Aerosol Observability Workshop: Validation and Verification of Aerosol Products for Operational Use

Jeffrey S. Reid, Jianglong Zhang, Edward J. Hyer, James A. Hansen, Peng Xian, and James R. Campbell

jeffrey.reid@nrlmry.navy.mil/ 1 (831) 656-4725
http://www.nrlmry.navy.mil/aerosol/
http://www.nrlmry.navy.mil/flambe/
There are a lot of products out there. It would be good to use all of them, but the shear numbers is getting difficult to manage.

Products are not what they appear. It does not matter if a product is uncertain, we just need to know what that uncertainty is per retrieval.

We don’t have the benefit of 30 years of operational history, and due to propagation of error we can’t directly apply their methods.

Need standardization for error stats and hopefully develop a system.

For discussion tomorrow: What products and error stats do we want generated?
Aerosol Relevant product lines
(there is a lot to keep track of)

Current:
- CNES: Parasol/POLDER
- ESA: ATSR, AATSR
- EUMETSAT: METEOSAT dust, fire, etc..
- JAXA: GOSAT, TRMM
- NASA: MODIS (Col 5 Land & Ocean, Deep Blue, fire), MISR, OMI/TOMS, CALIOP, Cloudsat, AVHRR/GACP, SeaWiFS (standard and Deep Blue)
- NOAA: AVHRR, GOES-GASP, Geo-ABBA fire constellation

Next generation (that I know of):
- JAXA: GPM, GCOM-C, EarthCARE-CPR
- ESA: EarthCARE-ATLID/MSI
- JPSS: NPP/JPSS VIIRS
- NASA: GLORY, MODIS MAIAC.
- NOAA: GOES-R aerosol and fire
Product Efficacy: Should you be concerned?

- Historically product developer paint a pretty rosy picture about product efficacy.
- Even so, in many parts of the world, assimilating as is will improve scores. Others, it will degrade the model.
- AERONET, while wonderful, can lead to representativeness bias in assessments.
- No developer has ever published a prognostic error model for their aerosol or fire product. At best you get a slope, Y-intercept, and r value. If lucky, get a mean bias. This leads to “worst case” usage=no impact.
Relative Levels of Efficacy Required
(Approximate and not meant to offend…)

Operational Agencies Focus on the Extremes
Historically imagery rules the day for operational requirements

- Imagery/Contextual “Advantage of Human Eye”
- Seasonal Climatology
  Basically want to know where stuff is. Can do one-up corrections
- Model Aps, V&V, Inventory
  Have stronger time constraints and need spatial bias elimination.
- Data Assimilation
  Quantify bias & uncertainty everywhere and correct where you can.

- Parametric Modeling and Lower Order Process Studies
  Correlations de-emphasize bias
- Trend Climatology
  Need to de-trend biases in retrieval and in sampling
- Higher Order Process Studies
  Push multi-product and satellite data

• Inverse modeling is sensitive to spatially and temporally correlated error.
• Forecasting is even more sensitive, as anomalously high data will create a “plumes” in the forecast fields. Forecasters are not used to this.
• Non-linear transfer function between AOD and model mass complicates error propagation, particularly at low AODs.
Types of Bias
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?????
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(θ)/g, ω₀, AOD

- Biases are often folded into “random” error models. If they are known, why not correct for them?

- Radiance Calibration: Individual wavelengths propagate non-linear through retrievals and are not easy to incorporate.
Where do we get validation data?

- Other AOD sources: higher uncertainty poses a problem with validation. High overhead of bringing in additional regional data sets?
- **PM10/2.5**: Can be trusted at most US/European data networks, but that is about it. In developing world data has uncertain error characteristics. If you want mass, use gravimetry….
- **Isolated super sites**: help on model processes and parameterizations but not validation/verification.
Direct Comparison: MODIS
Need years of Global Data (Zhang’s and Hyers papers)

MODIS bias versus ocean wind speed
Key to data assimilation: Quality Assurance
Can clear out a lot of junk through spatial tests

• Southern ocean aerosol anomaly: Fact or cloud bias?
• Northern oceans have same problem, but people quickly attributed it to China and CONUS.
• Spatial tests get rid of it.

Zhang et al., 2005
Impact of QA process

Impact of bad pixel

Natural

DA

Ratio: Spatial and cloud filter

Ratio: Empirical corrections
Another cost of QA:

Sampling/Contextual Biases
Clear sky, Scale/Amplitude, Species, Land Surface, Dynamical

• Analyses now require a number of “qualifiers” to describe what you are really seeing—especially for model and high level process studies.

• Clear sky bias for MODIS was calculated during 2 year data assimilation run by comparing 24 hour forecasts to that next day MODIS sampling.

• As expected, positive clear sky biases in tropics, negative bias in the mid-latitude due to storm track (usually—see Pacific).

• Individual events have bigger biases.

• This is why we need data assimilation
Spreading information on aerosol properties and error?

- Need ways to spread information and uncertainty from isolated observations, such as lidar and surface.
- Ensemble data assimilation/EnKF may help.
- Able to reproduce satellite method work on aerosol correlation length scales in 1 week.
- Very promising, but data intensive. We cannot do climatology runs efficiently except perhaps as consensus or multi model ensemble.
Why NOT Radiance Assimilation?

- Historically, centers go to radiance assimilation and away from product assimilation as soon as possible.

- Why? Products are at times inconsistent with model fields, resulting in low impact, 2) With radiances you can spread information more consistently, as well as cope with spatial/temporal correlated error, 3) With radiances, you can be consistent across sensors.

- Why not? 1) very data and computational intensive in the shortwave; background performance is poor and lower boundary condition not an output variable. 2) Hence, for the most part, need to go through the retrieval process anyway; 3) If so, take advantage of community product development and cal/val; Meet half way? Cleaned up radiance product at level 2 resolution and joint parameterizations?
Conclusion: What we need (I think)

• We need to work towards a consistent validation data set Even if it is for one year.
• Product developers should develop products which are tailored to model uncertainty requirements.
• We need to provide specifications on error models and agreed upon baselines.
• Retrievals: Program offices need to hold developers feet to the fire:
  – Science teams are responsible to their customers.
  – Prognostic error models need to be a deliverable for any retrieval proposal intended for public consumption.
• Model corollary: No easy way to validate the models. Need an equivalent to 500 hPa anomaly? Yes and no.
  – Test the super-ensemble hypothesis—all add skill to the ensemble?