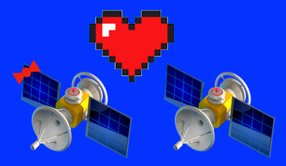
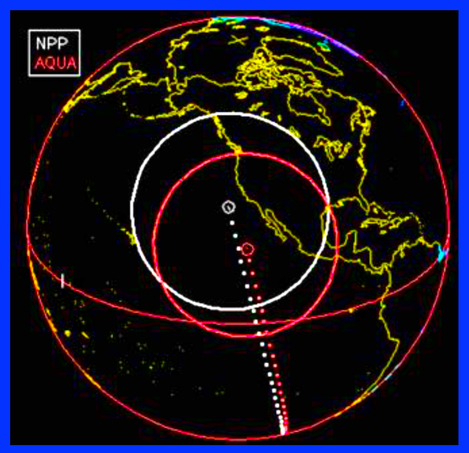


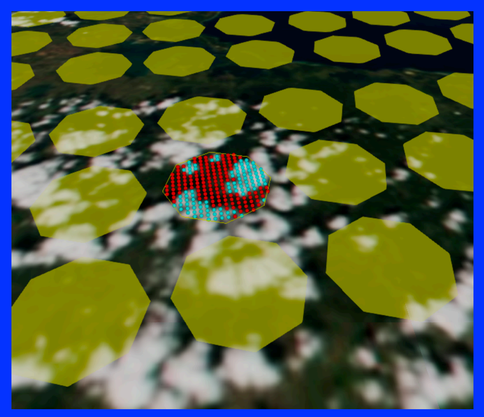
Matchmaker

A Toolkit for **Combining Satellite Observations** from Multiple Sensors



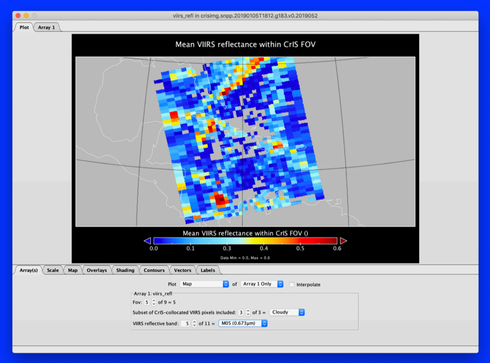
ACT 1

Orbital simulation identifies opportunities for coincident observations



ACT 2

Physical collocation finds individual observations that overlap spatially



ACT 3

NetCDF **match files** are created containing data from multiple sensors, **aligned** and **aggregated** for easy combined use

Matchmaker is powered by the **SciPy ecosystem**:

dask

- Enables Matchmaker to conveniently operate on arrays in fixed-sized chunks and avoid needing to fully load large datasets into memory
- Check out Dask GitHub PR #10237 (Array.vindex performance improvements) developed as part of work on Matchmaker

Numba

- Sensor data is aggregated to a common resolution before output, but NumPy functions only aggregate along fixed-size axes on rectilinear arrays
- Numba's JIT compilation accelerates specialized Python code that performs aggregation on Matchmaker's ragged data arrays
- We are exploring alternative SciPy tools to help flexibly aggregate results: Awkward Array, flox, numpy_groupies, …

xarray

- Xarray's Dataset object is the primary data structure supporting match file assembly
- NetCDF-based data model is a natural fit for existing satellite data collections
- Provides users with an established API for manipulating input data before collecting and aggregating for output

Cython

- The orbital simulation and physical collocation steps use a large collection of legacy tools written in Fortran
- Cython is used as a two-way bridge to enable function calls between the Python and Fortran code bases

SSEC

Greg Quinn
University of Wisconsin - Madison
Space Science & Engineering Center

Scan for additional details & resources