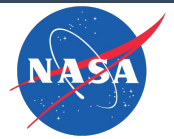


Analysis Ready Satellite Data

Chris Lynnes*, NASA

*U.S. Govt. Civil Servant

CARD4L* Definition

...satellite data that have been processed to a minimum set of requirements and organized into a form that allows immediate analysis with a minimum of additional user effort and interoperability both through time and with other datasets.

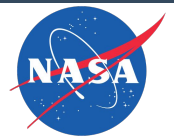
Analysis type unspecified, except...

...time series analysis emphasized

*CARD4L = CEOS Analysis Ready Data for Land

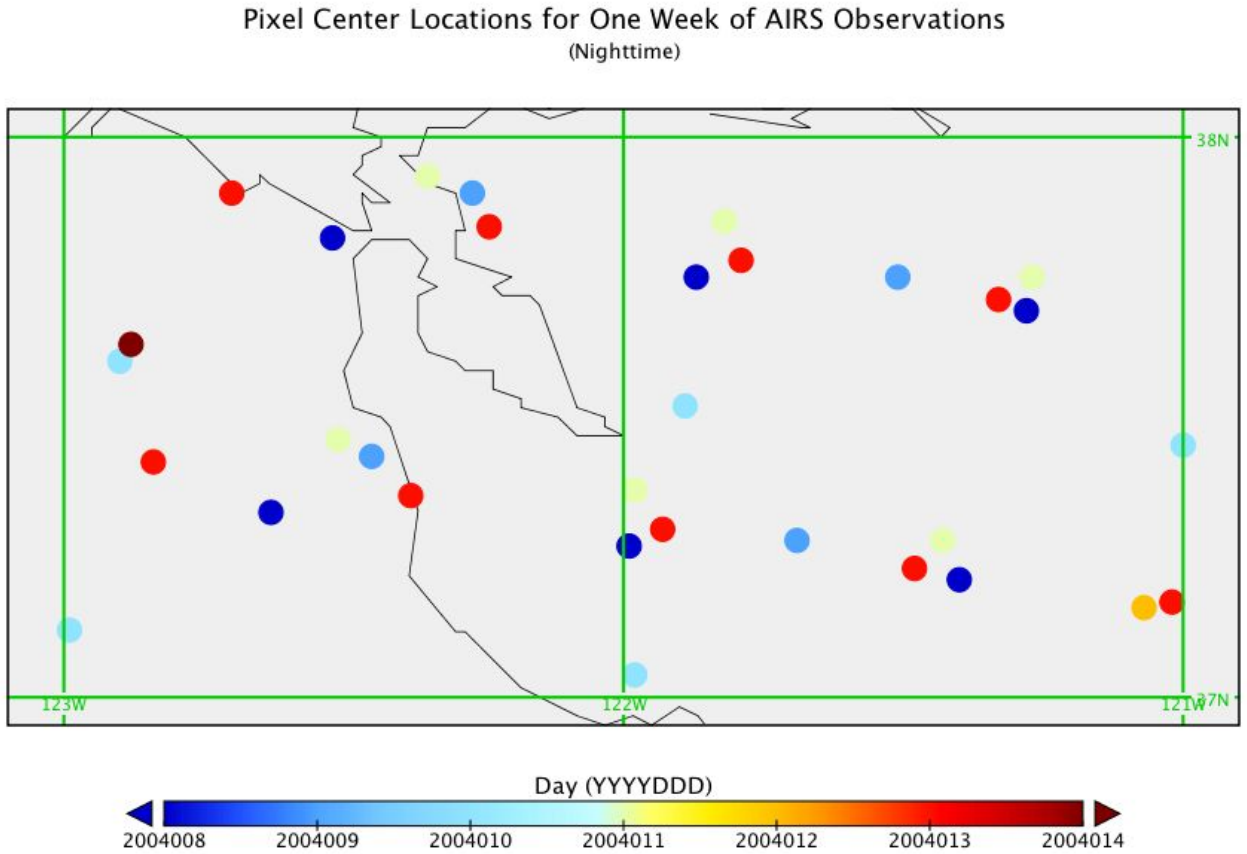
Analysis Ready Data (ARD) Benefits

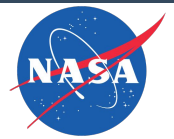
- ARD requires:
  - Important metadata
  - Easy to use spatial characteristics
  - Corrections for confounding factors (e.g., atmospheric)
- ARD promises to make time series analyses easier
- Especially useful for:
  - Applications



Can Level 2 Data Be Analysis-Ready?

L2 Time Series Analysis is...hard

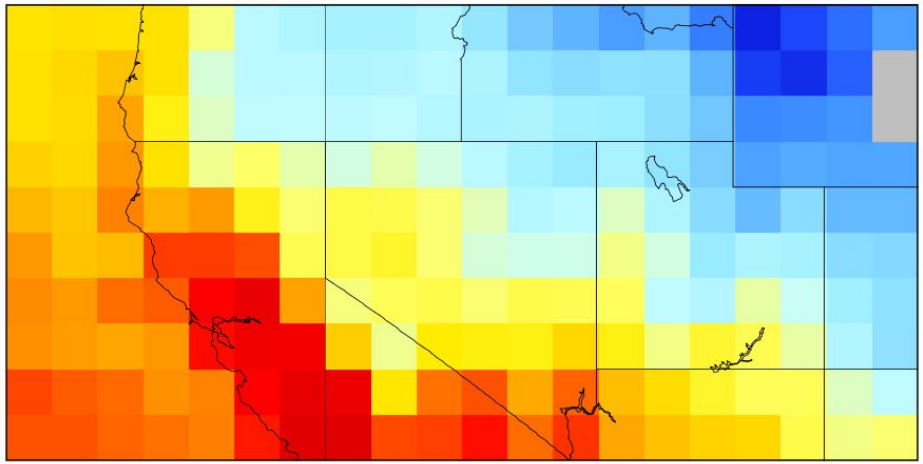




Aggregation to Level 3 (L3) gridded data regularizes geolocation, but...

...at the cost of spatial resolution...

Level 3 Surface Air Temperature

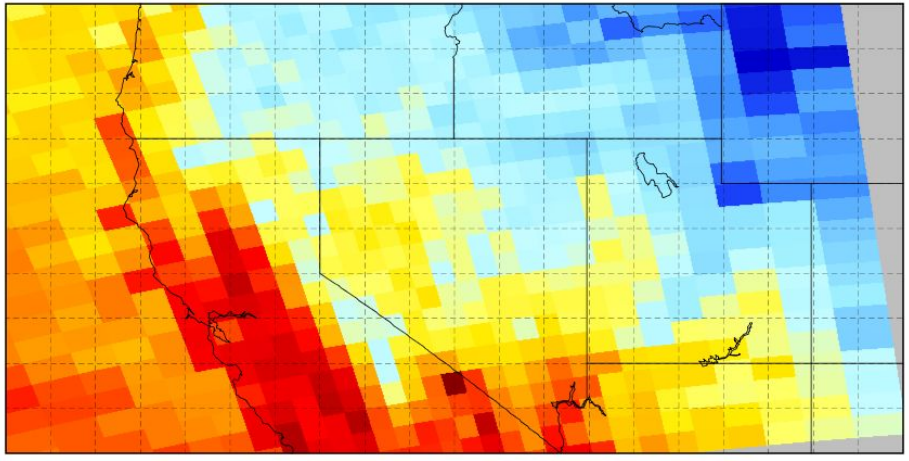


Temperature (K)

267.0 274.2 281.4 288.6 295.8 303.0

Data Min = 268.4, Max = 299.2

Level 2 Surface Air Temperature

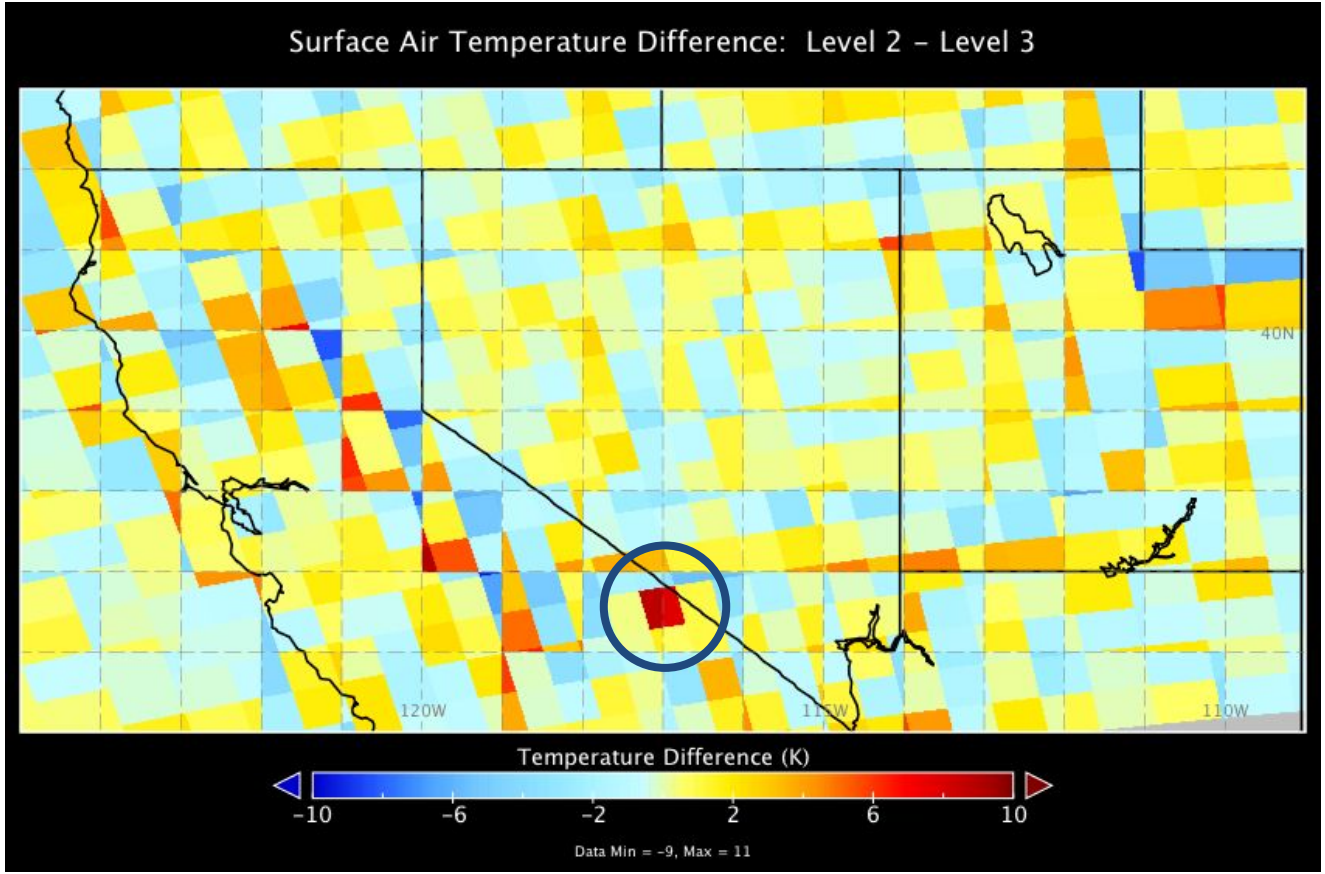


Temperature (K)

267.0 274.2 281.4 288.6 295.8 303.0

Data Min = 266.8, Max = 302.7

...and suppresses extreme values



Caveats of Level 3 Data

1. Spatial resolution degradation
2. Suppression of extreme values
3. Systematic bias over time in heterogeneous pixels
4. Day boundary artifacts
5. Difficulty of assigning numerical uncertainty to pixels
6. “Beating” of fast moving phenomena (e.g. dust storms) in long-duration aggregates (e.g., MISR monthly)
7. Irretrievable information loss when reprojecting from original L3 projection
8. Difficulty of assigning pixel level quality

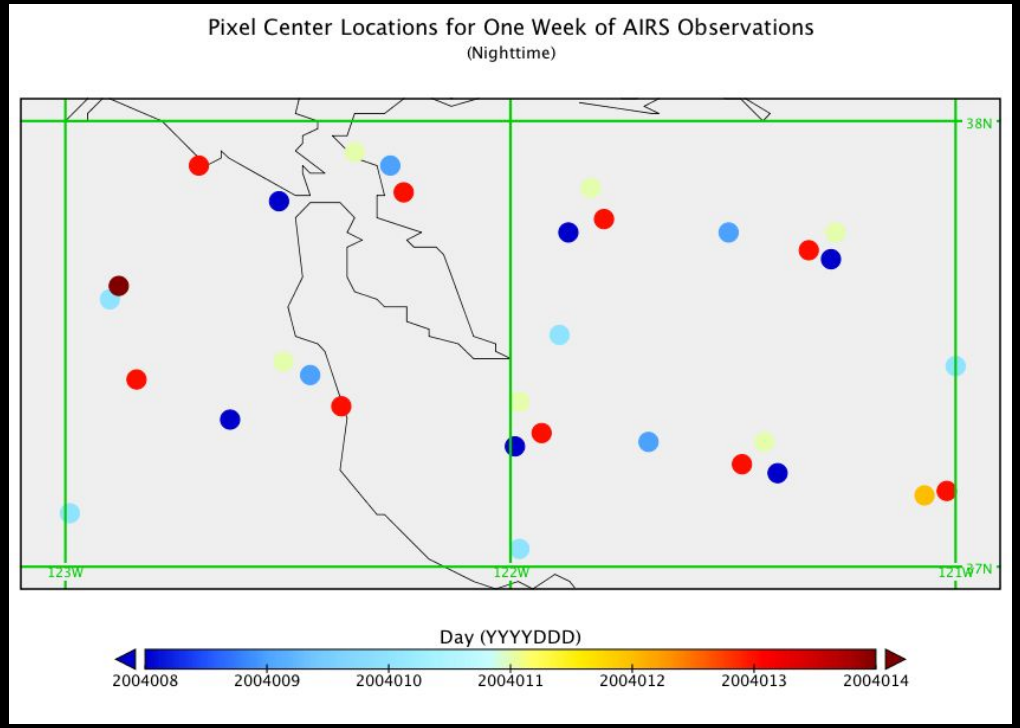
NASA Workshop 11/25/2020

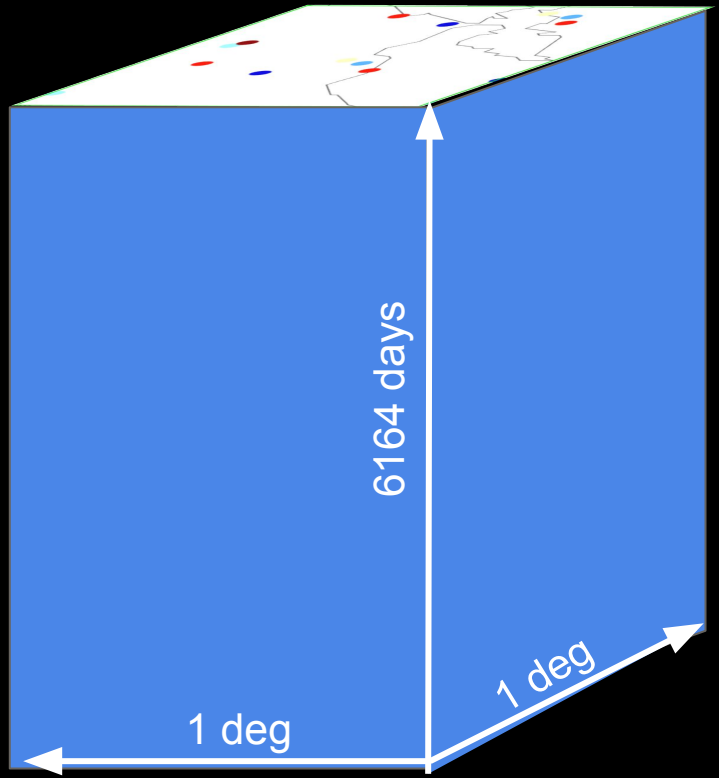
1. Try for quick wins with Level 3 / Level 4 data
2. Engage with CEOS w.r.t. ARD in other disciplines
3. Apply a more nuanced approach with L1/L2...

Making L1/L2 data analysis-ready

1. Make L1 and L2 data easier to analyse
   a. Tooling
   b. Modifications to data organization
   c. Capacity Building: examples and tutorials
2. Support on-the-fly gridded ARD from L1 and L2 data
   a. Defer resampling as long as possible to limit information loss
   b. Quantify information loss

L2G Time Stacks





Columnar bins of L2 data: lat_lon


```
$ ls -x TSurfAir.D
+24_-065   +24_-066   +24_-067   +24_-068   +24_-069
+24_-070   +24_-071   +24_-072   +24_-073   +24_-074
+24_-075   +24_-076   +24_-077   +24_-078   +24_-079
+24_-080   +24_-081   +24_-082   +24_-083   +24_-084
+24_-085   +24_-086   +24_-087   +24_-088   +24_-089
...
```

Simple python code for time series extraction from AIRS L2G Time Stacks


```
# Time series of data closest to a given point
ds['xdist'] = (ds['Longitude'] - my_lon)* \
                m.cos(my_lat * m.pi/180.)
ds['ydist'] =  ds['Latitude']-my_lat
ds['sq_dist'] = xr.ufuncs.square(ds['ydist']) \
              + xr.ufuncs.square(ds['xdist'])
df = ds.to_dataframe()
closest_idx = df.groupby("dataday")['sq_dist'].idxmin()
closest = df.loc[closest_idx,['dataday','TSurfAir']]
...
```