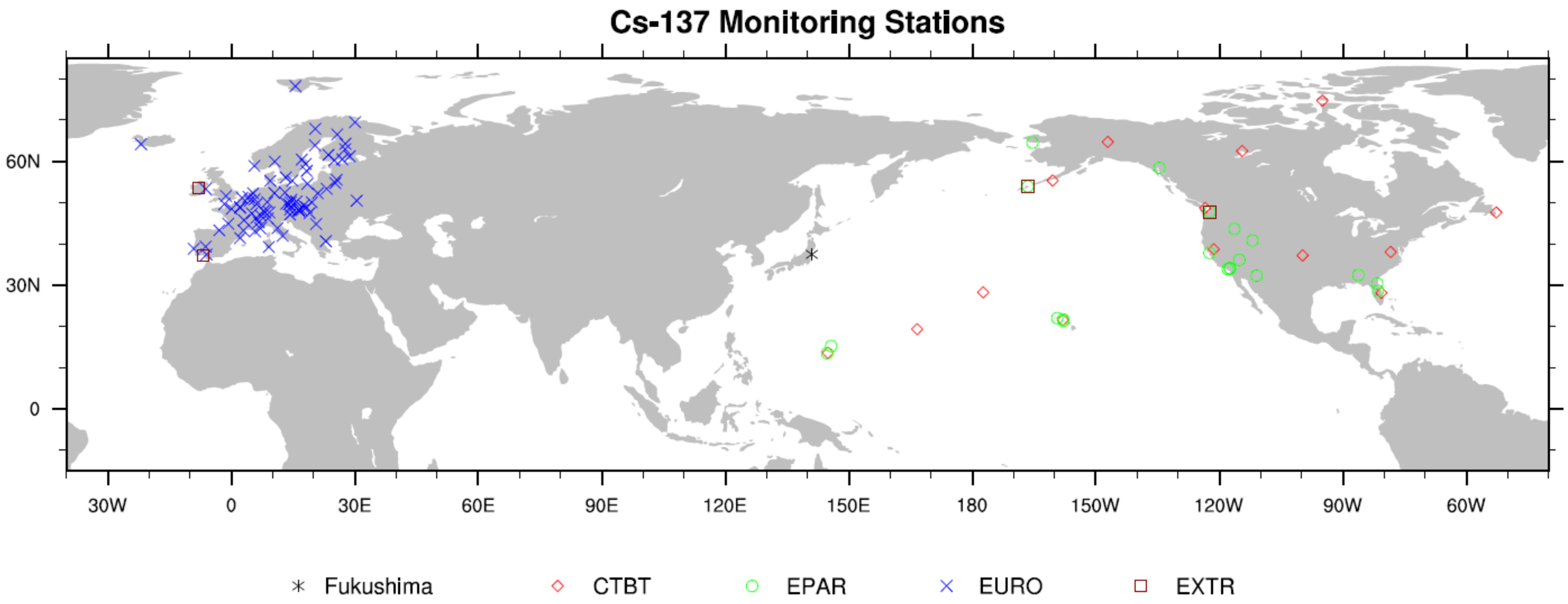




Cs-137 air concentrations



Global air Concentration measurements

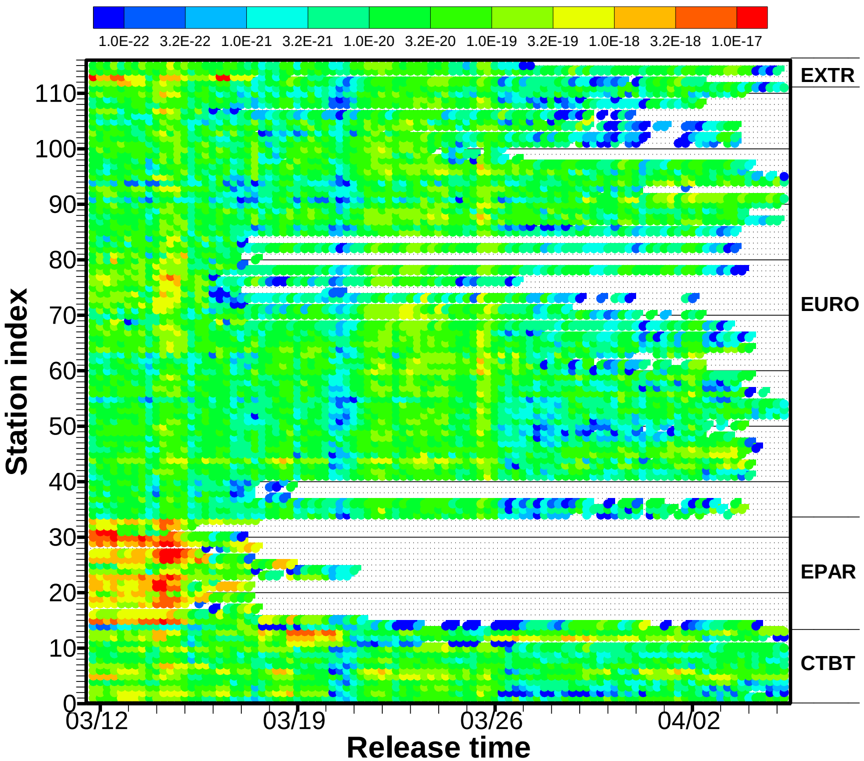


Data source	Number of monitoring stations	Count of total samples
CTBT	14	417 (421)
EPAR	19 (20)	35 (39)
EURO	78 (80)	785 (797)
EXTR	4	59 (61)
Total	115 (118)	1296 (1318)

Methodology

- A general data assimilation approach: minimizing a cost function that mainly measures the differences between observations and predictions, for the entire time period and all monitoring sites (4D-Var)

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



$$\mathcal{F} = \frac{1}{2}\sum_{n=1}^{N}\frac{(q_n - q_n^b)^2}{\sigma_n^2} + \frac{1}{2}\sum_{m=1}^{M}\frac{(c_m^h - c_m^o)^2}{\epsilon_m^2}$$

$$+ \quad c_{sm} \cdot \sum_{n=2}^{N-1}\frac{1}{2}\left[\frac{(q_{n-1} - q_{n-1}^b) - 2\cdot(q_n - q_n^b) + (q_{n+1} - q_{n+1}^b)}{q_c}\right]^2$$

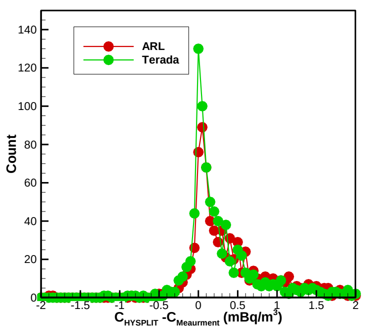
Twin experiment

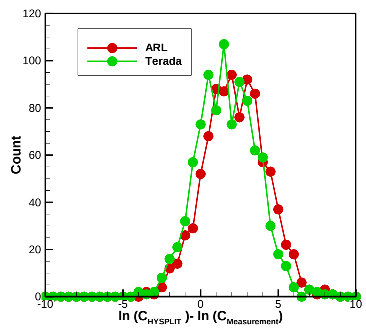
Pseudo-observations are generated with known sources $\mathbf{c^p} = \mathbf{H} \cdot \mathbf{q} + \varepsilon$

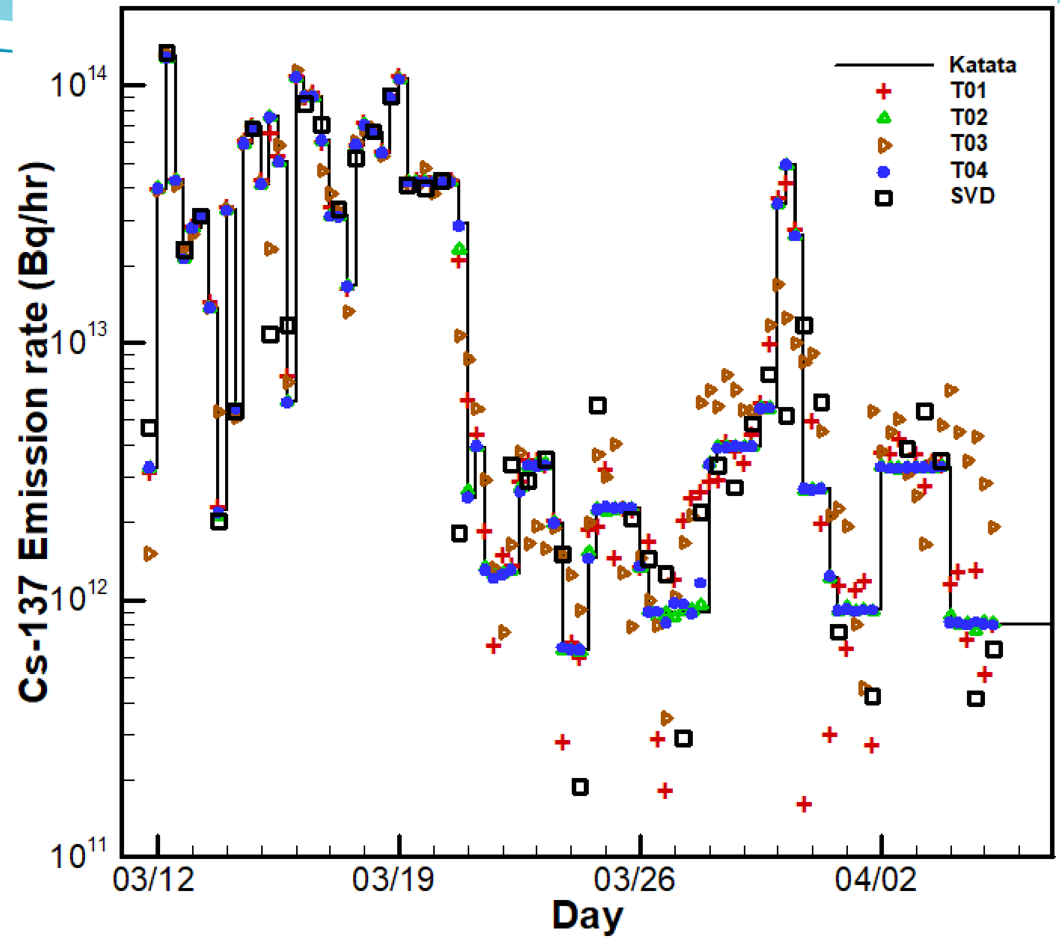
$$\varepsilon_m^p = (f_m^p \times c_m^p + a_m^p) \times r_m, \quad m = 1, \cdots, M$$

Control and metric variables

Case	x	y
T01	q	c
T02	q	ln(c)
T03	ln(q)	c
T04	ln(q)	ln(c)



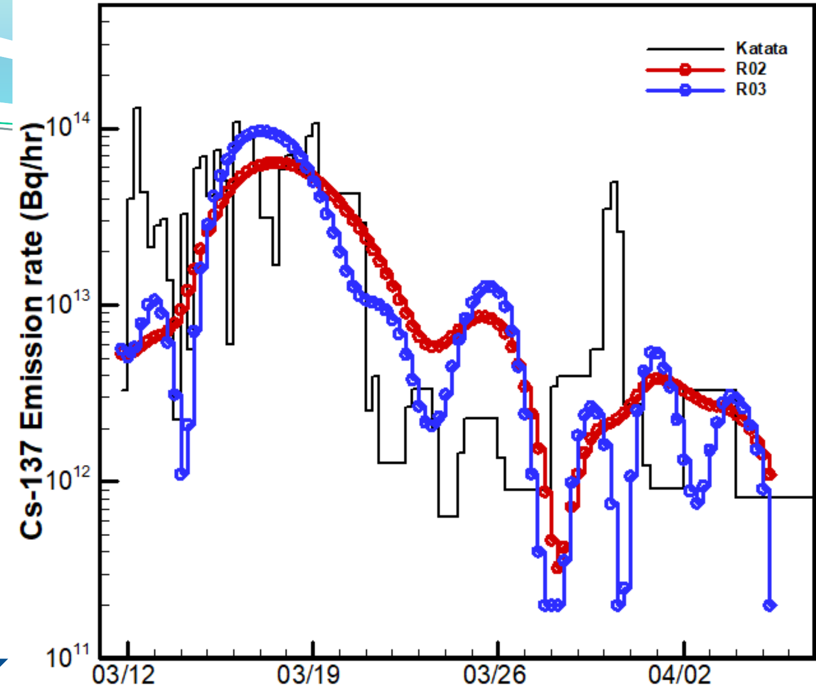




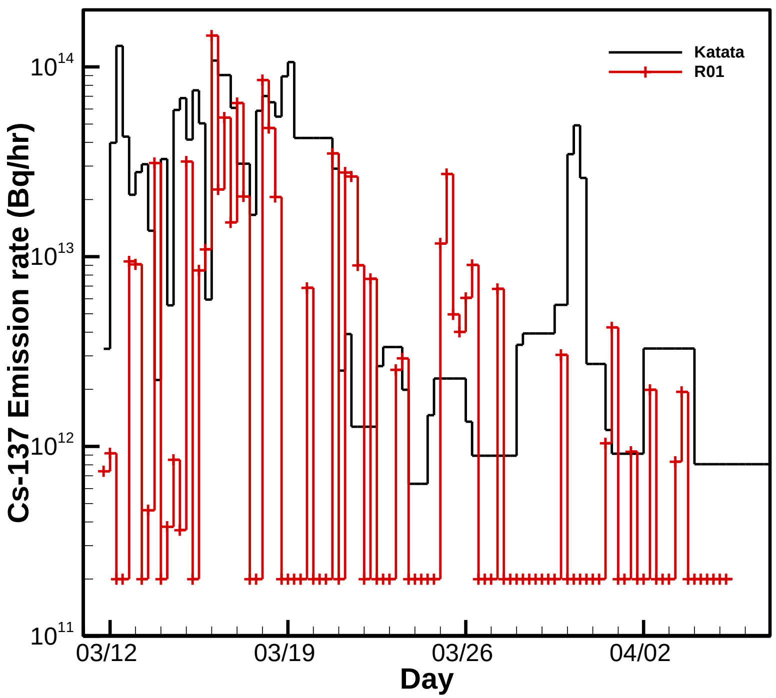
Case	Cs-137 release			
	MAE	RMSE	MRE	R
T01	0.90×10^{12}	1.88×10^{12}	0.229	0.998
T02	0.11×10^{12}	0.61×10^{12}	0.012	1.000
T03	3.32×10^{12}	7.66×10^{12}	0.740	0.965
T04	0.03×10^{12}	0.08×10^{12}	0.011	1.000
SVD	3.57×10^{12}	9.06×10^{12}	0.588	0.952

MAE: mean absolute error;
RMSE: root-mean-square error;
MRE: mean relative error;
R: linear correlation coefficient;
SVD: Direct Singular-value-decomposition solution.

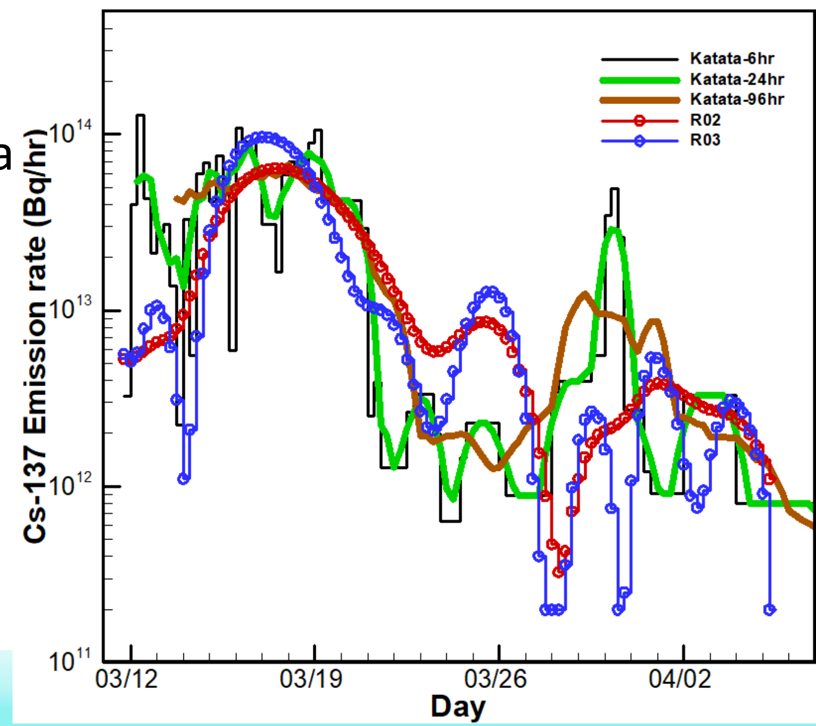
Adding Smoothness penalty



Real data



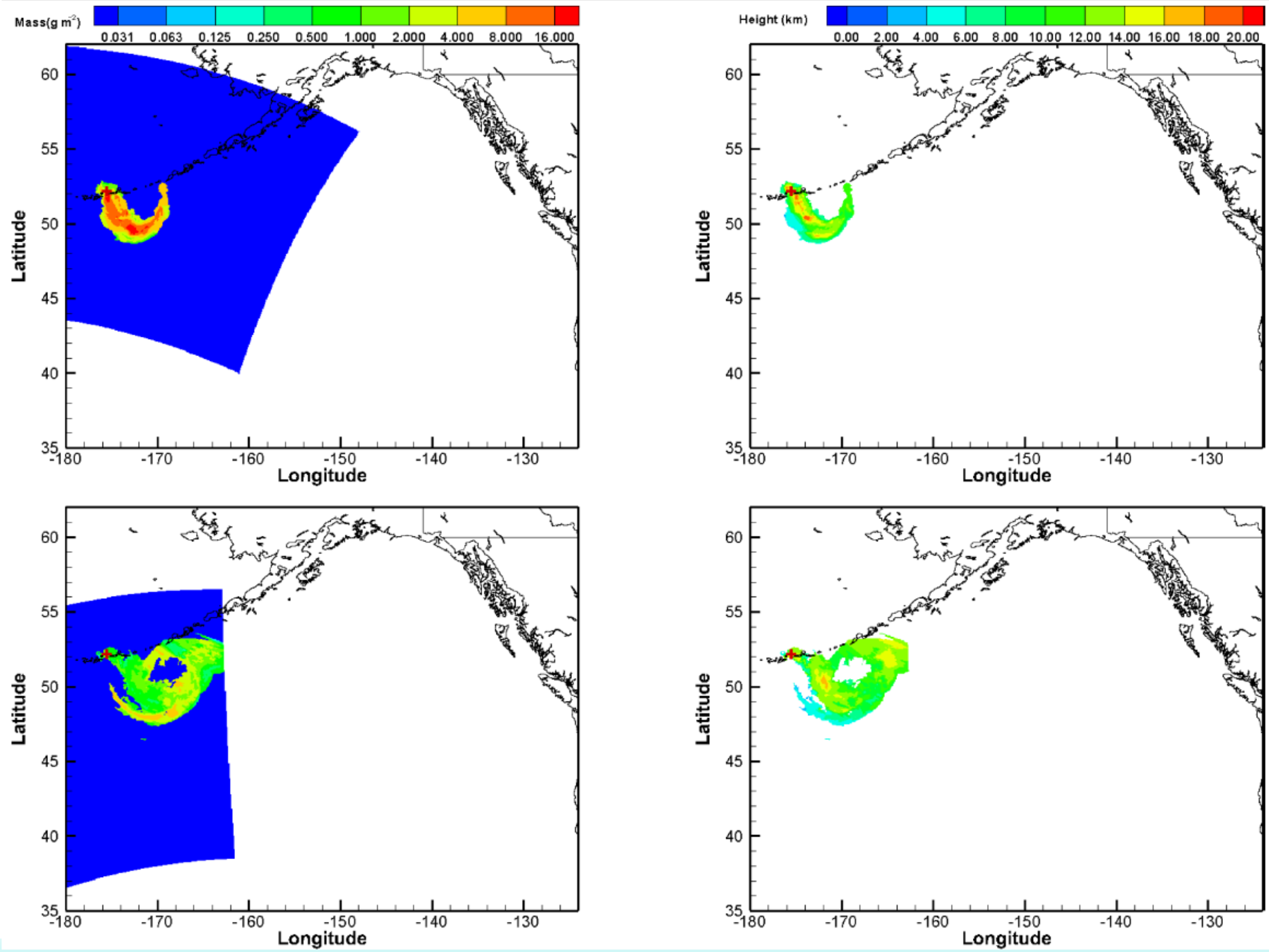
Running-average of Katata data



Source term estimation using air concentration measurements and a Lagrangian dispersion model–Experiments with pseudo and real cesium-137 observations from the Fukushima nuclear accident, T Chai, R Draxler, A Stein, Atmospheric Environment 106, 241-251, 2015

Volcanic ash - Kasatochi eruption in 2008

Satellite observations include mass loadings and ash cloud top height

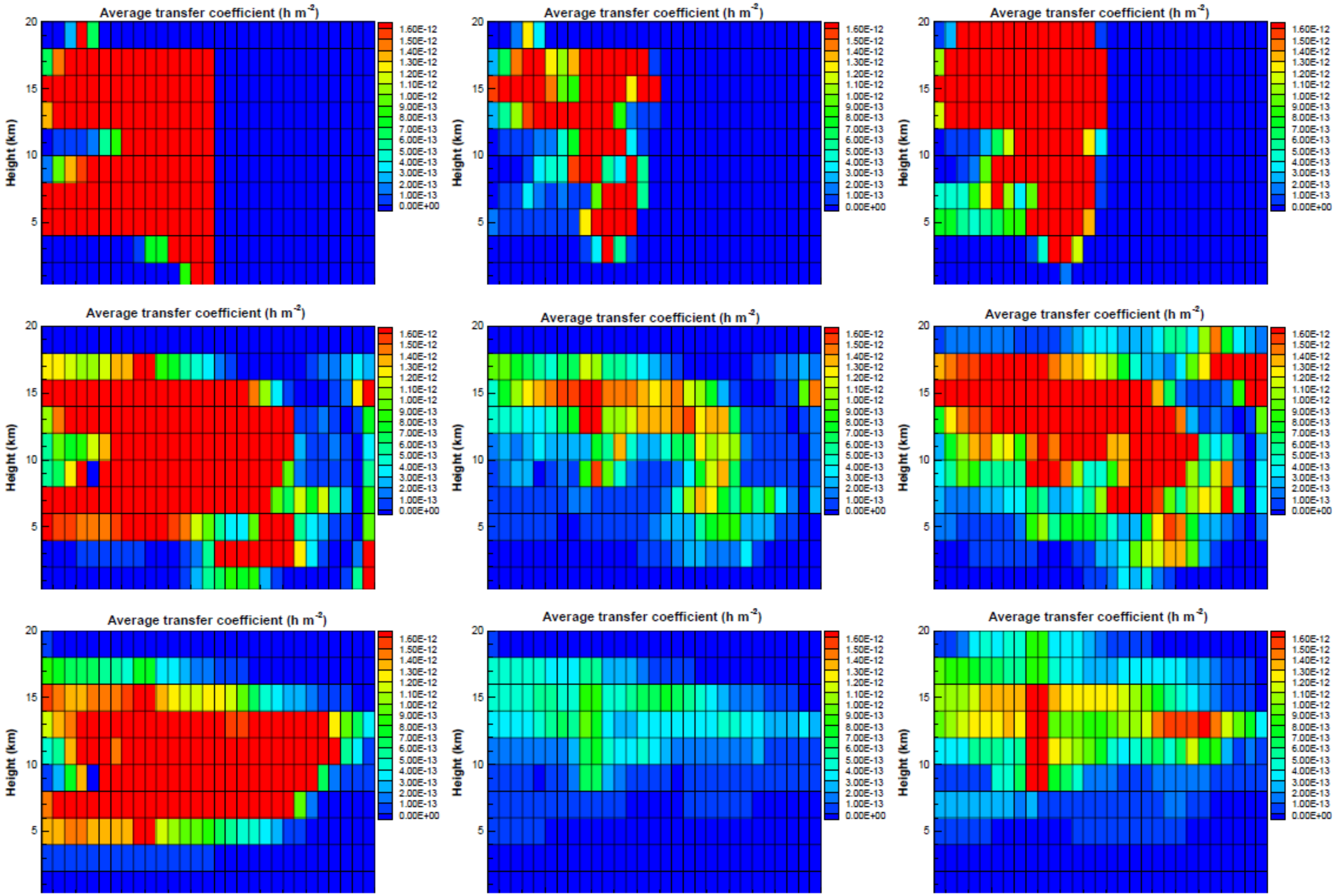


Granule1
13:40 UTC
August 8, 2008

Granule 2
00:50 UTC
August 9, 2008

Sensitivities of observations to 2-D Source terms

Model mass loadings are calculated in three different ways



Granule1
13:40 UTC
August 8, 2008

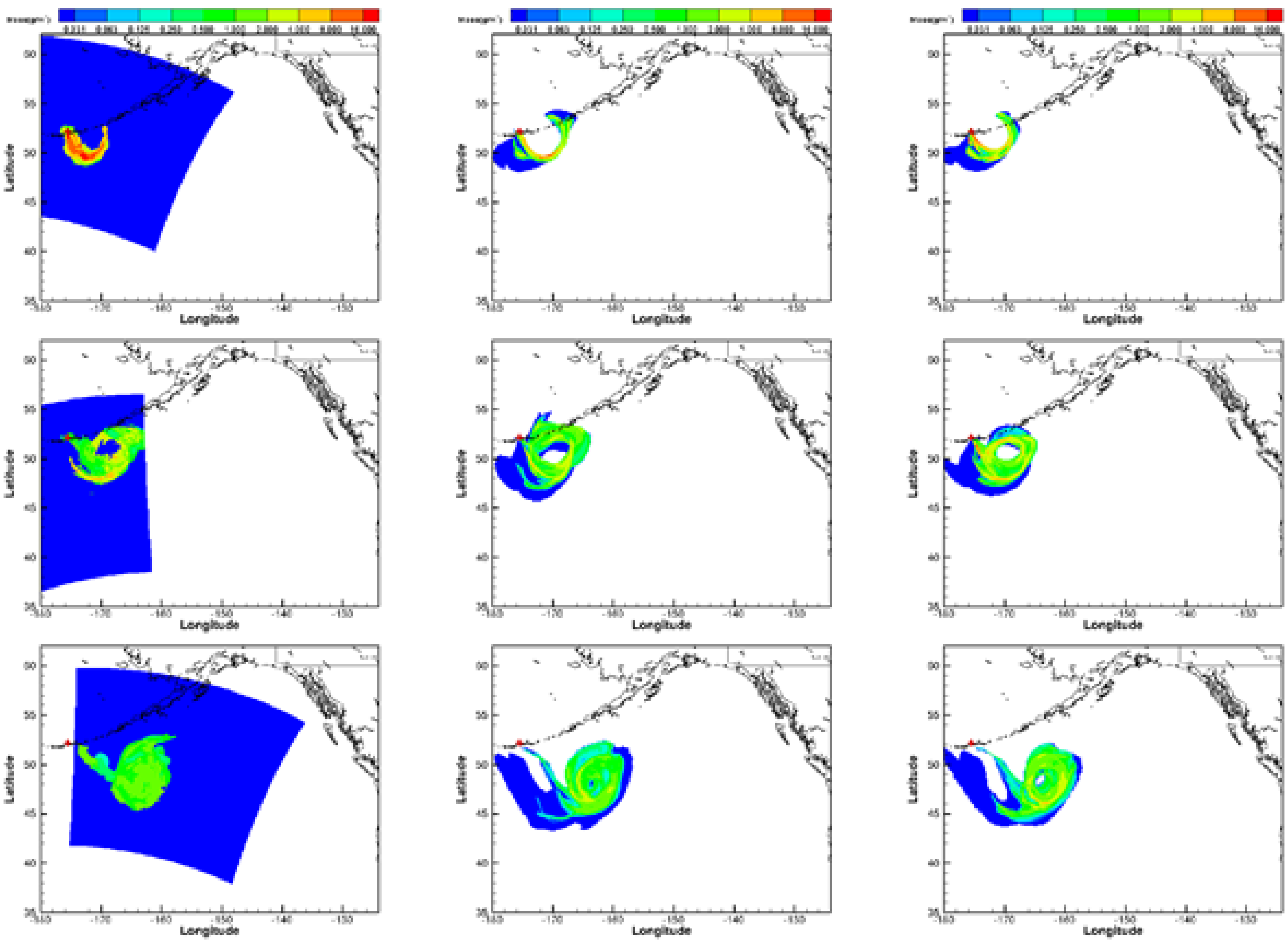
Granule 2
00:50 UTC
August 9, 2008

Granule 3
12:50 UTC
August 9, 2008

Surface to cloud top; 1-km layer; 3 layers (3km)

Simulated ash mass loadings with estimated emissions

MODIS | HYSPLIT w/ GDAS | HYSPLIT w/ ECMWF | 3 layer (3km)

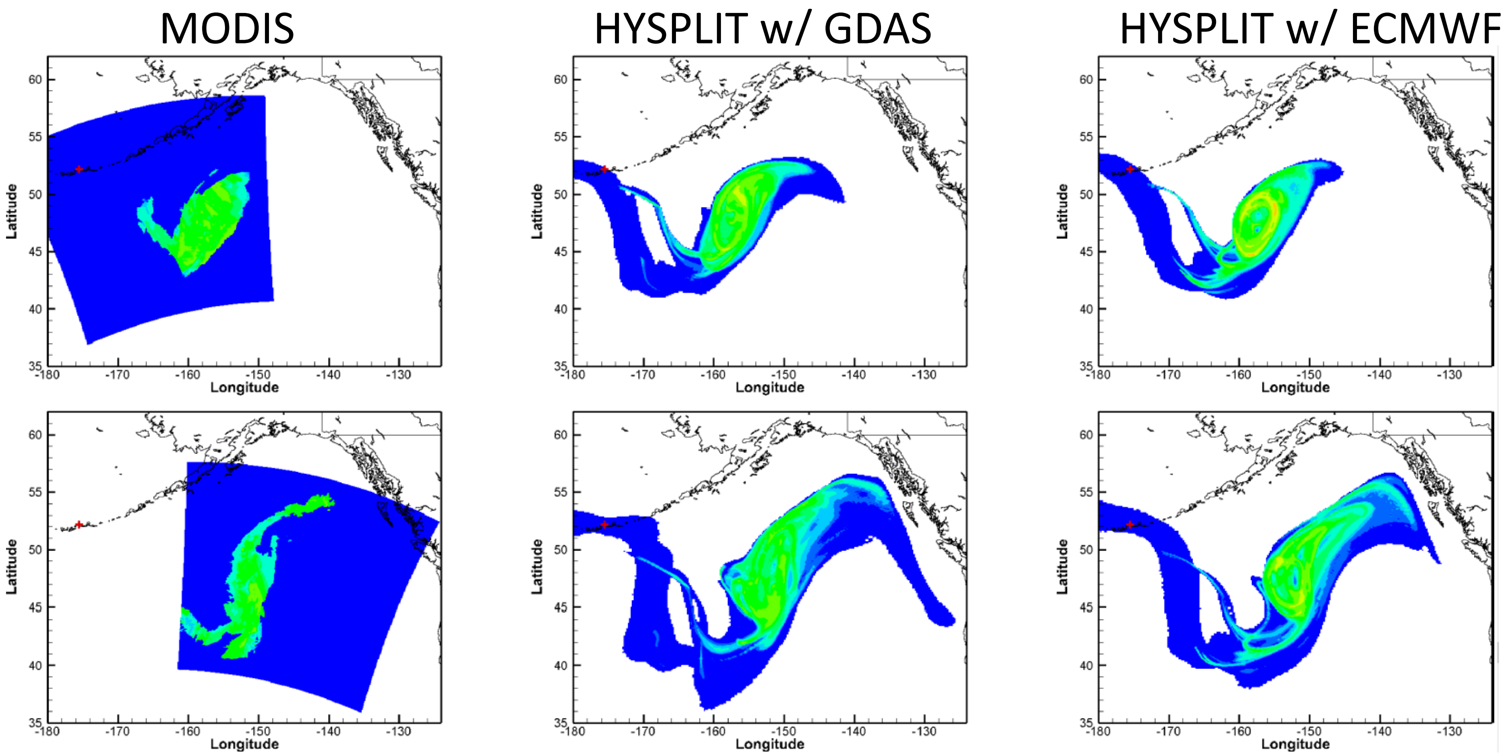


Granule1
13:40 UTC
August 8, 2008

Granule 2
00:50 UTC
August 9, 2008

Granule 3
12:50 UTC
August 9, 2008

"Forecasts" with the estimated volcanic ash emissions



Granule 4
00:00 UTC
August 10, 2008

Granule 5
11:50 UTC
August 10, 2008

Inputs	GDAS				ECMWF			
	G2	G3	G4	G5	G2	G3	G4	G5
G2	(2.70)	2.69	2.86	2.27	(2.90)	2.76	2.76	2.29
G1,G2	(2.66)	2.77	3.02	2.32	(2.90)	2.80	2.78	2.28
G3	2.59	(3.16)	2.89	2.20	2.43	(3.07)	2.78	2.10
G2,G3	(2.69)	(2.94)	2.94	2.26	(2.76)	(2.91)	2.81	2.23
G1,G2,G3	(2.61)	(2.93)	2.96	2.28	(2.77)	(2.98)	2.86	2.20

$$\mathrm{Rank} = R^2 + \left(1 - \frac{|\mathrm{FB}|}{2}\right) + \mathrm{CSI} + (1 - \mathrm{KSP}).$$

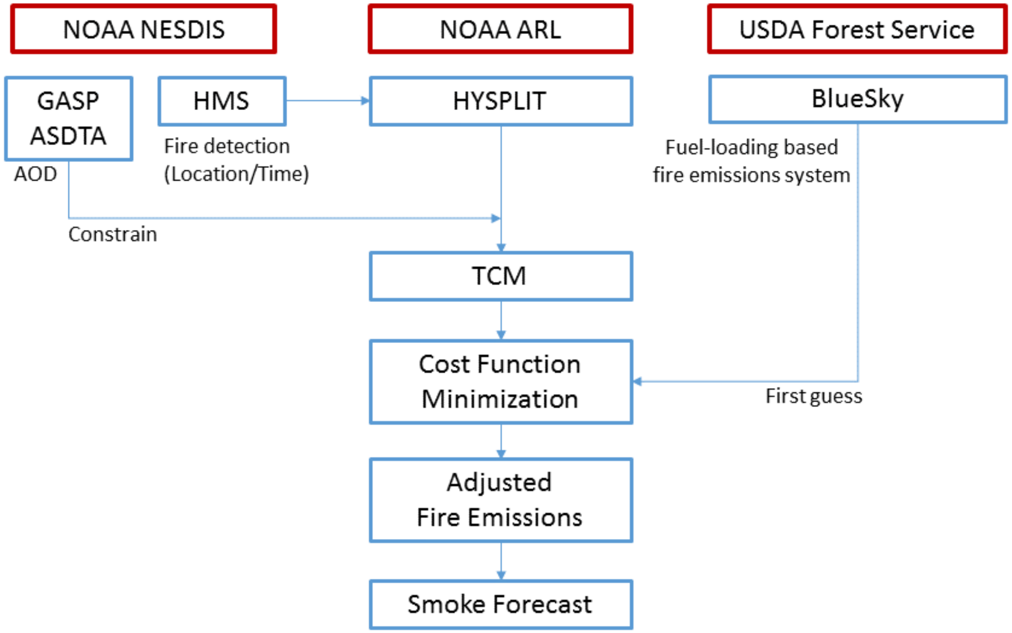
$$\mathrm{CSI} = \frac{N_{\mathrm{Hit}}}{N_{\mathrm{FalseAlarm}} + N_{\mathrm{Hit}} + N_{\mathrm{Miss}}},$$

Kolmogorov–Smirnov parameter (KSP): largest difference between the cumulative distribution functions of the model predictions and the observations

Chai, et al Atmos. Chem. Phys. 17, 2865-2879, 2017
Crawford, et al Atmos. Chem. Phys., 22, 13967–13996, 2022

Wildfire application

- Wildfire estimated using the satellite observations of the fire plumes in a 4D sense using HYSPLIT dispersion model.



HYSPLIT wildfire inverse modeling system

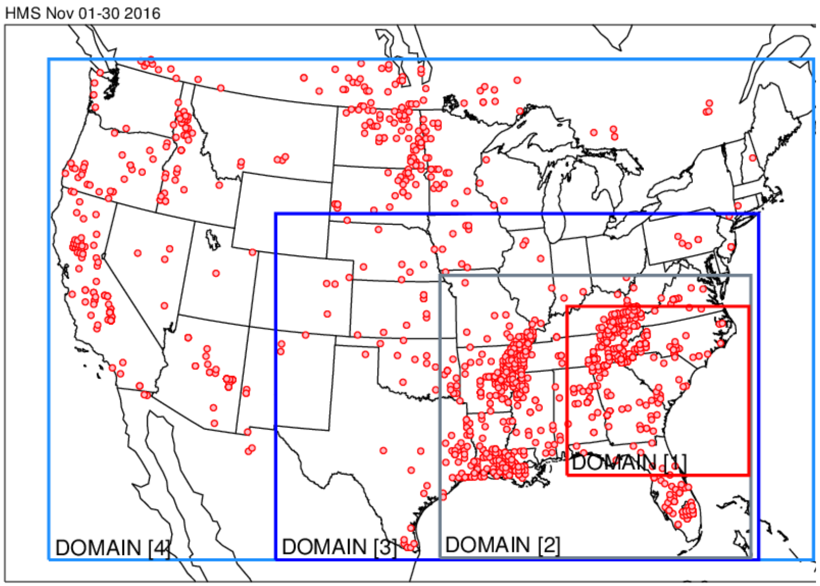
A helicopter makes a water bucket drop as it flies through smoky air while fighting a wildfire that flared up in the late afternoon near Omak, Wash., on Thursday (Aug 27, 2015). (Ted S. Warren/AP)

http://www.seattletimes.com/seattle-news/northwest/washington-wildfires-update-2/

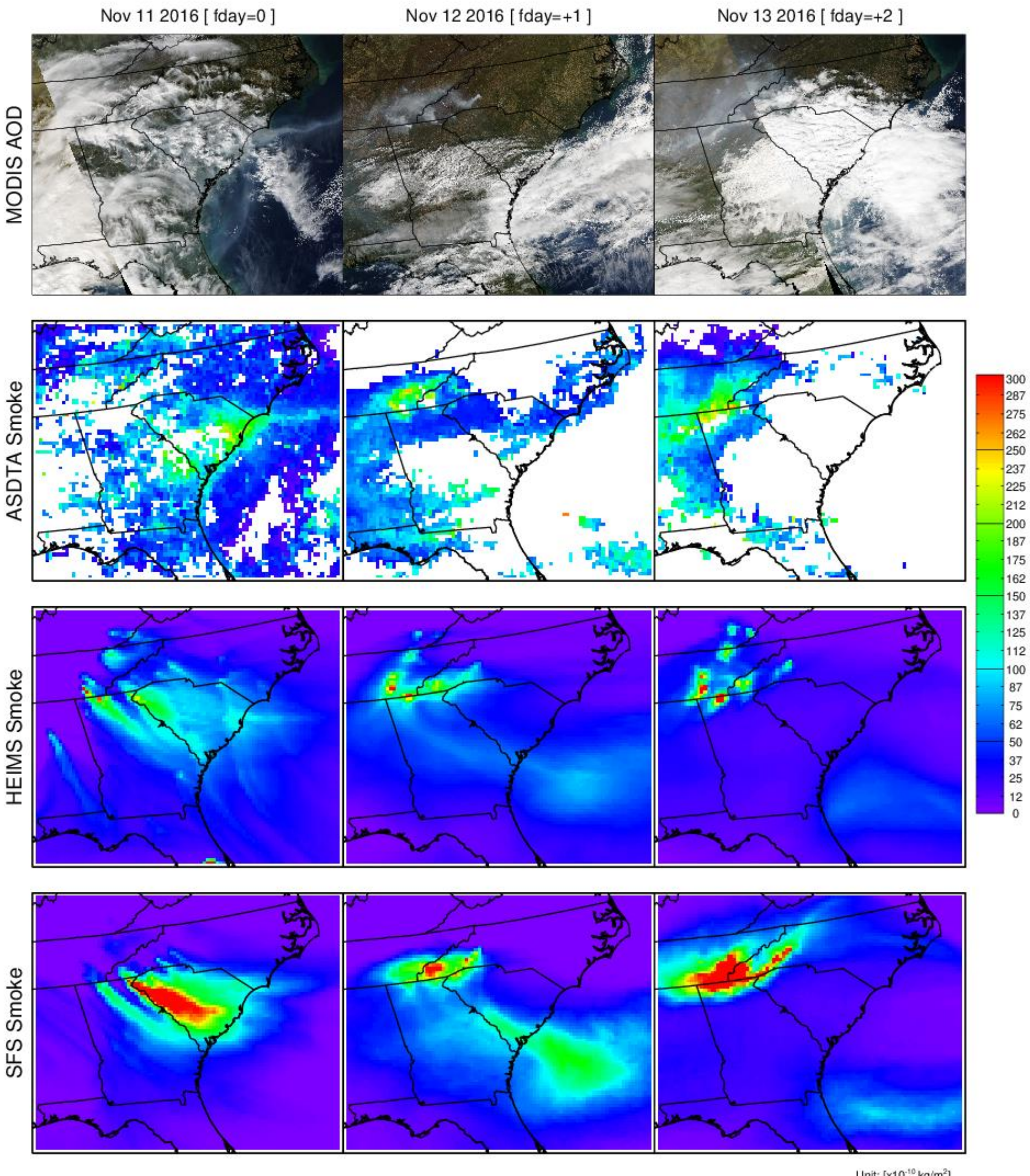
Meanwhile since Friday, more than 1,000 firefighters have struggled with a blaze started by lightning in the Chelan, Wash., area, where at least 49 buildings have been destroyed and authorities have issued evacuations that affect some 3,000 people.

http://news.discovery.com/earth/weather-extreme-events/will-more-wildfires-combust-our-health-150828.htm

A case study



Red dots indicate HMS detected fire locations in November, 2016.



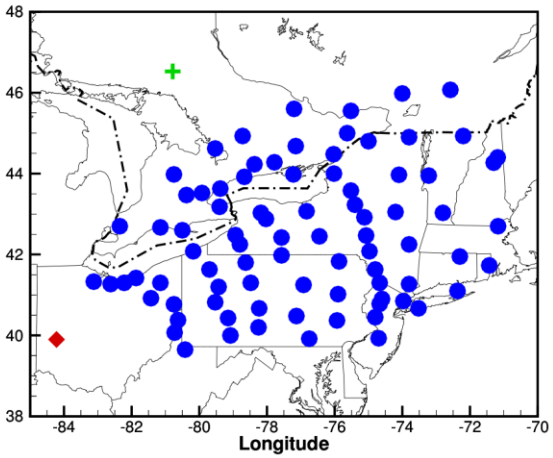
Inverse modeling of fire emissions constrained by smoke plume transport using HYSPLIT dispersion model and geostationary satellite observations, Kim, H. C., Chai, T., Stein, A., and Kondragunta, S., Atmos. Chem. Phys., 20, 10259–10277

True color image from MODIS, ASDTA smoke, HEIMS smoke hindcast, and SFS smoke forecast (from operation).

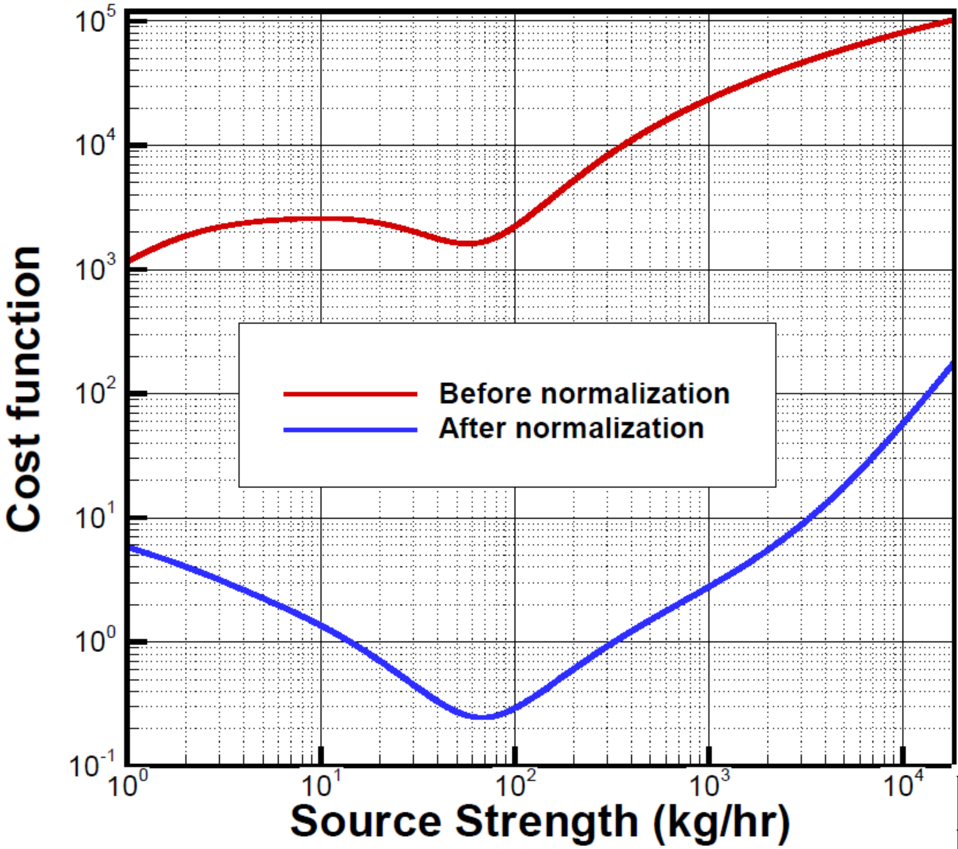
Cross Appalachian Tracer Experiment (CAPTEX)

- 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



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



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.

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

Estimates of R2 (actual release 67 kg/hr). Logrithmic concentration is used as the metric variable. $f^o = 20\%$, $a^o = 20$ pg/m^3.

Emission (kg/hr)	$a^h = 10\ pg/m^3$	$a^h = 20\ pg/m^3$	$a^h = 50\ pg/m^3$
$f^h = 0$	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

Inverse results for all releases

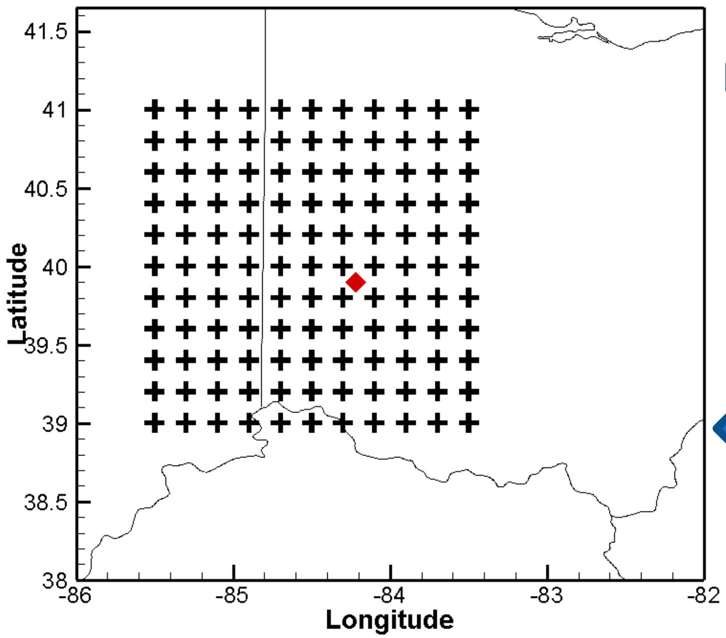
The source location (latitude, longitude) and release rate q_{min} identified by the minimal normalized cost function F_{min} for each CAPTEX releaseΔ is the distance between the point with F_{min} and the actual release site. q' is the estimated release rate by assuming that the actual release location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

Chai, T., Stein, A., and Ngan, F.: Weak-constraint inverse modeling using HYSPLIT-4 Lagrangian dispersion model and Cross-Appalachian Tracer Experiment (CAPTEX) observations – effect of including model uncertainties on source term estimation, Geosci. Model Dev., 11, 5135–5148, https://doi.org/10.5194/gmd-11-5135-2018, 2018.

Forward and backward source-receptor sensitivities

- Sensitivity runs are based on CAPTEX release 2 (3-hour release from Dayton, Ohio, USA, 17-20Z, Sep 25, 1983).
- 121 forward HYSPLIT dispersion runs are carried out from different candidate locations ("+" in the left figure). The forward sensitivities to 400 measurements from 84 monitoring sites (circles in right figure) are calculated.
- 400 backward runs starts from each measurement site and measurement ending time. The backward sensitivities at each of the 121 candidate locations at the release time are calculated.

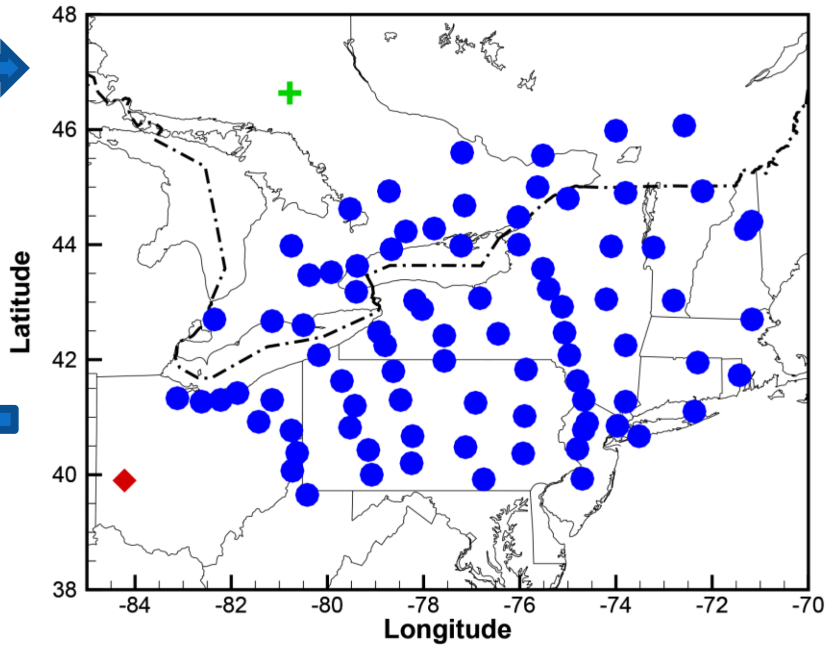


121 forward runs

48400 source-receptor sensitivities are calculated from both forward and backward runs.

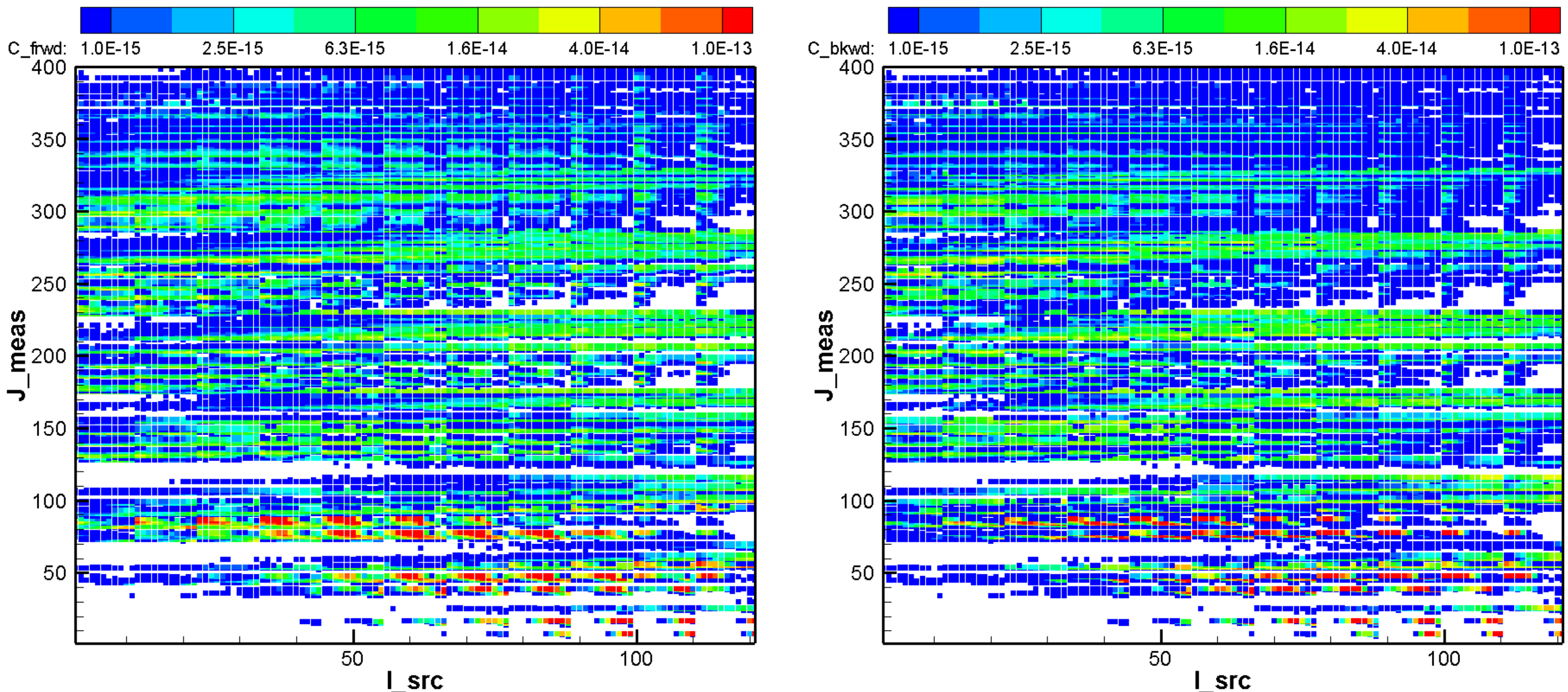
400 backward runs

Dayton, Ohio, U.S. shown as a red diamond, and Sudbury, Ontario, Canada shown as a green cross



Forward and backward TCMs

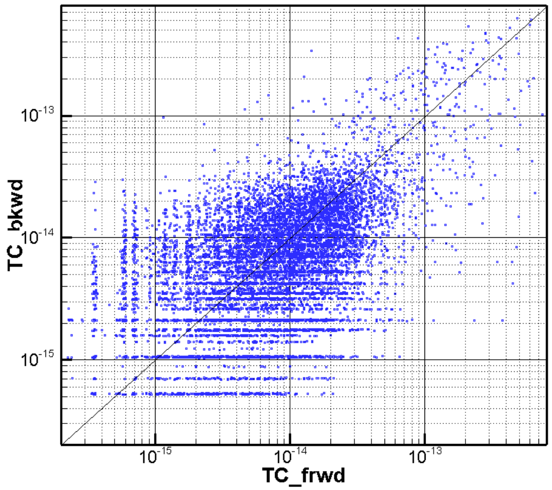
Number of particles: 160,000



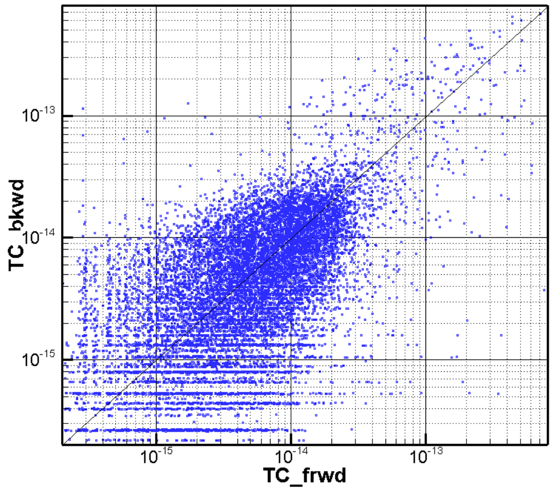
For comparison, the backward TCM is rotated to be in the same shape as the forward TCM.

Scatter plots of forward and backward sensitivities with three different number of 3D particles (Np) in HYSPLIT

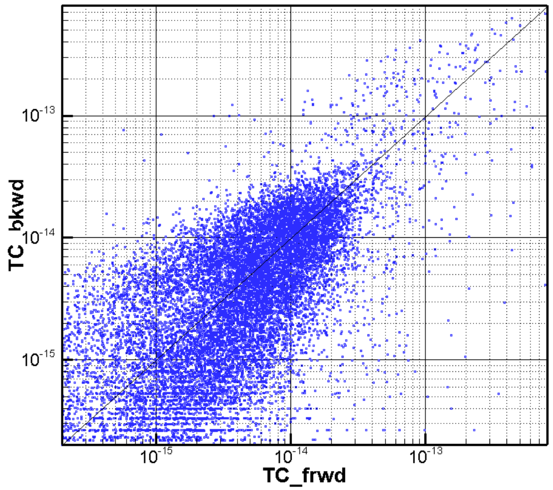
Np= 2,500, R_{ln}= 0.52



Np= 10,000, R_{ln}= 0.60



Np= 40,000, R_{ln}= 0.68



Increasing number of 3D particles used in the HYSPLIT model improves the agreement between forward and backward source-receptor sensitivities. Np=40,000 will be used in the later tests.

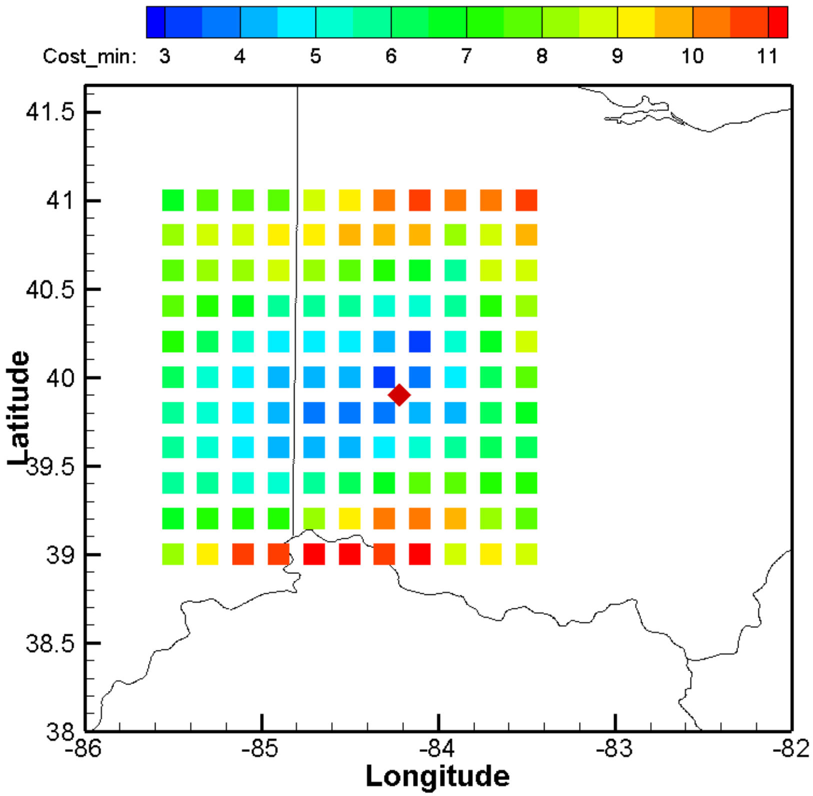
Note: Correlation coefficient R_{ln} is calculated after taking logarithm operation of the sensitivities.

Inverse modeling results using forward and backward TCMs

Minimal cost function associated with an optimal emission at all candidate locations

Actual release site:39.90ºN, 84.22ºW, marked as a red diamond); Actual release rate: 67 kg/hour.

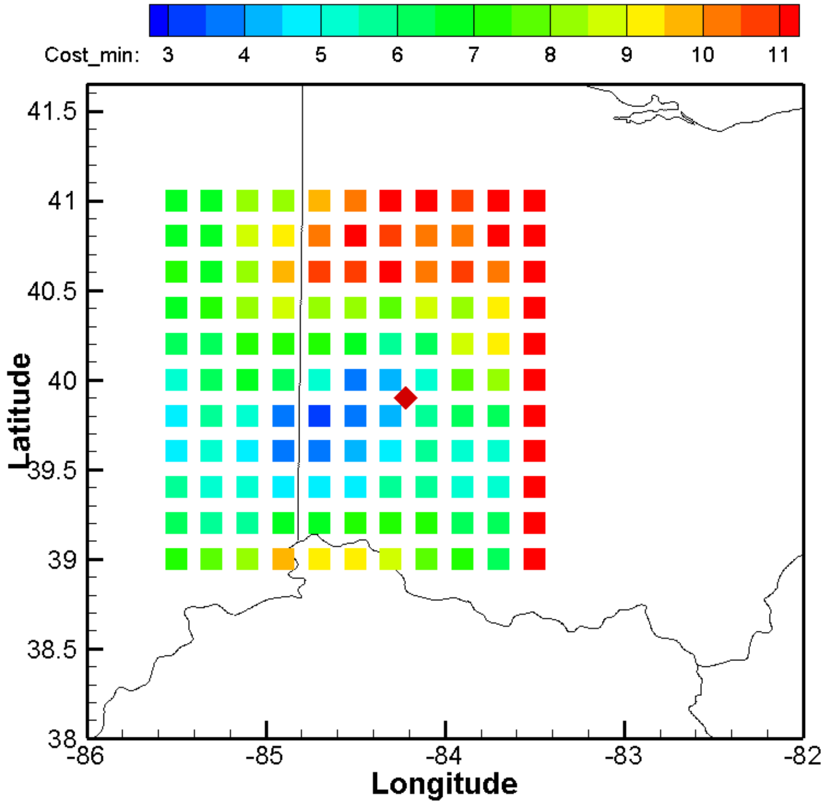
Results using Forward TCM



Estimated location: 40.0ºN, 84.3ºW, Δ= 13.0km

Estimated emission: 65.3 Kg/hour

Results using Backward TCM

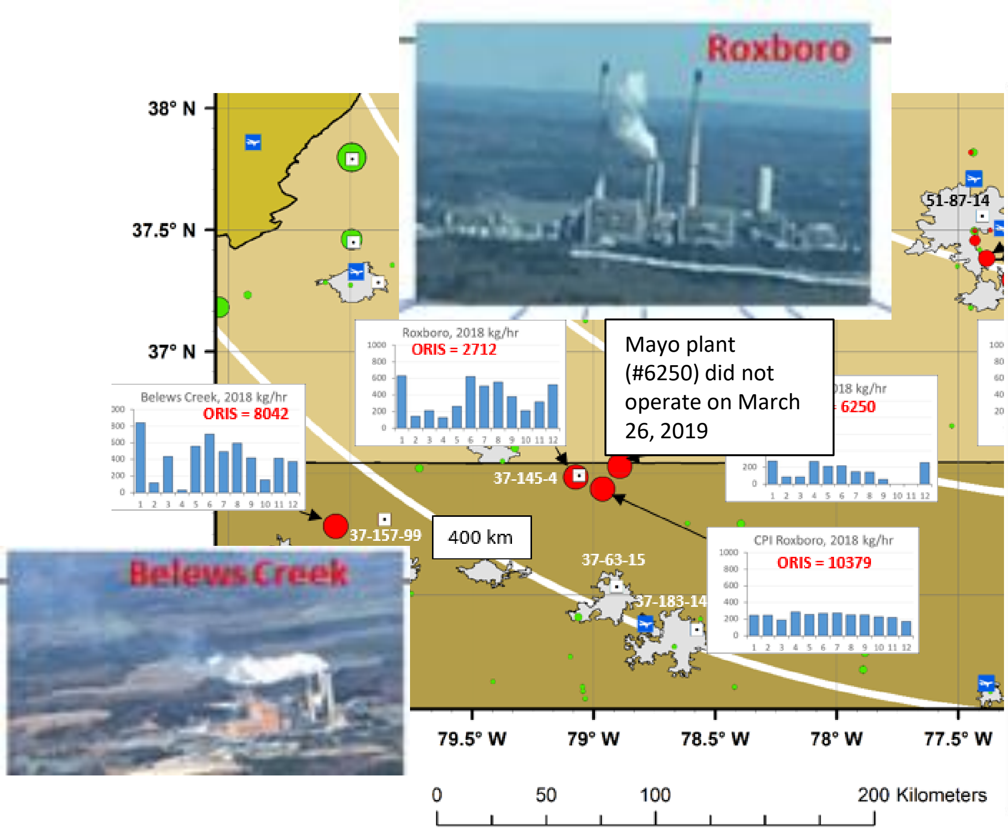


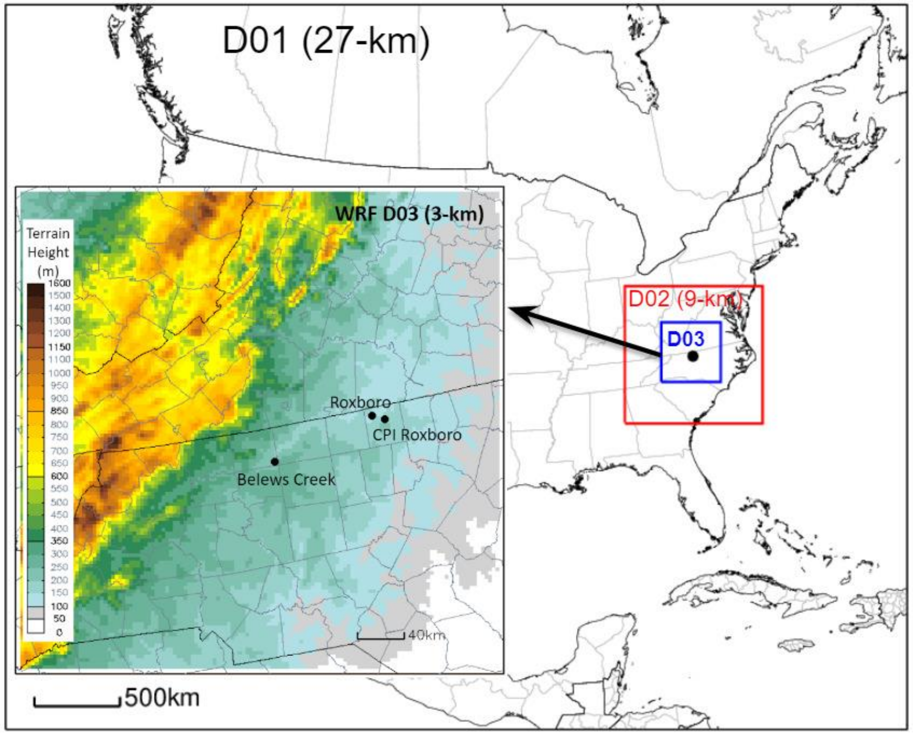
Estimated location: 39.8ºN, 84.7ºW, Δ= 42.5km

Estimated emission: 80.5 Kg/hour

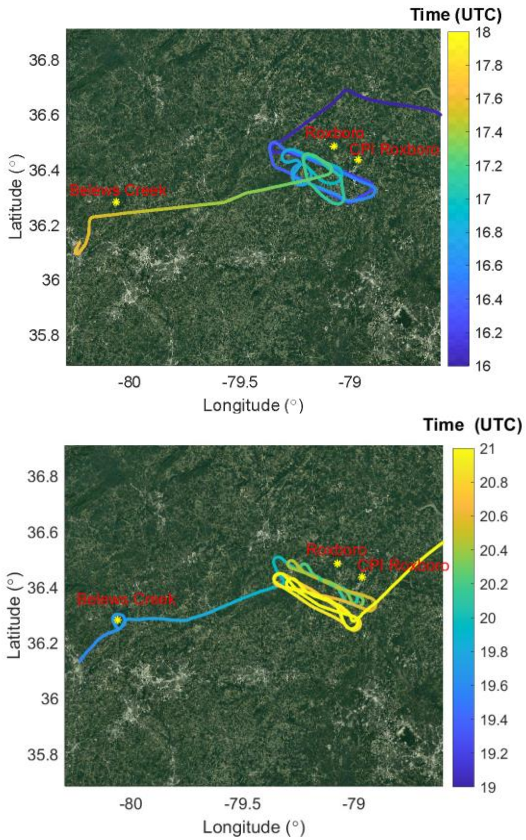
Power plant SO_2 emission estimation using aircraft observations with plume rise ensemble runs

Power plant locations and WRF domains

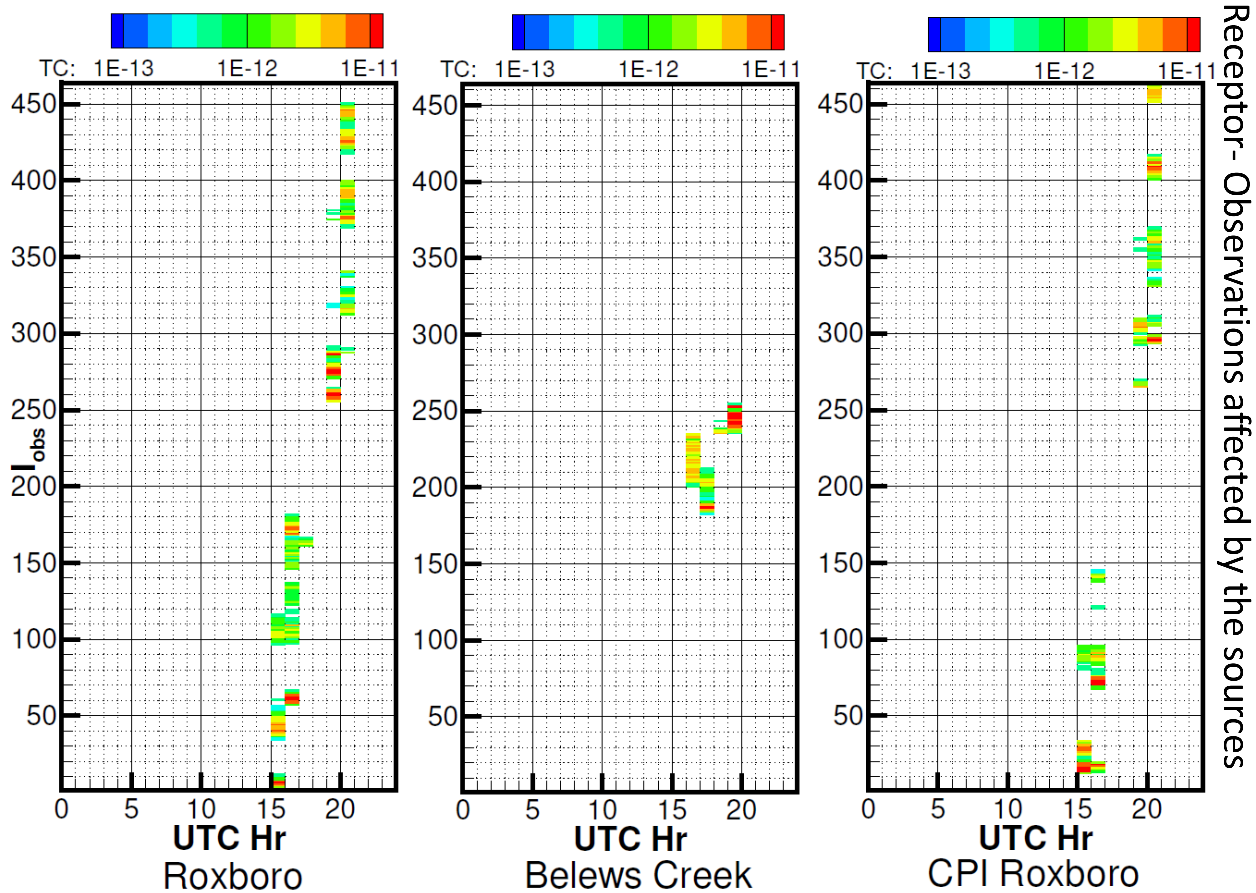




Transfer Coefficient Matrix (TCM) – SO_2



Morning and afternoon flight tracks



Plume rise – Briggs (1969)

$$\Delta H = \begin{cases} 1.3 \frac{F_b}{\bar{u} {u_*}^2}, & neutral,\ unstable \\ 2.6 {F_b}^{1/3} \bar{u}^{-1/3} s^{-1/3}, & stable, \bar{u} > 0.5\ m/s \\ 5.3 {F_b}^{1/4} s^{-3/8}, & stable, \bar{u} \leq 0.5\ m/s \end{cases}$$

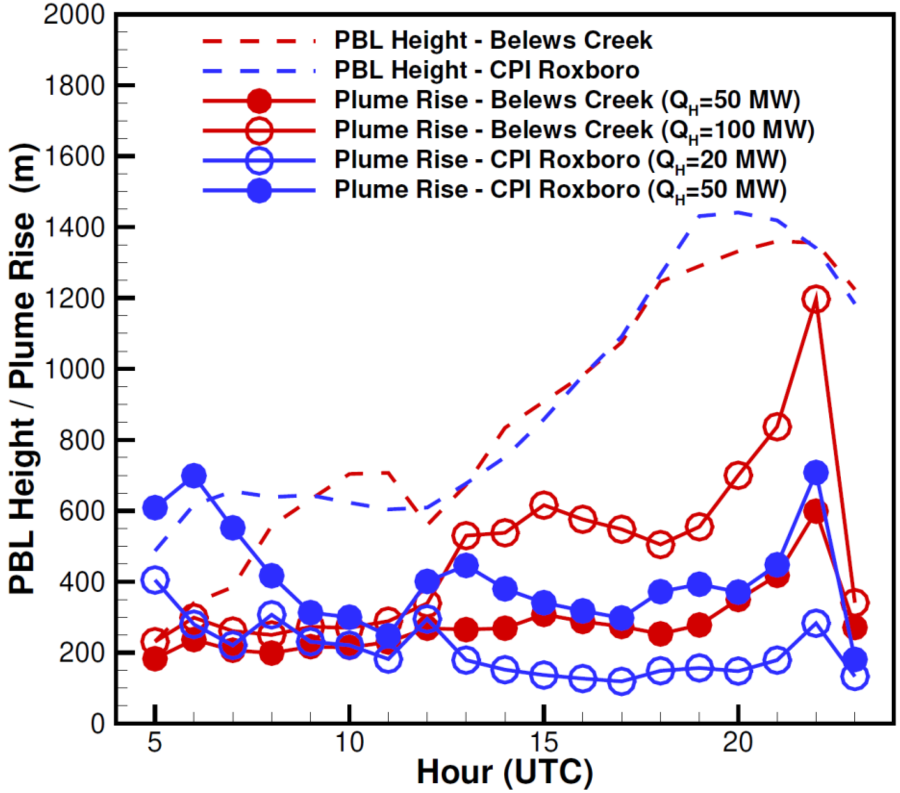
F_b is the buoyancy flux term, $\bar{u}$ is the mean wind speed, u_* is the friction velocity, and s is the static stability parameter

$$s = \frac{g}{T_v} \frac{\partial \bar{\theta}_v}{\partial z}$$

g is gravitational acceleration. T_v is the moist air virtual temperature. $\bar{\theta}_v$ is the mean virtual potential temperature.

$$F_b = \frac{g Q_H}{\pi c_p \rho T} \approx 8.8 \times 10^{-6} [\frac{m^4/sec^3}{watts}] Q_H [watts],$$

c_p, ρ, and T are the specific heat at constant pressure, average density, and temperature of ambient air, respectively. Q_H is the heat emission from the stack.



PBL heights and the plume rise calculated with different Q_H at Belews Creek and CPI Roxboro on Mar. 26, 2019.

RMSEs with estimated emissions using with different Q_H

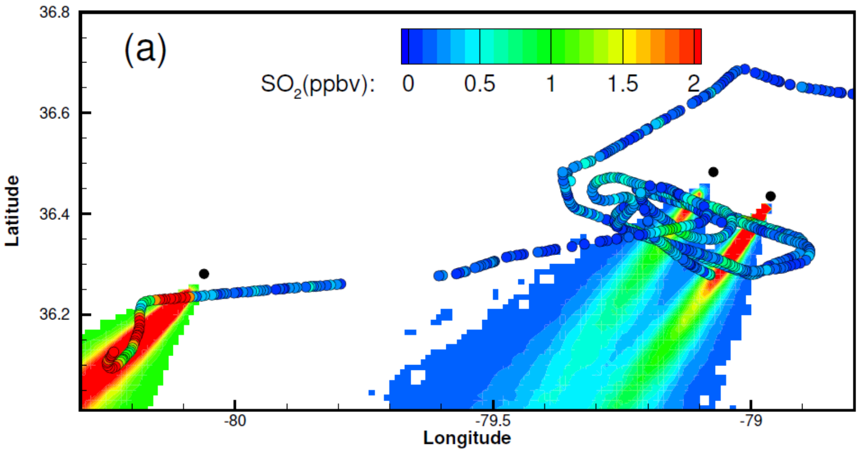
The Q_H that yields the highest correlation coefficient may not generate the minimal RMSE

SO_2 RMSE (ppbv) /	Roxboro		Belews Creek		CPI Roxboro	
Assumed heat emission (MW)	morning	afternoon	morning	afternoon	morning	afternoon
10	0.635	0.429	1.409	2.590	0.612	0.538
20	0.640	0.469	1.368	2.140	0.525	0.564
30	0.635	0.486	1.242	2.079	0.438	0.566
40	0.684	0.681	2.706	2.212	1.299	0.984
50	0.539	0.442	1.106	2.520	0.509	0.470
60	0.444	0.478	0.916	2.681	0.464	0.527
70	**0.434**	0.476	0.918	2.412	0.470	0.559
80	0.471	0.431	**0.859**	2.222	0.451	0.522
90	0.455	0.428	1.031	1.916	**0.496**	0.527
100	0.481	**0.456**	1.040	1.905	0.586	0.630
110	0.511	0.488	1.290	*2.334*	0.871	0.630
120	0.563	0.589	1.299	2.879	1.777	0.699
130	0.725	0.652	1.362	2.120	2.679	0.665
140	0.766	0.838	1.590	1.874	4.701	**0.553**
150	0.893	0.866	1.630	1.903	2.956	0.563

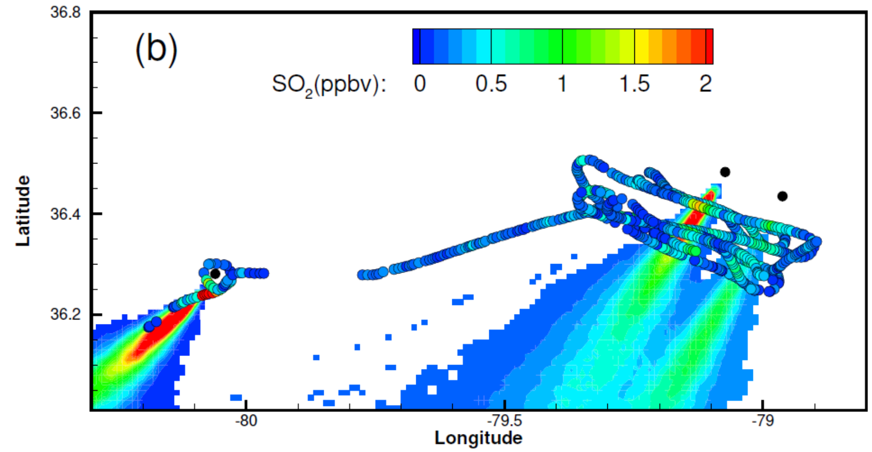
Snapshots of SO_2 predictions at 800 m AGL

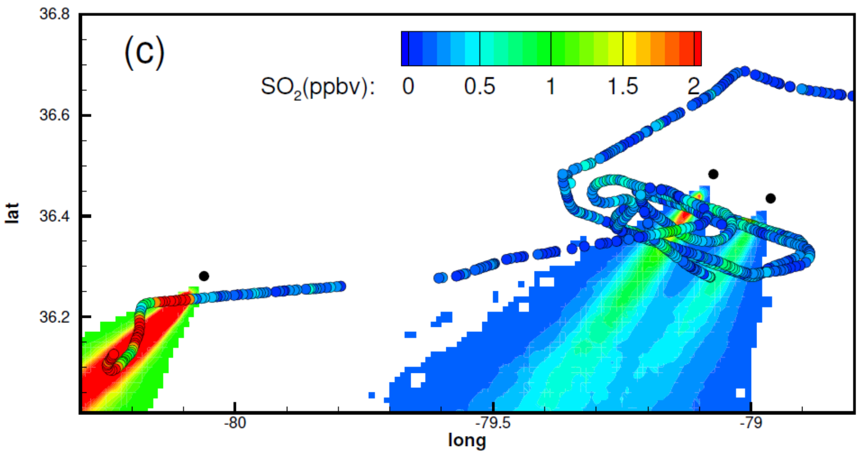
17:00Z

19:00Z

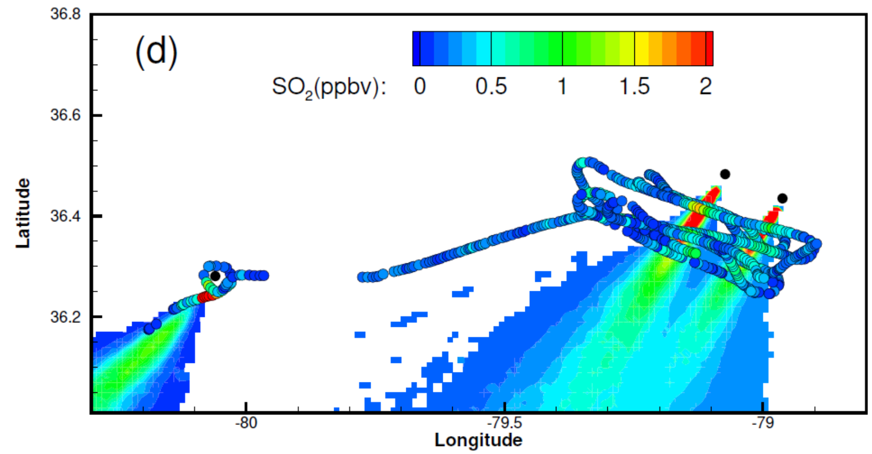


r-based





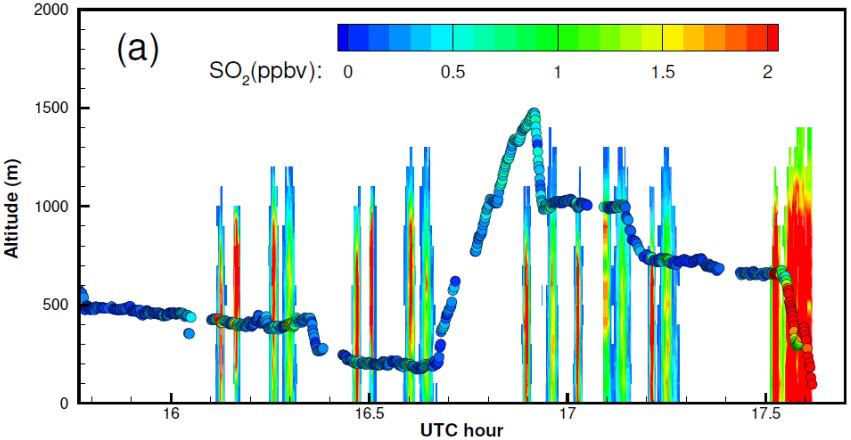
RMSE-based

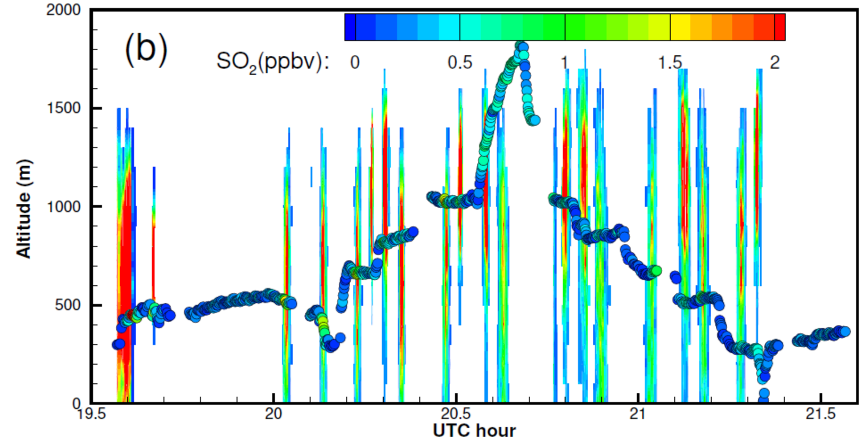


Curtain plots of SO_2 predictions along the

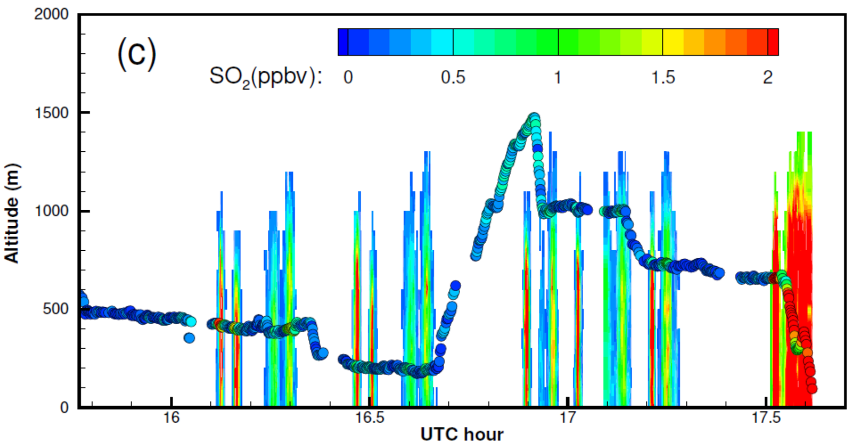
Morning flight

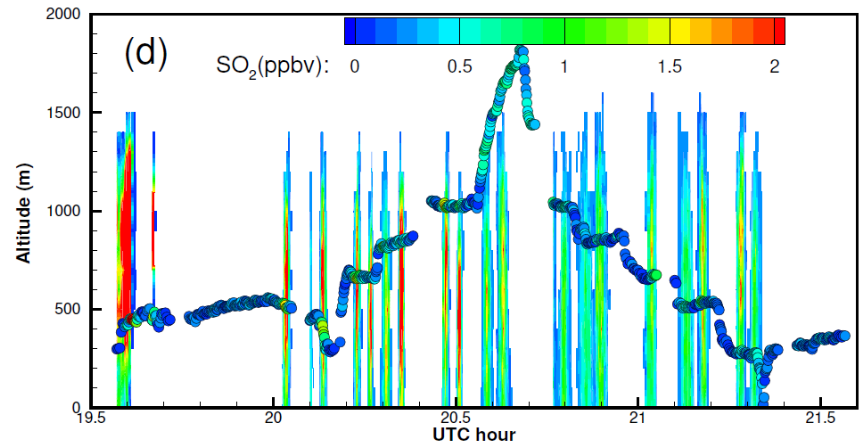
Afternoon flight





R-based



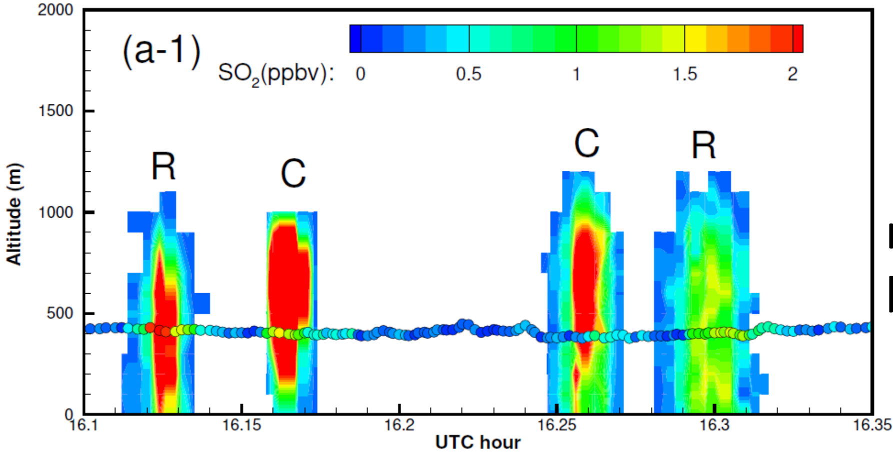


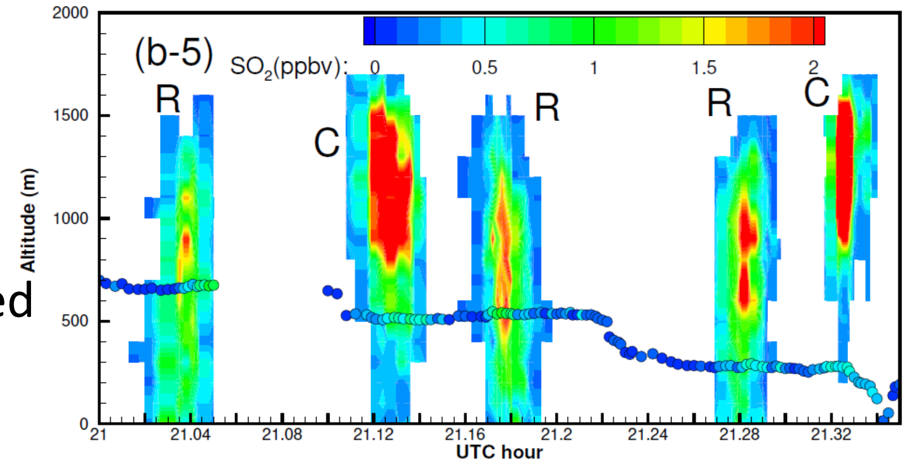
RMSE-based

Sections of curtain plots

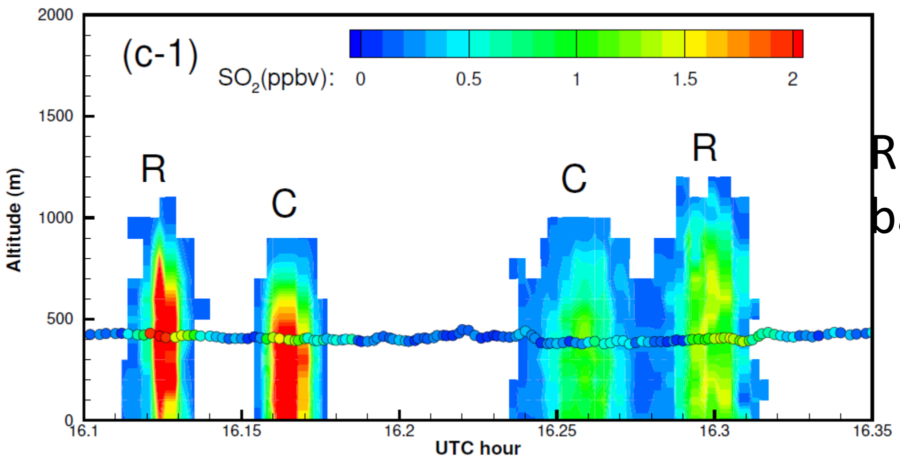
Morning flight

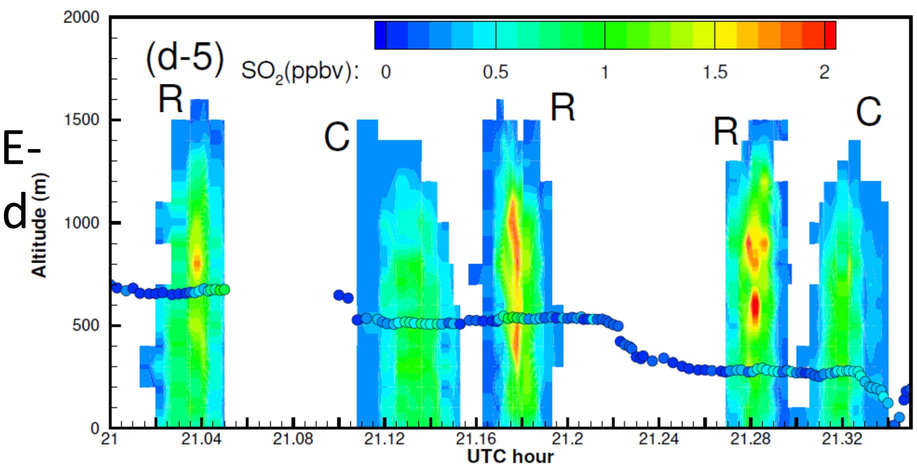
Afternoon flight





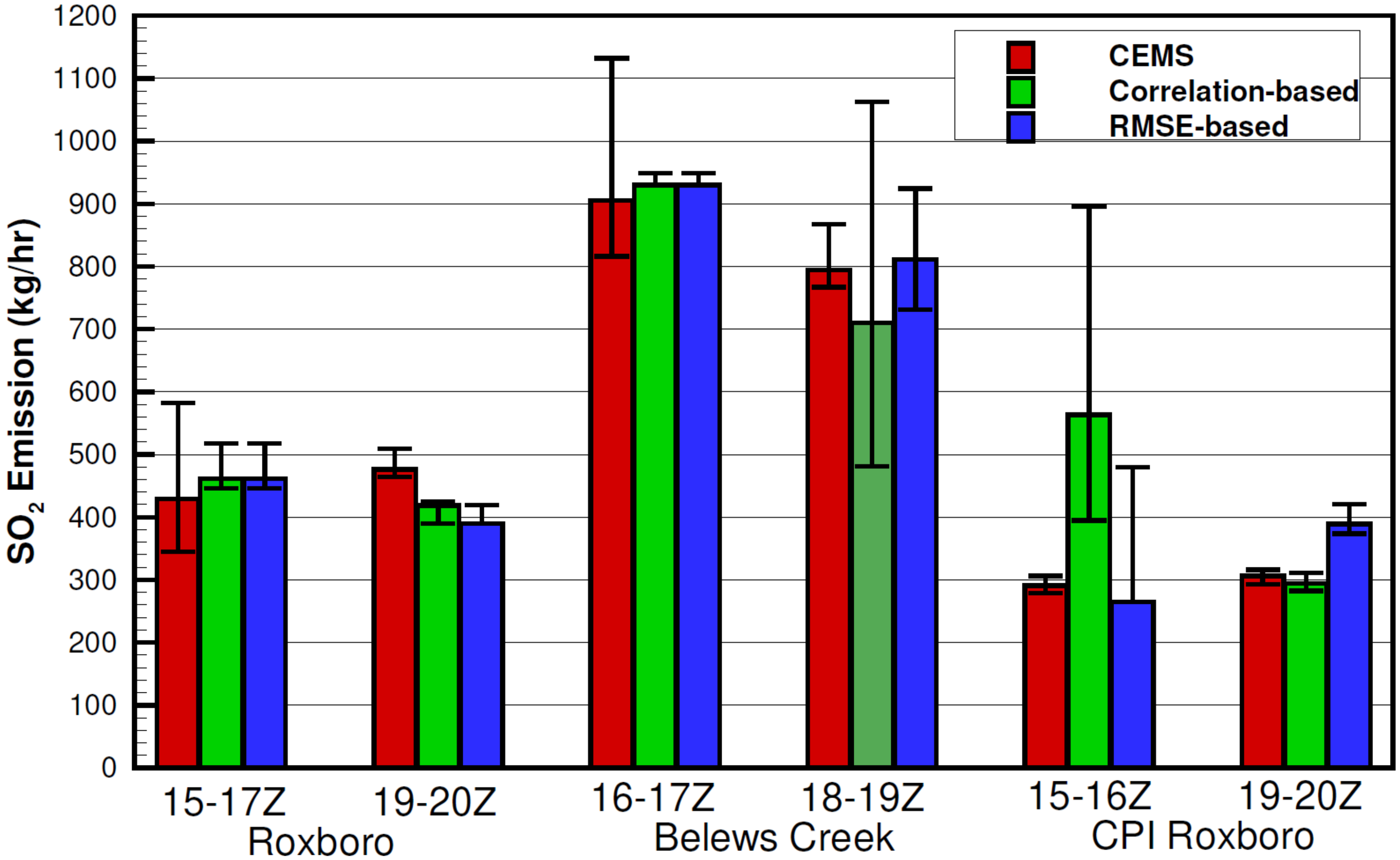
r-based





RMSE-based

Estimated and CEMS SO_2 emissions



Chai, T., Ren, X., Ngan, F., Cohen, M., and Crawford, A.: Estimation of power plant SO2 emissions using HYSPLIT dispersion model and airborne observations with plume rise ensemble runs, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2023-329, 2023.

Summary/conclusion

- Emission inversion based on the assimilation of 4D observations using HYSPLIT model, its TCM, and a cost function is developed;
- When comparing model predictions and observations, it is better to use original variables for satellite observations, but logarithmic conversion has advantage for more accurate observations;
- Emission-dependent model uncertainties and cost function normalization are introduced to avoid emission underestimation;
- Power plant SO_2 emissions are estimated using aircraft observations with exclusive plume rise ensemble runs and two criteria for the "optimal solution" based on correlation coefficients and RMSEs.
- Remaining problems :
  - Not enough observations to constrain the model, such as the SO_2 case;
  - Causes of the differences between the sensitivities calculated using forward and backward dispersion runs needs to be investigated.

Thank you!