Aerosol Verification for Global Systems: Where are we really?

J.S. Reid[1], J. Rubin[2], P. Lynch[3], J. Zhang[4], E. Hyer[1]. W. J Campbell[1] , J. Hansen[1], C. Sampson[1], W. Sessions[3], Y. Shi[4], D. L. Westphal[1] w/liberal use of the GEWEX aerosol report draft and products from people in this room.

[1]Naval Research Laboratory Monterey, CA
[2]NRC Fellow, Monterey CA
[3]CSC, Monterey CA
[4]University of North Dakota, Grand Forks, ND

October, 2014

http://www.nrlmry.navy.mil/aerosol/

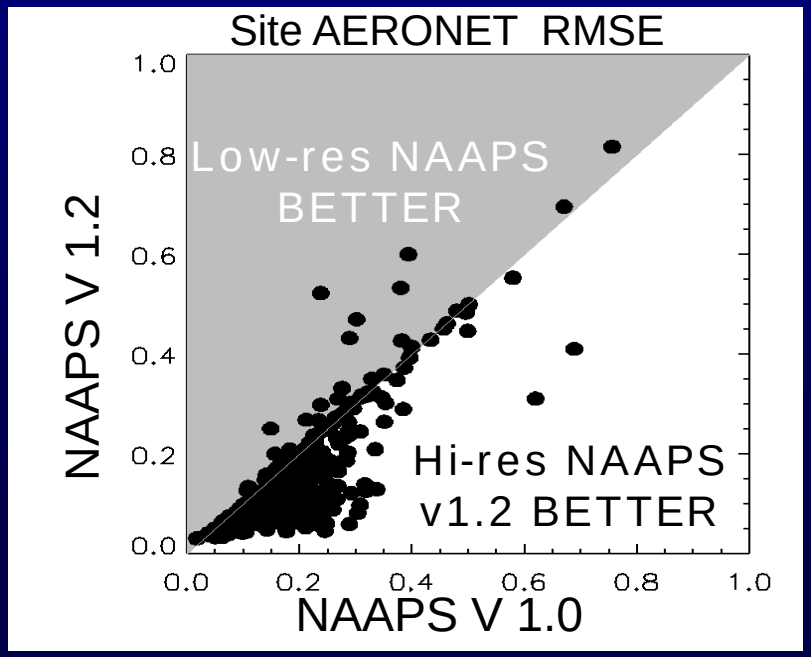
The Good…..

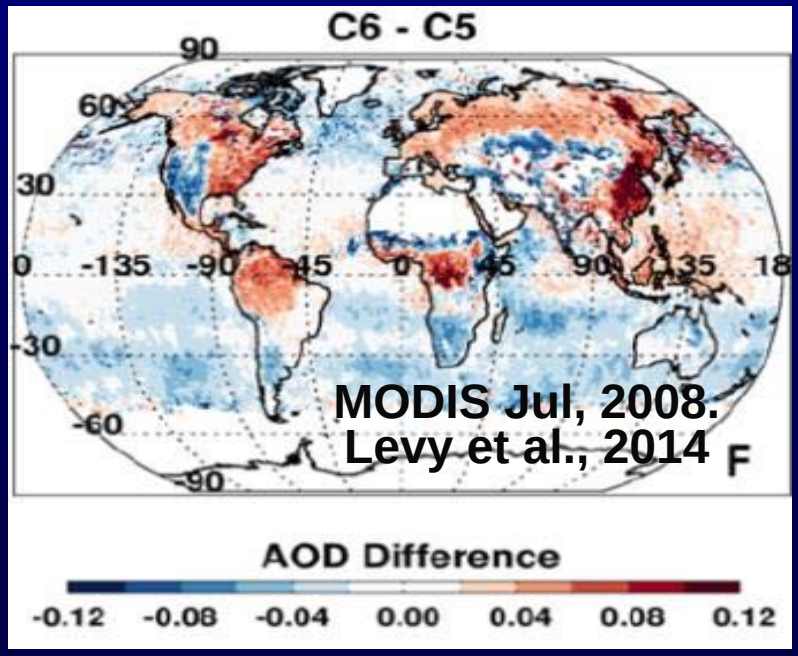
- Terra and the A-train kicked in the door for global aerosol observations giving a good qualitative view of the world.
- From a modeling point of view, AEROCOM and ICAP have been able to consistently evaluate models, resulting in the development towards tractable baselines of performance.
- Verification of remote sensing products is clearly getting traction at programmatic levels, and is trickling down to science teams. There are several external verification groups now out there (GEWEX, AEROSTAT, AEROSOL CCI).
- There are more model, satellite and verification products available, and data assimilation technology is ever improving.

More good!

Things are getting better, and uncertainties are better understood if not being reduced (I think).

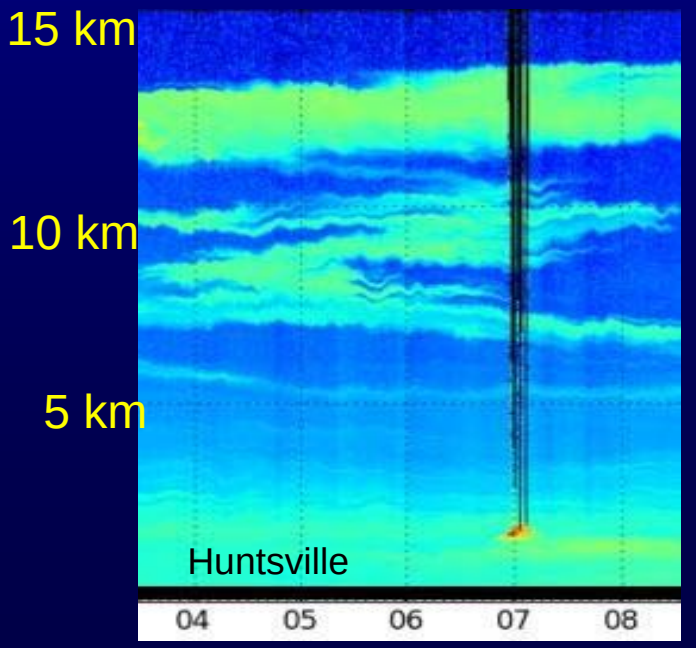
- Model biases are being incrementally improved, based on self reporting. But, with ICAP MME we can now look at a consistent time element.
- Satellite products for data assimilation are also incrementally improving, and data is *much* easier to get.
- Verification datasets are becoming more available, and there is a recognition that “observations” have their own set of errors.
- Everything points to rapid progress in the next 5 year. Lidar will be key.





The Bad

- Scientists organize as well as cats. Thus community wide verification is largely regional and inconsistent.
- It is difficult for external verification to directly impact development.
- We need way more verification data (particularly PM)! But it is not just volume, but consistently collected, formatted and characterized. We also need the tools in an NWP framework to crunch it.
- Verification ob errors are typically ignored.
- From AOT to the vertical: Surface data is in needed, as we know such data has a short correlation lengths. What do we do about the middle?



The Ugly

- Verification also has many rationales, many fundamentally economic. Thus there is an element of subjectivity. Framing bias is a concern.
- Even the most basic metrics are not reported consistently between groups. Beware of baseline fudging….
- Product expectations are beyond what can be delivered.
- And now, with all of the products that are out there (good), we need to reconcile clear differences in the most promising (bad). Nobody want to pay for that (ugly).

We as customers, are in a position to make a difference here. But we need to be consistent ourselves.

So what do we want from verification?

From Anna Ghelli 2010: Consistency, value, and common methods of formulations.

Additional emphasis from CSIRO website: Quality, value, accuracy, bias, association, skill, reliability, sharpness, discrimination, uncertainty.

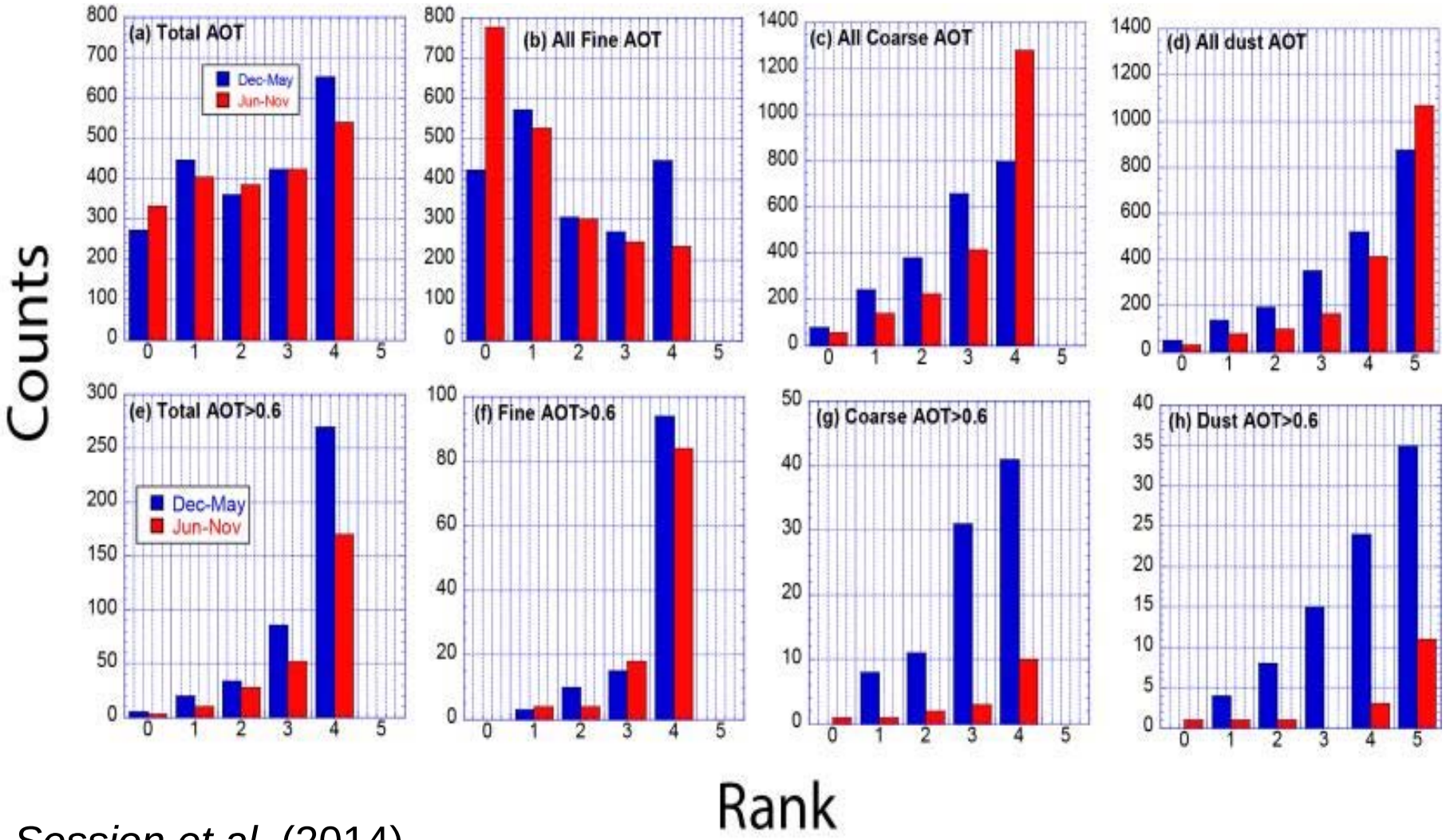
Methods: Visual, continuous, dichotomous(e.g., hit rates), multi-category, probabilistic, spatial, ensemble, etc…

We need:

a) Actionable information from point-wise error models accounting for their own intrinsic error.

b) Recognition that our verification data has uncertainty and representativeness issues.

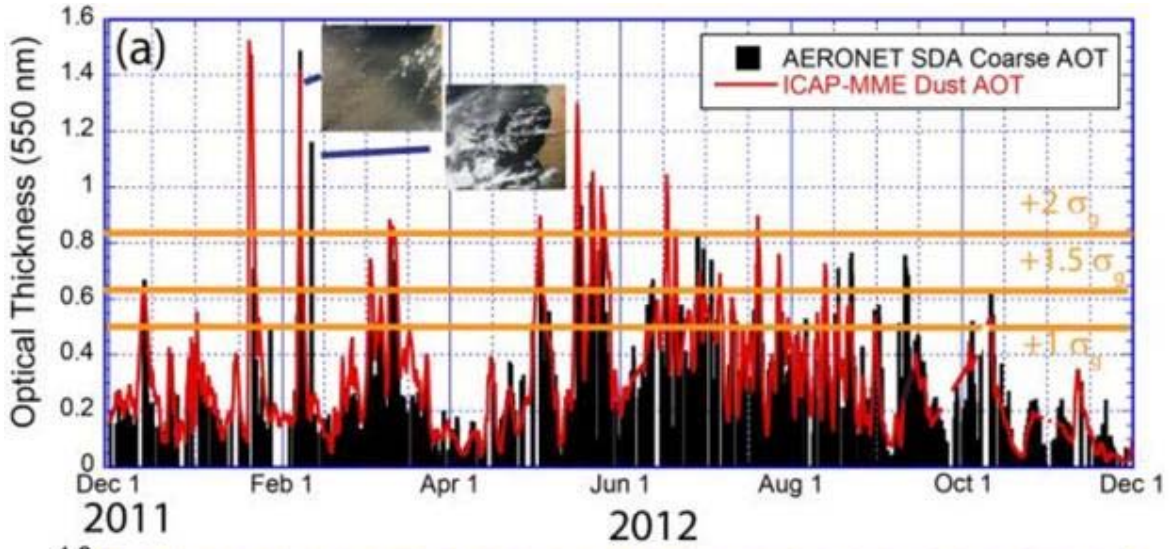
Part 1: Let's tell as story
Under predicting severe events
A common problem

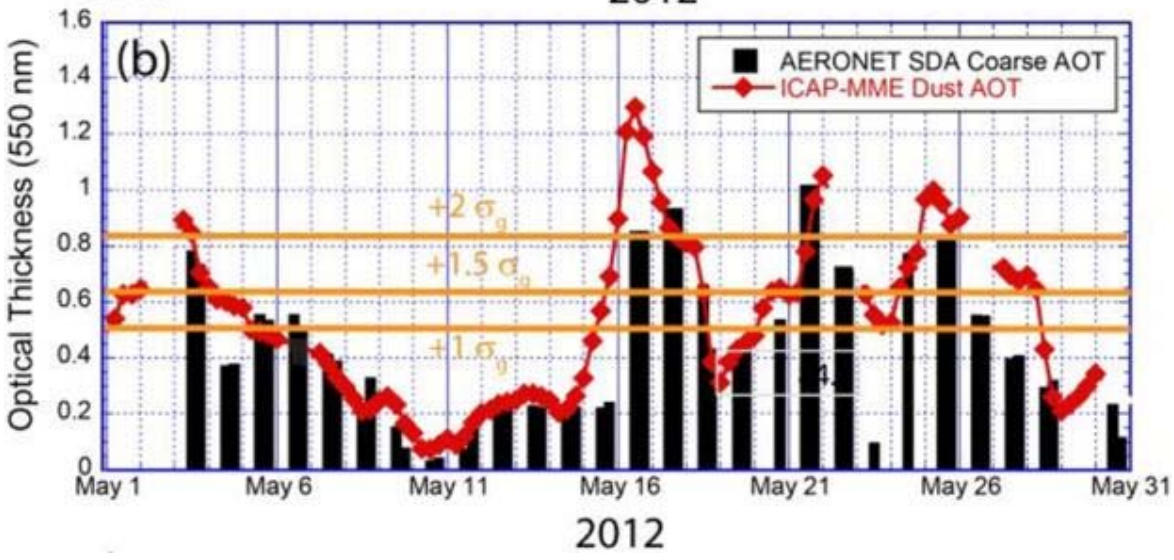


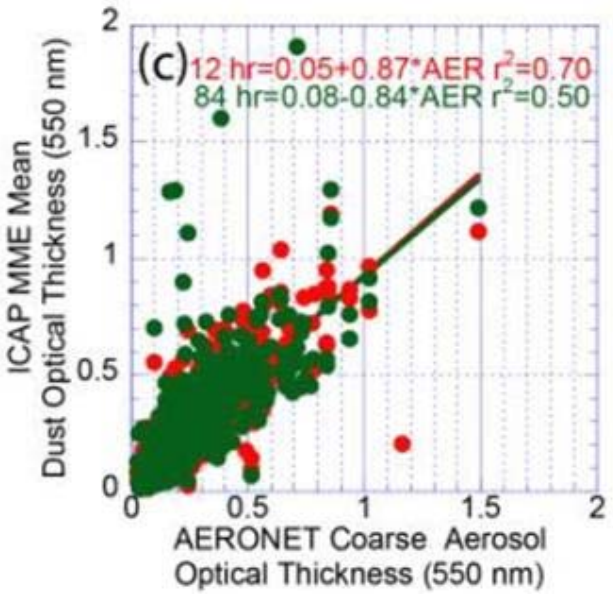
Session et al. (2014)

An easy case
Good News: Cape Verde

- Cape Verde is the benchmark site for dust.
- All model do well in analysis, and pretty good 4 days out.
- It is a great dataset to play with if you want to figure out how to score model and obs alike.



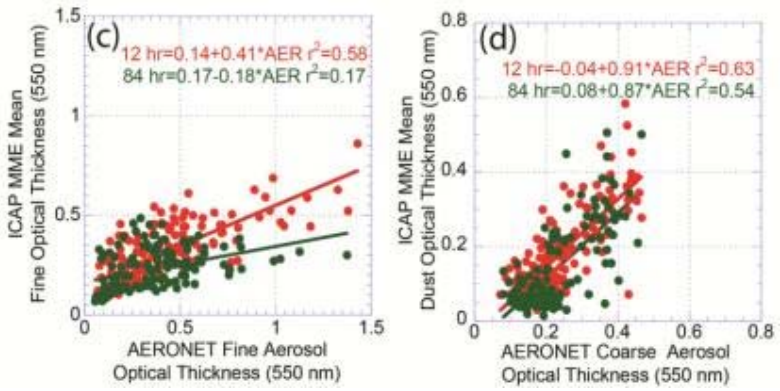


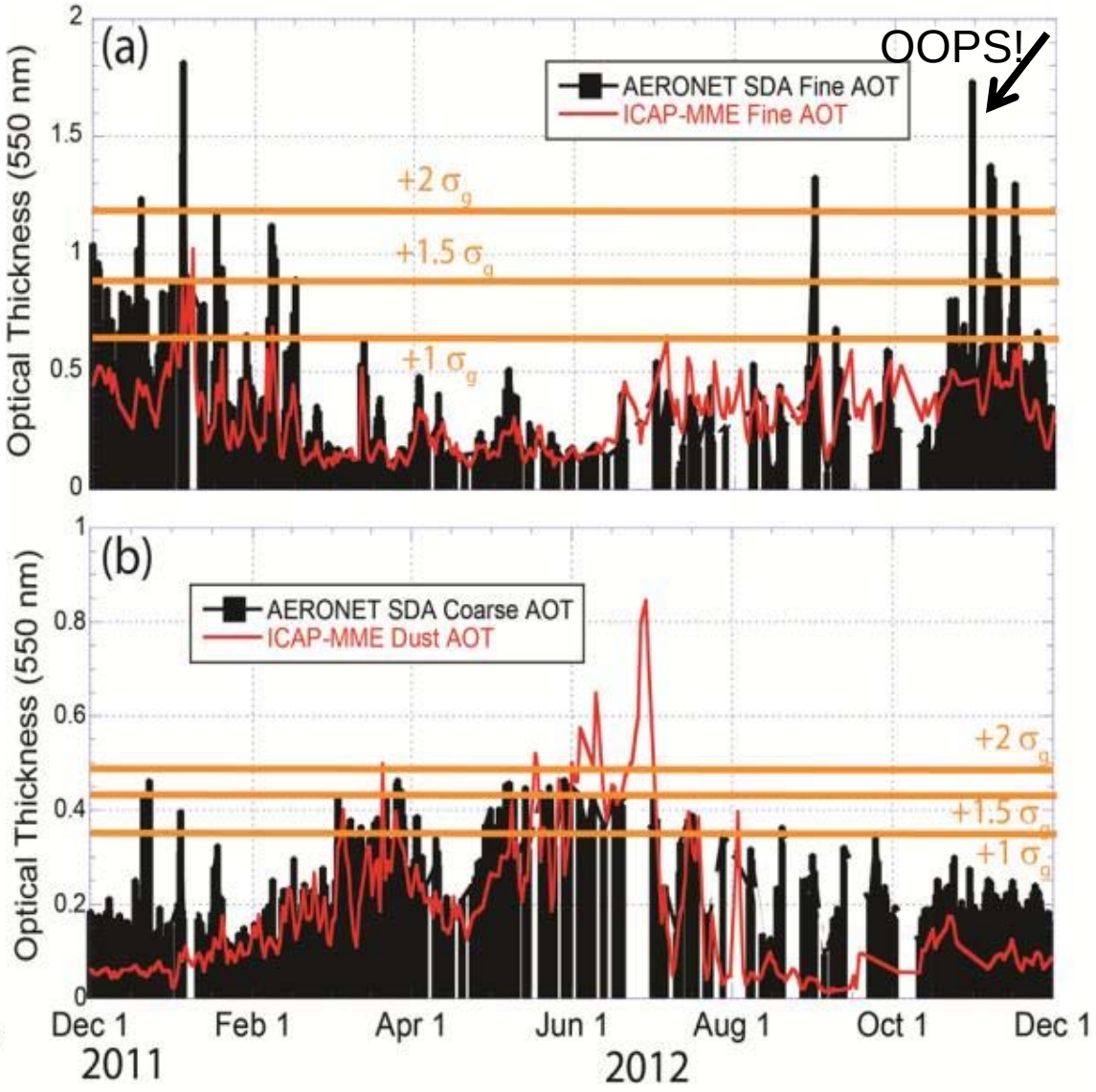


Not so easy-Kanpur: RMSE 50%

Has more AERONET data so lets have a look

- At first glance, Kanpur is not as bad as stats may suggest. But everything compounds to a bad score.
- But, embedded are clear bias and event errors.
- Forecasting of the winter is downright grim, likely due to boundary layer effects and stratocu.

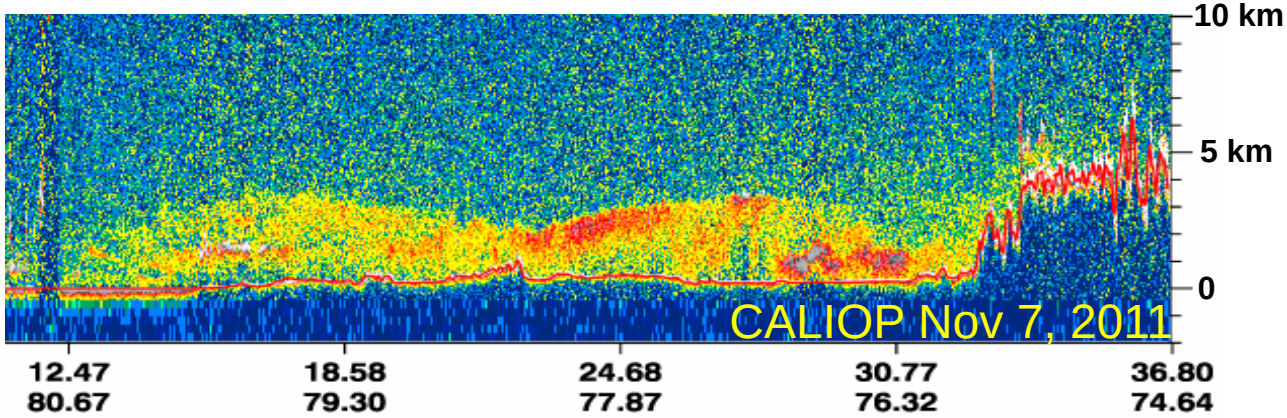


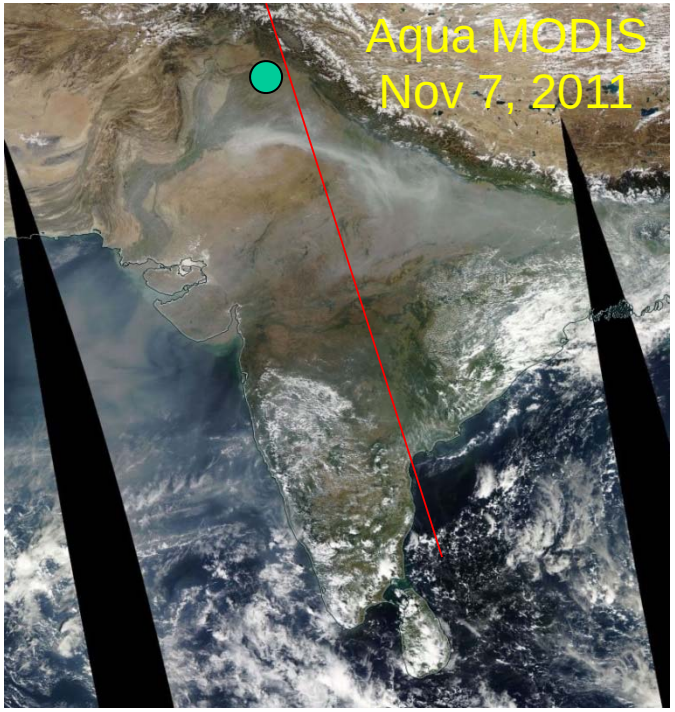


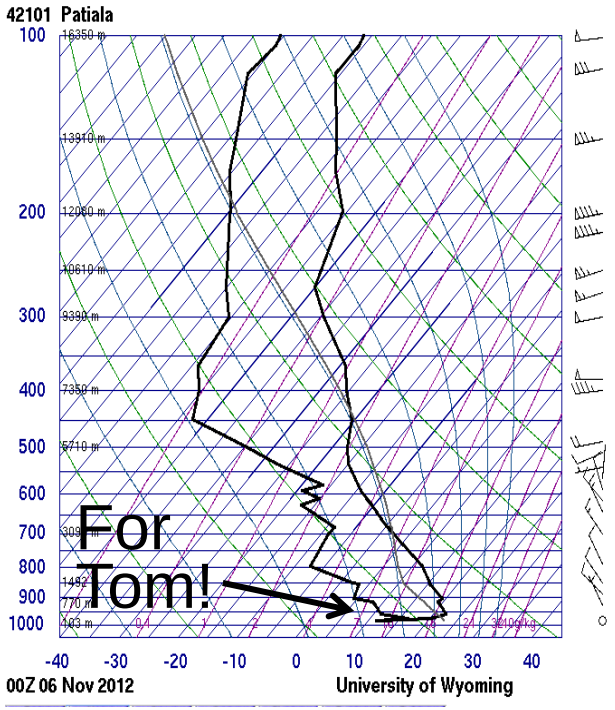
What is happening in November?

PBL dynamics lead to intense surface events, with winter haze, fog and stratocumulus lead to high hygroscopicity and secondary particle production.

So how do we parse out error?

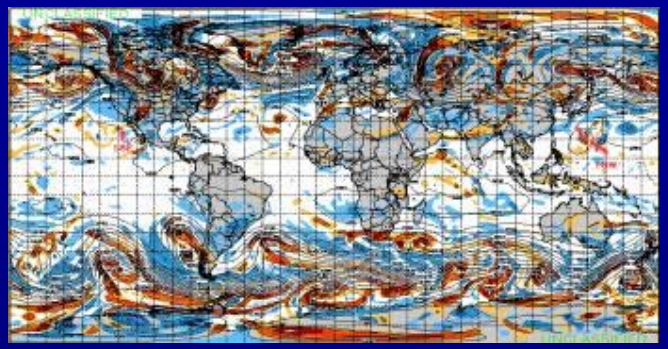


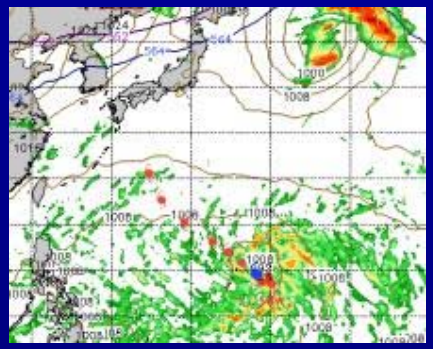




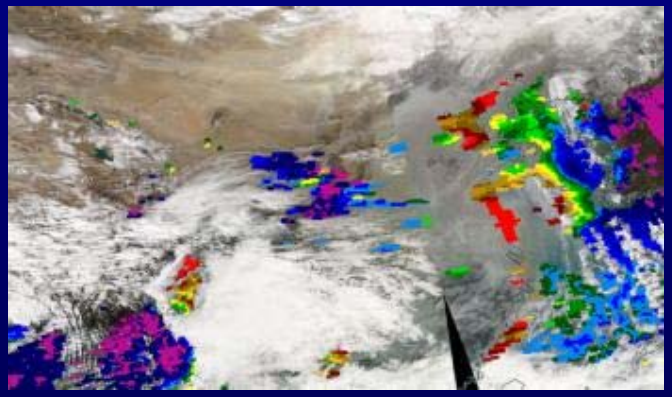
So what went wrong? Some broad components of verification we need to keep track of

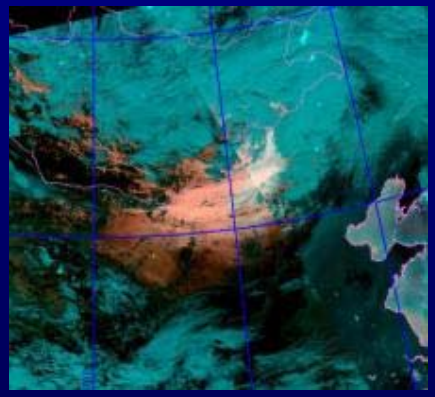
Meteorology
Bias, scale, busts…



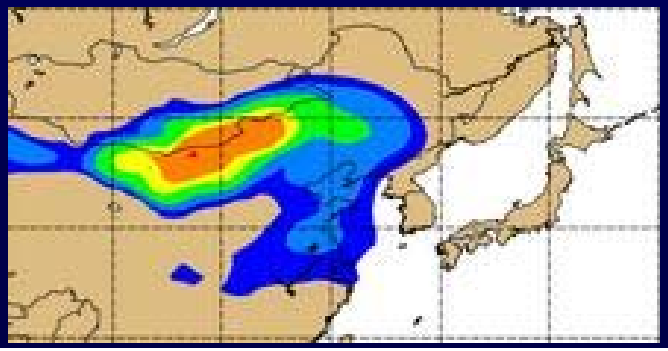


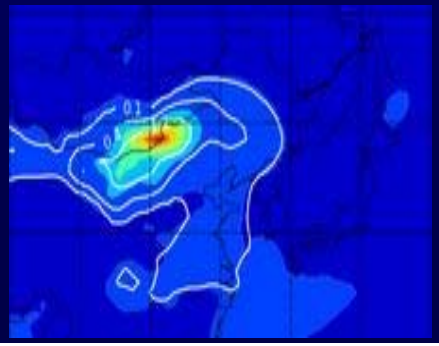
Aerosol input data
Initialization, data assimilation, masks & verification data too.





Aerosol model
Species, source/sink, microphysics, transformation, advection





#1 Meteorology Parameters of Concern
It all matters, but some more than others for global aerosol prediction

Current BIG three

- Surface wind/friction velocity
- Relative humidity
- Precipitation (esp convective)

In the future, the next BIG three:

- Boundary layer winds & height
- Energy/radiation balance
- Ocean wave fields

Example

Problem: What we care about is not always what NWP cares about.

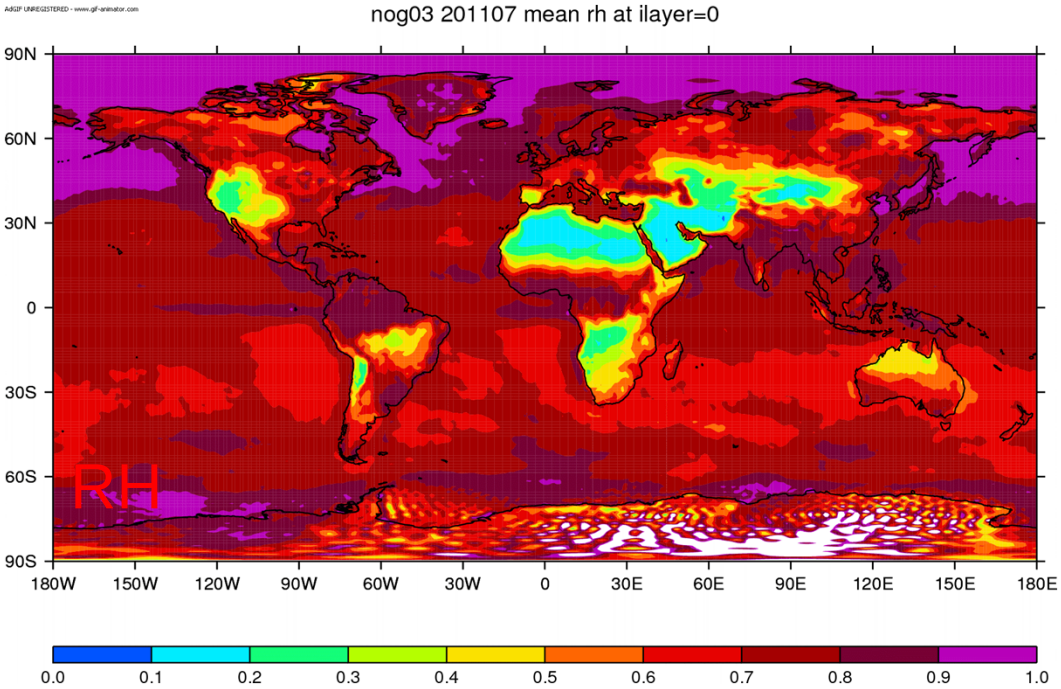
NAVGEM: 500 mb AC, TC track, low level ocean winds.

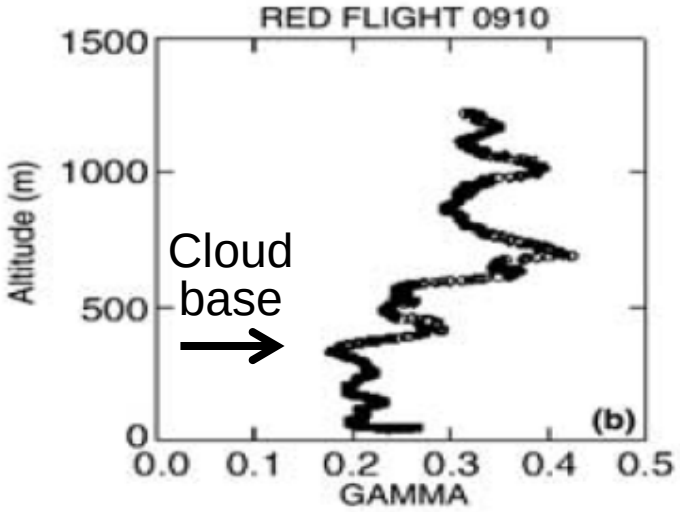
Hygroscopicity Example
Something simple-like the MBL

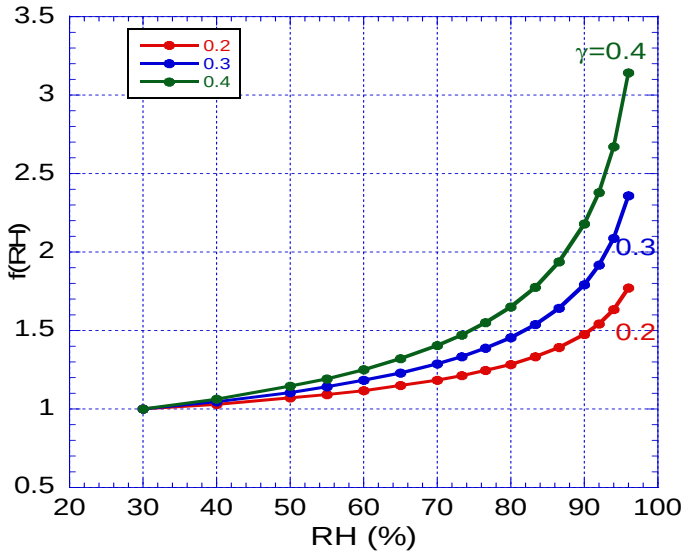
RH bias has been a moving target for us.

But it is not just the RH, we have to consider hygroscopicity.

Hygroscopicity systematically changes within the boundary layer, creating a complex error propagation model.

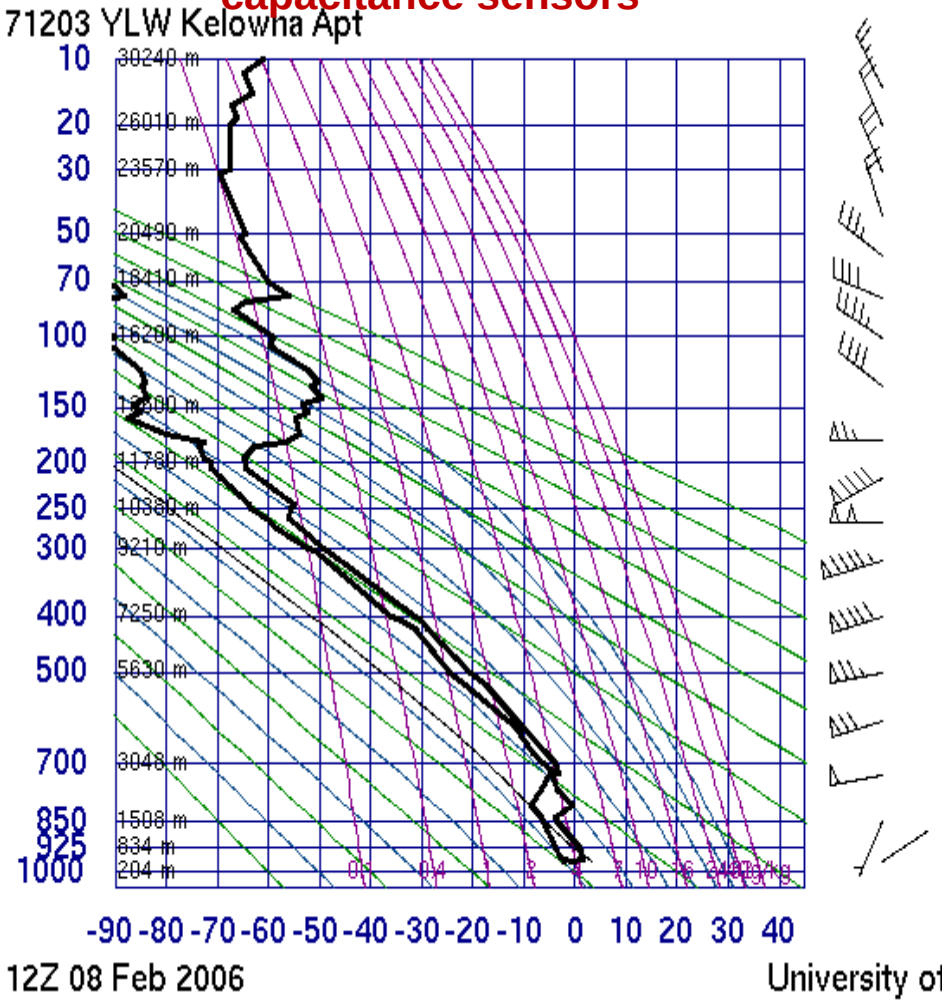




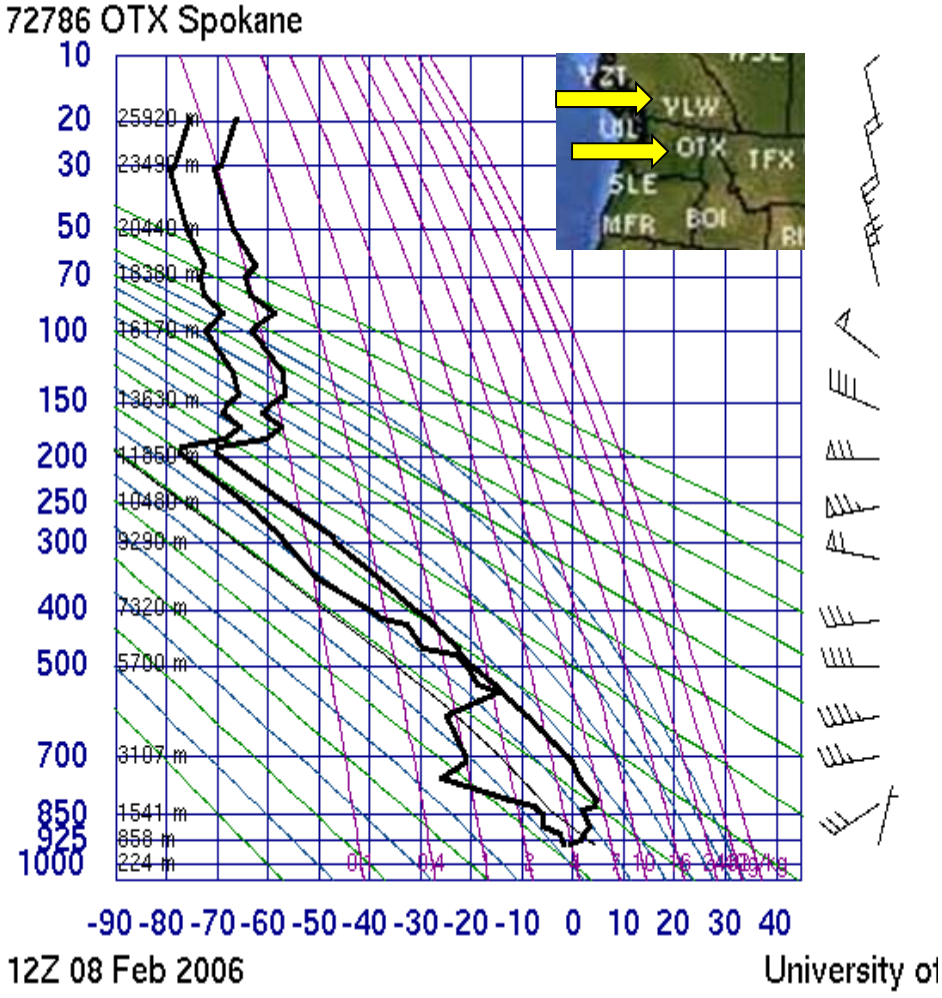


And errors are not just in the model

71203—Vaisala RS90 Twin capacitance sensors



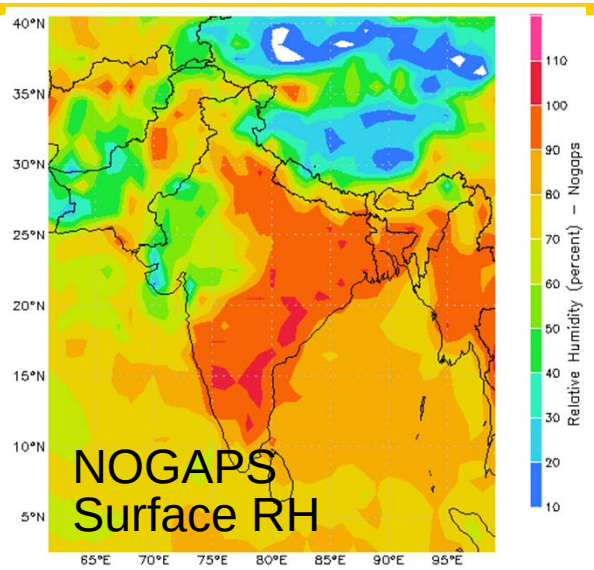
72786—VIZ-B2 Carbon Hygristor sensor

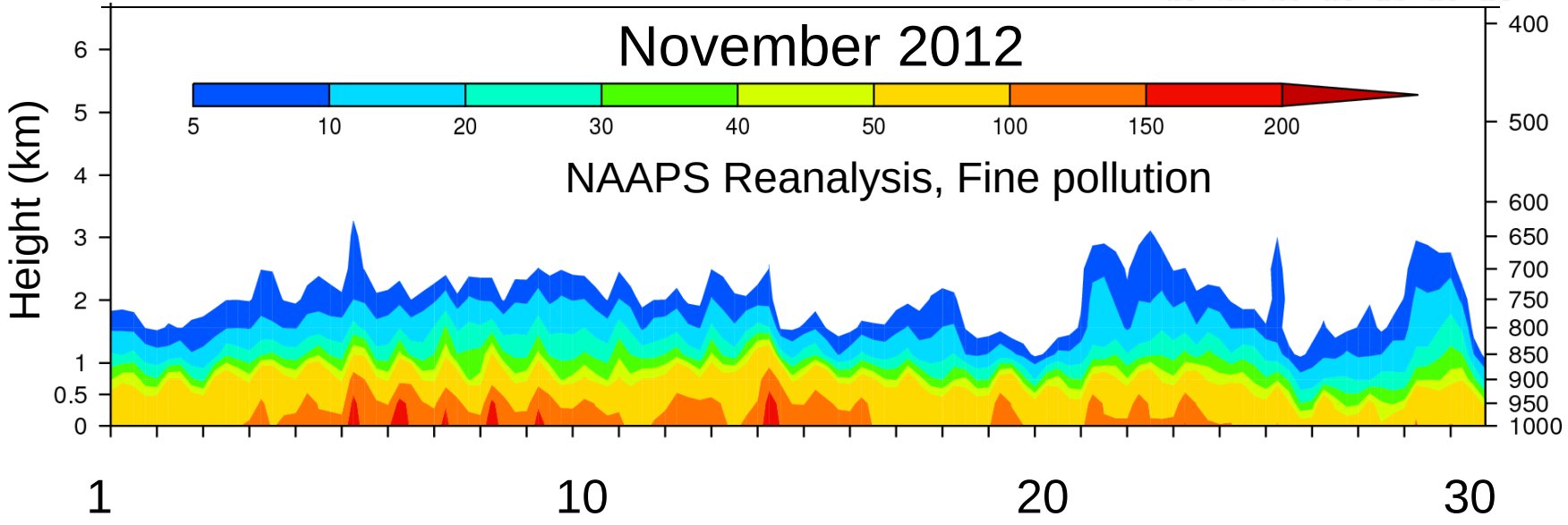


(Courtesy Pat Pauly, NRL)

So how about Kanpur?

- Aerosol particles are in the right place.
- Slight low bias in RH. But what does this do to PBL dynamics and overall AOT?
- Not much QAed weahter data to analyze.

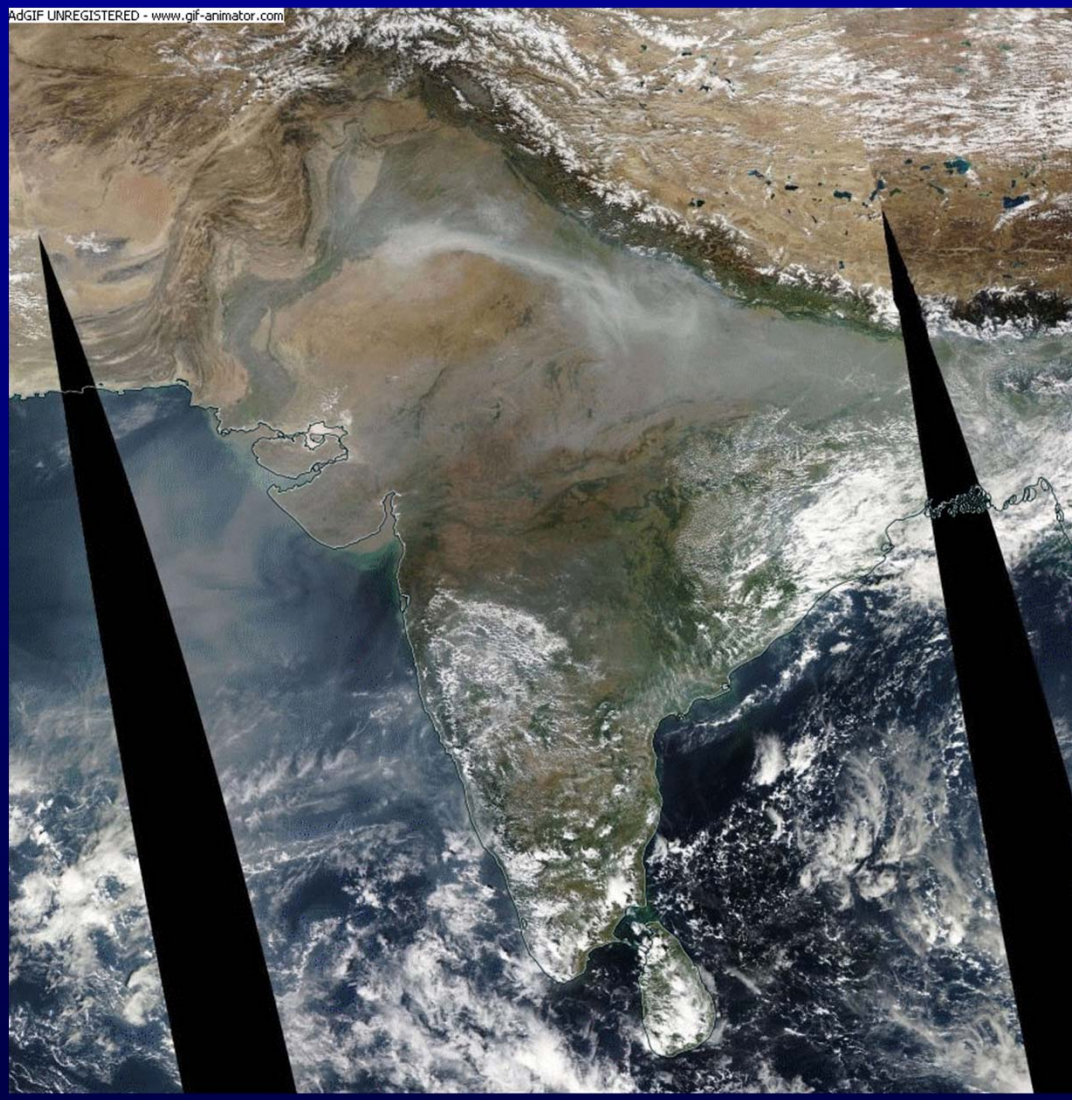




#2 Aerosol Input Data: Why did DA not fix this?

In the age of data assimilation, the models are at the mercy of input data

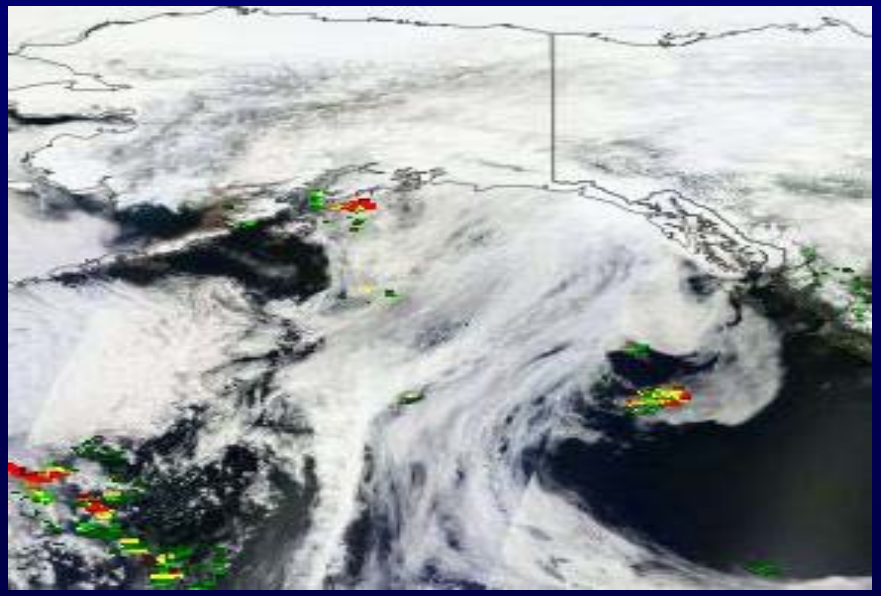
- Have a look! You can find many problems with the C5 MODIS retrieval.
- We often concentrate on AERONET matchups, but this has an inherent sampling bias to the benefit of the retrieval.
- Data availability and location are key. Depending on how the DA system is constructed, adding “good data” to the wrong place can degrade model performance.

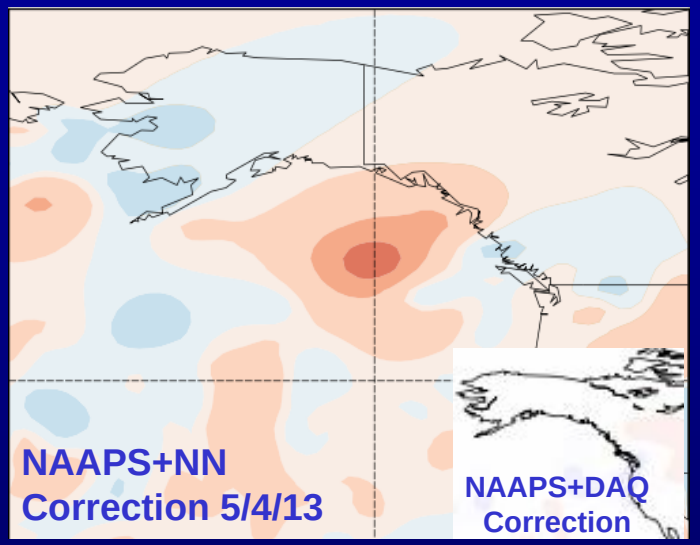


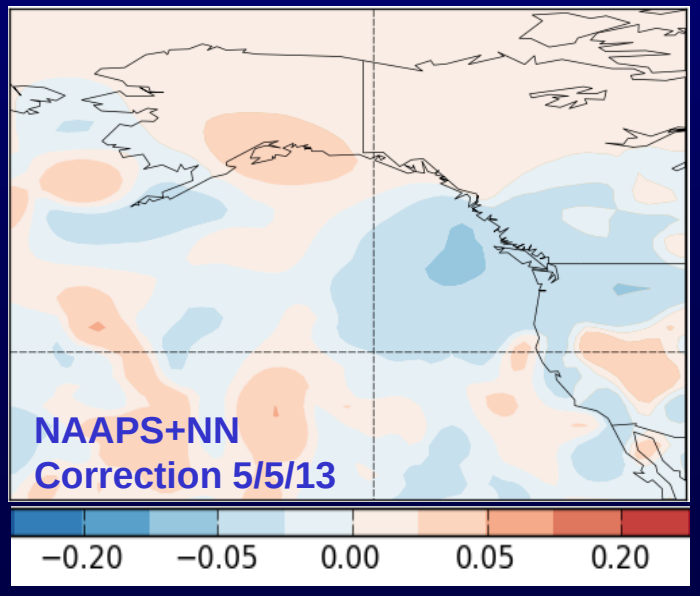
Error in DA Space Example: 5/4/13

Rubin tested the GMAO NN product for potential inclusion in NAAPS. Maybe this will clear the air

- Here is a classic problem, isolated high biased retrievals due to clouds resulting in dipoles.
- These dipoles do not show up so strongly in bulk stats, but make a big difference in local forecasts.
- How to deal with it? GMAO does cloud clearing bias within DA system, NRL does it outside.
- So for the question do we want more coverage with less quality, we say, we want all data with realistic error bars.

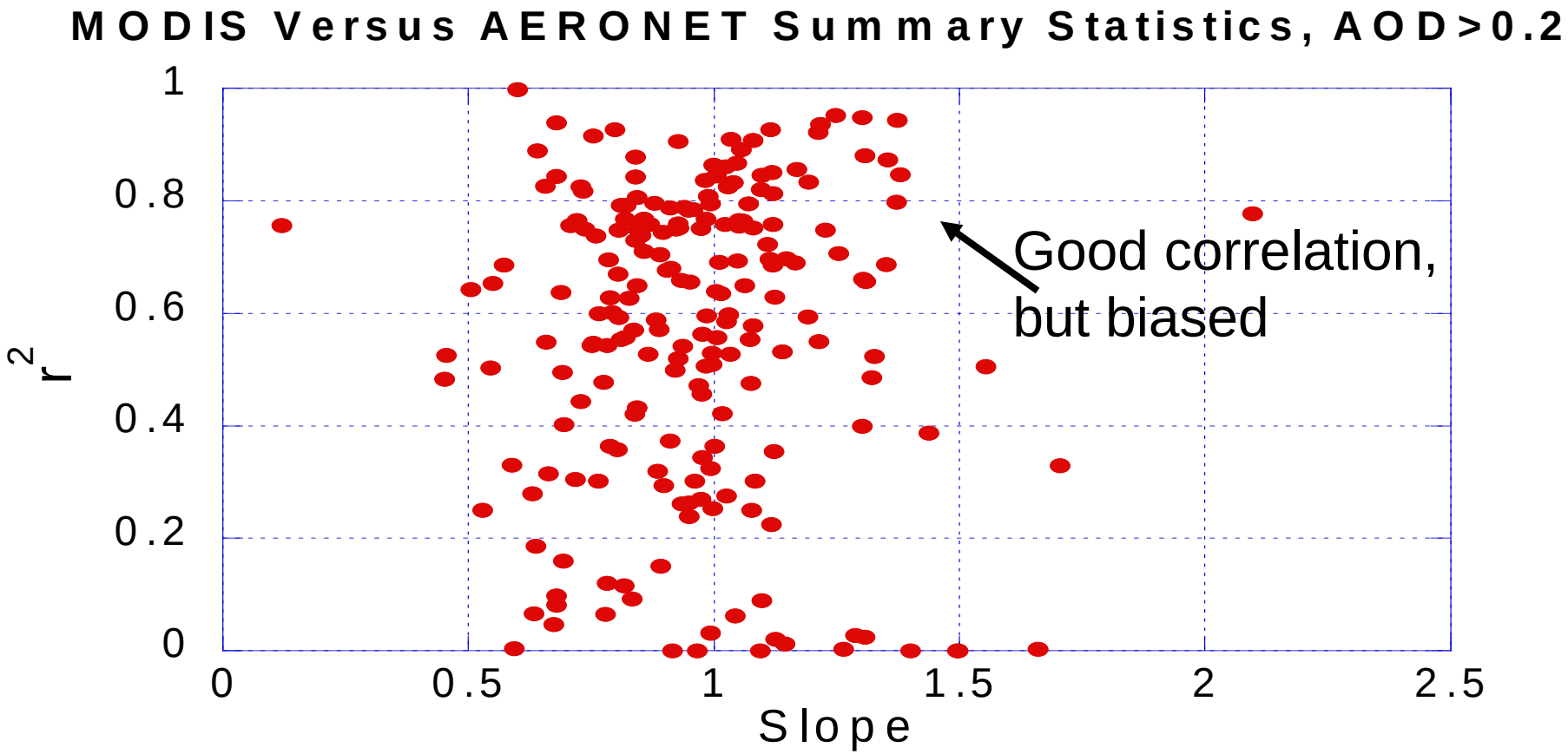






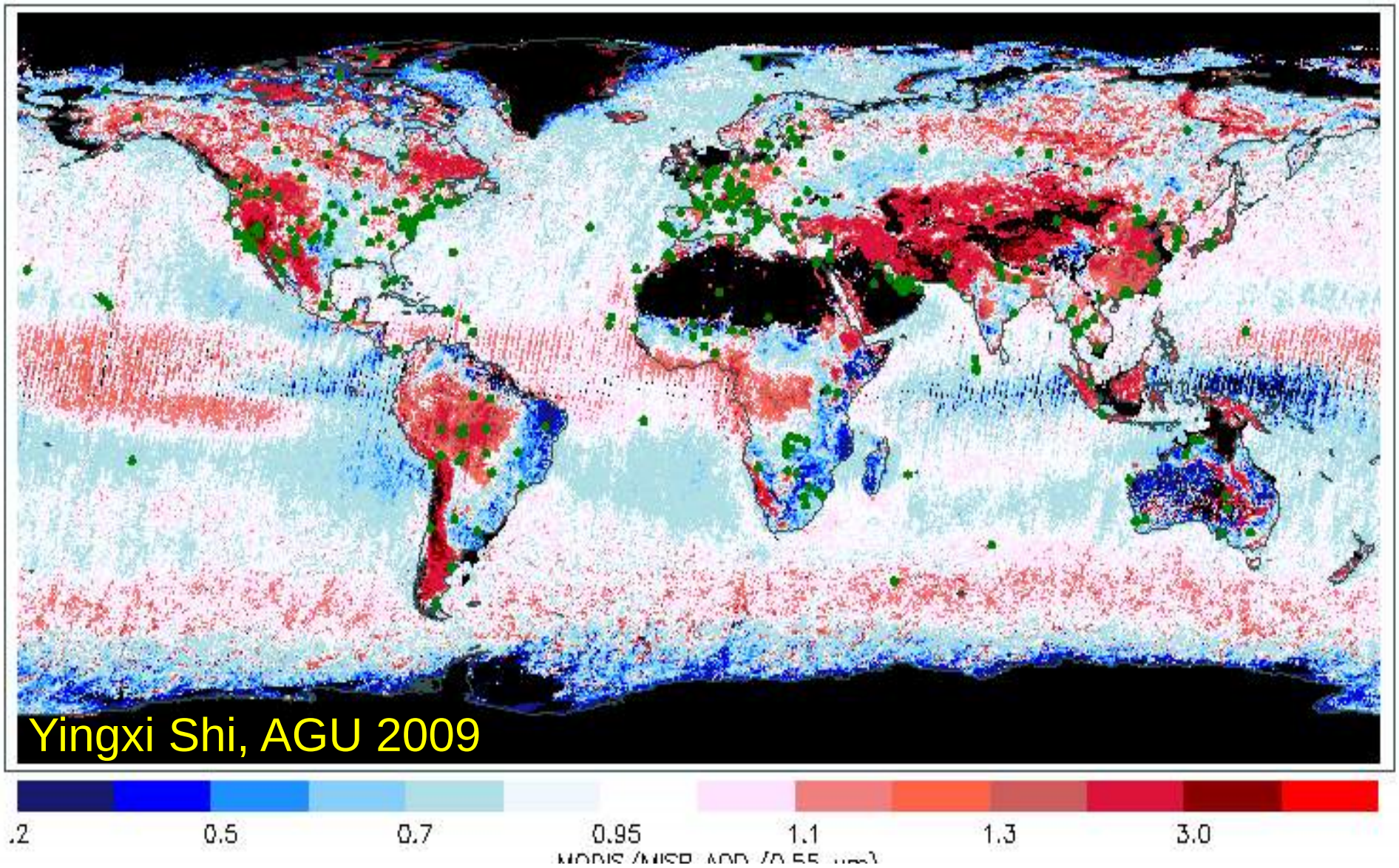
Errors to model:
Correlated bias-particularly MODIS C5 over land

•The core retrieval biases related to clouds, lower boundary condition, and microphysics are non-random, but spatially and temporally correlated-invalidating most commonly used V&V methods and data assimilation assumptions.



Correlated bias in lower boundary condition, microphysics, cloud masking.

Ratio of MODIS to MISR. These features dominate innovation vectors and hence any inverted quantity.



Progress? MODIS Col5->6 Example

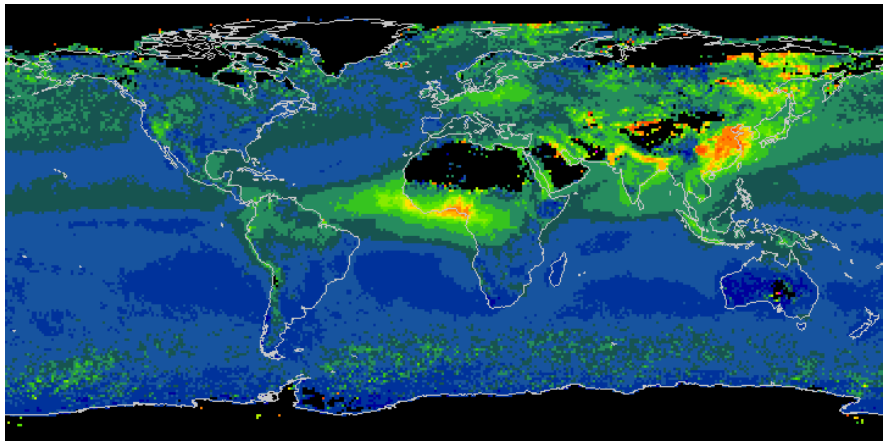
2006: Courtesy of Yingxi Shi

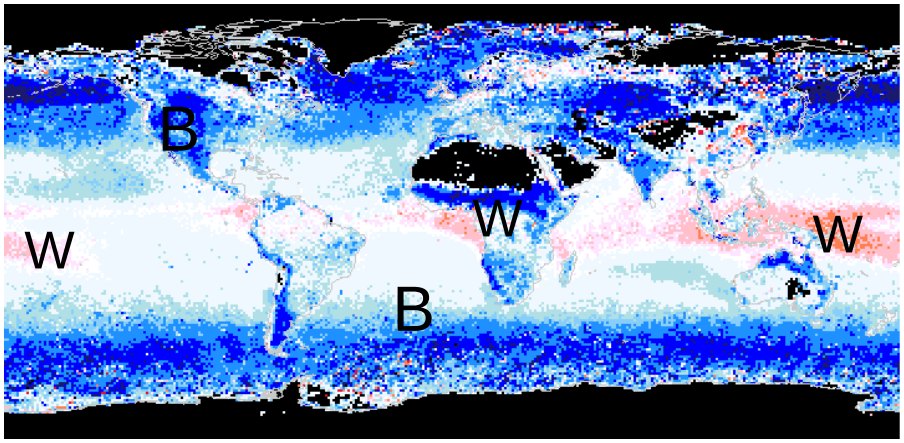
Average AOT

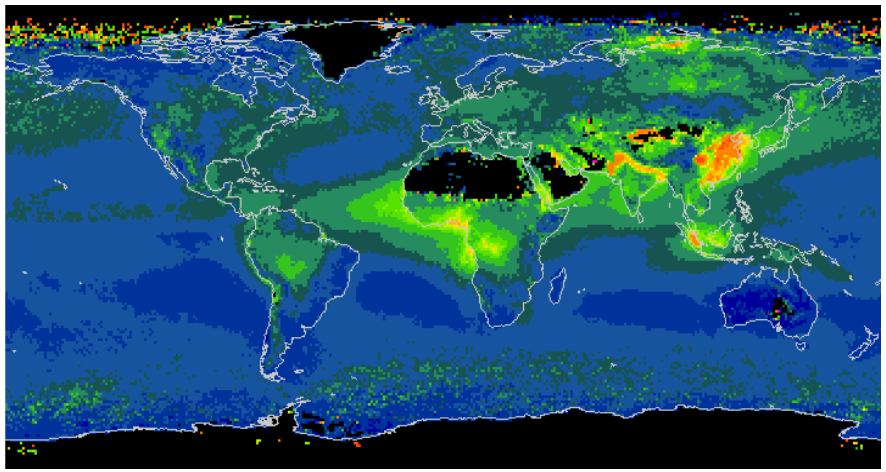
C6-C5 AOT

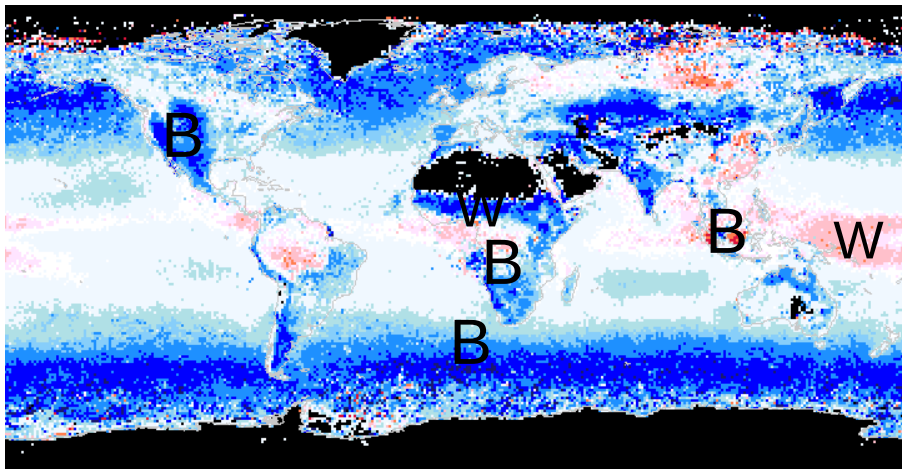
Dec-May

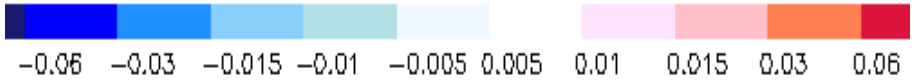
Jun-Nov





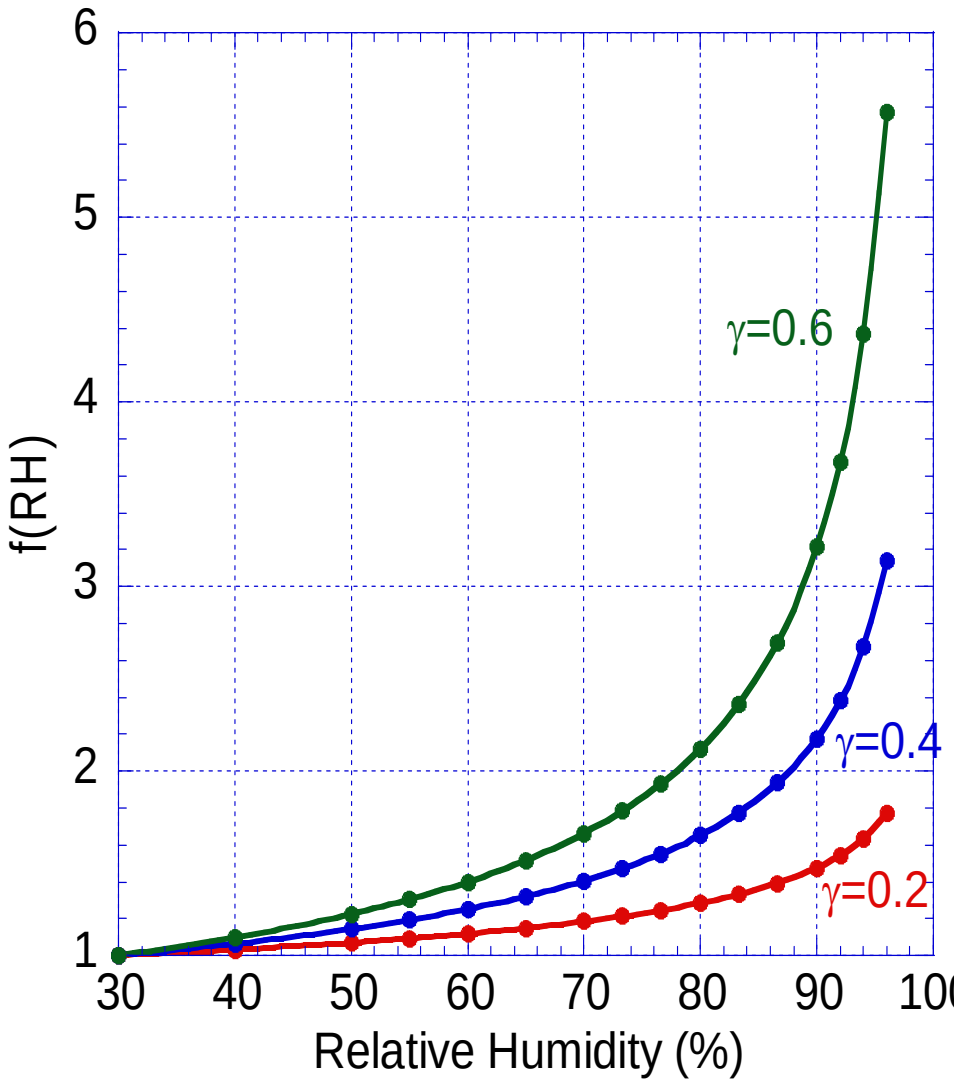






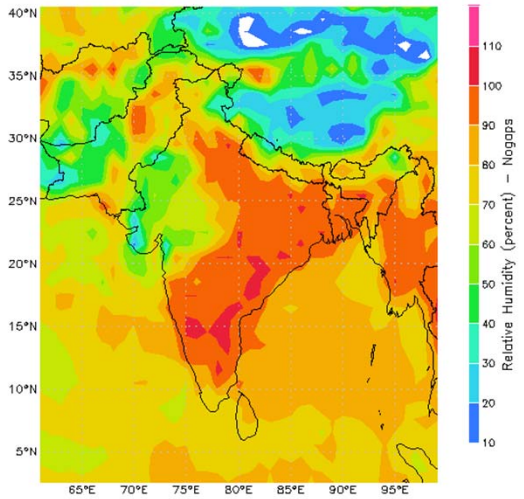
#3 Model Deficiencies
(I am sure people do not want to dwell on this)

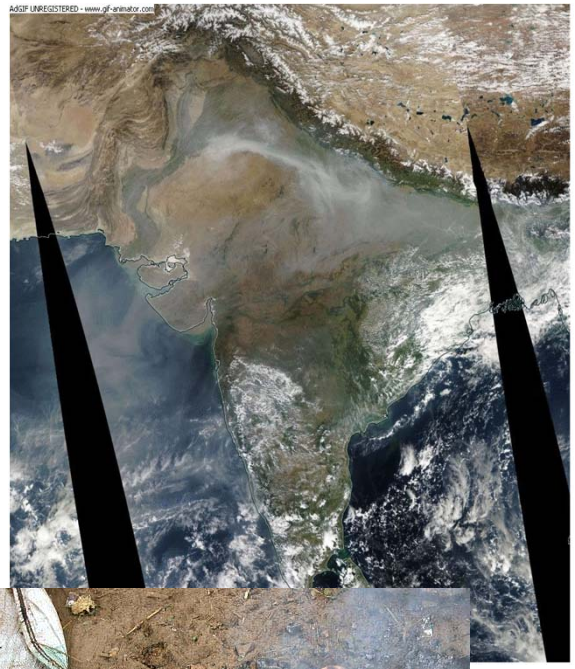
- Lets face it, aerosol world is a complicated place, without the benefit of Navier Stokes equations and “simple” thermodynamics.
- Add all NWP errors plus:
  - Sources
  - Transformation
  - Sinks
  - Microphysical simplifications
  - Numerics
- The predictability wall is much higher than traditional NWP.
- Modelers tune to compensate for for meteorological and observational error.
- But, there are very few sites to tune, and many regional effects. Beware of the water balloon effect

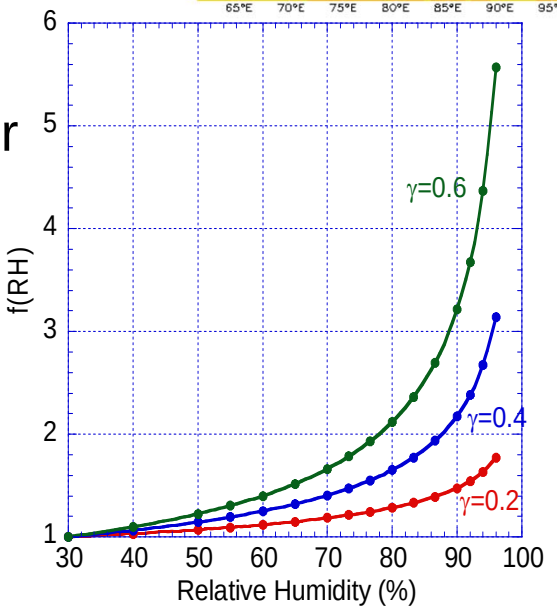


So back to Kanpur: Plenty of blame to go around

- Tough meteorological environ. and PBL dynamics. Few meteorological observations.
- Satellite obs have their set of retrieval and sampling issues, impacting models differently based on their DA systems.
- The region is a microphysical bear, with big uncertainties on sources & microphysics coupled with the above biases.







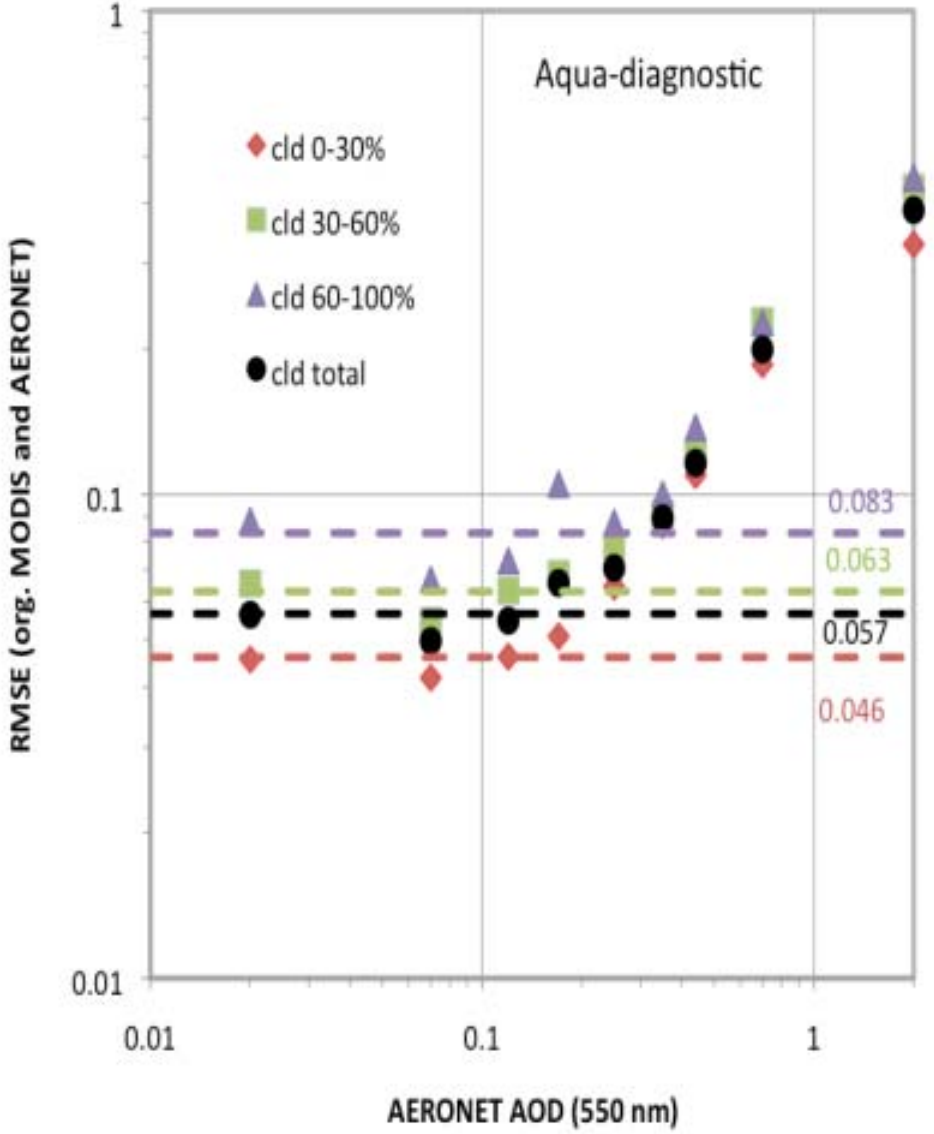
Part 2: Where do we go from here?

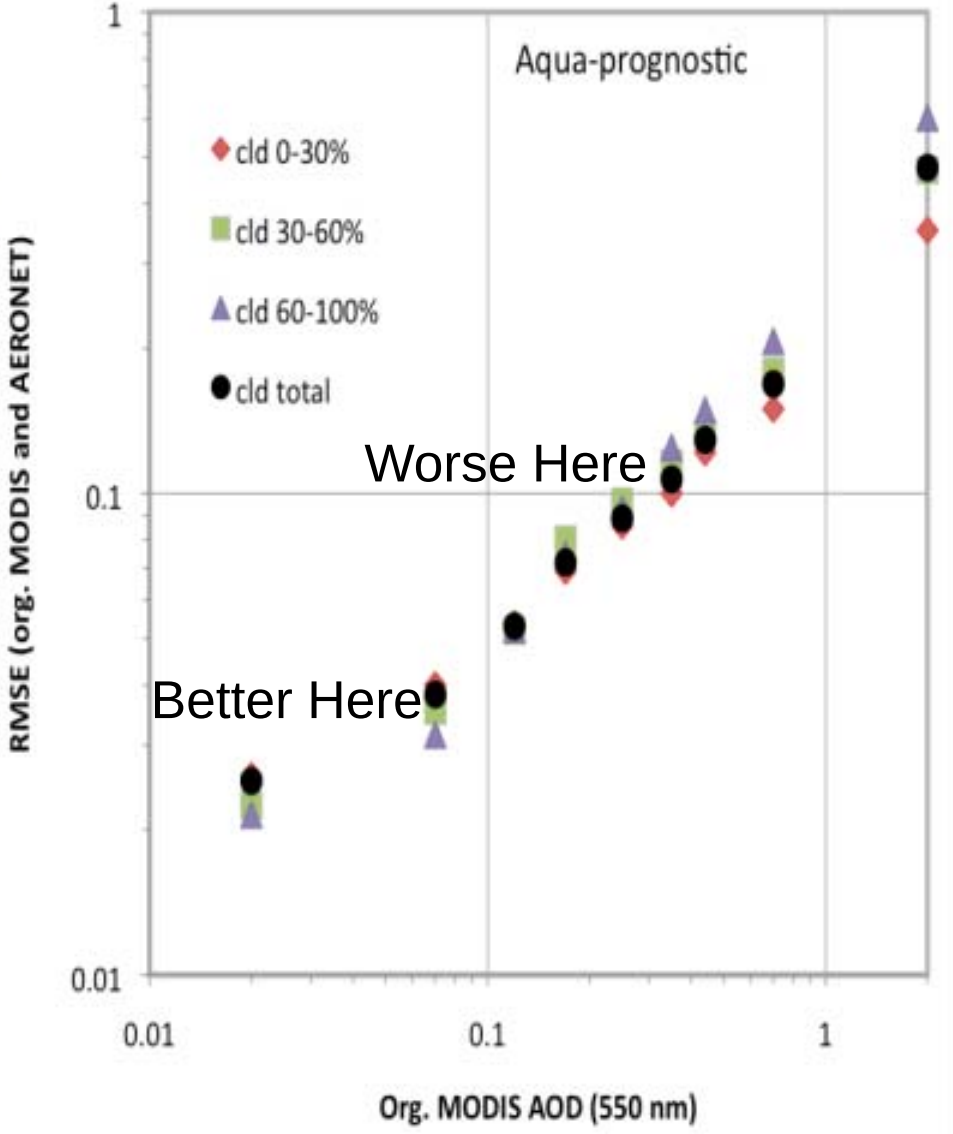
- Learn from NWP. Much of the good, bad, and ugly is right there.
- Recognize models and the environment are moving targets. Verification needs to be continuous and simple. This limits data sources.
- We need to move community wide to vertical verification.
- As a community, we need to clearly define our metrics and baseline, as simplistic as it may be.
- For ICAP-MME the next step is to study model independence.
- We need to similarly push the observational community for tractable improvements and real errors with a common voice.

Lets focus on the last topic here (its what Mike wants):

Diagnostic versus prognostic error models
MODIS over ocean example in the simplest form

(Shi et al., 2011)





Types of bias for which we must contend

Each a talk in themselves

- **Radiometric Bias:** Calibration/characterization at the sensor level.
- **Retrieval Bias:** Biases related to shortcomings in the retrieval itself.
- **Sampling/Contextual Bias:** Biases related to where a retrieval is/is not performed or contextually related uncertainty in a scene. This leads to a skewed data population relative to what is thought to have been collected.
- **Aggregation/Data Reduction Bias:** Loss of required information during conversion to level 3 or during analysis.
- **Cognitive Bias:** We, the investigators, misinterpret, withhold, or frame data/results without consideration of the full nature of the data.
- **Other Considerations for multi-sensor work: a) Correlated error-** "Independent" products that share similar biases; **b) Tautology -** Circular reasoning or treating non-independent data as independent during tuning.

And we wonder why modelers want to assimilate radiances?????

Considerations

Simple AERONET comparisons are a good start. But…

- We cant use bulk regressions, or compliance stats. We want point wise RMSE
- One way or another, it is best if we can de-bias the data
- Everything we do to the data has a consequence.
- Sole AERONET verification games errors in favor of the satellite product through sampling in many forms (cloud screening, support availability).
- Tuning to AERONET does not get at error covariance.
- And AERONET has its own errors, particularly in association with perceived coarse mode.
- AOT is simple, tractable and generally has an ob error much less than the model. SDA give us a good fine/coarse partition too. AAOT or ω_o not so, and the error bars are large on all fronts. So how do we want error information delivered?

Components of Level 2 Error Model (requires lots of data to pull out)

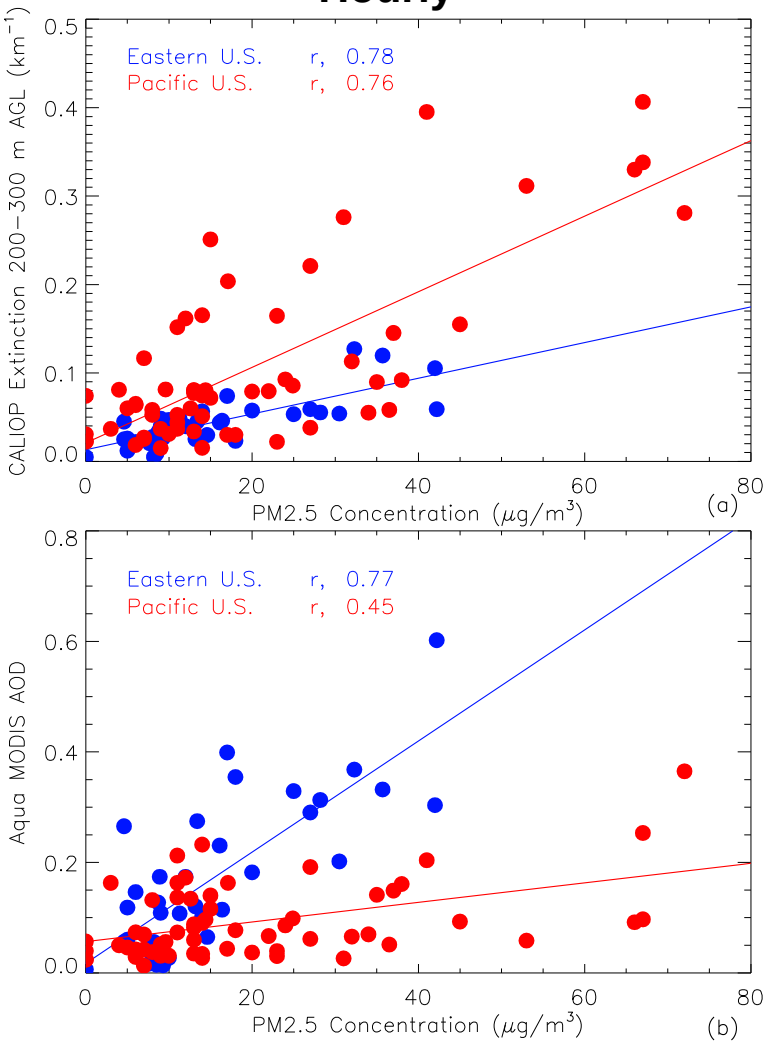
- **Can be as simple as RMSE as a function of AOD**
  - AOD can be from AERONET (diagnostic) or own AOD (prognostic).
  - But, RMSE is symmetric nor does it address massive outliers which are often the problem.
- **Terms include:**
  - Differential Signal to Noise: Lower boundary minus total, including view angle/optical path length.
  - Lower Boundary Condition:
    - Ocean: Wind/glint/whitecap, class 2 waters, sea ice
    - Land: Surface reflectance model, snow, view angle/BRDF/hotspot
  - Cloud mask
  - Microphysical: Fine coarse/partition, $P(\theta)/g$, ω_o, AOD
- **Errors are not symmetric, so you need to debias the data.**
- **Radiance Calibration: Individual wavelengths propagate non-linear through retrievals and are not easy to incorporate.**

Moving to the vertical

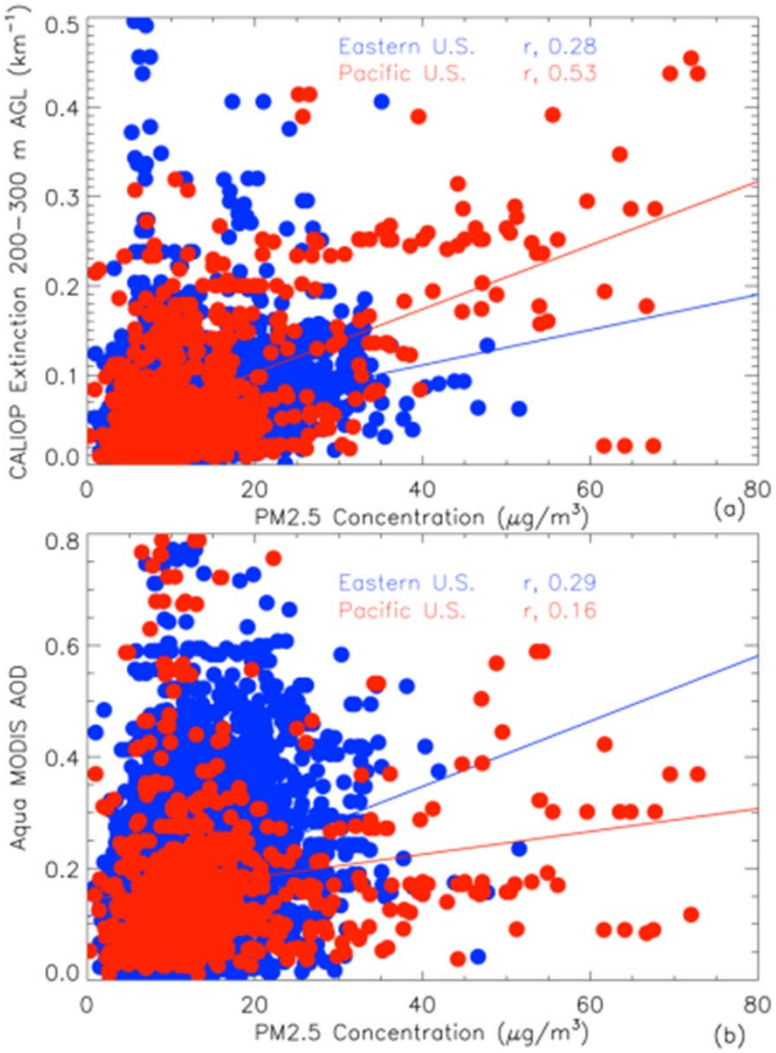
Clearly we need to move towards lidar and ground stations, but his is an order of magnitude more work.

CONUS matchups from Toth et al., 2014

Hourly



Daily

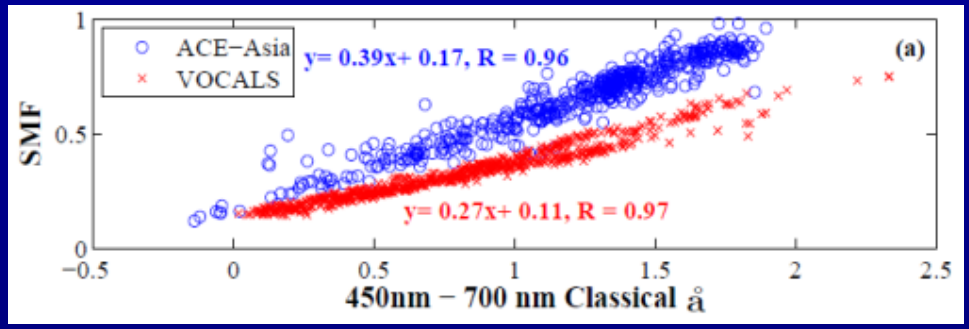


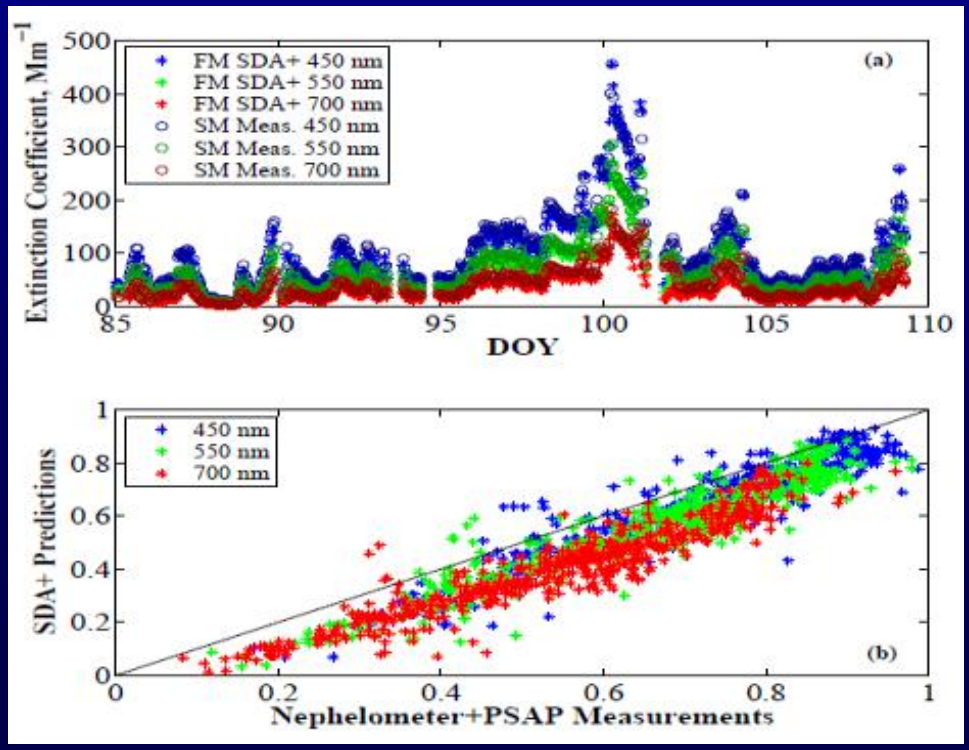
Opening up more surface obs
Spectral Deconvolution Algorithm verifies
(Kaku et al., 2014; AMT)

Aerosol size verification is typically through the Angstrom exponent, although it is not really the tool for the job.

Kaku et al ported SDA to nephelometer and PSAP data from the Ron Brown, thus opening a new avenue to in situ verification, and verifying the method itself.

Caveats: Assume fine and coarse mode behavior and VERY sensitive to calibration error.





Immediate things we can do

We should exercise the bully pulpit of NWP. The hardest part of the problem is to define what it is we want. This is followed by funding 8^)

Baseline development has been slow, but with the latest updates to satellite data and ICAP-MME I think we will have more progress in the coming year.

Don't think too big, but in tangible and consistent datasets.

Be thoughtful about site selection!

If you have an idea, just do it.