

Jet Propulsion Laboratory
California Institute of Technology

CMFTagger Flightline + Plume Candidate QC: Quickstart + Labeling Guidelines

Brian Bue (bbue@jpl.nasa.gov)

Last updated: April 07, 2022

Starting + Accessing the CMFTagger on the G3000

- Connect to G3000 (g3k.cminc.io) and start an instance of the cmftagger server

```
bbue@laptop:~$ ssh bbue@g3k.cminc.io
bbue@dl:~$ /shared/anaconda3/bin/python /shared/srcfinder/cmftagger/cmftagger.py
```

- Note your **port** number (ports assigned by username) in the cmftagger.py output

```
bbue@dl:~$ /shared/anaconda3/bin/python /shared/srcfinder/cmftagger/cmftagger.py
SRCFINDER_ROOT: "/shared/srcfinder/"
335 total products listed in AVng_2019_Permian_ch4mf_plumes_20210122_flightlines_simplified
56 products assigned to user bbue
Searching /localstore/ang/y19/cmf/ch4/ort
Found 334 total ('cmf', 'ch4mf') products
bbue: 0 of 56 assigned flightlines complete
ang20190921t165138_ch4mf_v2x1_img @ native resolution load time elapsed = 0.01077711868286133s
Loaded ang20190921t165138_ch4mf_v2x1_img image shape (2106, 17468) with 4 candidates (0 plumes, 4 non-plumes)
Launching server at http://localhost:5006
```

- Set up local ssh tunnel to map cmftagger **port** from G3000 to your local machine

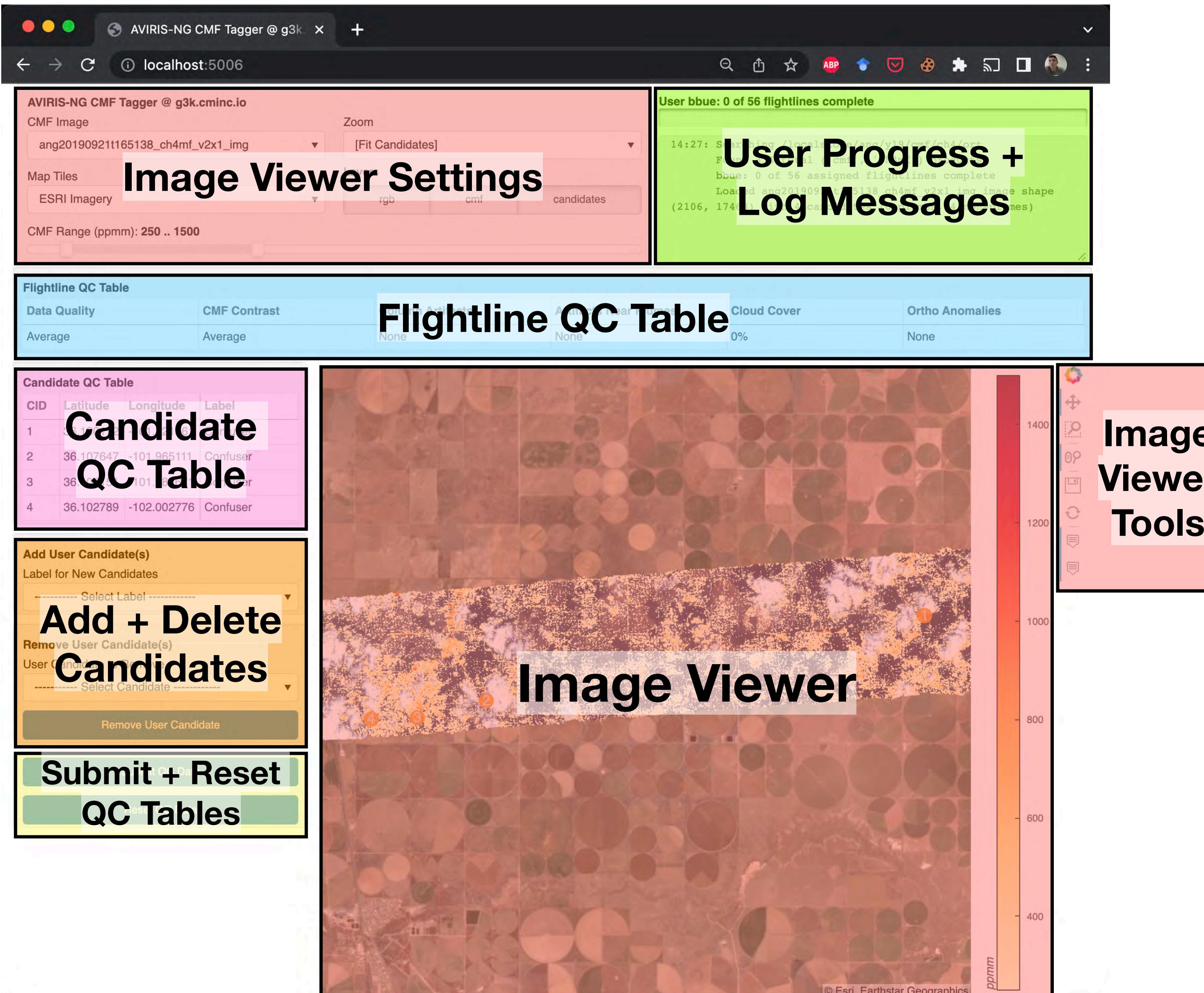
```
bbue@laptop:~$ ssh -v -N -L 5006:localhost:5006 bbue@g3k.cminc.io
...
Authenticated to g3k.cminc.io ([66.215.80.101]:22).
debug1: Local connections to LOCALHOST:5006 forwarded to remote address localhost:5006
...
```

- Open <http://localhost:port> in a web browser; cmftagger app should appear in a few seconds.
- When finished, [control-C] to stop cmftagger.py instance on G3000 + close local ssh tunnel

Nomenclature

- **ROI:** a group of spatially adjacent CMF pixels with $\text{enhancements} > \text{minppmm}$ that are visibly distinct from the CMF background enhancement.
- **CID:** alphanumeric Candidate ID value of a labeled ROI
(e.g., Candidate ID=ang202000830t203348-B → CID=B)
- **LID:** Flightline ID of an image product or Candidate ID
(e.g., image=ang202000830t203348_ch4mf_v2n1_img → LID=ang202000830t203348)
- **Column artifacts/Systematics:** linear artifacts with artificially high enhancements in downtrack flight direction resulting from statistical matched filter errors, typically caused by outlier radiances that sometimes occur in pushbroom sensor image columns.
- **False enhancement:** CMF ROI not generated by an anthropogenic CH₄ source.
- **Background enhancement:** average enhancement of positive, non-plume CMF image pixels.

CMFTagger Interface: Overview



- **Image Viewer:** displays selected CMF images + candidate locations + WMTS map tiles
- **Flightline QC Table:** user-updated summary of flightline
- **Candidate QC Table:** user-updated labels for individual candidates
- **Add + Delete Candidates:** adds new candidates with specified label + removes candidates by id
- **Submit + Reset QC Tables:** submits user updates to flightline and candidate QC tables as complete; resets QC tables to original values in campaign plume list.

CMFTagger Interface: Image Viewer + Viewer Settings/Tools

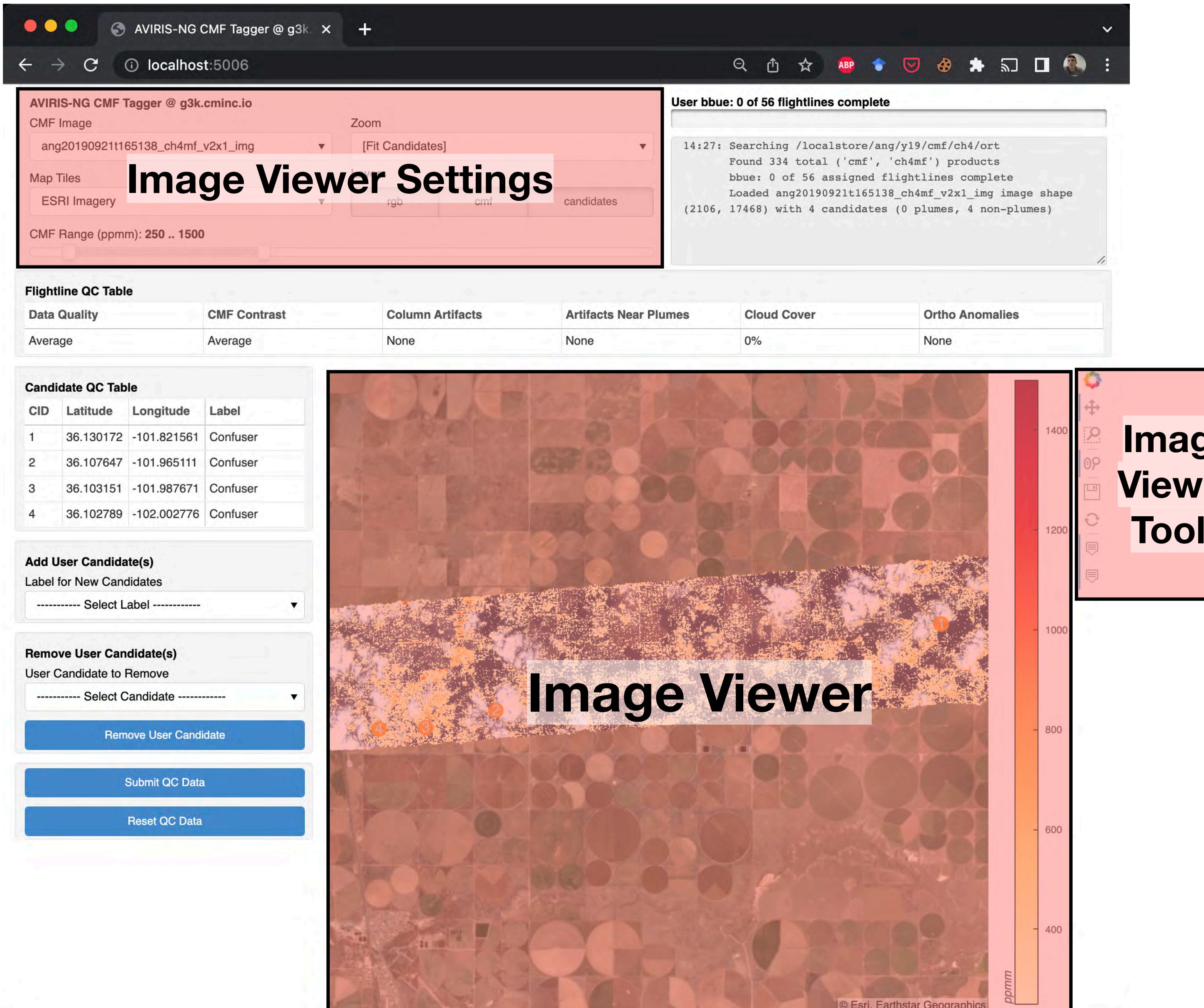
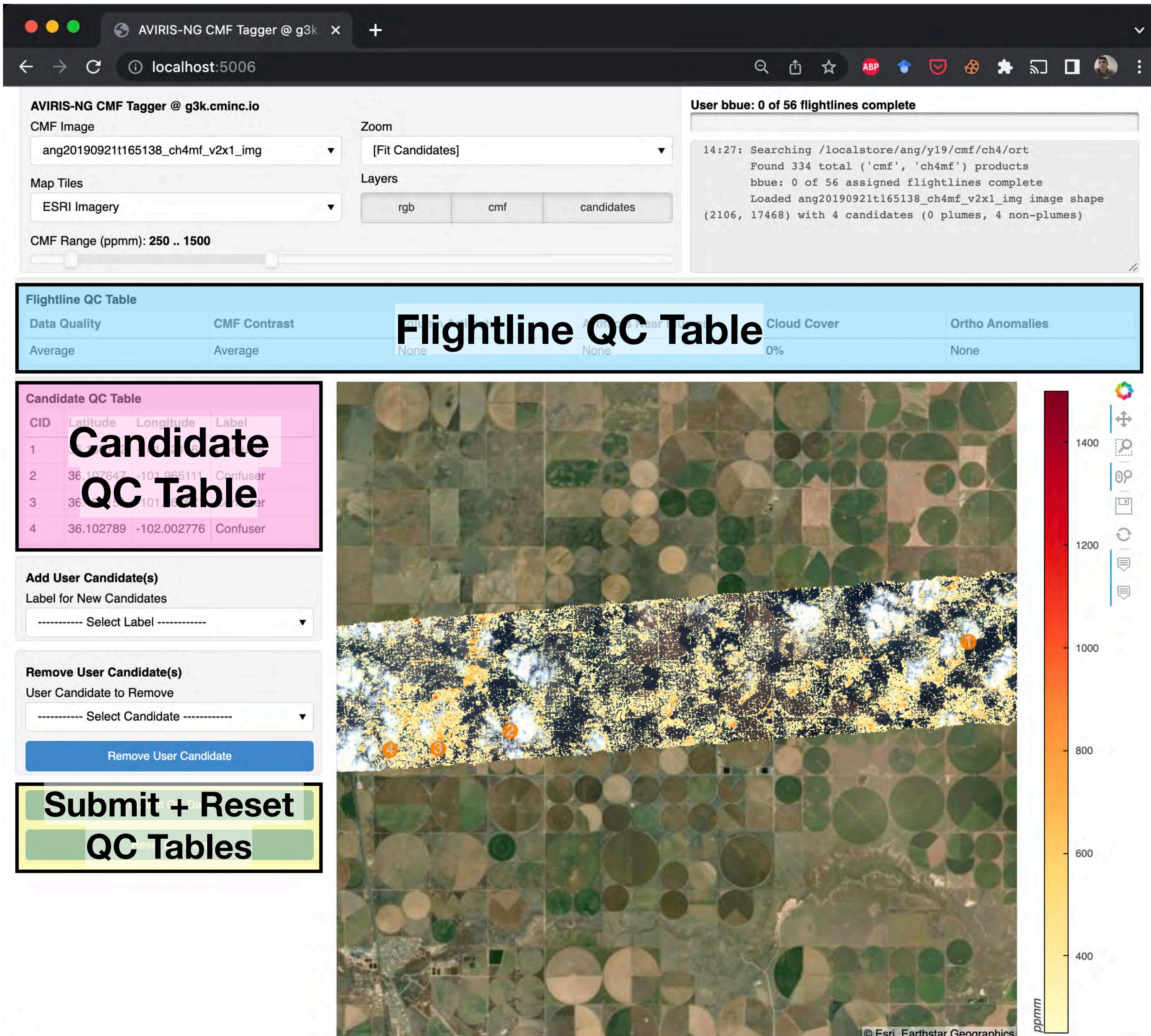


Image Viewer

- Viewer Settings
 - select CMF to view
 - change background map tiles
 - clip CMF range
 - set image/candidate zoom level
 - toggle visibility of CMF + RGB images + candidate locations
- Viewer Tools (click to toggle)
 - Defaults
 - Pan with mouse drag
 - Zoom with mouse wheel
 - Mouse Hover Tools
 - CMF value @ mouse coords
 - Candidate metadata
 - Save Viewer Image

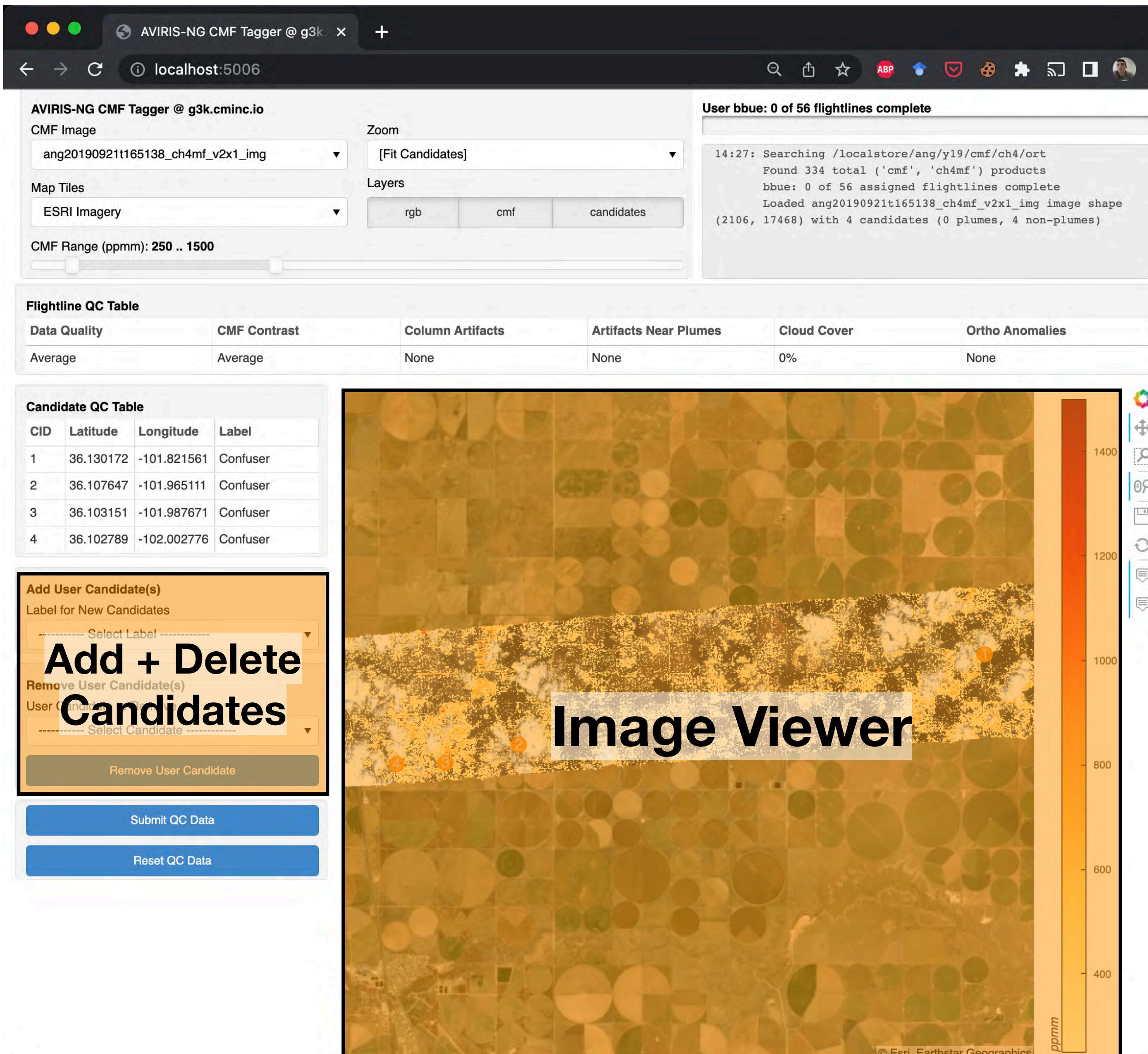
CMFTagger Interface: QC Tables



QC Tables

- Submit
 - Submits QC tables as complete for current flightline
 - Moves current flightline to “done” list from “todo” list
 - Selects next flightline in queue
- Reset
 - Moves current flightline to “todo” list if in “done” list
 - Resets Flightline QC table to default values
 - Reverts Candidate QC table to values in campaign plume list

CMFTagger Interface: Add + Delete Candidates



Add + Delete Candidates

- Add User Candidate
 - Select label for new candidate(s) from menu
 - Double click location(s) for new candidate(s) with specified label in image viewer

- Delete User Candidate
 - Select candidate id from menu
 - Press remove user candidate button

- Does **not** permit removal of candidates from campaign plume list
 - users required to qc all assigned candidates in campaign plume list

Flightline + Candidate QC Instructions

1. Zoom in to perform an initial scan of the flightline

- Assess overall flightline data quality and noise characteristics
- Note locations of unlabeled ROIs that might be plumes
- Identify which types of false enhancements are present and where they occur

2. Update entries in the Flightline QC table based on flightline scan

3. Inspect each candidate and update labels in Candidate QC Table

- Select a row in the Candidate QC Table to zoom to that candidate
- Assign “plume” labels conservatively — if you aren’t sure, label the candidate “ambiguous”

4. Add new candidates to any unlabeled plume ROIs

- This step is essential — all obvious plume ROIs must be labeled to ensure they will not be accidentally included as (randomly selected) background samples to train the CNN

5. [optional] Add new candidates for unlabeled ROIs of ambiguous enhancements

- Useful but not essential — flagging ROIs ambiguous will exclude them from training/validation and may improve CNN accuracy. Don’t bother adding candidates for small or weak ROIs.

6. Press “Submit QC Data” button when you’ve finished updating the QC tables.

Flightline QC Table: Definitions

Column	Details	Default
Data Quality	Flightline average, high or low data quality. <u>High quality</u> flightlines have no observable quality issues, and contain only visibly distinct plumes and unambiguous false enhancements.	Average
CMF Contrast	Foreground (plumes) + background enhancements difficult to separate due to either high or low background enhancement.	Average
Column Artifacts	Downtrack linear artifacts visible in CMF. Severe artifacts cover large regions of the flightline and are visibly obvious at low zoom levels.	None
Artifacts Near Plumes	Artifacts visible nearby one or more labeled plume candidates.	None
Cloud Cover	Percentage of visible clouds covering RGB quicklook.	0%
Ortho Anomalies	Flightline contains ortho issues that may produce false enhancements due to spatially varying pixel footprints.	None

Candidate QC Table Label Definitions

Plume Candidates (positive class):

- **Plume:** candidate CMF enhancement clearly distinguishable as a (segment of a) methane plume emitted from an anthropogenic “point” source, candidate enhancement visibly separable from any nearby false enhancements.
- **SuperPlume:** candidate CMF enhancement clearly distinguishable as an anthropogenic methane plume but emissions generated by multiple point sources or an area source (e.g., manure ponds, landfills).

False Enhancement Candidates (negative class):

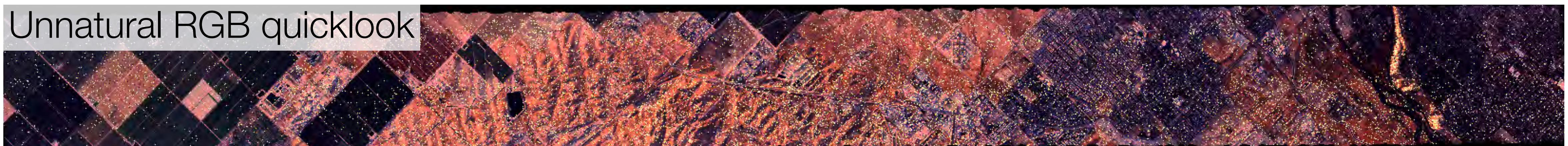
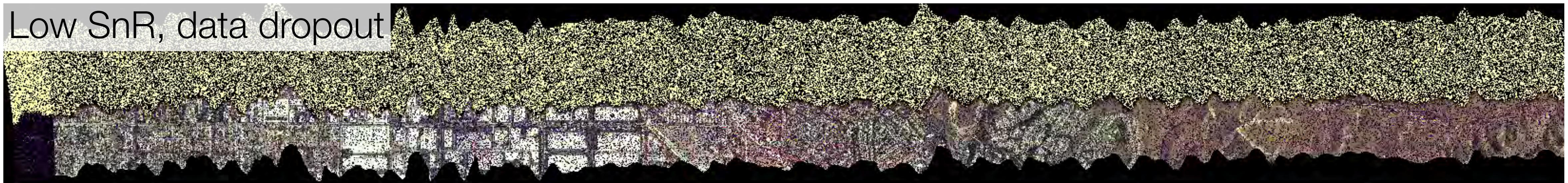
- **Artifact:** candidate within extent of / adjacent to columnwise systematic(s) or other likely imaging artifacts (e.g., high enhancements in “tails” of bad ortho flightlines, “unnatural” looking RGB image values)
- **Cloud:** candidate within extent of / adjacent to clouds/cloud shadows visible in RGB image
- **False enhancement:** candidate ROI corresponds to a non-plume enhancement of unspecified type (e.g., surface spectroscopy, solar panels, flare stacks, confuser surface materials).

Uninformative Candidates (class uncertain/undefined; redundant or conflicting candidates):

- **Ambiguous:** cannot confidently determine if candidate is a plume based on CMF enhancement.
- **Duplicate:** candidate is a duplicate of one or more neighboring candidates representing the same ROI.
- **Background:** all enhancements near candidate location near or below minppmm=250ppmm.

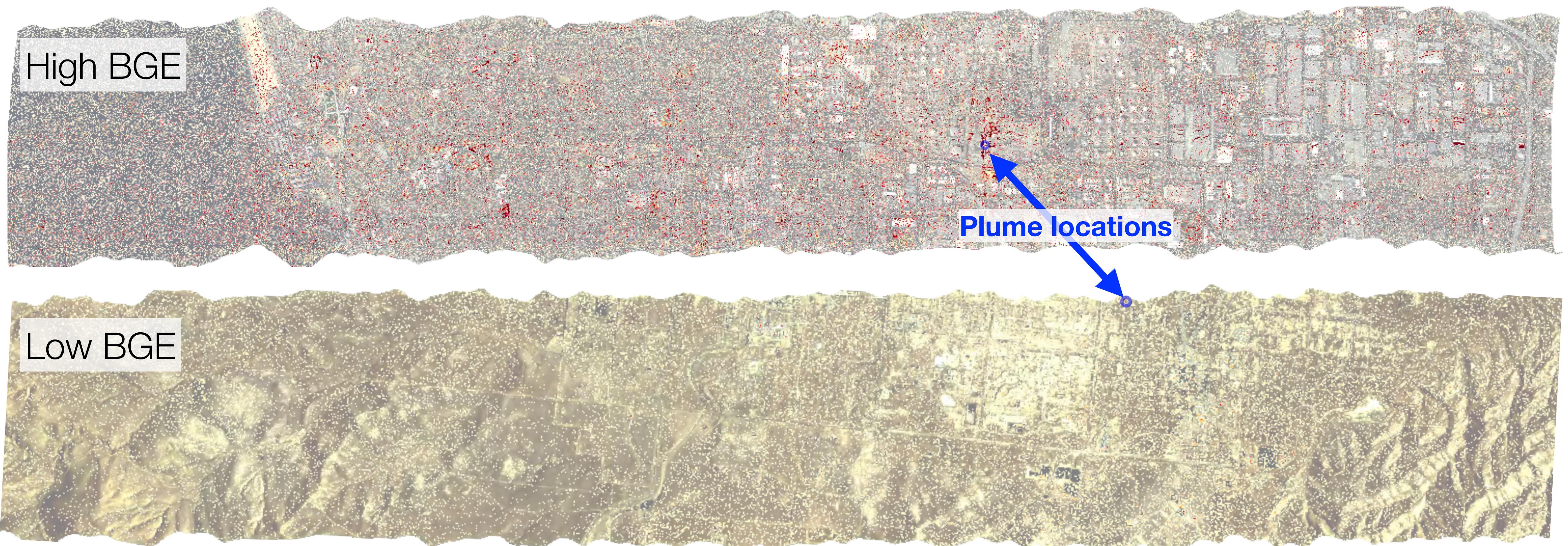
Flightline QC Issues: Data Quality

- Applying CNN to bad data will either produce no detections (wasting compute resources) or bad detections (reducing accuracy)
- Example CMF data quality issues include: low SnR (e.g., cannot identify surface features in CMF or RGB quicklook) or “unnatural” looking RGB quicklooks (typically calibration issues), overexposure/saturation, or data dropout.
- Note: significant data quality issues are rare in AVIRIS-NG flightlines captured after 2017.



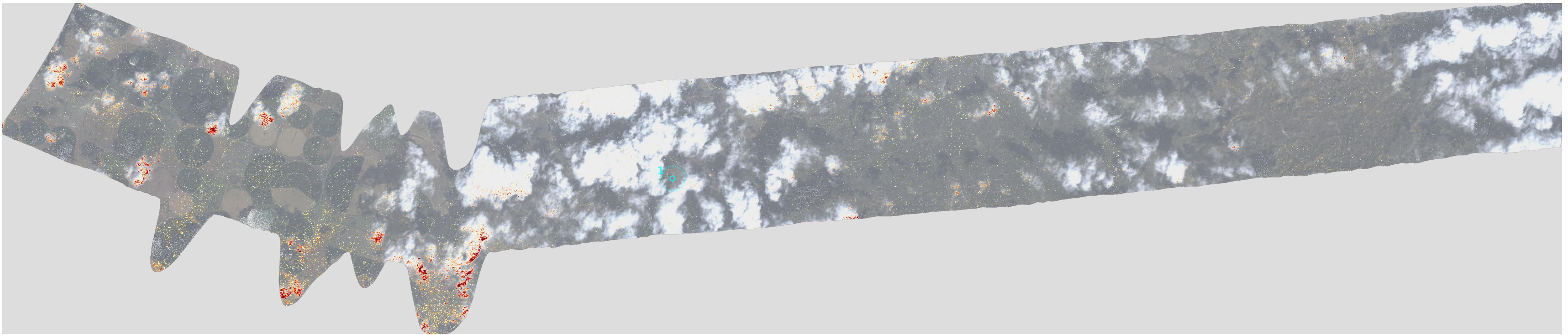
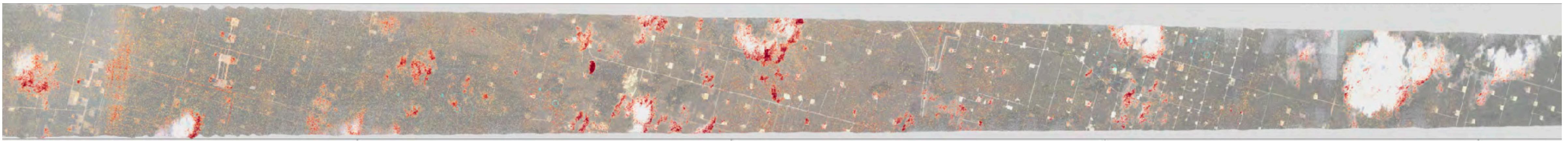
Flightline QC Issues: CMF Contrast

- CMF Contrast = difference (ppmm) between foreground enhancements (plumes) and background (not plumes) enhancement. Hard to detect plumes in low contrast CMFs.
- Low BGE = FG + BG enhancements low. High BGE = FG + BG enhancements high.
- Note: CMF contrast issues rare in AVIRIS-NG flightlines captured after 2017.



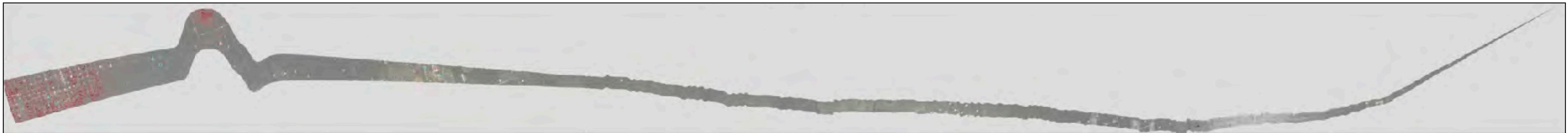
Flightline QC Issues: Clouds

- Clouds + cloud shadows produce false CMF enhancements that can be indistinguishable from plumes in CMF images.
- Candidates overlapping or adjacent to clouds / cloud shadows should be relabeled “Cloud” accordingly.



Flightline QC Issues: Ortho (Viewing Geometry) Anomalies

- Ortho issues caused by viewing geometry changes over flightline can produce false enhancements and inaccurate IME/flux estimates due to varying pixel footprints
- False enhancements typically occur at head or tail end of flightline.
- Candidates overlapping or adjacent to false enhancements in bad ortho flightlines should be relabeled “**Artifact**”.



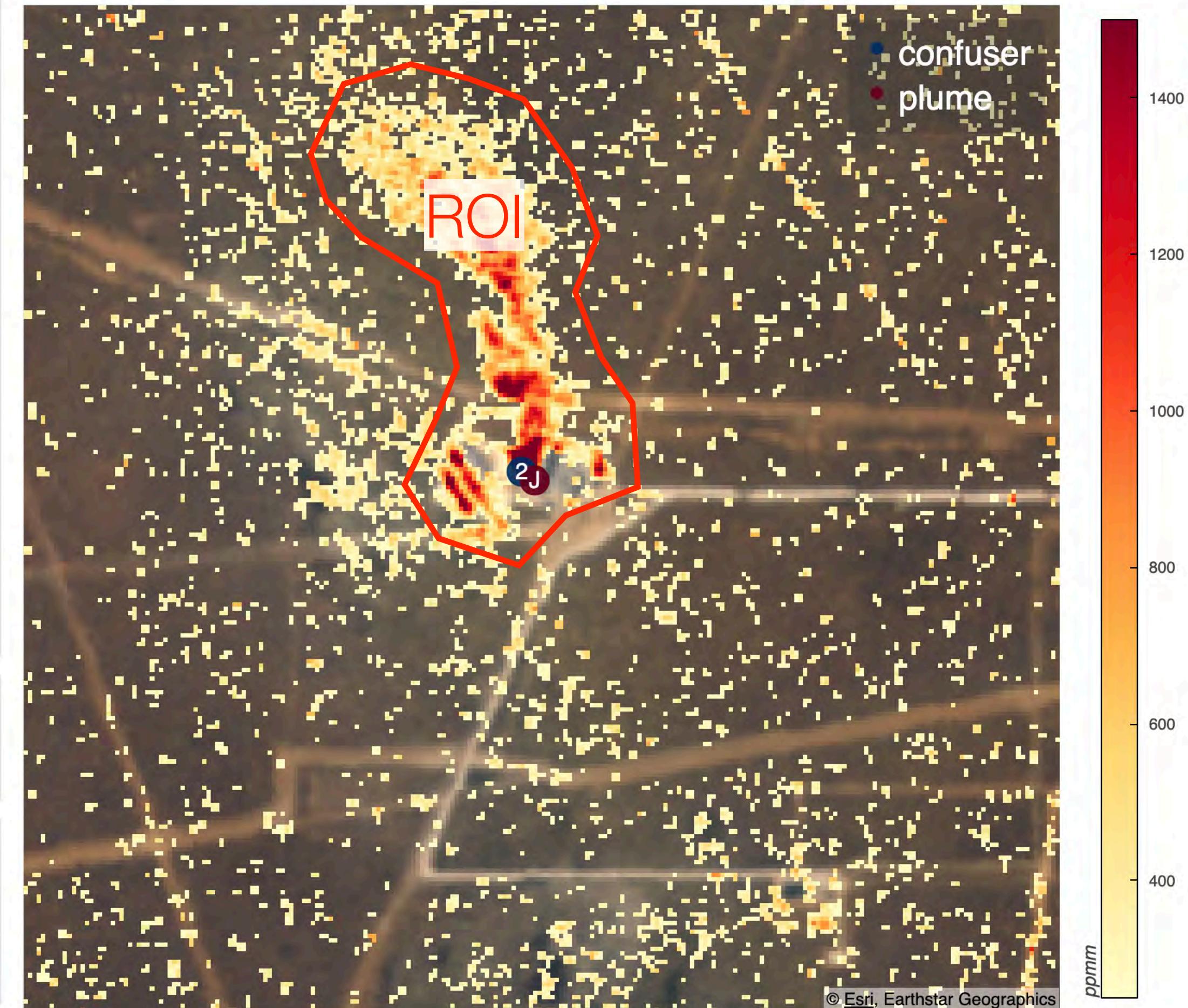
Candidate QC Labels: Duplicate

- **Duplicate** CIDs = multiple spatially adjacent CIDs that refer to a single ROI.
- If multiple spatially adjacent CIDs refer to the same ROI...
 - change the label of all of the duplicate CIDs to “**Ambiguous**” if you’re not sure if the ROI is a plume or false enhancement.
 - otherwise keep a single CID with the correct label and all of the remaining duplicate CIDs to “**Duplicate**”

Candidate QC Table			
CID	Latitude	Longitude	Label
1	32.212354	-103.697188	Confuser
2	32.20063	-103.684375	Duplicate
3	32.116367	-103.626792	Confuser
4	31.734044	-103.282694	Confuser
A	31.746281	-103.28727	Plume
B	31.733738	-103.28271	Plume
C	31.844338	-103.37374	Plume
D	32.089579	-103.56726	Plume
E	32.116028	-103.62661	Plume
F	31.784417	-103.32045	Plume
G	32.212117	-103.69709	Plume
H	32.283928	-103.74186	Plume
I	32.29695	-103.73814	Plume
J	32.200507	-103.68416	Plume

Add User Candidate(s)
Label for New Candidates
----- Select Label -----

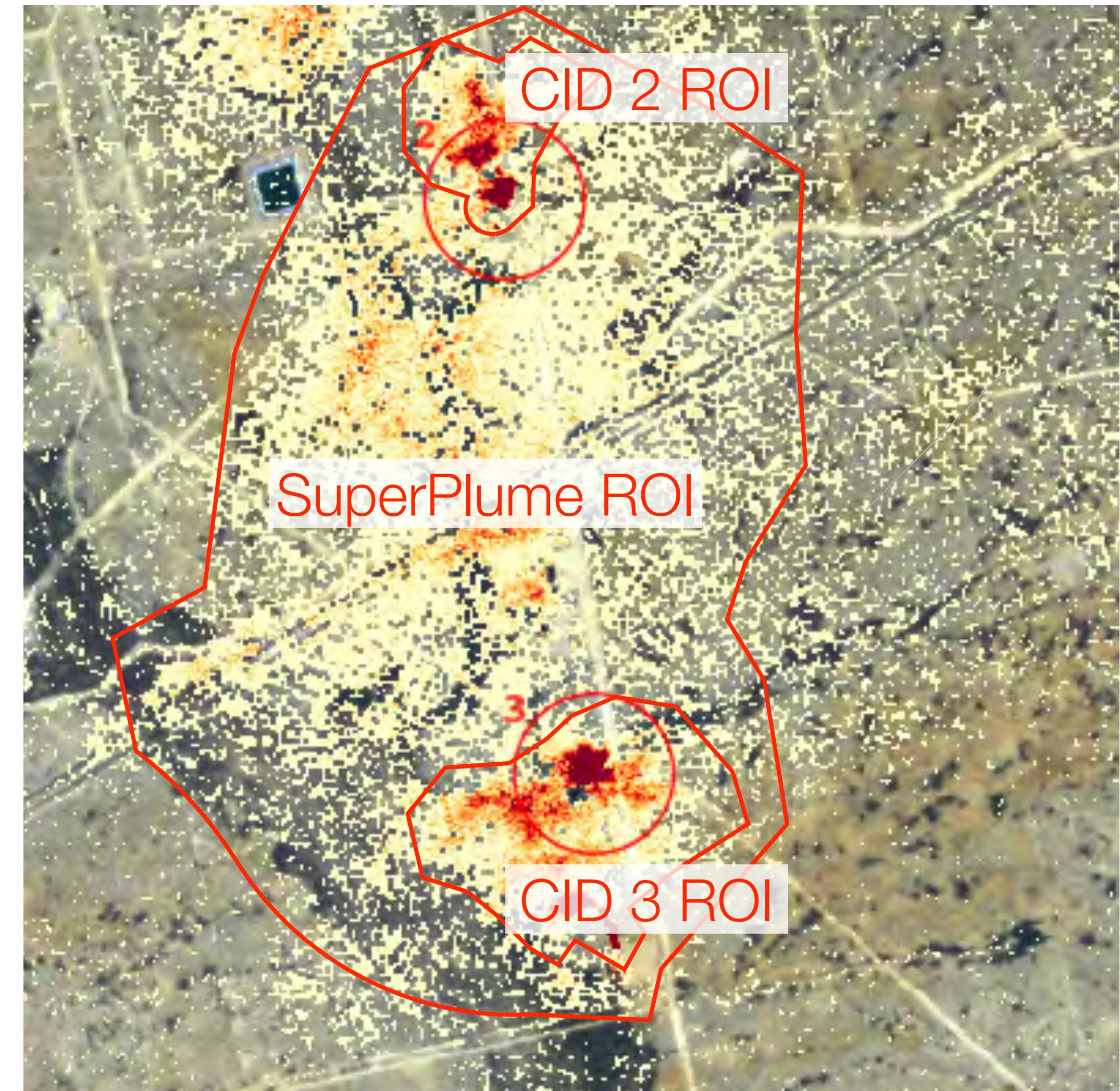
Remove User Candidate(s)
User Candidate to Remove
----- Select Candidate -----
Remove User Candidate



Above: CID 2 (label=Confuser) and CID J (label=Plume) both refer to the same ROI (outlined in red). Since this ROI is likely a real plume, we keep the correctly-labeled CID J unchanged and change the label of CID 2 to “Duplicate.”

Candidate QC Labels: Super Plumes

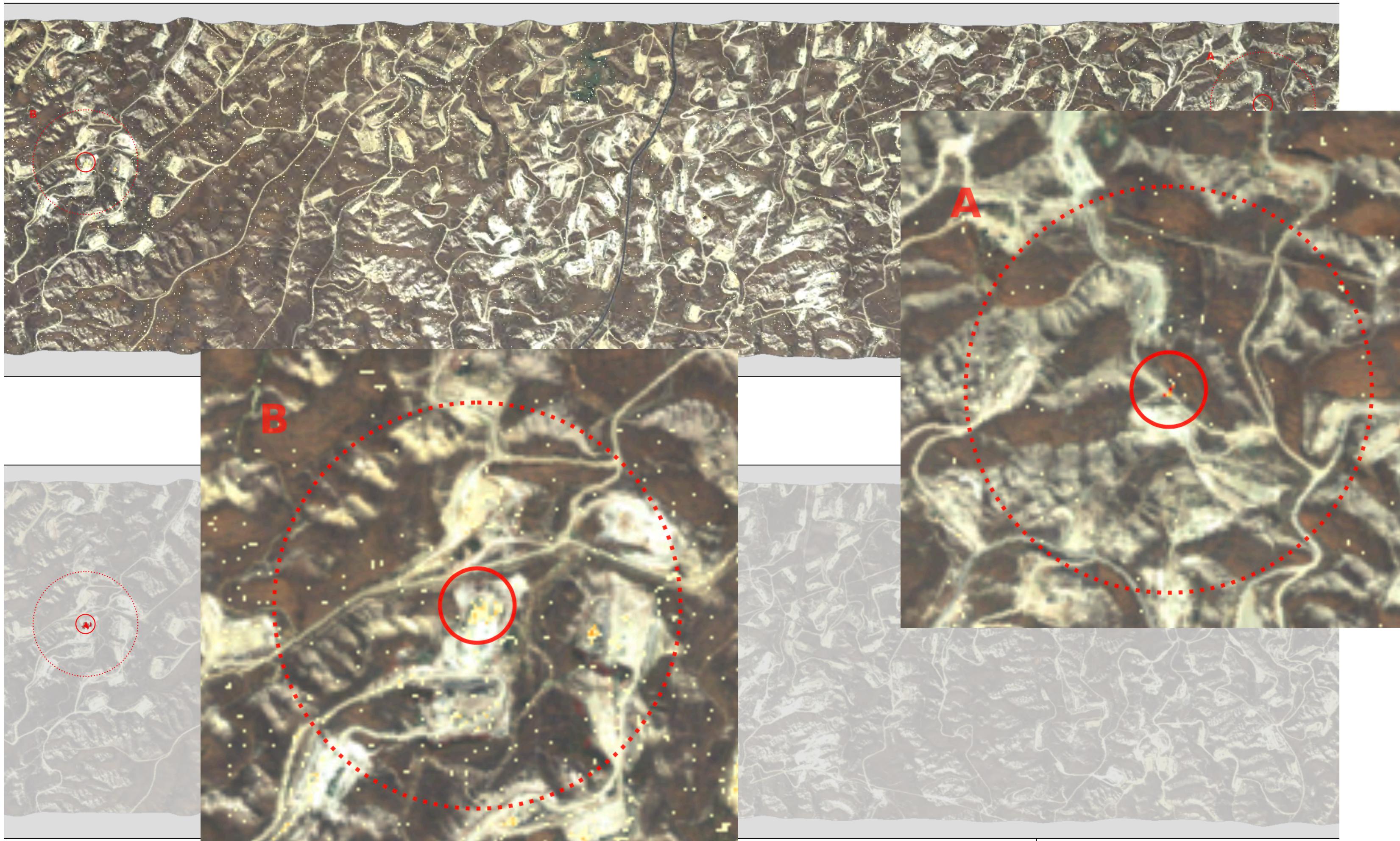
- Nontrivial detection, quantification and source attribution issues occur when...
 - spatially-converging plumes from neighboring but distinct sources or
 - plumes emitted from one or more “patch” sources (e.g., landfills or manure ponds)
- ...are present.
- We use the “**SuperPlume**” label when...
 - ...the ROIs of spatially distinct CIDs merge into a single plume, or
 - ...the ROI of a single CID refers to a large, spatially diffuse plume lacking a clear “point” source of emissions.



Above: the plume ROIs of CID 2 and CID 3 merge into a single large plume, so we change the label of each CID to “SuperPlume.”

Candidate QC Labels: Background

- Candidate ROIs with weak enhancements (near or below minppmm=250ppmm) are indistinguishable from the CMF background enhancement.
- These should be relabeled “**Background**” accordingly.



Candidate QC Labels: Plumes vs. Systematics

- False enhancement CIDs adjacent / overlapping systematics should be labeled “Artifact.”
- Plume CIDs that...
 - ...overlap / are adjacent to systematics or occur in roughly the same column as systematics

...should be labeled “Ambiguous” rather than “Artifact.”
- Be careful: small disconnected segments of systematics often look like small plumes.

Candidate QC Table			
CID	Latitude	Longitude	Label
1	32.212354	-103.697188	Confuser
2	32.20063	-103.684375	Confuser
3	32.116367	-103.626792	Confuser
4	31.734044	-103.282694	Confuser
A	31.746281	-103.28727	Plume
B	31.733738	-103.28271	Plume
C	31.844338	-103.37374	Ambiguous
D	32.089579	-103.56726	Plume
E	32.116028	-103.62661	Plume
F	31.784417	-103.32045	Plume
G	32.212117	-103.69709	Plume
H	32.283928	-103.74186	Plume
I	32.29695	-103.73814	Plume
J	32.200507	-103.68416	Plume

Add User Candidate(s)

Label for New Candidates

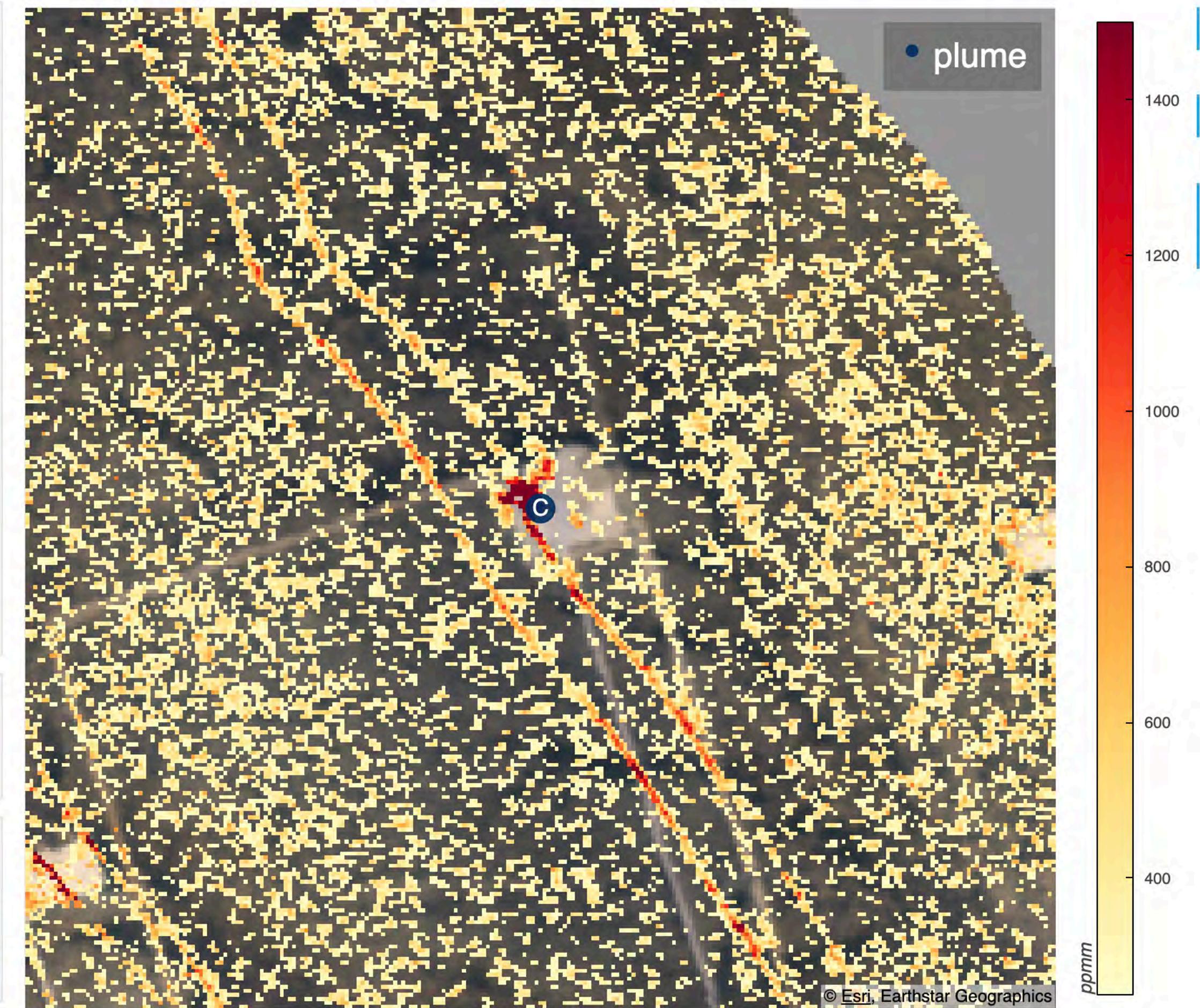
----- Select Label -----

Remove User Candidate(s)

User Candidate to Remove

----- Select Candidate -----

Remove User Candidate



Above: the ROI of CID C (label=Plume) overlaps a columnwise artifact, and should be relabeled as “Ambiguous”

Candidate QC Labeling Suggestions

If you are...	...use this label
...not confident that a candidate is a methane plume based primarily on it's CMF enhancement	Ambiguous
...not confident candidate is a methane plume without context from RGB quicklook, satellite imagery and/or map tiles	Ambiguous
...confident candidate is a false enhancement but category (e.g., cloud, artifact, confuser, etc.) uncertain or unspecified	False Enhancement
...confident candidate is a methane plume, but emissions generated by multiple neighboring point sources or a patch source (e.g., landfills, manure ponds)	SuperPlume
...confident candidate is a methane plume generated by a single point source, but location of source / source attribution details uncertain	Plume

Usage Guidelines

- Don't bother adding new candidates in flightlines with...
 - major data quality issues
 - very low contrast (high / low BGE)
 - severe ortho issues
 - overcast cloud cover
- If image rendering issues occur → pan/zoom scene (to force redraw)
- If you can't connect to cmftagger on the G3000 via ssh tunnel, make sure that...
 - only one instance of cmftagger.py is running on the G3000 ('ps aux|grep cmftagger.py')
 - cmftagger.py output ends with “Launching server at http://localhost:port”
 - ssh tunnel is running on the port specified in the cmftagger.py output (above)
 - only one ssh tunnel is running on your local machine ('ps aux|grep ssh')