O2.1-677

Status of radionuclide source location exercise for the automatic data fusion tool at the CTBTO

Joshua Kunkle, Hossein Hassani, Ian Hoffman, Monika Krysta, Mark Prior, Robin Schoemaker, Anne Tipka

Head, Software Integration Unit, IDC, CTBTO

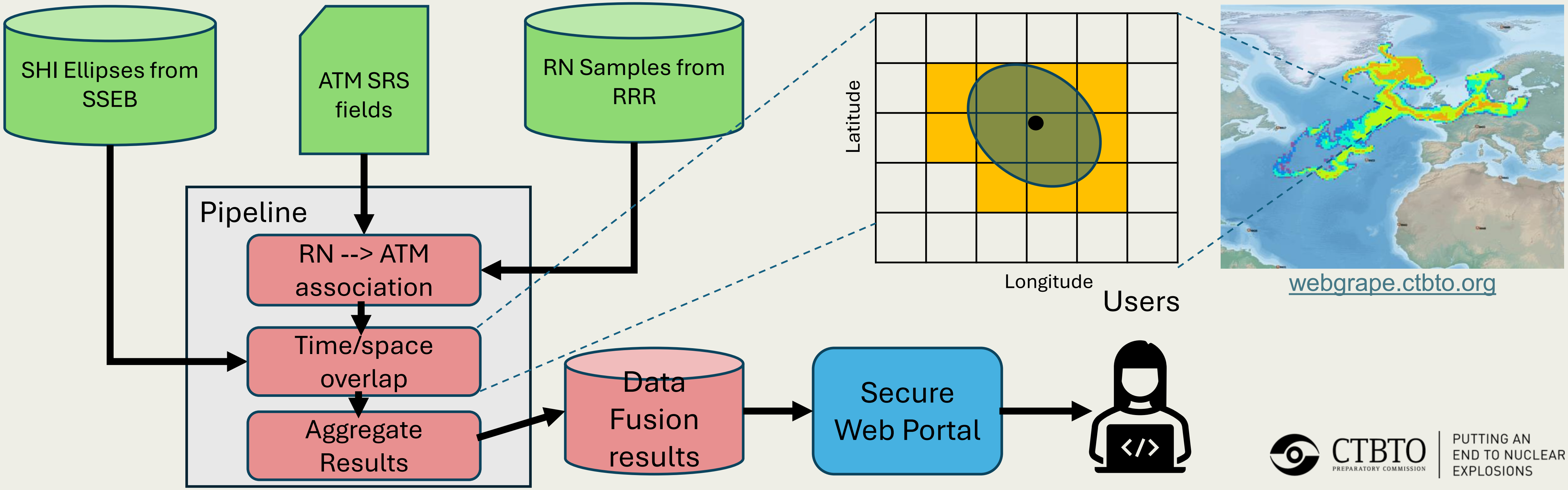
CTBTO PREPARATORY COMMISSION | PUTTING AN END TO NUCLEAR EXPLOSIONS

11 September 2025

Joshua Kunkle et. al

Data Fusion at CTBTO

The automatic data fusion tool at the CTBTO aims to produce the most comprehensive view of IMS data by combining **SHI events** with **radionuclide measurements** using **atmospheric transport modeling (ATM)**



Joshua Kunkle et. al

ATM at CTBTO

Analyze transportation of radionuclides through the atmosphere by propagating passive tracers following wind fields

From this we can extract two pieces of information:

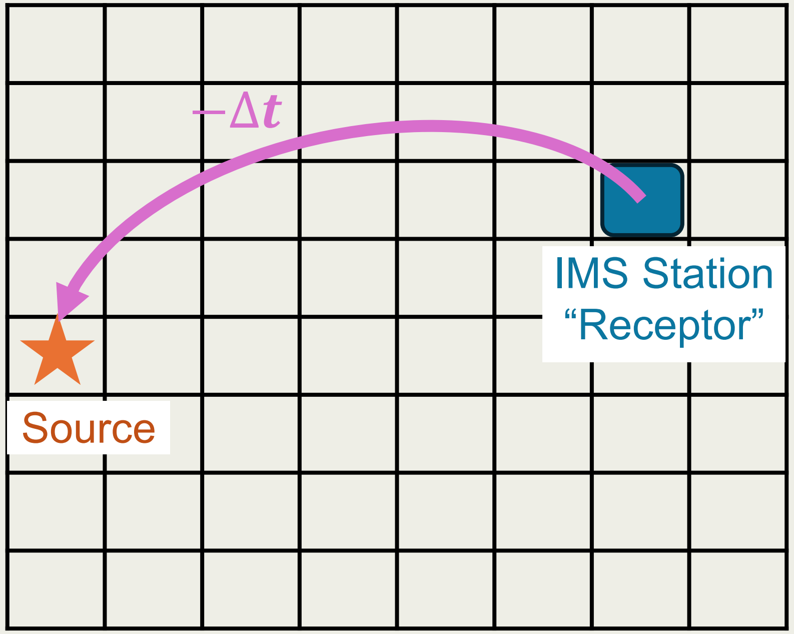
➔ Does the plume arrive to the station at a given time?

➔ What fraction of released particles arrives at the station?

  ➔ Dilution Factor, D [1/m^3]

$$A_m = RDe^{-\lambda t}$$

- A_m ⟶ Measured activity concentration [Bq/m^3]
- R ⟶ Release [Bq]
- λ ⟶ Half-life of isotope
- t ⟶ Time between release and measurement



<u>Operational setup:</u>

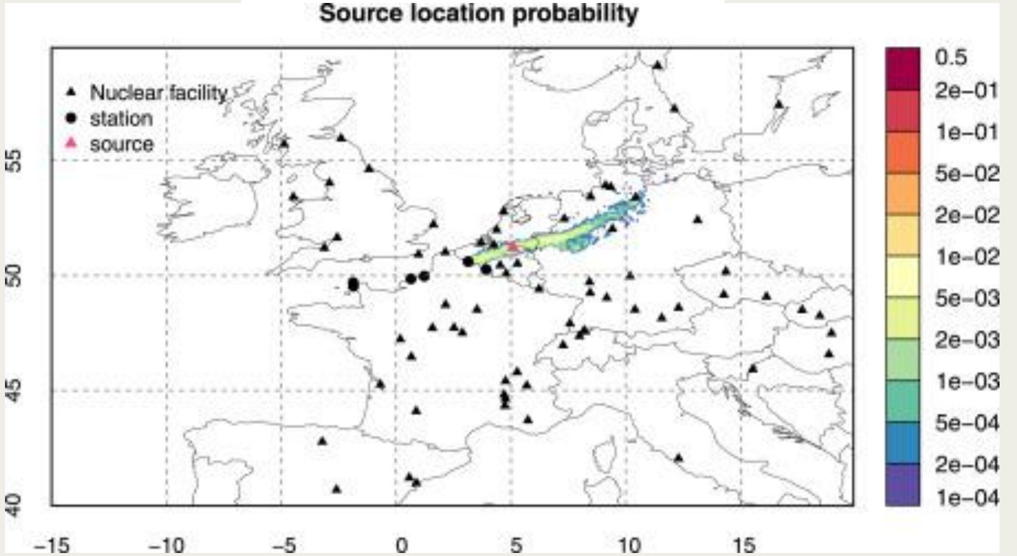
- Propagate backward from station at sampling time. For each RN sample collection produce a Source Receptor Sensitivity (**SRS**)
- SRS results produced on a grid of 0.5 (lat) x 0.5 (lon), in 1 hour intervals up to of 336 hours (14 days)
- Separate simulations using ECMWF and NCEP wind fields

Joshua Kunkle et. al

Automatic RN source location as improvement to Data Fusion

- **Current solution produces too many matches, making review difficult**
- RN source location would focus data fusion matches on areas of likely radionuclide release
- Radionuclide source location has been extensively studied with multiple methods available
  - PSR, Cost function, Bayesian inference, etc
- These options show promise as an expert tool, but they do not immediately lend themselves to automatic processing
- **Investigate a new "global reconstruction" method aimed towards automatic processing**

Example of published studies

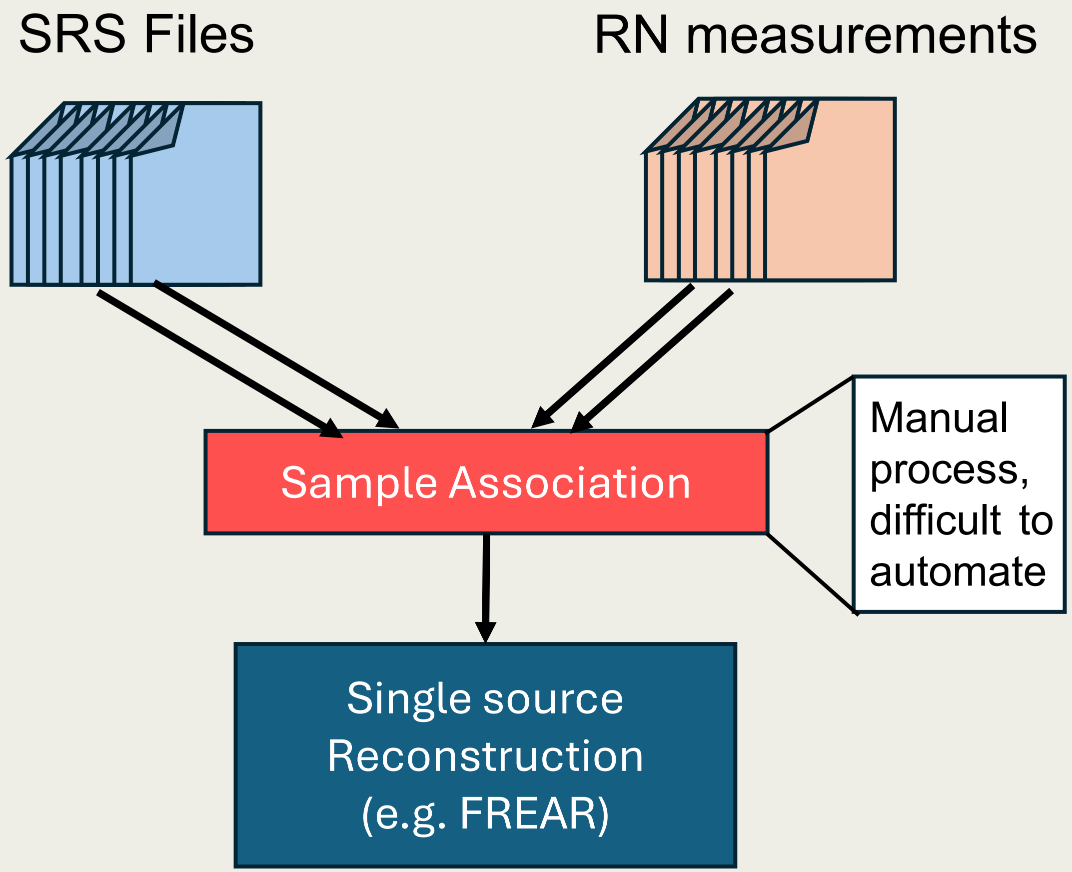


P. De Meutter, I. Hoffman, "Bayesian source reconstruction of an anomalous Selenium-75 release at a nuclear research institute"

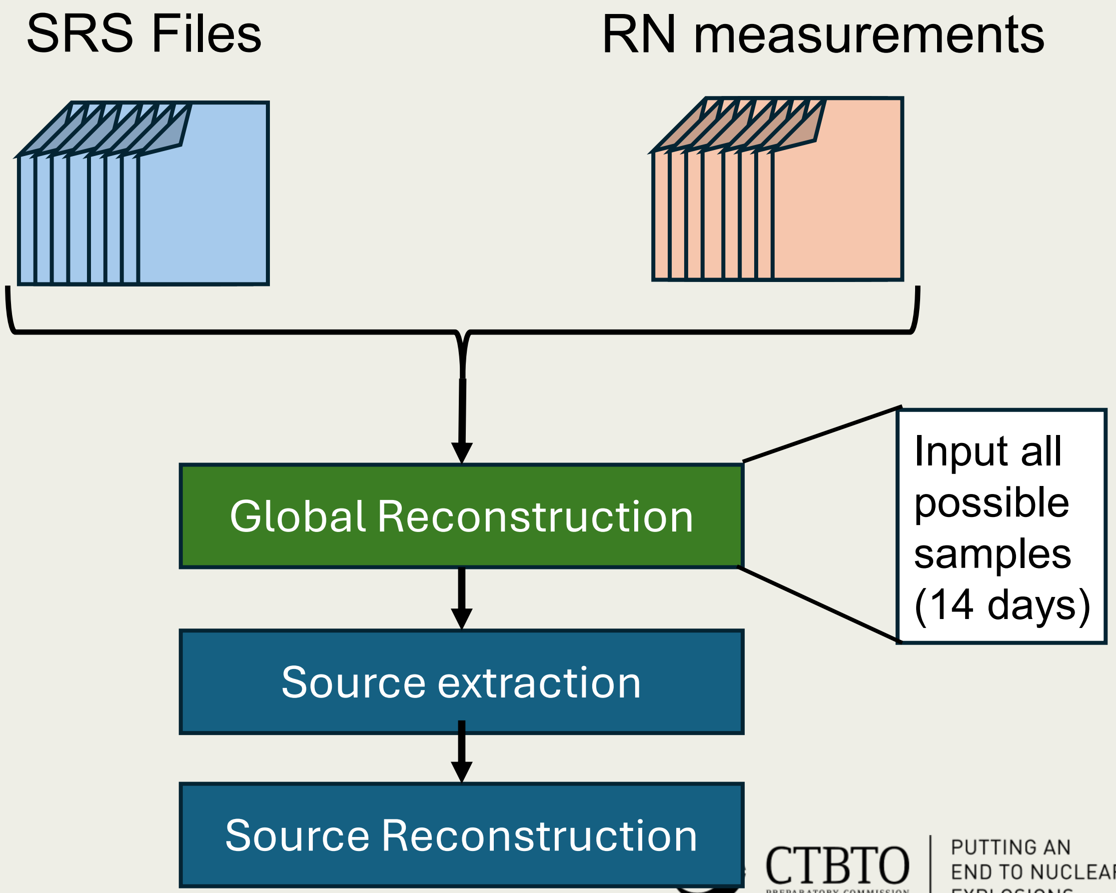
Joshua Kunkle et. al

Methodologies for RN source location

Local Reconstruction



Global Reconstruction



Joshua Kunkle et. al

Log Likelihood Ratio (LLR)

Goal : To develop a **score** that represents the probability of a radionuclide release for a given **time** and **location**

Possible Solution : Utilizing a maximum log likelihood ratio (LLR)

log-likelihood ratio (LLR) score to compare:

- **Null (H_0):** no release (likelihood maximized inside the constrained “no-release” parameter space).
- **Alternative (H_1):** release allowed (likelihood maximized over the full parameter space).

$$\mathrm{LLR} = -2\,\ln\left(\frac{L(\hat{\theta}_0; x)}{L(\hat{\theta}; x)}\right) = 2\left[\ln L(\hat{\theta}; x) - \ln L(\hat{\theta}_0; x)\right].$$

- If **LLR is large**, the data fit the release model much better than the no-release model → evidence *for* a release.
- If **LLR is small**, the no-release model explains the data about as well as the release model.
- **Log scale is good:** sums logs instead of multiplying small probabilities; it’s **numerically stable**
  - **Easily interpretable:** as it is additive across observations

Joshua Kunkle et. al

Probability Model

What is the **likelihood** of a release at a particular **time** and **location** given the measurements at IMS stations based on historical data?

Model : joint probability of compatibility with release over,

- **detections** that result from the release
- **non-detections** where release detection is expected

$$\mathcal{L}(R|\boldsymbol{I}, \boldsymbol{lat}, \boldsymbol{lon}, \boldsymbol{t}) = \prod_{det} P_{det} \prod_{!det} (1 - P_{det})$$

ATM simulations provide a **probability of arrival**, and **compatibility with RN measurement**

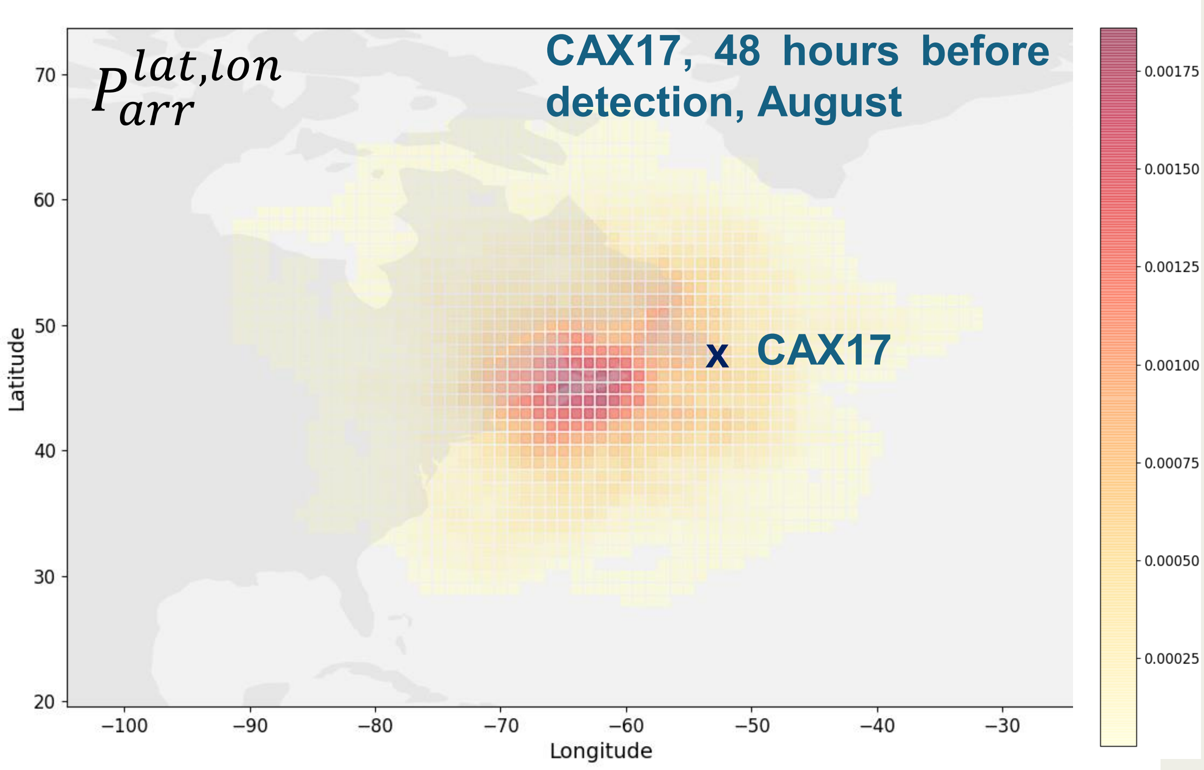
$$P_{det}(R\,|I, lat, lon, t) = P_{arr}^{lat,lon} \int_{M-\varepsilon}^{M+\varepsilon} RDe^{-\lambda(I)t}$$

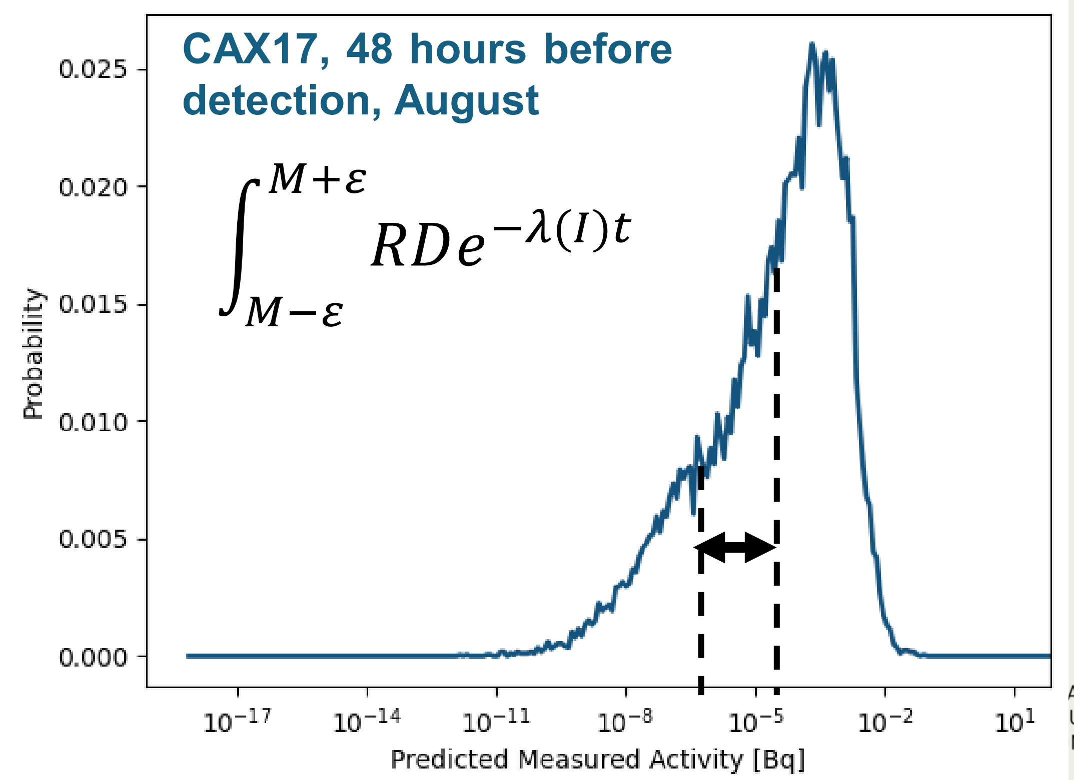
Joshua Kunkle et. al

Building the model

Generate hypothetical releases over 10 years of ATM simulations (ECMWF) at all spatial points

- Record fraction of detections vs location and time ($P_{arr}^{lat,lon}$)
- Record dilution factor (D)
- Segment model into stations, and month in year (account for yearly weather patterns)

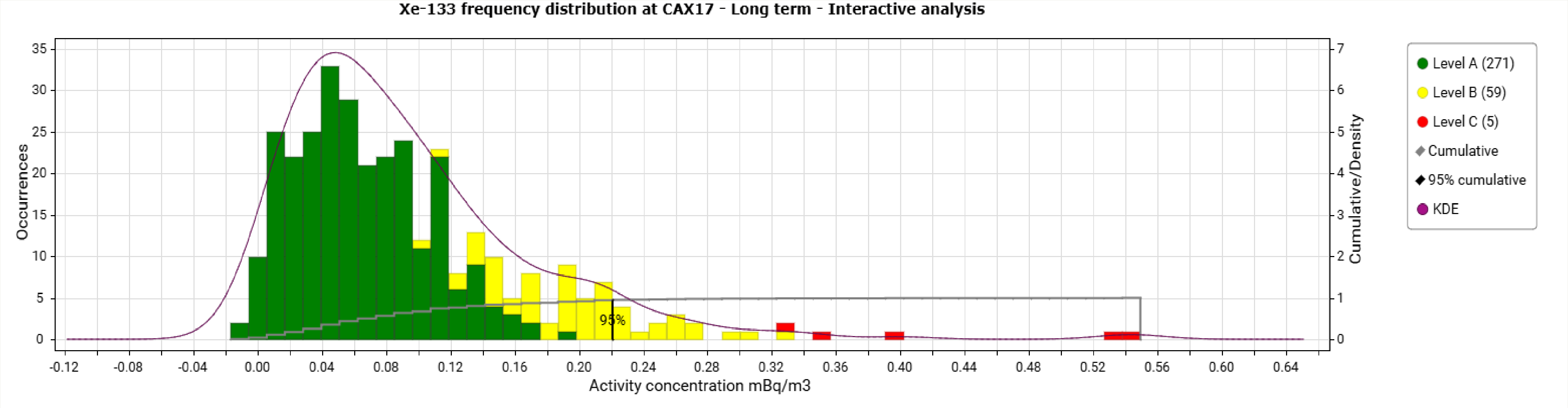




Joshua Kunkle et. al

Applying the model

- Focus on **Radioxenon network** for this study
- Generate synthetic sources using ATM simulations on NCEP data
  - **Detections – Level B and C samples** : synthetic activity concentration exceeds critical limit
  - **Non-Detections – Level A samples**

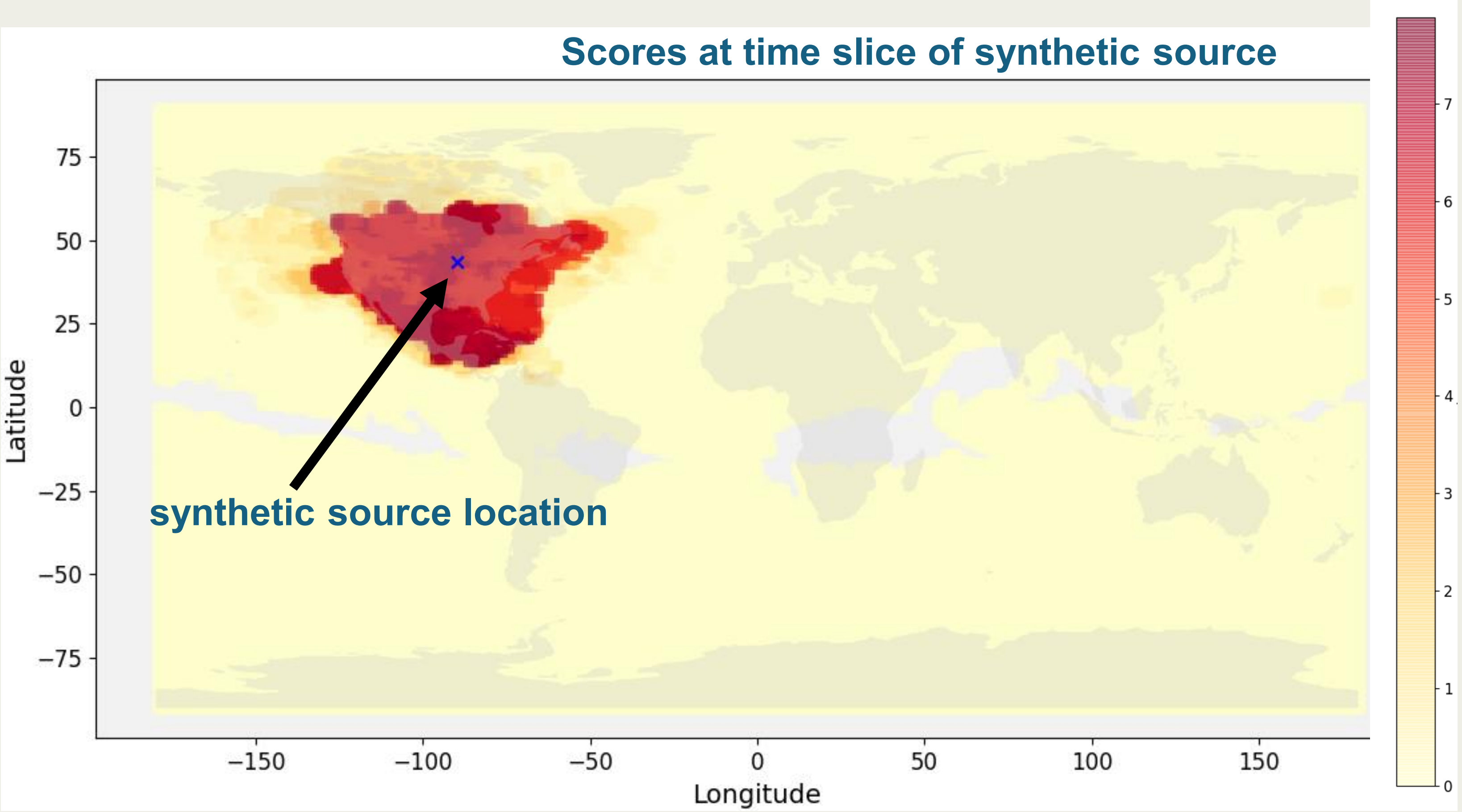


- Evaluate model in time slices, inputting all possible samples up to 14 days in the future
- Maximize likelihoods over possible released activities (1e10 – 1e18 Bq)
- Null hypothesis : Release activity <= 1e12 (hyper-parameter)

Joshua Kunkle et. al

Result on test scenario

- In synthetic test cases, the model shows reasonable agreement
- **Regions of Interest** (ROIs) can be selected by applying a **threshold on the score** **(hyper-parameter)**
- It is possible therefore to identify multiple ROIs (performance to be tested)



Investigating the scores

For each ROI, break down the score into individual contributions for each sample (average over ROI)

Each sample contributes to the total score,

$$\text{score(sample)} = \begin{cases} \ln \hat{P}^0_{det} - \ln \hat{P}_{det} & \textit{detected} \\ \ln(1-\hat{P}^0_{det}) - \ln\left(1-\hat{P}_{det}\right) & \textit{not detected} \end{cases}$$

Selection of samples above a specific score achieves **sample association**

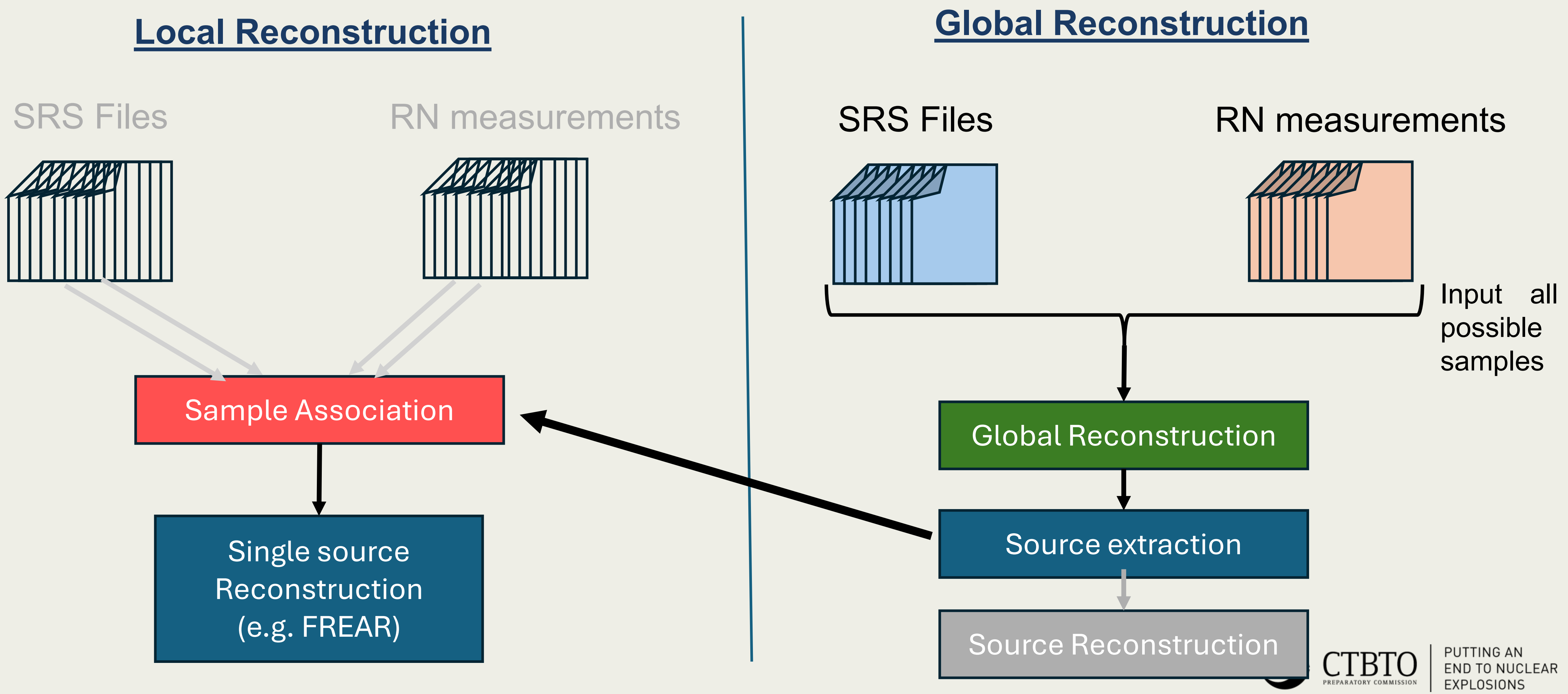
Sample scores within ROI, sorted

sample ID (synthetic)	station	detection?	Mean score
1	USX75	YES	5.87
2	USX75	YES	2.09
3	USX75	YES	1.75
4	USX75	YES	1.52
5	USX75	YES	1.29
8	USX75	YES	0.31
339	USX74	NO	0.04
10	USX75	YES	0.04
12	USX75	YES	0.02

Sample Association

Joshua Kunkle et. al

Bringing methods together

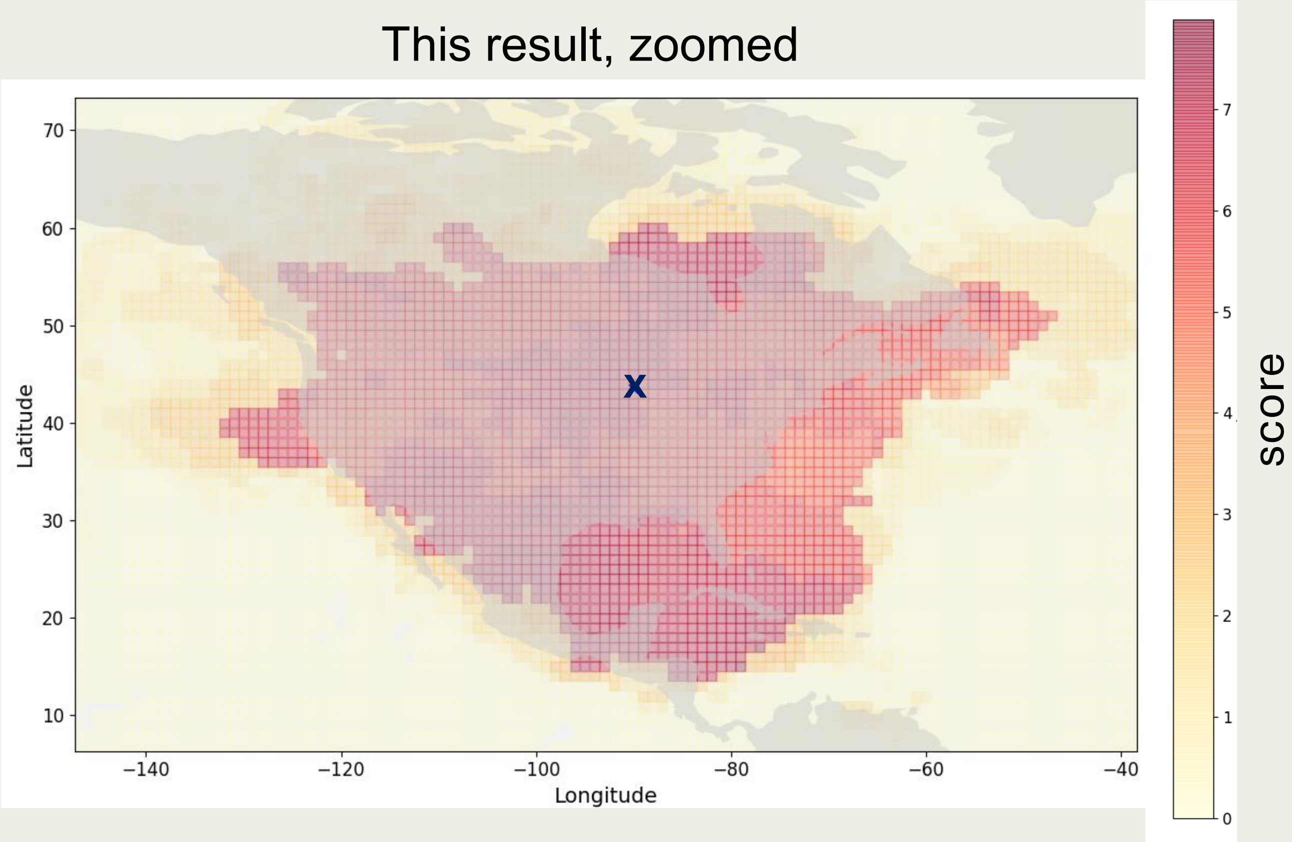


Joshua Kunkle et. al

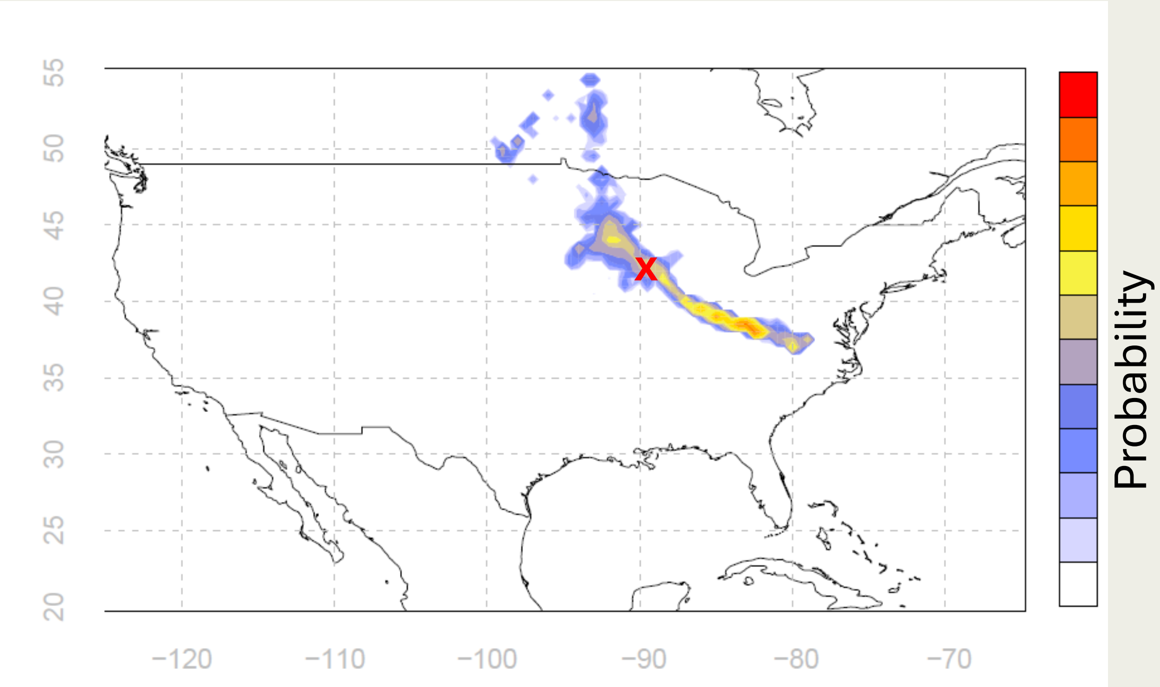
Comparison of results with Bayesian inference

Bayesian inference based on the ATM files specific to the scenario can produce more precise results than the result using historical data

This result, zoomed



Bayesian Inference using FREAR



x = synthetic source location

Joshua Kunkle et. al

Conclusions

- An effort to improve the automatic data fusion tool at the CTBTO is ongoing, **radionuclide source reconstruction** identified as key next step
- A global reconstruction model based on historical ATM data was investigated
- Model results are a sum over probability-based scores, **ensuring explainability**
- Model shows promise in **accomplishing sample association, removing a key blocker to automatizing radionuclide source location for data fusion**
- Next step is to generate large-scale synthetic data to evaluate reconstruction methods and tune hyper parameters