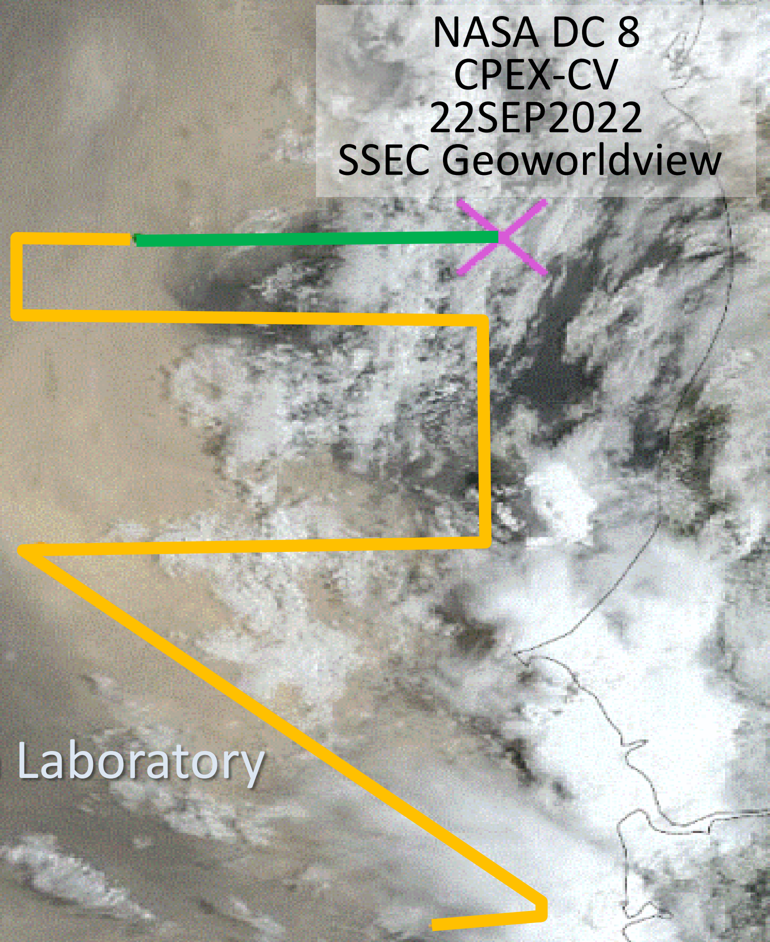
Making the best use of airborne field campaign data

12th ICAP Working Group
Monterey CA
October, 2022



Send comments
Jeffrey S. Reid, US Naval Research Laboratory
jeffrey.reid@nrlmry.navy.mil

U.S. NAVAL RESEARCH LABORATORY

SSEC

DISTRIBUTION STATEMENT A: Approved for public release. Distribution unlimited.

Field Observations: It can be fun in the field but then what?

Field collection is expensive, and cooperation between agencies during large “capital field campaigns” offers quantity, quality, & location value to the community.

“The need to improve models” is often used as a rationale for campaign work. But how much data is really used in a meaningful way?

A man with one watch, or two? Do all of the different measurements makes sense together?

Challenge: The data is great, but the archive is diffuse, and the products nuanced. Often this results in a series of one ups for each mission. Missions lack the infrastructure for searchable metadata so the community can benefit.

DISTRIBUTION STATEMENT A: Approved for public release.   Distribution unlimited.

From a 2020 OSTP Report On Improving Earth System Prediction

Existing observations are underutilized for providing constraints on both physical processes and phenomenology in ESMs. In part, this is due to the differences between the spatial-temporal scales of in-situ observations, satellite observations, and model-derived predictions of aerosol and cloud properties. Additionally, the quantities that are directly observed by in situ and satellite observations do not always provide useful constraints on fundamental aerosol and cloud processes that are important to predictability. Opportunities include new techniques (such as machine learning) to derive parameters that better constrain predictability from coordination of existing observations, as well as to identification of new technologies or observations that can fill the gaps in characterizing aerosol-cloud processes important to predictability of precipitation. Programmatic roadblocks include the lack of a mechanism for the consolidation, synthesis, and mining of the diverse observations collected by different agencies. The goal of this activity is to incentivize creative thinking in bridging the gaps (satellite simulators, direct use of spectral data, nudged model simulations, use of new ML/AI approaches, etc.). Of particular relevance are past efforts utilizing NASA remote sensing data, DOE ARM measurements, and field campaigns supported by NASA, NOAA, NSF, and ONR (TC-4, SEAC4RS, ORACLES, PISTON, YMC, CAMP2Ex, MARCUS, SOCRATES, ATOMIC, etc.) for this purpose.

DISTRIBUTION STATEMENT A: Approved for public release. Distribution unlimited.

Field Measurements of Aerosols, Clouds and their Interaction for Earth Systems Models (MACIE)

- Every agency is facing similar challenges in achieving, processing and applying observations to models.
- As part of an OSTP effort, an interagency working group was established on data and program planning with US fed and lead university partners.
- This is distinct from other interagency efforts (WMO, ICAMS), in that it is populated with working scientists fro bottom up. Exposure prevents decisions that "seemed like a good idea at the time."
- And people are pitching in. We utilize good ideas where we can find them, while balancing what WMO is up to on CF compliance.
- Our focus in the last year has been on data harmonization through templates, and working with NASA LaRC/GSFC to develop next gen ASCII and NetCDF convections.
- Potential for adding European datasets?

DISTRIBUTION STATEMENT A: Approved for public release. Distribution unlimited.

Hmm, in some ways MACIE sounds a bit like ICAP.

But, MACIE has in implicit interoperability mandate.

- Data needs to be 1) Accessible; 2) Timely; and 3) Well characterized.
- Everybody does things just differently enough to make importation of a field data set a significant job.
- Goal is to generate templates that can guide data files. Gao Chen of LaRC is the hero of the year. We should have a side discussion on what the modeling community thinks is appropriate.
- ICARTT ASCII format, while it has a long legacy, needs an update. A real timestamp will help.
- A heavier lift is NetCDF. Can be easy to generate, but to be fully “self describing” there are lots of flags and parameters that need to be set

```
netcdf file:/C:/Users/reidj/Documents/Telework/FTAC-MACIE/MACIE/netcdf/CAMP2EX-mrg01-P3B_
  dimensions:
    time = 32458;
    tbnds = 2;
    SMPS_Bin = 30;
    FIMS_Bin = 30;
    LAS_Bin = 26;
    APS_Bin = 13;
    WL_RSP1 = 10;
    WL_SSFR = 12;
  variables:
    double time(time=32458);
      :units = "seconds since 2019-09-06 00:00:00";
      :long_name = "UTC mid time of the merge interval";
      :standard_name = "time";
      :VariableType = "Coordinates";
      :bounds = "time_bnds";
      :_ChunkSizes = 32458U; // uint

    double time_bnds(time=32458, tbnds=2);
      :VariableType = "bounds";
      :units = "seconds since 2019-09-06 00:00:00";
      :long_name = "start and end time of merge intervals";

    float lat(time=32458);
      :VariableType = "Coordinates";
      :standard_name = "Latitude";
      :MACIEstandard_name = "Platform_Latitude_InSitu_None";
      :units = "degrees_north";
      :long_name = "average gps latitude for the merge time interval";
      :source = "CAMP2Ex-MetNav";
      :SourceVarName = "Latitude";
      :_ChunkSizes = 32458U; // uint

    float lon(time=32458);
      :VariableType = "Coordinates";
      :units = "degrees_east";
      :standard_name = "Longitude";
      :MACIEstandard_name = "Platform_Longitude_InSitu_None";
      :long_name = "average gps longitude for the merge time interval";
      :DataSource = "CAMP2Ex-MetNav";
      :SourceVarName = "Longitude";
      :_ChunkSizes = 32458U; // uint

    float alt(time=32458);
      :VariableType = "Coordinates";
      :units = "m";
      :standard_name = "altitude";
      :MACIEstandard_name = "Platform_AltitudeMSL_InSitu_None";
      :positive = "up";
      :long_name = "average gps MSL altitude for the merge time interval";
      :DataSource = "CAMP2Ex-MetNav";
      :SourceVarName = "GPS_Altitude";
      :_ChunkSizes = 32458U; // uint
```

DISTRIBUTION STATEMENT A: Approved for public release. Distribution unlimited.

MACIE HDF and NetCDF File Requirements (I)

- **Background:** HDF and NetCDF files are widely used in airborne field campaigns to report remote sensing observations
- **Motivation:** use of HDF and NetCDF has NOT increased the data readability and usability, and the current CF conventions, while popular with the modeling community, are not sufficient to provide adequate descriptions of the files and the variables.
- **Objectives:** Create a community standard for HDF and NetCDF files that enables implementations of F.A.I.R. principles and open data/open-source policies
  - ✓ Be as CF compliant as possible, i.e., use applicable CF attributes and CF coordinate system
  - ✓ Incorporate “domain relevant metadata” specific to various types of instruments
  - ✓ Use variable standard names which can provide adequate measurement coverage
  - ✓ Use WMO WIGOS standards when applicable, e.g., measurement units

DISTRIBUTION STATEMENT A: Approved for public release. Distribution unlimited.

Gao Chen, NASA LaRC (gao.chen@nasa.gov)

MACIE HDF and NetCDF File Requirements (II)

- All data variables should be under root. Intermediate/ancillary data should be under subgroups
- Global attributes (CF compliant and/or ESIP recommendations)
  - General file information
  - PI and data submitter information
  - Project and platform information
  - Version control information
  - Information helpful to others to find your data as necessary
  - Instrument/measurement specific information as necessary
- Example: Use of CF compliant coordinate and time system
  - Time: use "time" for short name and make "time" a dimension scale

    time: long_name = "mid (or start, or stop) of the interval"
    time: units = "seconds since YYYY-MM-DD 00:00:00"
    Use "time_bnds" is start and stop times are needed to define the interval and add "bounds" attribute
  - Latitude: use "lat" for short name, attached to "time"

    lat: long_name =
    lat: units = "degrees_north"
  - Longitude: use "lon" for short name, attached to "time"

    lon: long_name =
    Lon: units = "degrees_east"

DISTRIBUTION STATEMENT A: Approved for public release. Distribution unlimited.

Gao Chen, NASA LaRC (gao.chen@nasa.gov)

MACIE HDF and NetCDF File Requirements (III)

- Data variables (basic attributes)
  - 1-D datasets: x(time = n)

    x: long_name =
    x: units = (http://codes.wmo.int/wmdr/unit)
    x: missing_FillValues =
    x: MACIE_standard _name =
    x: uncertainty = uncertainty or uncertainty variable name
    x: DectionLimit andDectionLimit_flag as necessary

  - 2-D datasets: x(time = n, z = m), e.g., z is scale of vertical profile

    x: long_name =
    x: units = (http://codes.wmo.int/wmdr/unit)
    x: missing_FillValues =
    x: MACIE_standard _name =
    x: uncertainty = uncertainty or uncertainty variable name
    x: DectionLimit andDectionLimit_flag as necessary

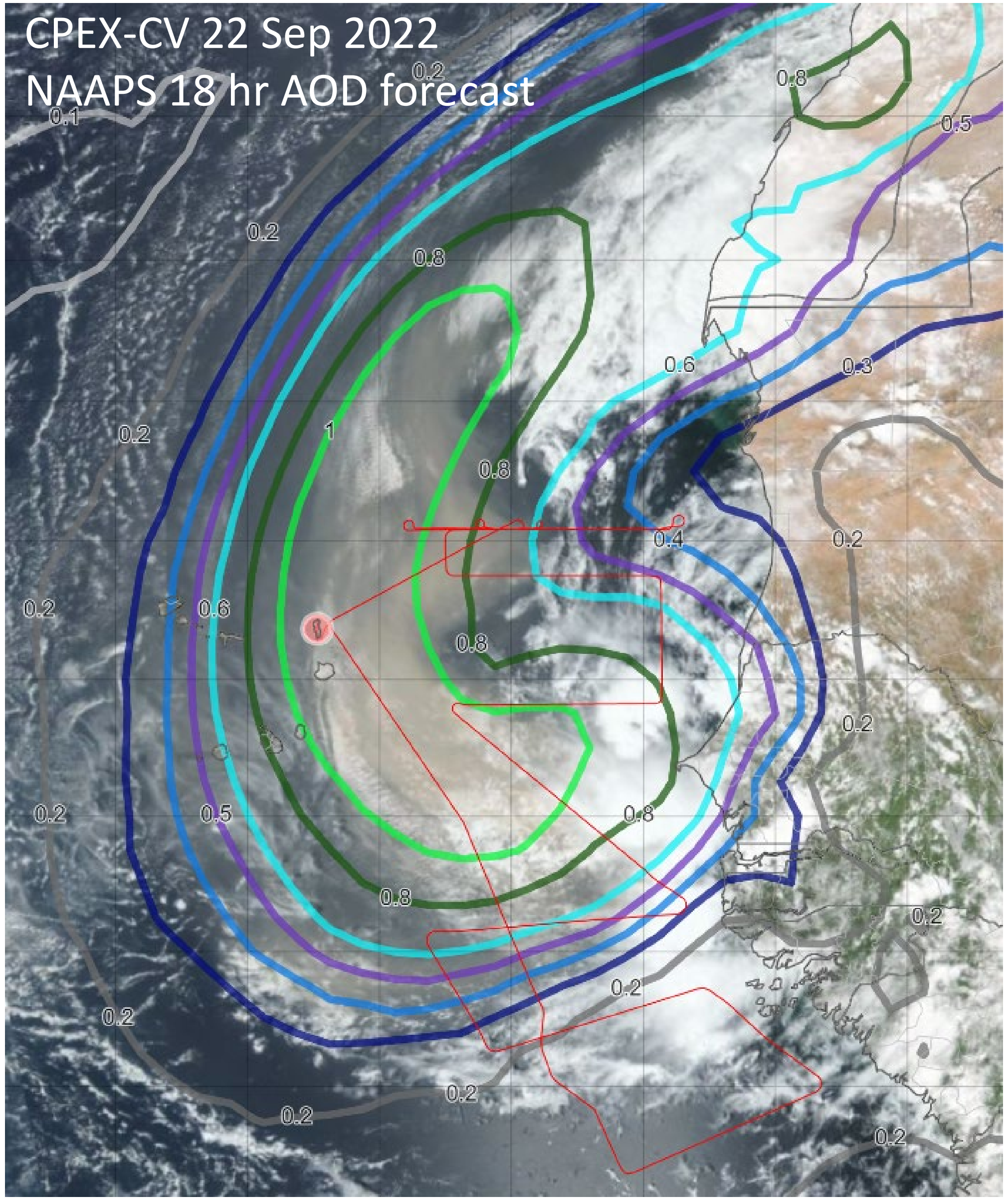
  - Flag(time = n) (data quality, measurement mode, or sampling flags)
    - x: long_name =
    - missing_FillValues =
    - flag_values or flag_masks =
    - flag_meanings =

DISTRIBUTION STATEMENT A: Approved for public release. Distribution unlimited.

Gao Chen, NASA LaRC (gao.chen@nasa.gov)

Keeping track of measurements

- Aircraft measurements are nearly always aliased in space and time. Thus, work needs to be done to address scale mismatches.
- Observation and model variables for particles often exist in different parameter spaces. Thus, there is near universal agreement on the need for instrument simulators.
- Once you get into particle properties and add an inlet, almost all bets are off as to what you are *really* measuring.
- If the field obs indicate poor model performance, is there enough data to determine why?
- All of this said, the size of the problem depends on the magnitude of error in the model and its application.



DISTRIBUTION STATEMENT A: Approved for public release. Distribution unlimited.

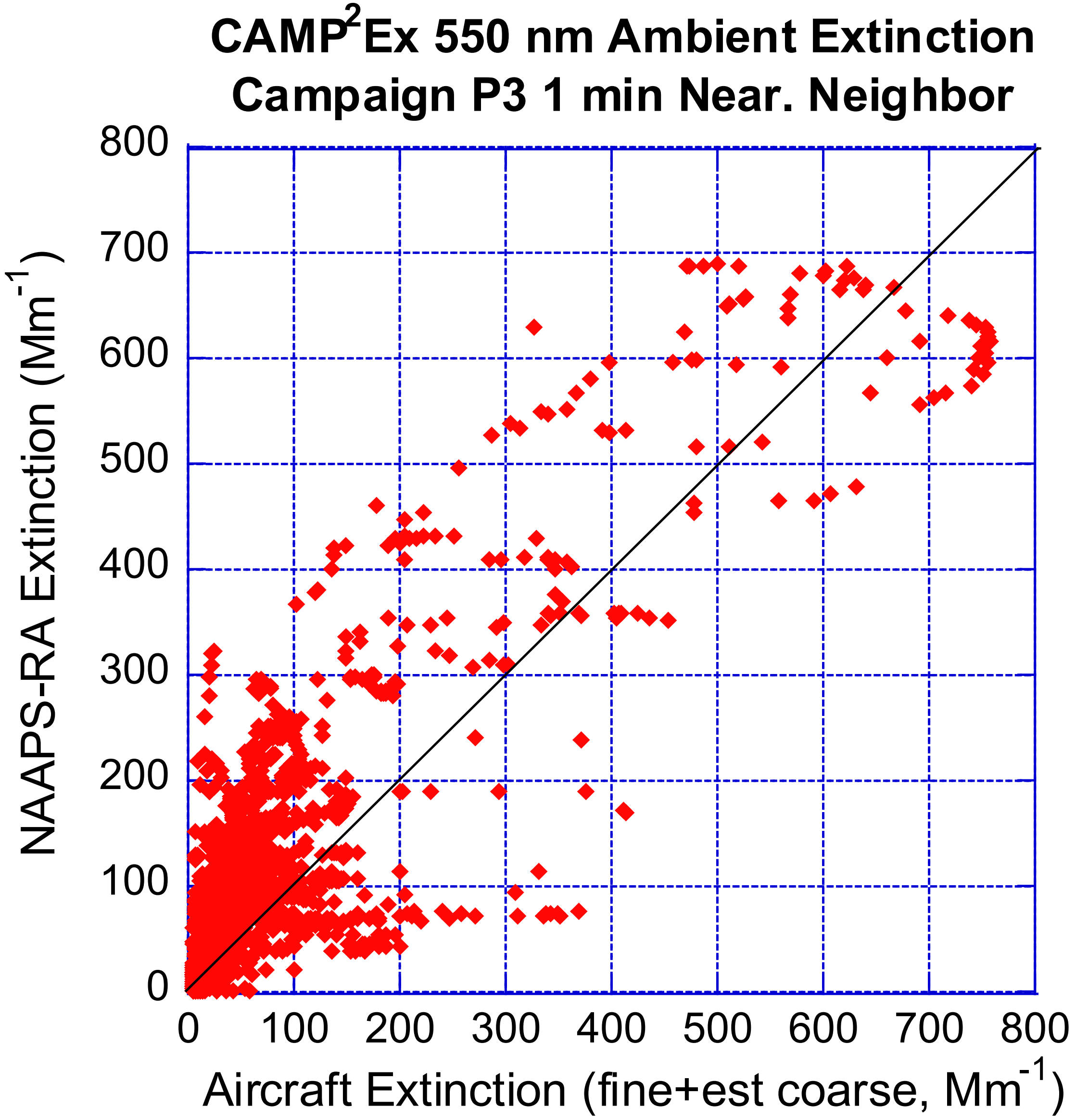
Data Management

•We are starting with pathfinder data: CAMP²Ex (WESPAC), KORUS-AQ (Korea), SEAC4RS (CONUS), and now CPEX-CV (Subtropical Atlantic). Why? Gao makes our life easy.

•CIRPAS: Code was written by U Arizona for the reprocessing of the decadal CIRPAS/Sorooshian dataset into ASCII.

•Data system such as Unidata THREDDS data server can be used to collocate model data. Python for slice/dice/online computations

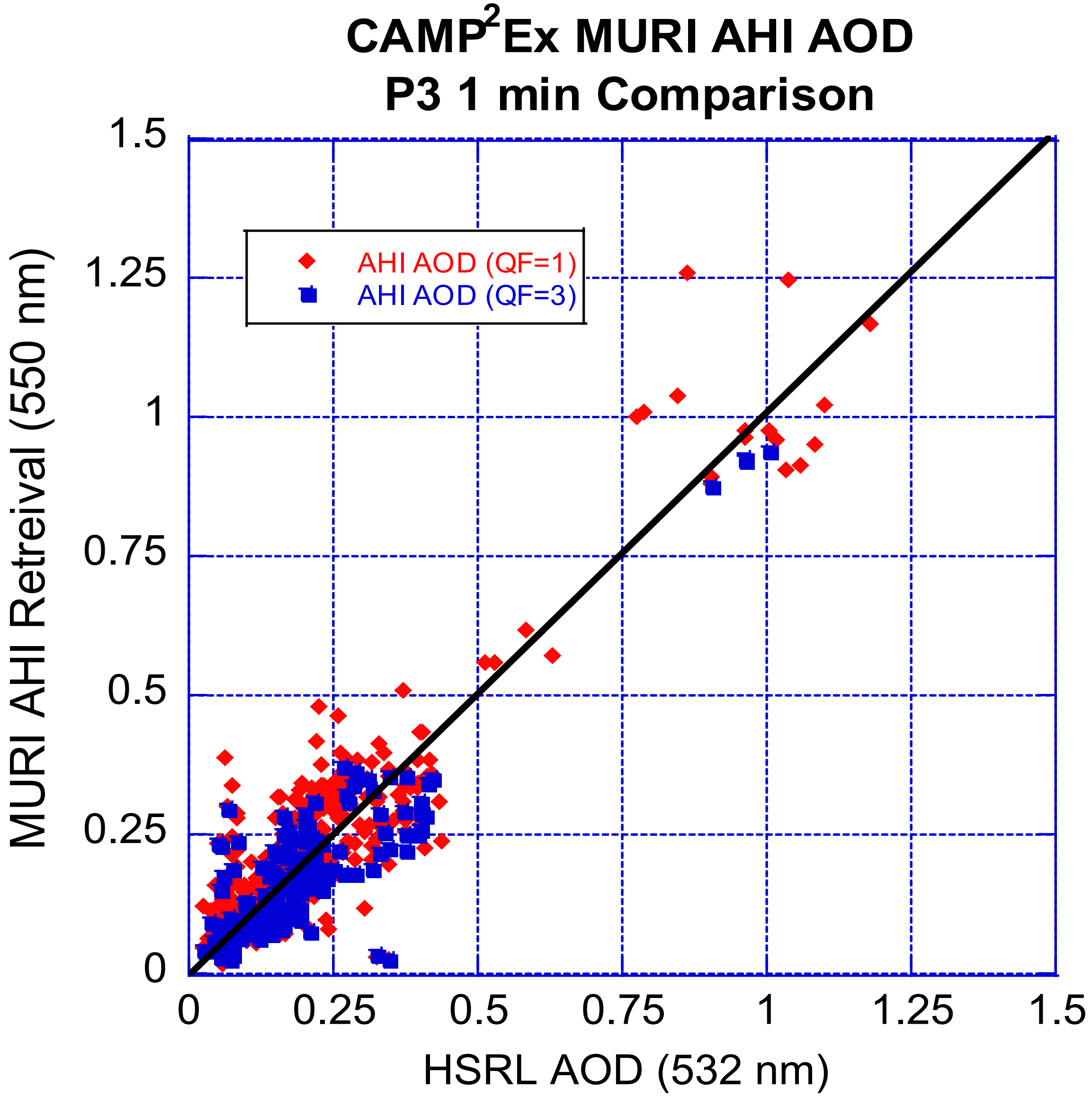
•We can now quickly collocate obs and models. So far have NAVGEM, NAAPS-RA & working on ENAAPS. Next on the list is COAMPS.



DISTRIBUTION STATEMENT A: Approved for public release. Distribution unlimited.

Add SSEC Remote Sensing: A mix of observations and model

- Not only do we want to assign model state vectors to obs, but also incorporate surface, air and satellite remote sensing products to allow for a coupled analysis.
- SSEC has developed remote sensing projection and regridding software to line up satellite obs to the aircraft. Transitioning to interface with GeoIPS.
- Starting with aerosol and cloud and moving to radiances.
- There is the possibility to use THREDDS to help integrate satellite data.



DISTRIBUTION STATEMENT A: Approved for public release. Distribution unlimited.

Quality Assurance : Just because it is measured does not mean that it is right, or even just right enough

- We need to face facts that data are of varying quality. E.g., “Facility parameters” are often reported, but not always well characterized.
- Need to spot data anomalies quickly. Derived quantities across parameters such as mass scattering and CCN efficiencies are canaries in the coalmine.
- Lots of aerosol analysis tools out there, but there is diffuse distribution
- Some big challenges that need community effort:
  - Lack of “real mass” values on aircraft
  - Organics
  - Coarse mode characterization
  - Relation to CCN and hygroscopicity
  - Best way to deal with “size?”

DISTRIBUTION STATEMENT A: Approved for public release.   Distribution unlimited.

We may need to revisit "closure"?

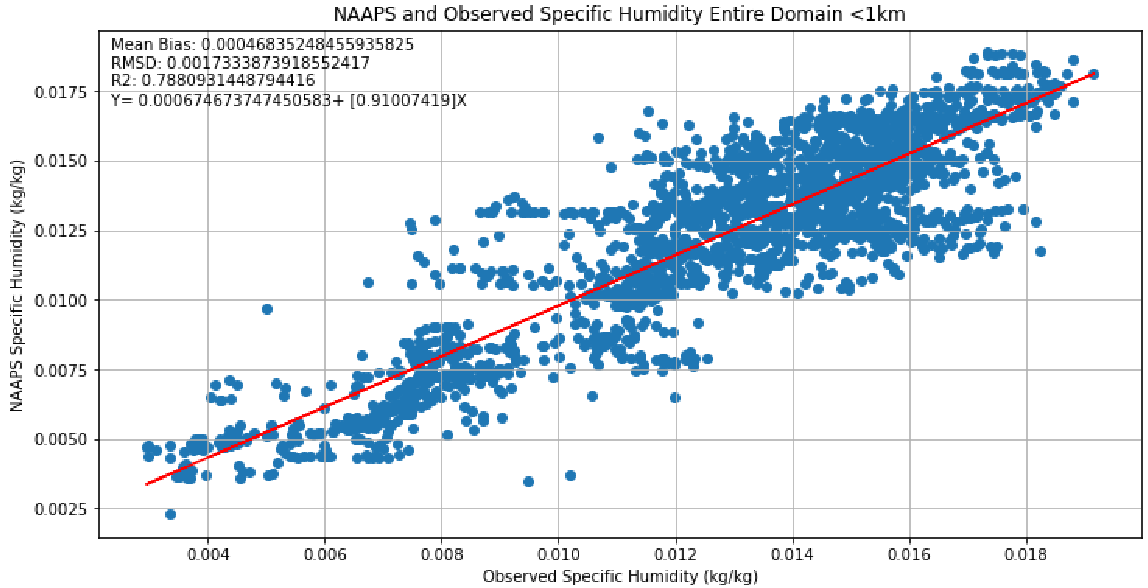
If you want all of those things from satellites and models, we need to know how well we can even measure it in the first place.

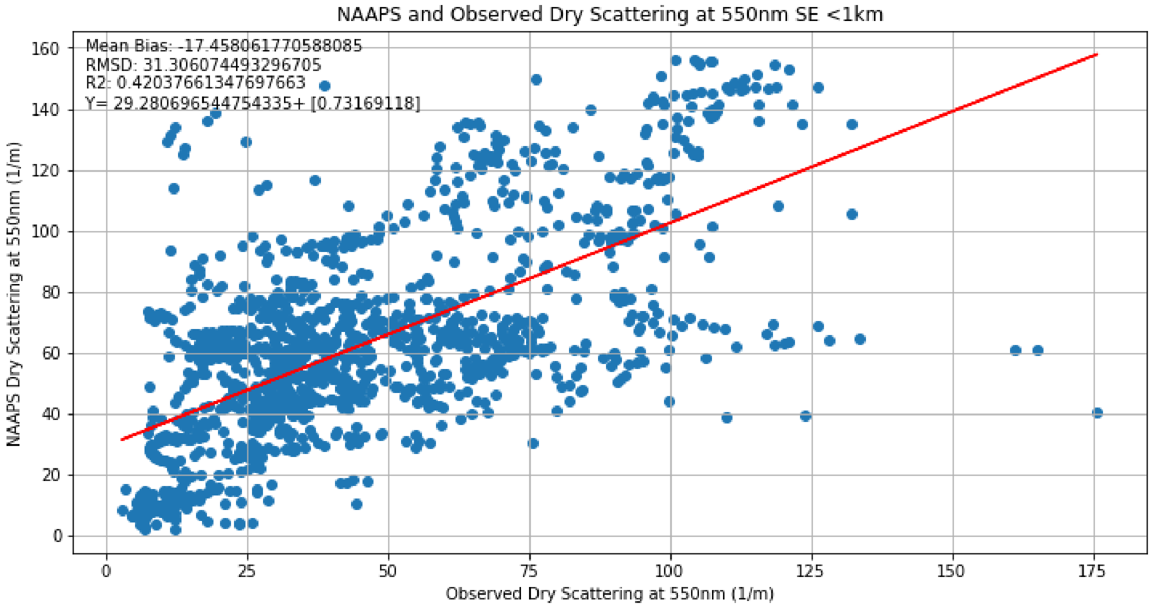
- "Closure" is the accounting and reconciliation of microphysical forward models to the observable.
- Observed vs. calculated scattering-absorption-extinction-AOD, from size, chemistry, density, mixing state, morphology, refractive index, hygroscopicity. Lots of degrees of freedom and assumptions…
- Density, morphology, & ref. index are often "quasi-free" parameters, constrained by closure residuals.
- Internal closure is, "local" typically sharing an inlet; External closure is to the column with sun photometer, satellite, and now HSRL + multi-angle polarimetry.
- Science need is driving closure requirements to include more parameters and shorter time/spatial scales.

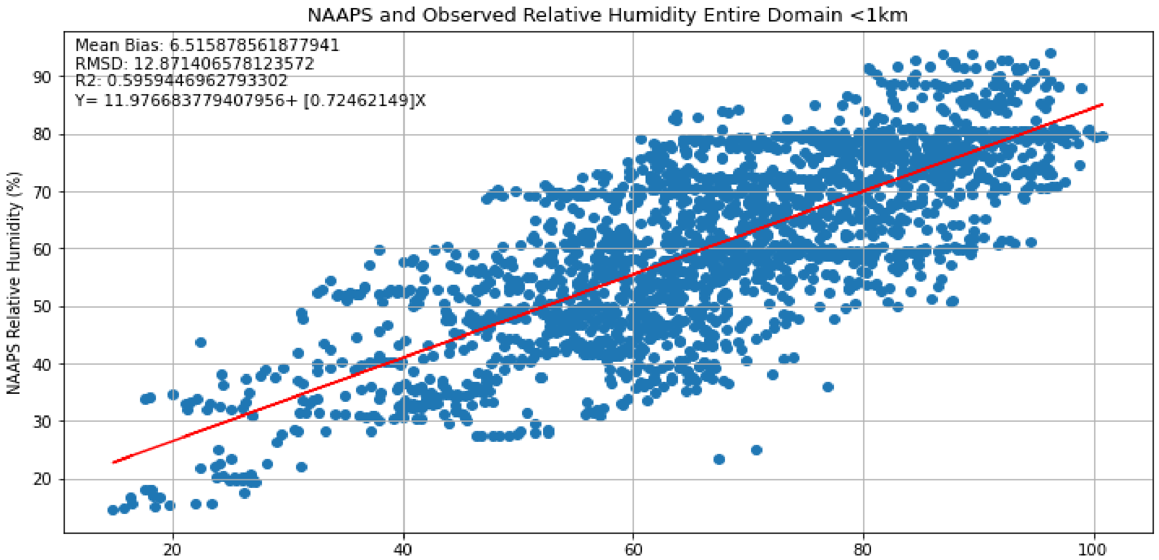
The trick is there is no perfect measurement of truth, only projections.

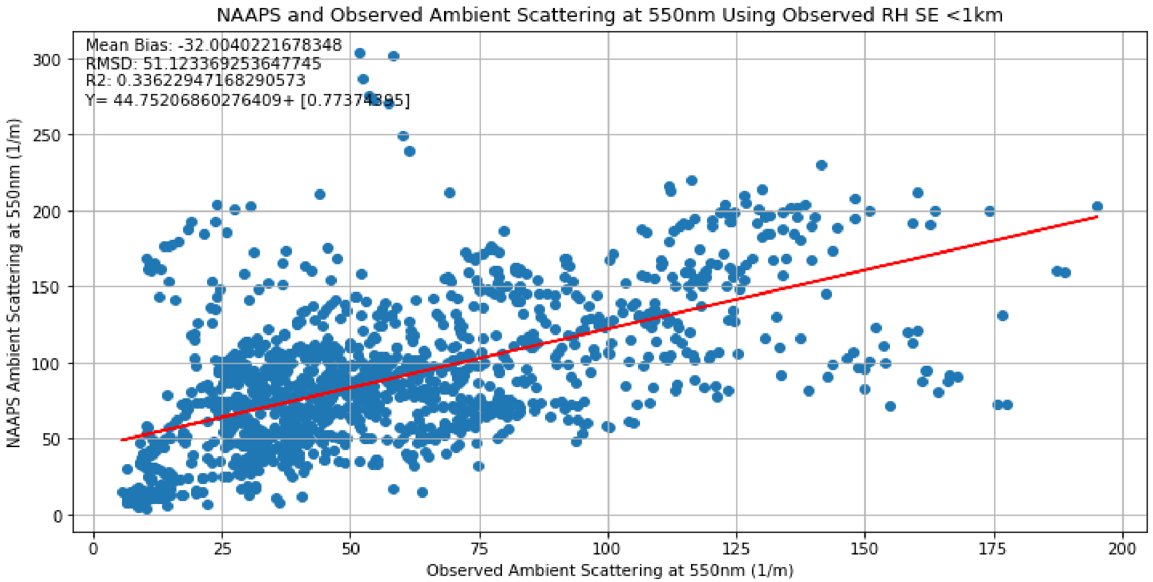
DISTRIBUTION STATEMENT A: Approved for public release. Distribution unlimited.

With care, we are now getting to a point where we can vectorize *local* error: SEAC4RS example









DISTRIBUTION STATEMENT A: Approved for public release. Distribution unlimited.

In closing:
Like the ICAP Consensus, many is almost as easy as one

- If the ICAP Consensus has taught us anything, once you get the data into the same format and matrix, the rest is easy.
- One of the hardest nuts for us to crack is scale and the implication of differences in meteorology, especially RH and hygroscopicity.
- We can generate navigation track files across many missions, others can add their own state vectors for a combined analysis.
- The real heavy lift is to go from "local error" to global error, including sources, vertical distribution, etc.   This will require complex adjoint and ensemble sensitivity analyses.
- Keep in mind little things do make a difference, like nearest neighbor versus interpolated values.

DISTRIBUTION STATEMENT A: Approved for public release.   Distribution unlimited.