Validating uncertainty estimates

Andrew Sayer
GESTARII-UMBC at NASA GSFC
andrew.sayer@nasa.gov
So your retrieval has an uncertainty estimate...

• How do we know if the uncertainty estimate is useful or realistic?

Statistics!
Reminder: uncertainty and error are different

• Uncertainty is an *expected level of dispersion*  
  • Typically one standard deviation (1σ, ~68% of the time as good as X)

• Error is an observed quantity  
  • A *draw from the distribution* of possible errors given the uncertainty level

You *expect* 3.5  
You *get* 1, 2, 3, 4, 5, or 6
Uncertainty distributions and error distributions

• Suppose AOD uncertainty is ±(0.05+15%). Then:
Solution: compare distributions of observed errors to expected discrepancy (ED)

- Observed error = (truth – retrieval)
- Expected discrepancy contains contributions from:
  - Retrieval uncertainty $\sigma_{\text{sat}}$
  - Reference uncertainty $\sigma_{\text{ref}}$
  - Spatiotemporal variability inherent in the comparison $\sigma_{\text{var}}$
- If assumed independent, then:

$$ED = \sqrt{(\sigma_{\text{sat}}^2 + \sigma_{\text{ref}}^2 + \sigma_{\text{var}}^2)}$$

Propagated or simulated

Documented and small (hopefully)

Can be a real pain to generalise
Spatial and temporal variations can both be important

- Look at *spatial statistics* of satellite data in the area, *temporal statistics* of the ground data
- *Generalise* based on a model
- Use *robust statistics* (e.g., median not mean)
- *Shrink* area aggregated or comparison window
Spatial and temporal variations can both be important

(a) Kanpur (26.51° N, 80.23° E), aerosol loading

Sayer, ESS (2020)
Look at distributions of normalised error

- Normalise as error/ED
- If unbiased and appropriate magnitude, distributions should follow $N(0,1)$

Sayer et al., AMT (2020)
Examples on real aerosol data

(a) Land sites

(d) GSFC, CDF

- ADV
- BAR
- CISAR
- DB
- DT
- MISR
- ORAC

Standard deviation of normalized error

Mean normalized error

Cumulative fraction

Normalized AOD error
Can we tell low-uncertainty situations from high ones?

- Stratify by ED
- Look at $1\sigma$ point (68th percentile) of absolute error
  - Additional percentiles can tell you more about *how* Gaussian the error distribution is for a given ED
- Similar to *forecast calibration, skill scores*, etc

![Graph](e) GSFC, $1\sigma$
Closing the loop

• Validating retrievals gives us diagnostic metrics of quality to report uncertainty

• Validating uncertainty estimates tells us if things are working as we expect
  • Regardless of how well the retrievals perform

• Figure out where your assumptions are at fault, then refine them!
References and resources


