SatVu

From space to GeoTIFF – creating remote sensing data products

Commercial in Confidence

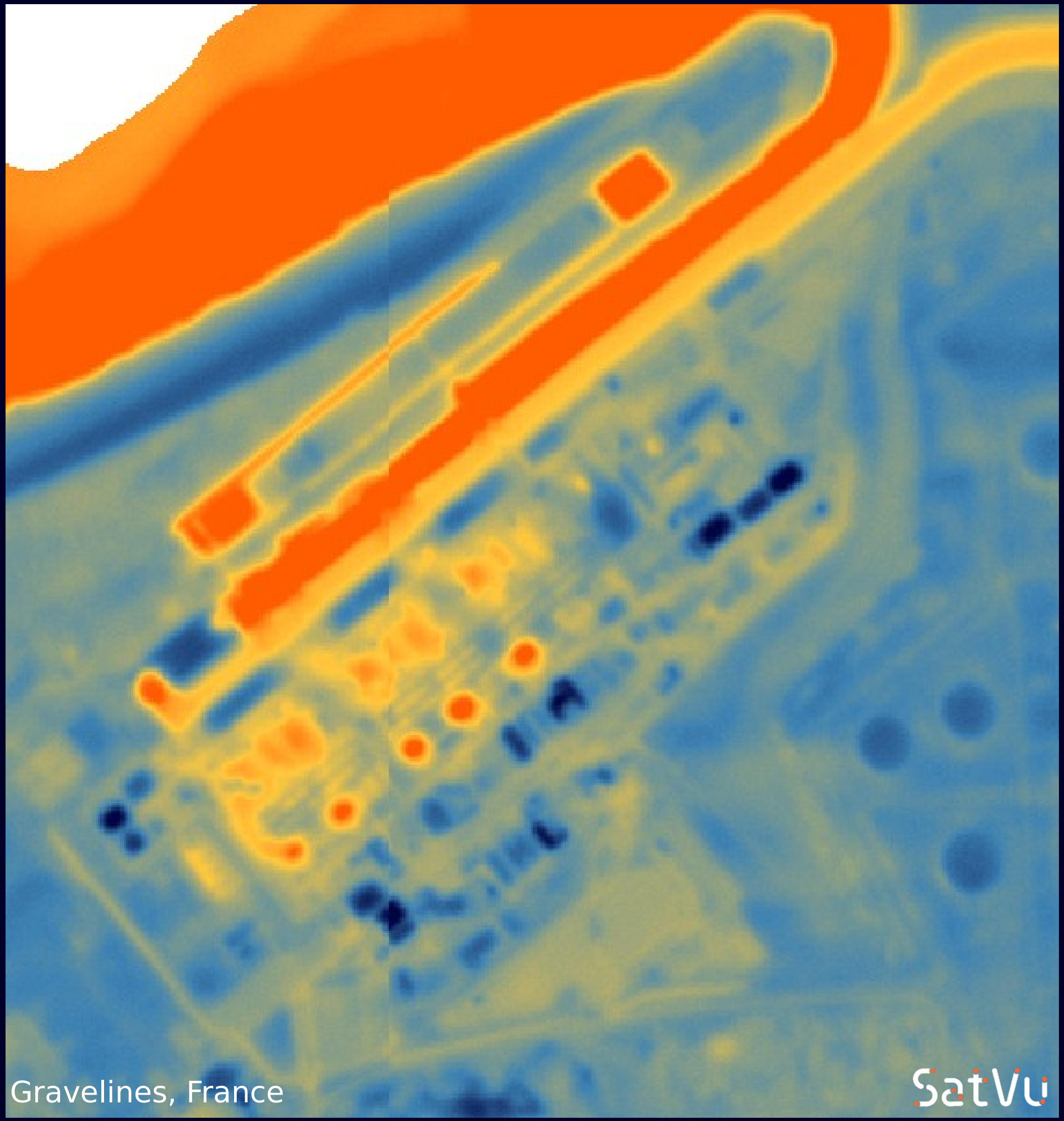
Intro - SatVu

Thermal imaging from space

Mid-wave IR satellite launched in June

3.5m GSD, offering a ‘visual’ product initially

Web APIs for tasking, searching, download



Intro – Image data workflow

- Raw image data received from satellite
- Correct and calibrate image
- Georeference image
- Collect image metadata (STAC)
- Quality assurance
- Publish image to catalogue

Intro - Architecture

Serverless infrastructure on AWS

Python 3.10, baked into Docker images

Rasterio does the heavy lifting for image data

Pystac for STAC work

STAC server implementation: customised STAC-FastAPI

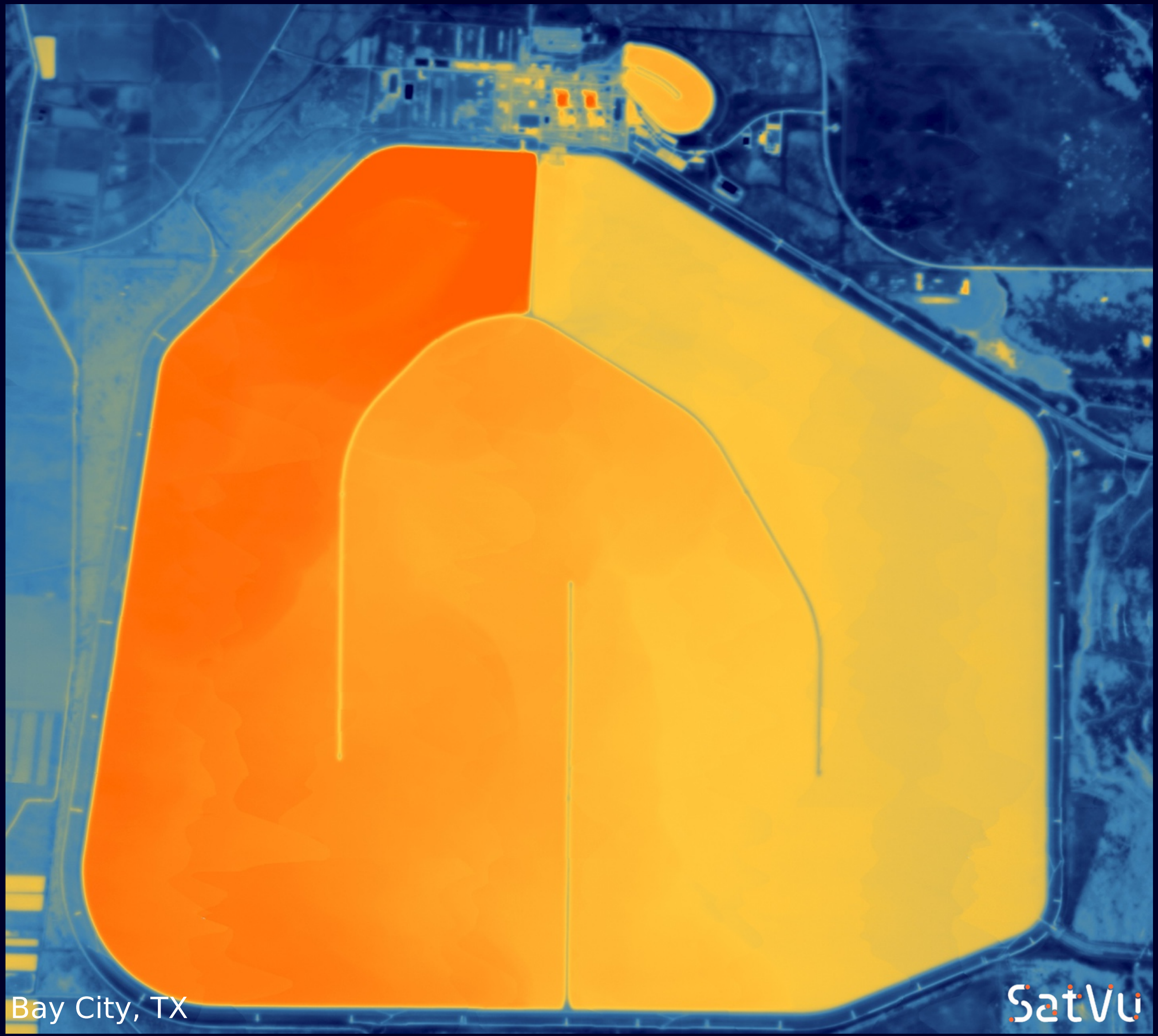
geojson-pydantic, GeoAlchemy2, odc-stac, rioxarray, ...

Images before launch

Aerial imaging campaigns over last three years

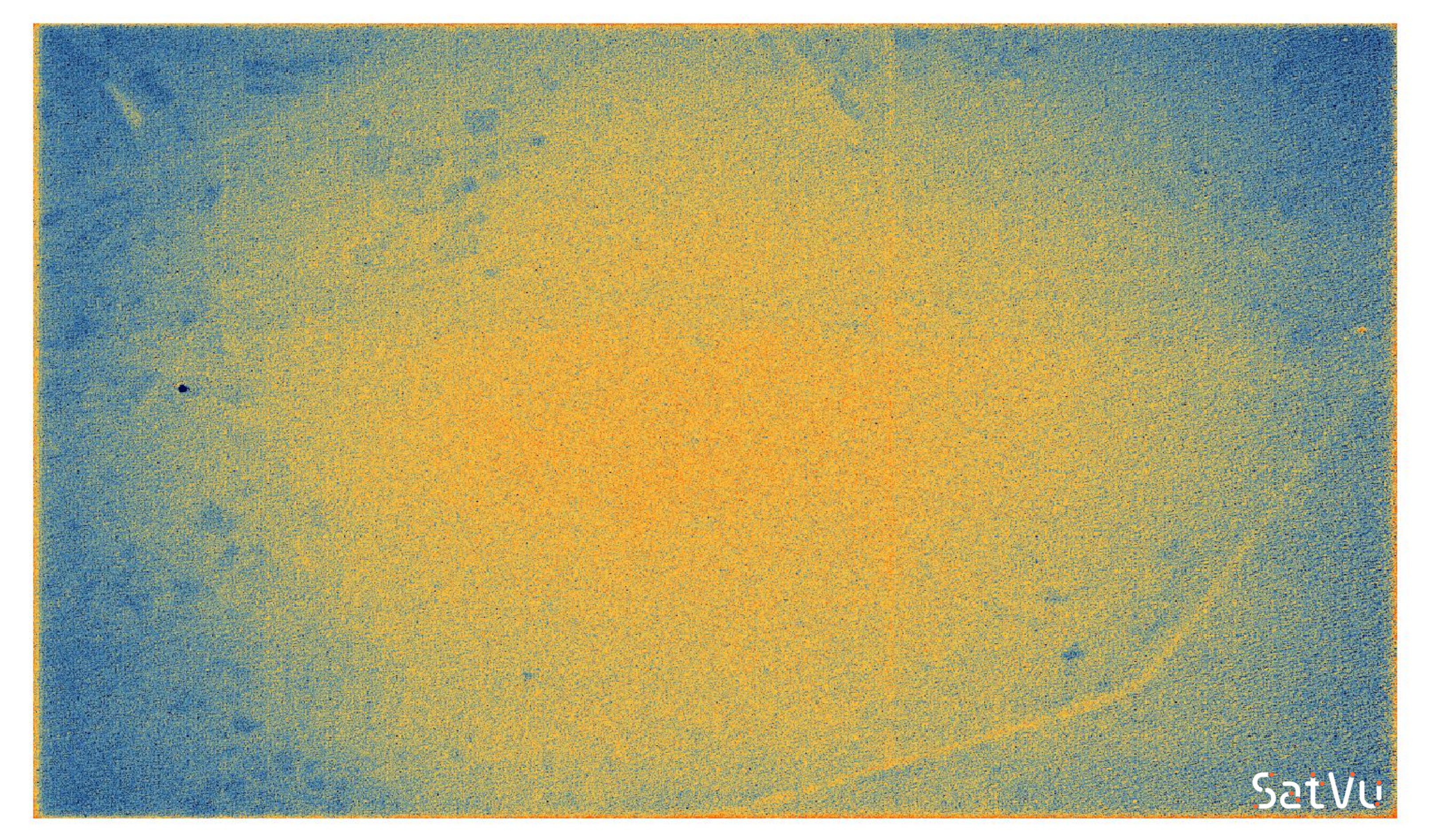
Degraded to simulate satellite conditions – get familiar with the data

Real images exist – but waiting for PR to go out!



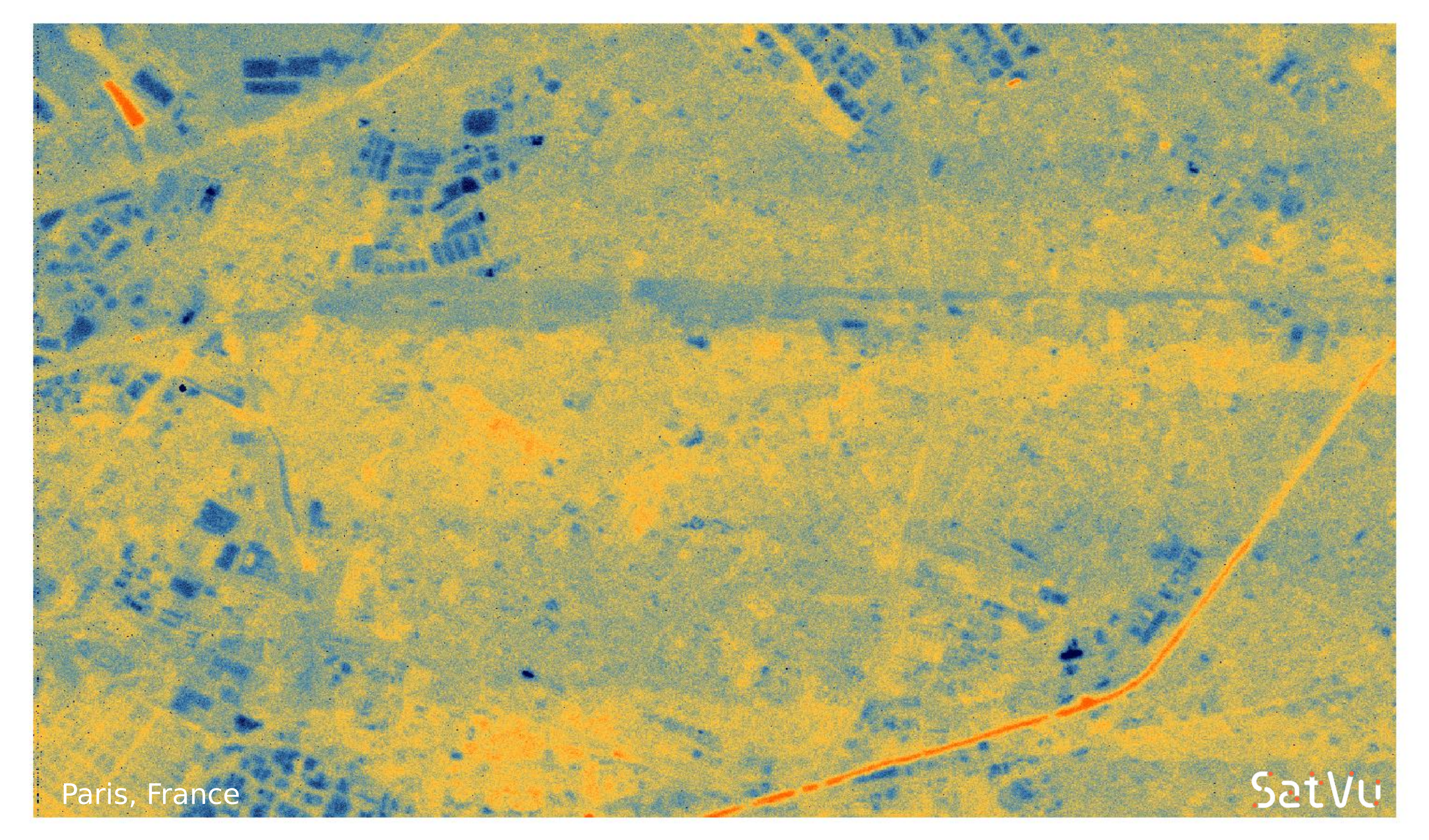
Processing – Calibration and correction

Raw MWIR images are pretty rough!

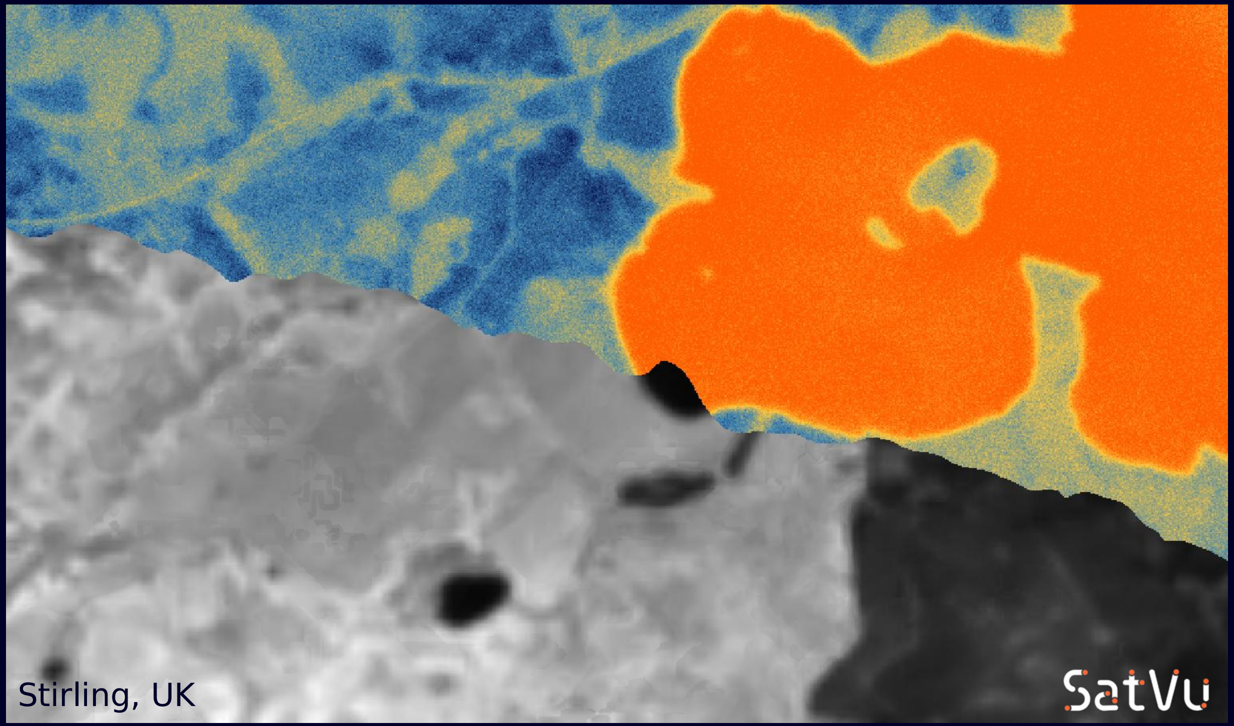


Processing - Calibration and correction

With proper characterisation, we can create useable data



Processing - Georeferencing

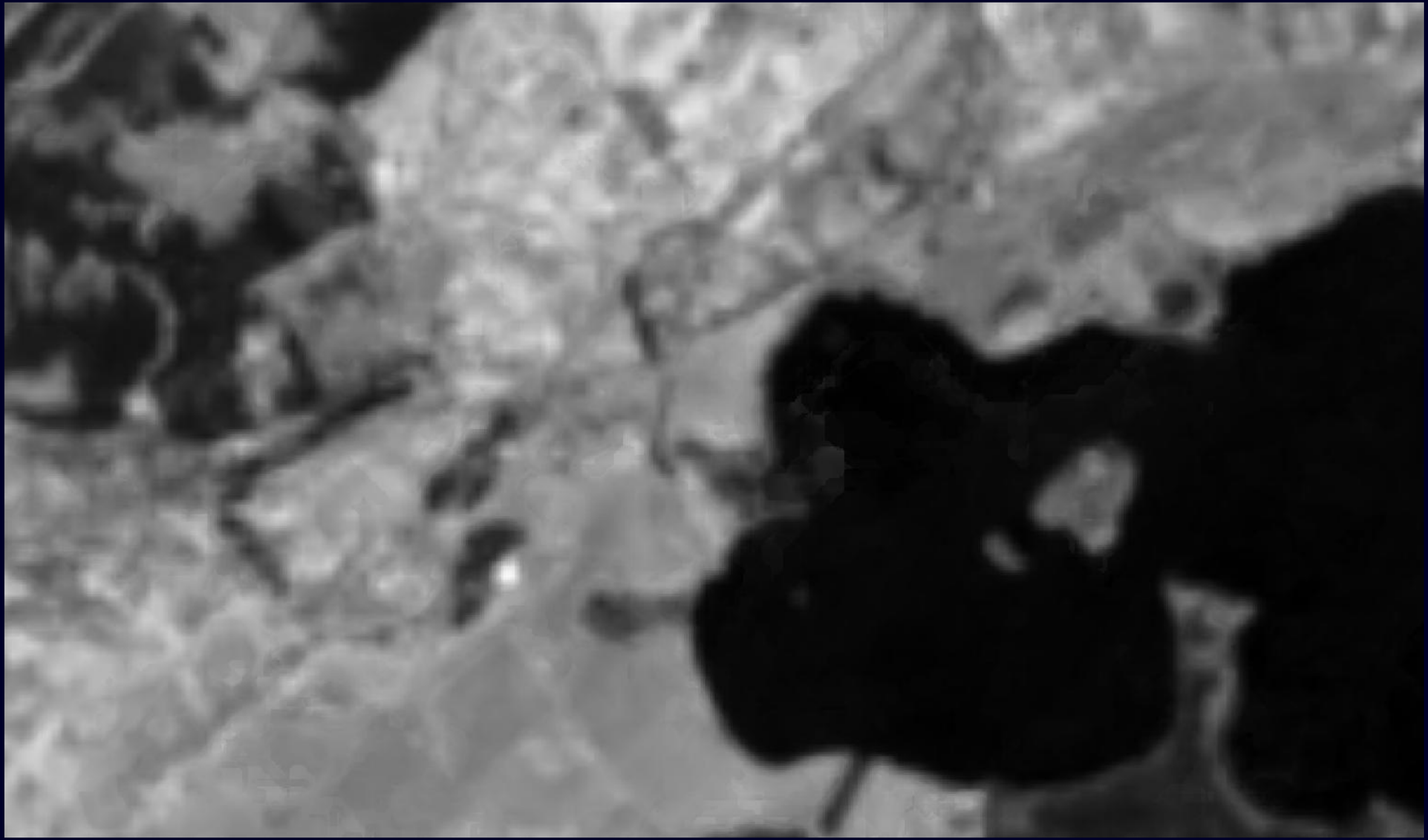


Georeference to Sentinel-2 NIR images

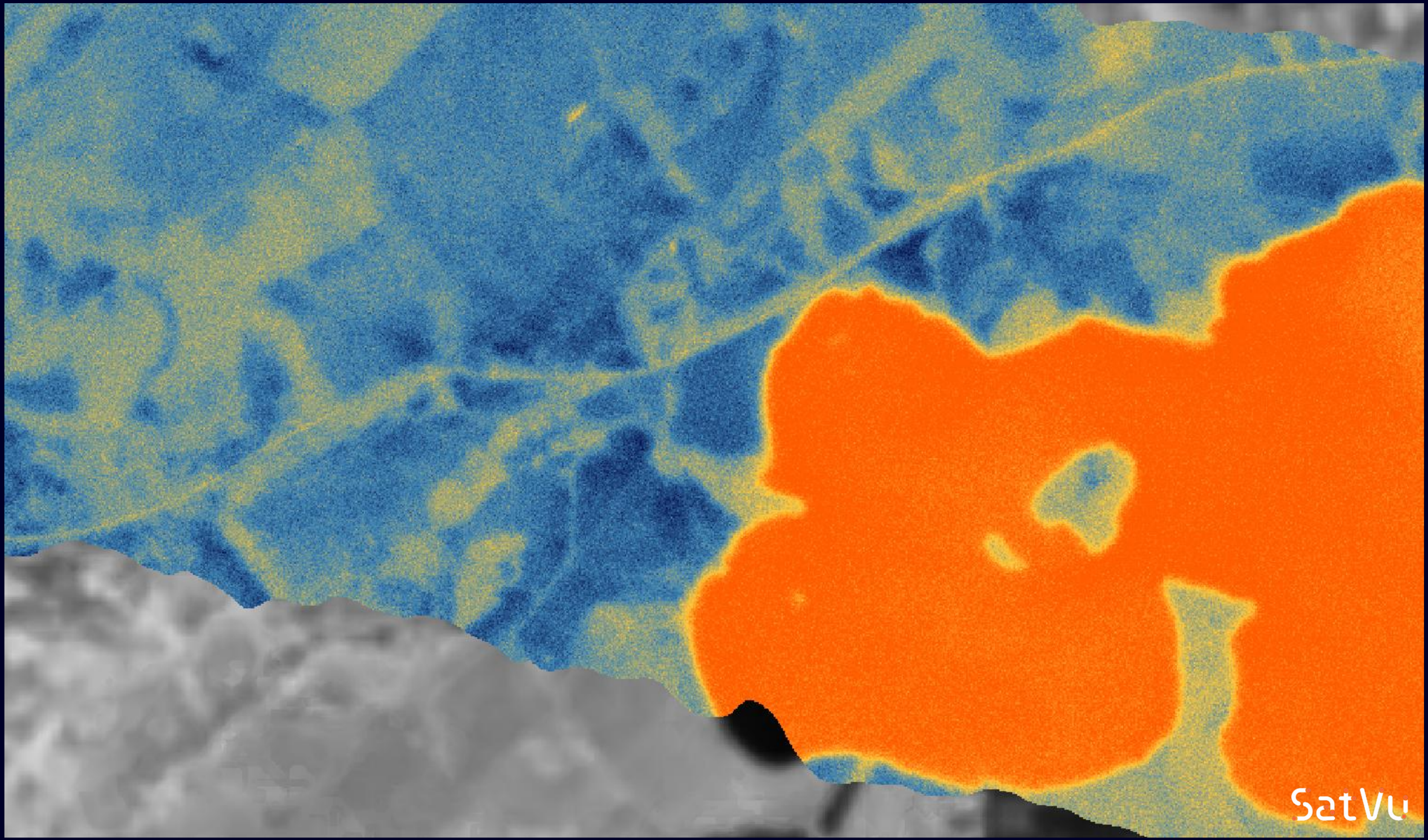
At a structural level, the world looks similar

Inverted brightness - reflective things don’t emit much thermal radiation

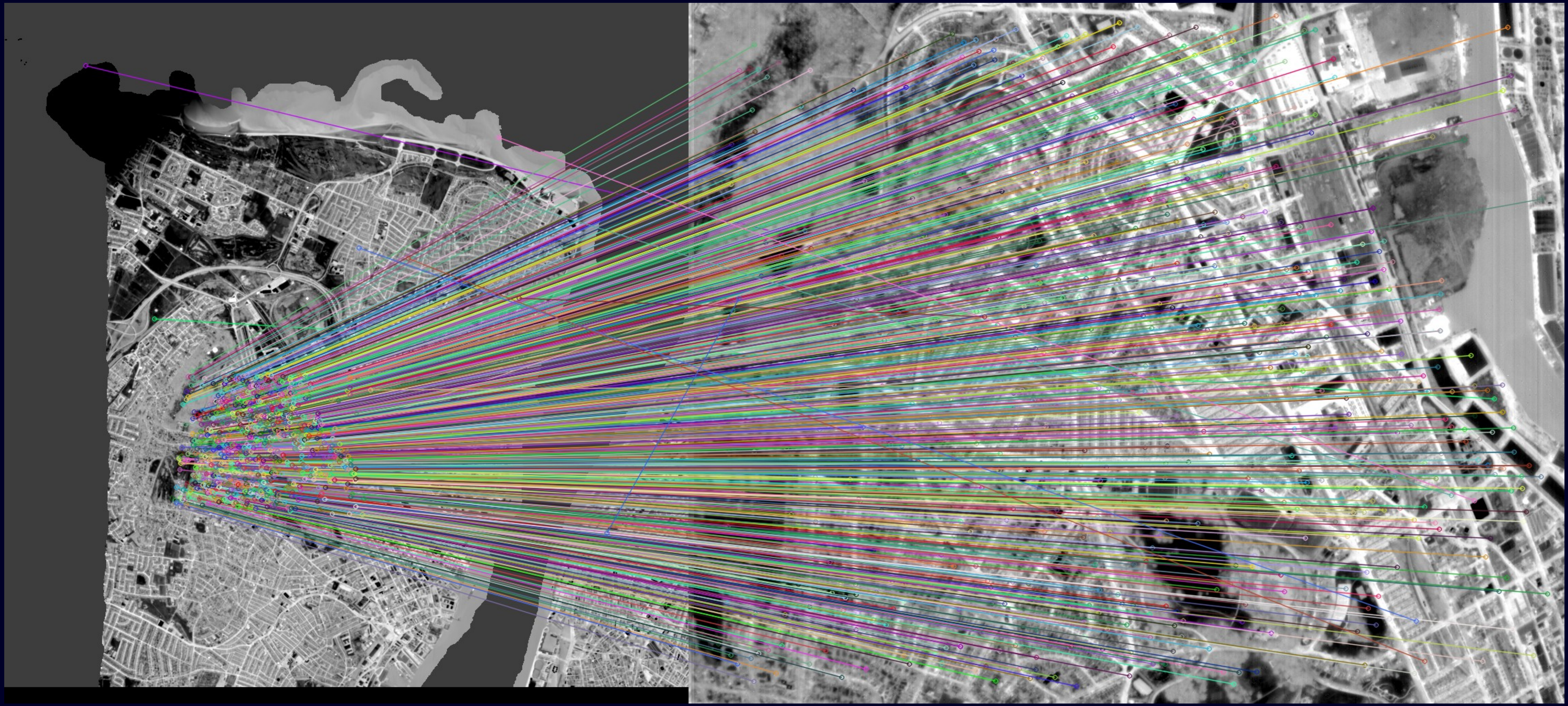
Processing - Georeferencing



Processing - Georeferencing



Processing - Georeferencing



Processing - Georeferencing

S2 data retrieved via Microsoft Planetary Computer or Element84 STAC catalogues

odc-stac does heavy lifting – get a single raster

Some pain points:

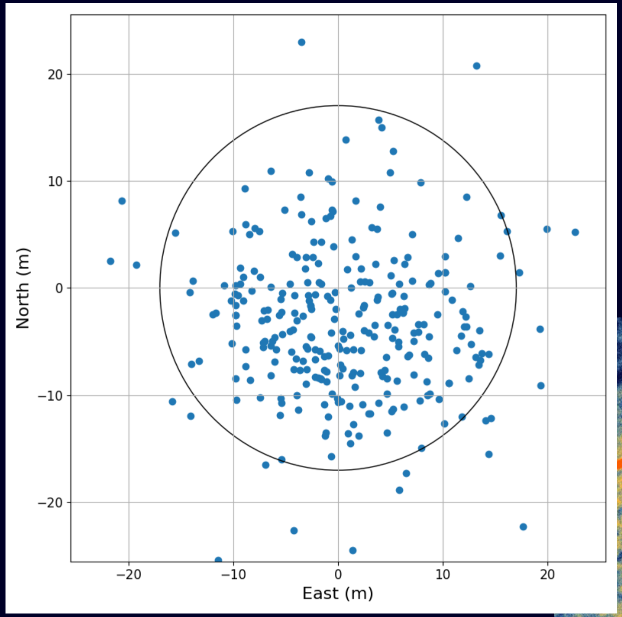
- Very different STAC item layouts
- Different amounts of source metadata included
- Poor reliability of MPC (improved over time)
- Inconsistent Sentinel-2 processing

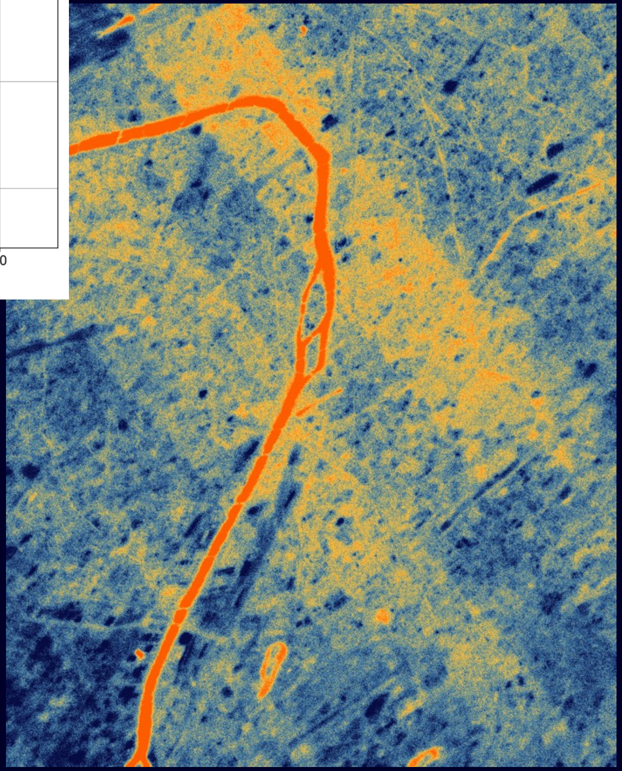
Processing - Georeferencing

Generated lots of synthetic data to test georeferencing

Off-nadir imaging angles

Realistic distribution of images to estimate CE90



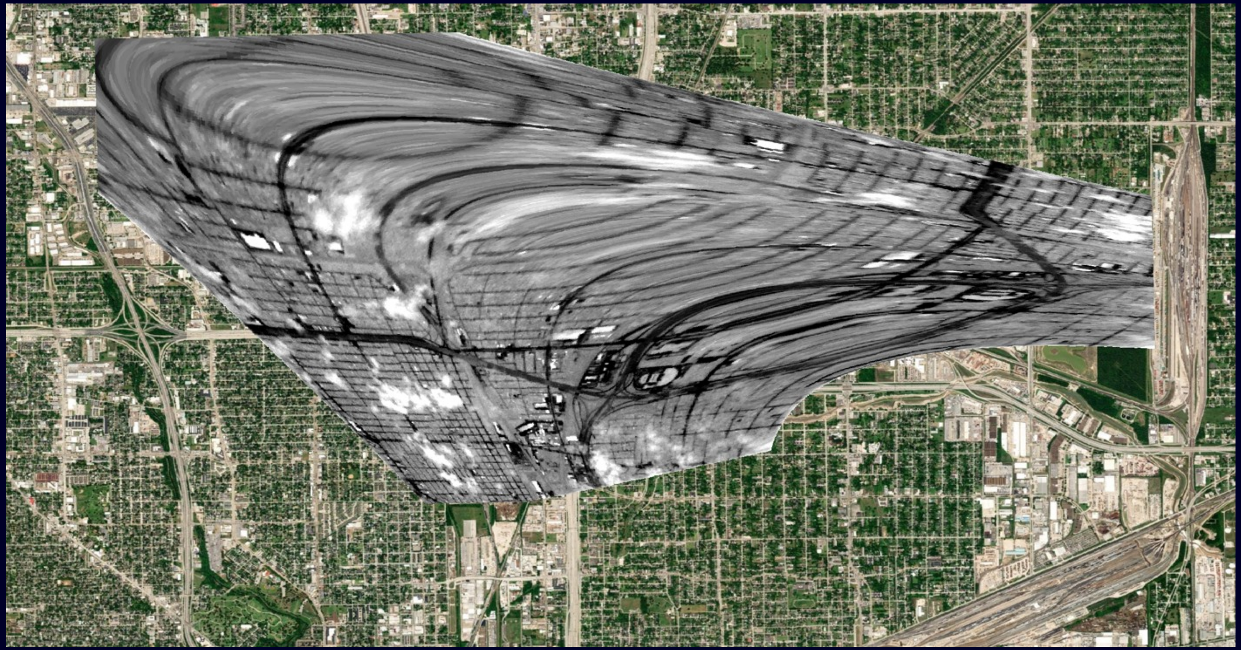


Processing - Georeferencing

Most georeferencing tools apply a polynomial warp

Generally, we expect a simple affine translation from sensor -> projected

Interested to hear about alternative tools!



Processing - Metadata

STAC – Spatio-Temporal Asset Catalog, OGC standard

Externally – standard interface for customers

Internally – provides almost everything we need to record about our data

STAC

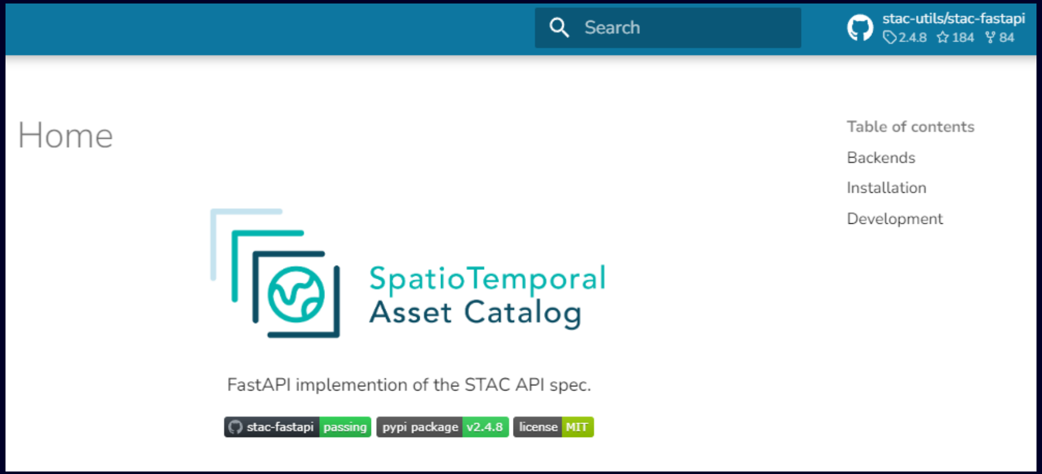
SpatioTemporal Asset Catalogs

The STAC specification is a **common language to describe geospatial information**, so it can more easily be worked with, indexed, and discovered.

Explore Tutorials

https://stacspec.org/en

Processing - Metadata



PySTAC Documentation

PySTAC is a library for working with SpatioTemporal Asset Catalogs (STAC) in Python 3. Some nice features of PySTAC are:

- Reading and writing STAC version 1.0. Future versions will read older versions of STAC, but always write the latest supported version. See STAC Spec Version Support for details.
- In-memory manipulations of STAC catalogs.
- Extend the I/O of STAC metadata to provide support for other platforms (e.g. cloud providers).
- Easy, efficient crawling of STAC catalogs. STAC objects are only read in when needed.
- Easily write “absolute published”, “relative published” and “self-contained” catalogs as described in the best practices documentation.

STAC ecosystem growing quickly

Don’t need to roll your own

STAC-FastAPI and underlying DB

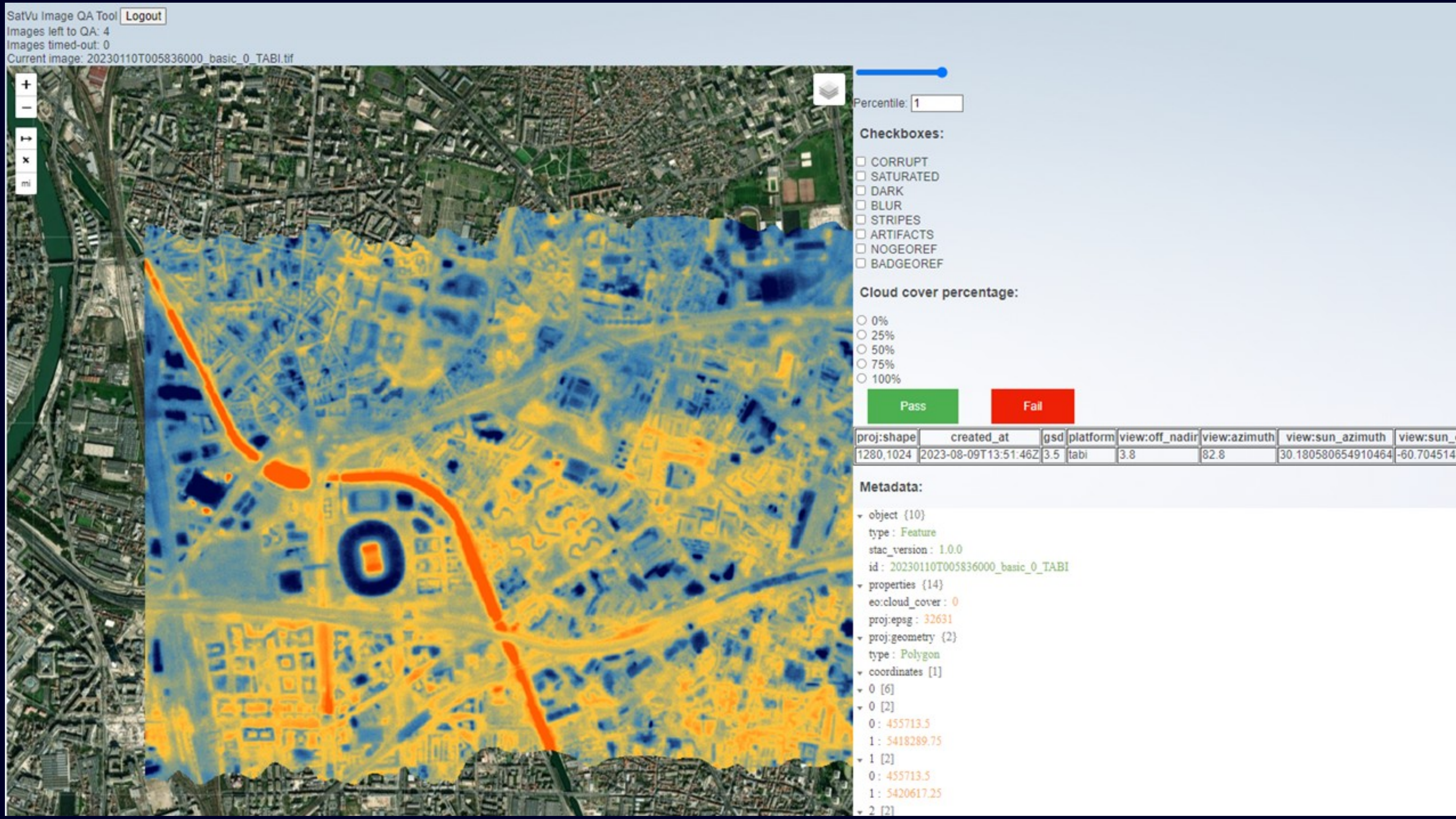
pystac + pystac-client

Processing – Quality Assurance

Lots of unknowns!

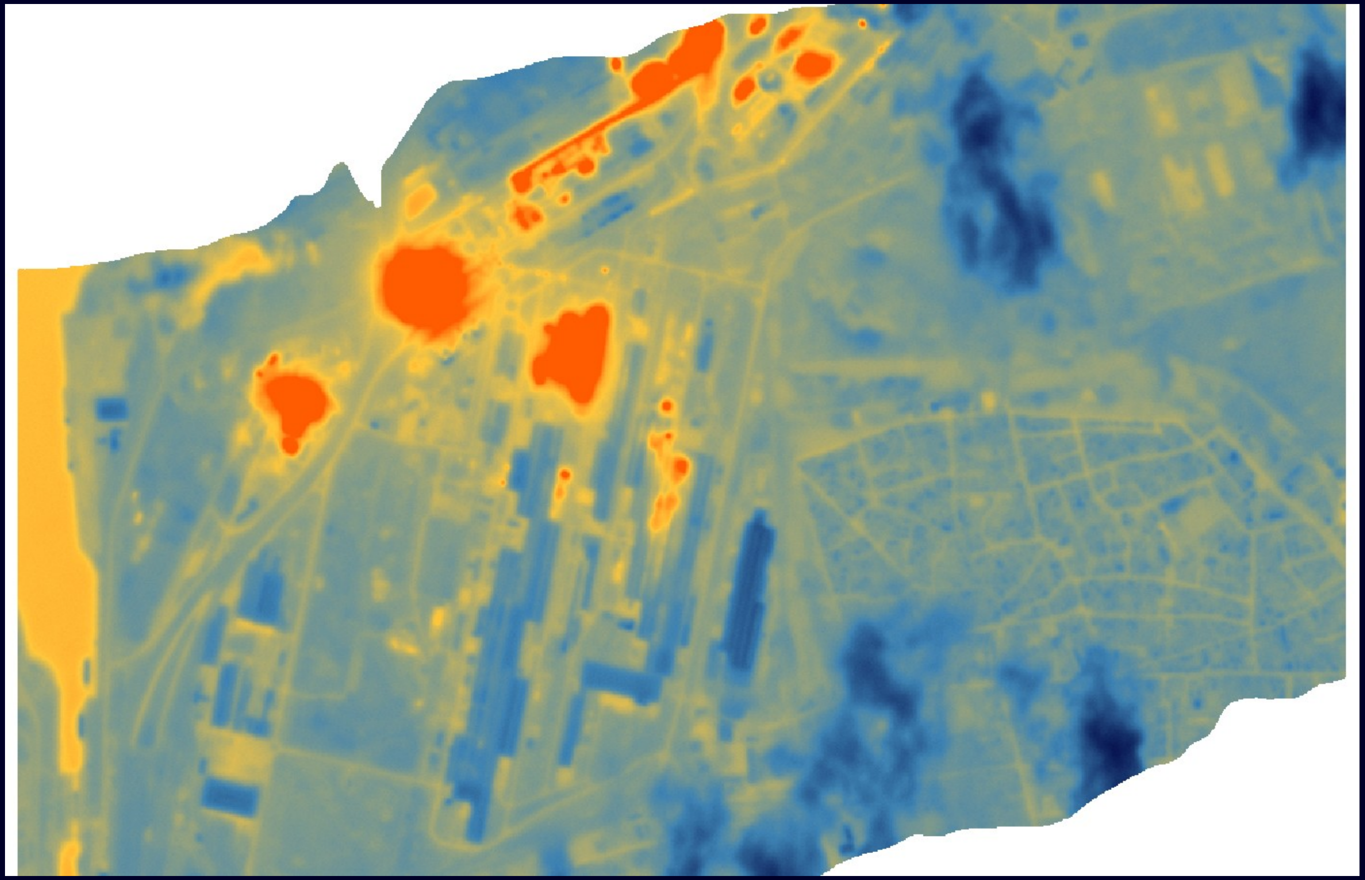
Will start with manual QA of data

Automate as we understand common issues



Processing – Quality Assurance

Cloud detection can be tricky
in single band images!



Cloud Architecture - Serverless

Serverless – AWS Lambda

Process individual images – works well

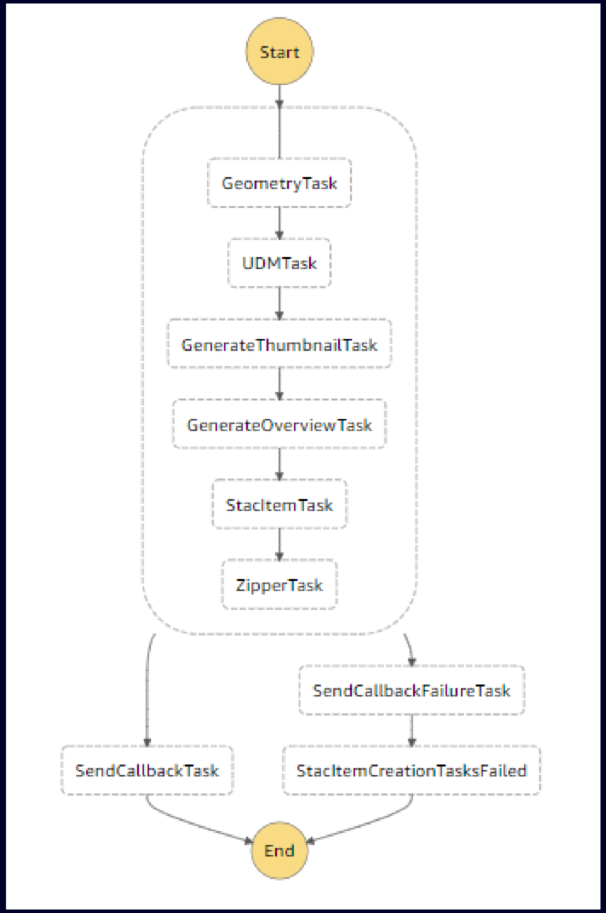
Python 3.10 in Docker Images

Abstract away from GDAL – compilation painful

Python GIS ecosystem mature

Just install rasterio, etc. on a standard Python Docker image

Cloud Architecture - Serverless



Orchestrating Lambdas can be hard

We love AWS State Machines, provide structured workflows

Summary

Going from raw data to geospatial – lots of variety!

Standards increasingly easy to adopt

Georeferencing – it's a pain

Geospatial in the cloud can be very painless