The GEOS-5 Neural Network Retrieval (NNR) for AOD

Patricia Castellanos[1,2] & Arlindo da Silva[2]

H34F: Machine Learning Applications in Earth Science and Remote Sensing II

AGU Fall Meeting 2017

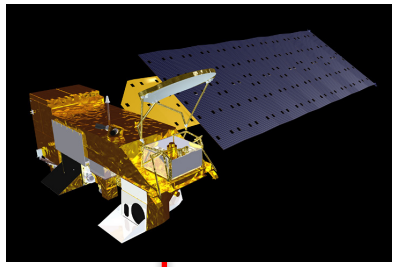
[1]GESTAR, Universities Space Research Association

[2]Global Modeling and Assimilation Office, NASA/GSFC

Remote Sensing of Aerosols

AOD

- Column integrated value (top of the atmosphere to surface)
- Optical measurement of aerosol loading – unitless.
- AOD is function of shape, size, type and number concentration of aerosols



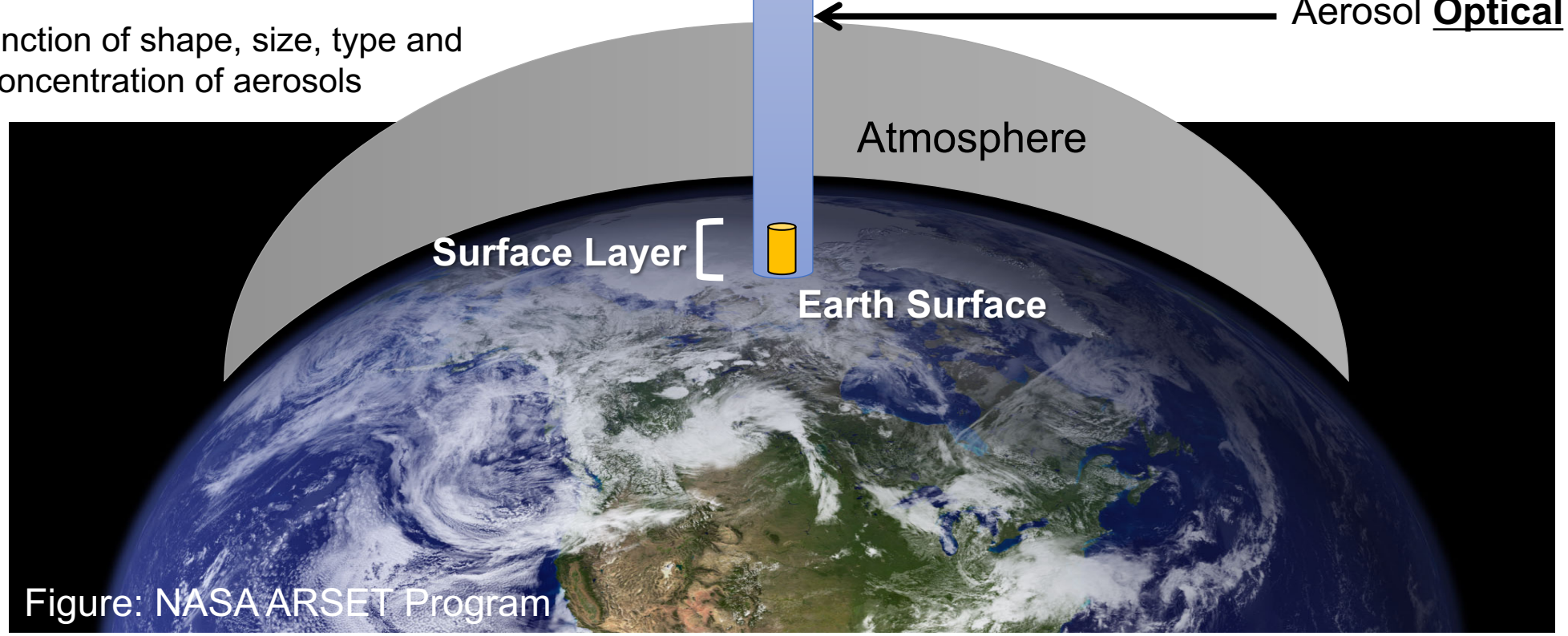
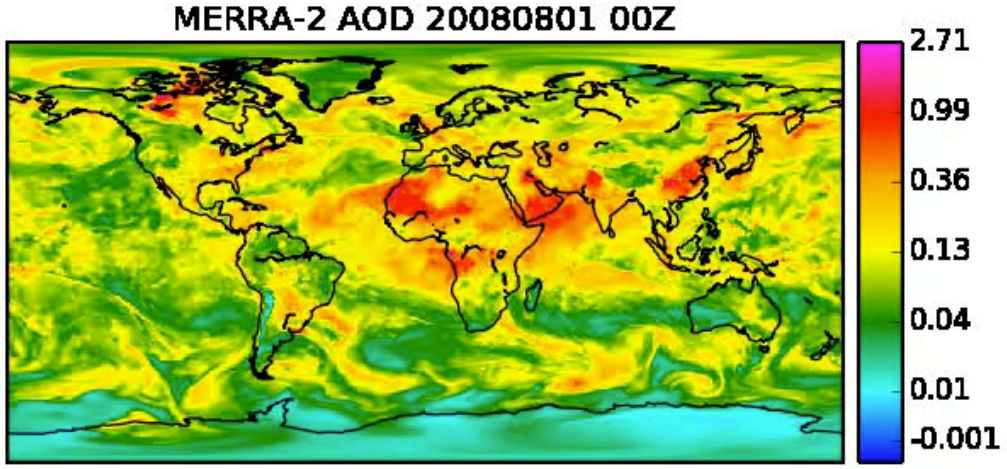
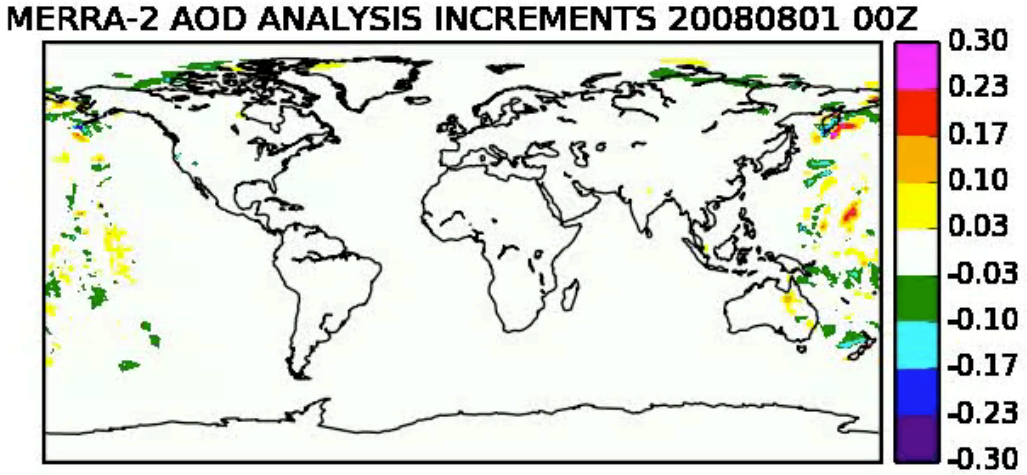
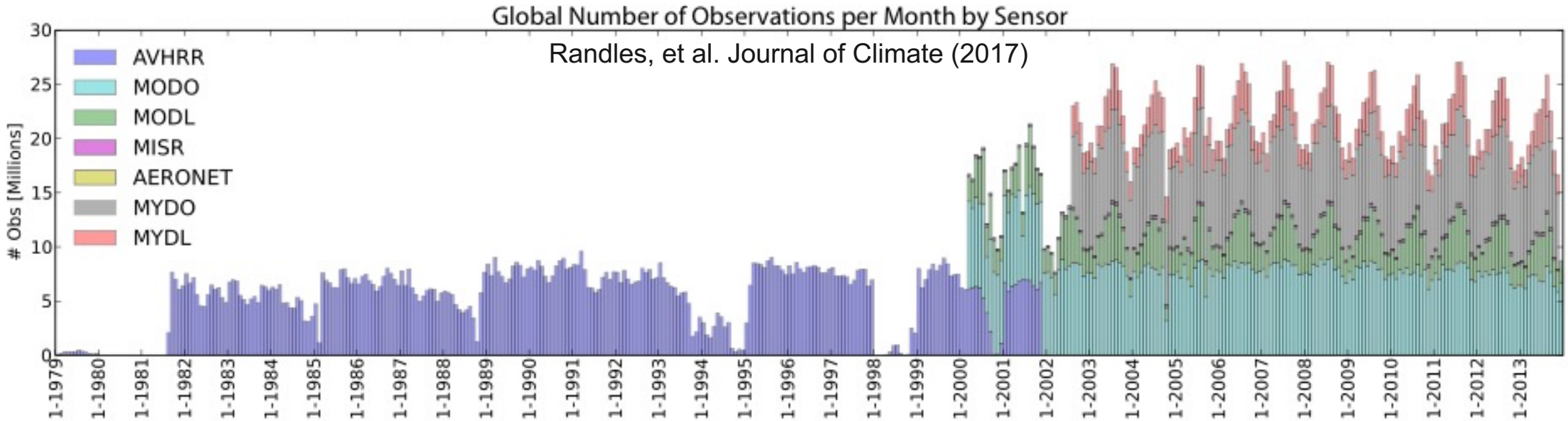


Figure: NASA ARSET Program

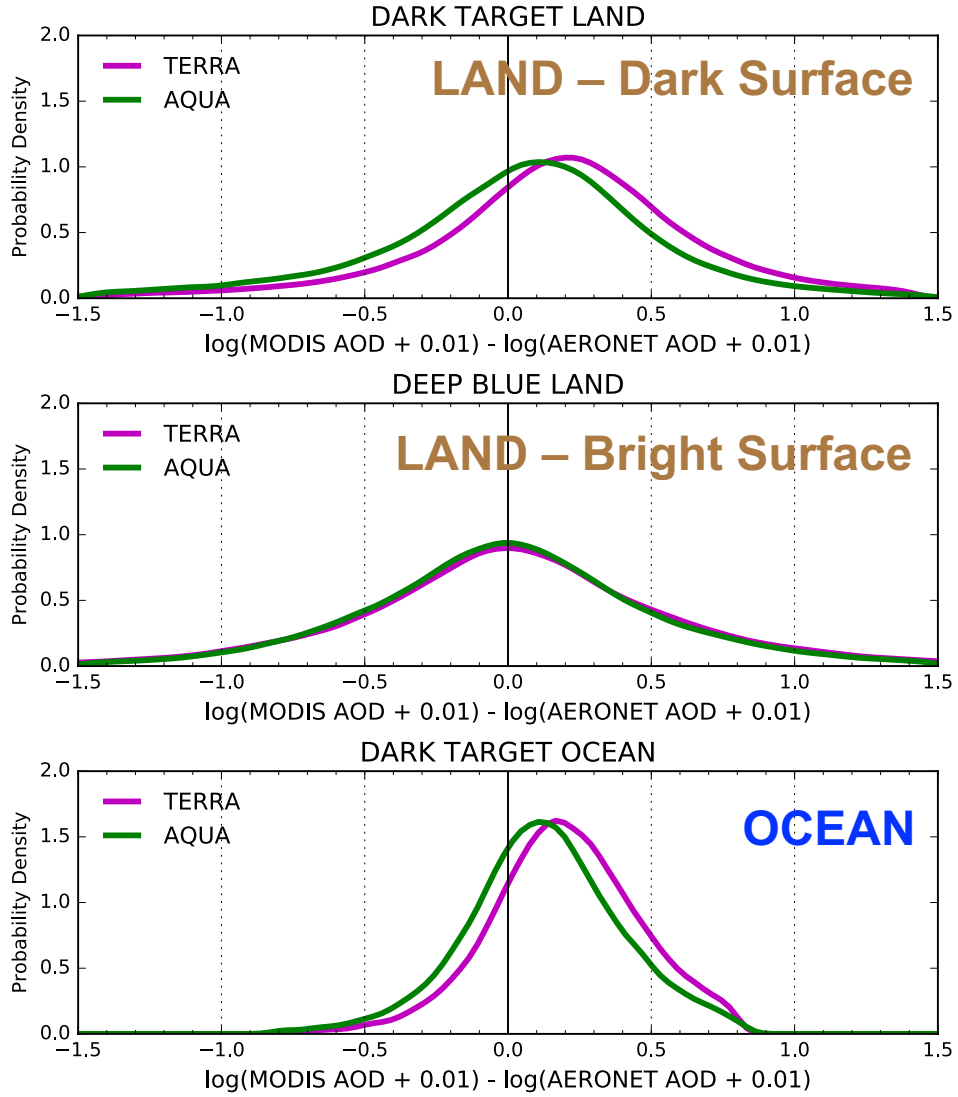
GEOS-5 Aerosol Data Assimilation





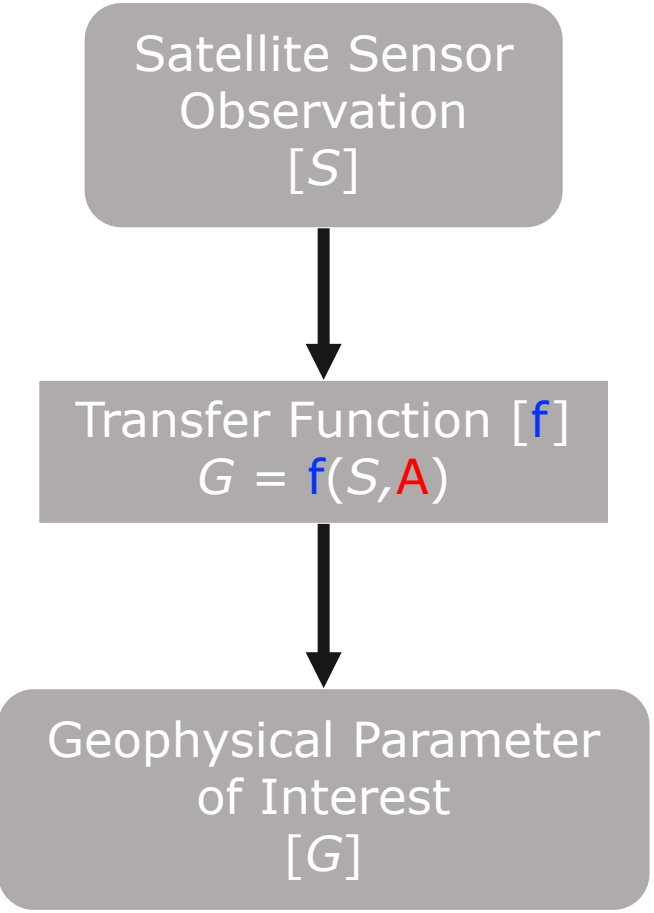


Comparison of MODIS Standard AOD Retrievals



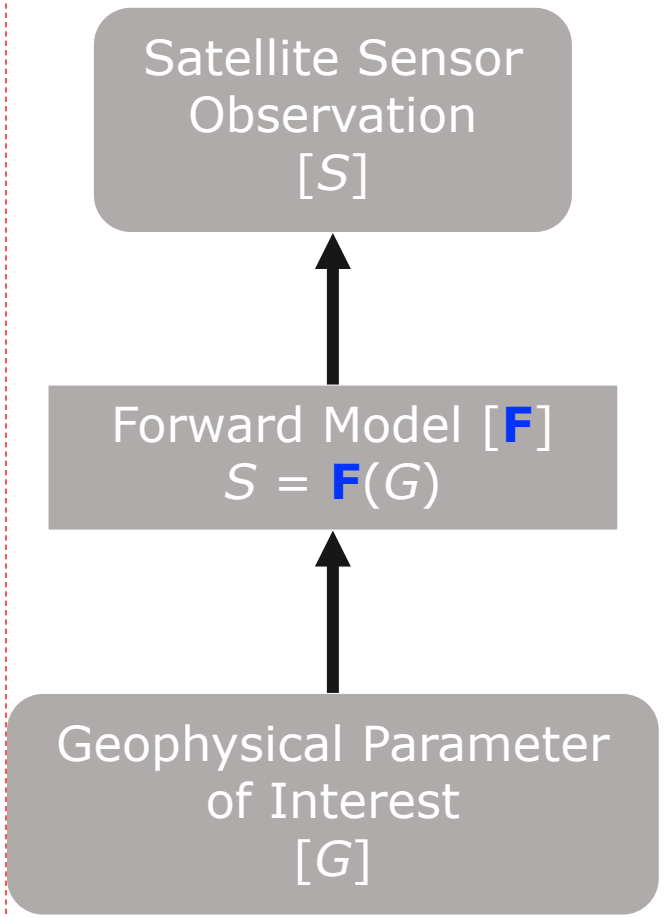
- The aerosol data assimilation problem requires a homogenized dataset of AOD across different platforms
- Biases between datasets can propagate in the model forecast

Empirical Retrievals



- f is a continuous function that maps *S* to *G*
- Represent *f* with a mathematical function that contains a set of empirical parameters, A
- A are determined from a training dataset of pairs of *G* and *S* observations.

Physically Based Retrievals

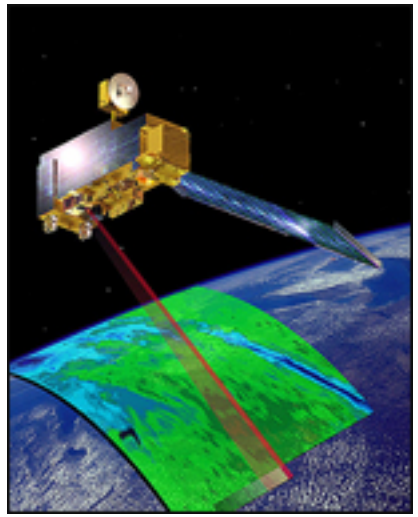


- **F** is a physical model derived from first principles (e.g. radiative transfer model)
- **F** is not easily inverted
- The objective of the retrieval algorithm is to search for a G* that minimizes ||*S* – **F**(*G*)||

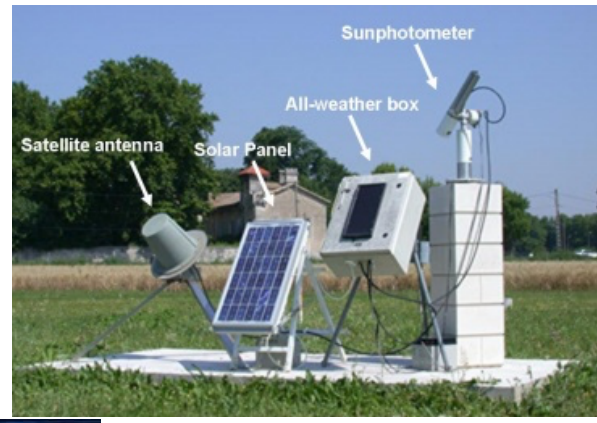
Observations

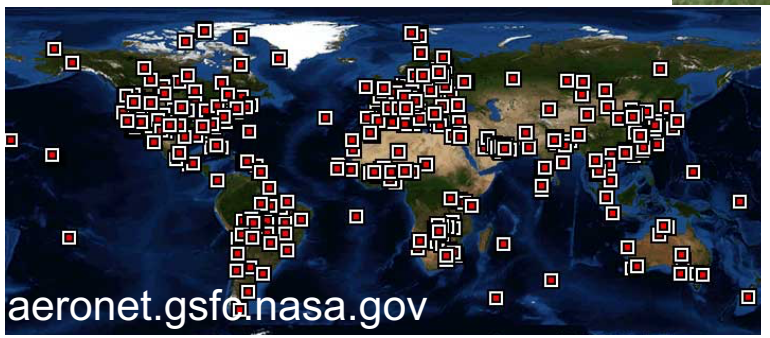
Satellite Sensor Observation [*S*]:	Geophysical Parameter of Interest [*G*]:
MODIS MOD04 /MYD04 Level 2 Reflectance	**550 nm AOD**

- Cloud masked, quality controlled, 10 km data
- Deep Blue Land
  - 3 channels over bright surfaces
  - 412 nm, 470 nm, and 670 nm
- Dark Target Land
  - 9 channels over dark surfaces
  - 412-2100 nm
- Dark Target Ocean
  - 7 channels over ocean
  - 470-2100 nm

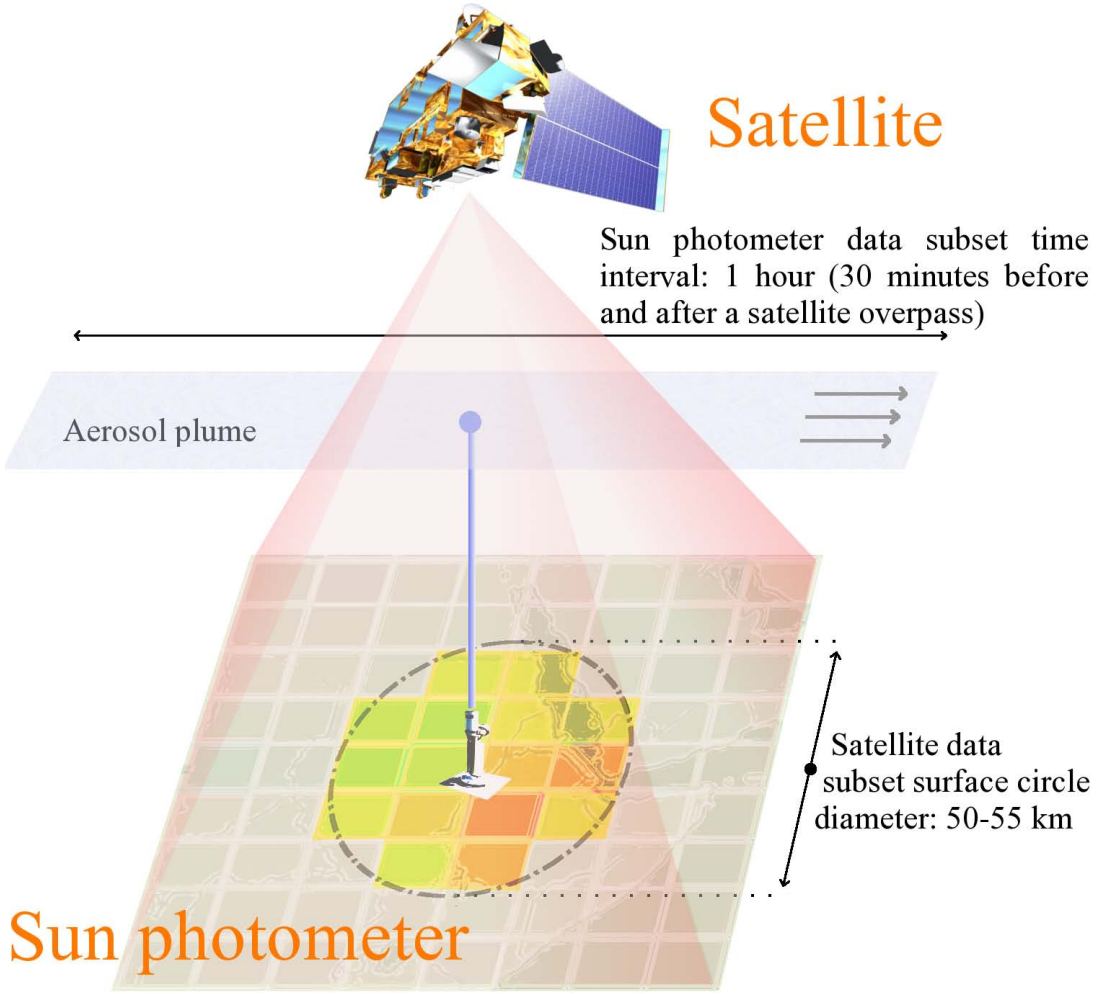


- Aerosol Robotic Network (AERONET) observations of AOD
  - Global network of sunphotometers
  - 15 minute sampling
  - Low uncertainty (±0.01)





MODIS-AERONET Data Pairs



Petrenko, et al. AMT (2012)

- 15 years of data (2000-2015)
- ~150K Deep Blue data pairs
- ~100K DT-Land data pairs
- ~40K DT-Ocean data pairs

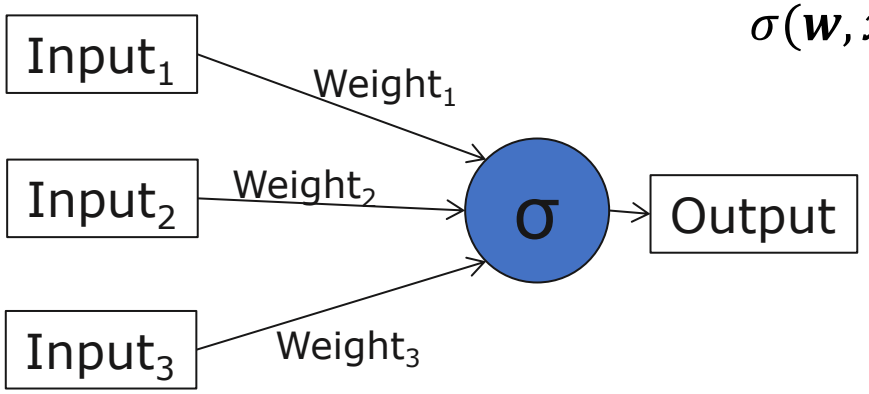
Additional Data Screening

- Outlier removal
- Cloud Fraction < 0.7
- Used MERRA-2 to “balance” the dataset by aerosol type

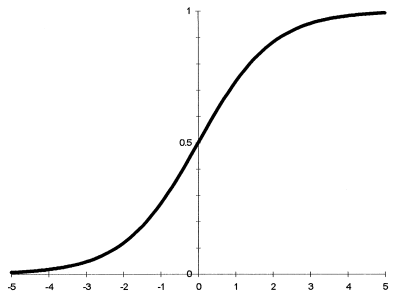
- Reduces number of data pairs by half

Neural Networks

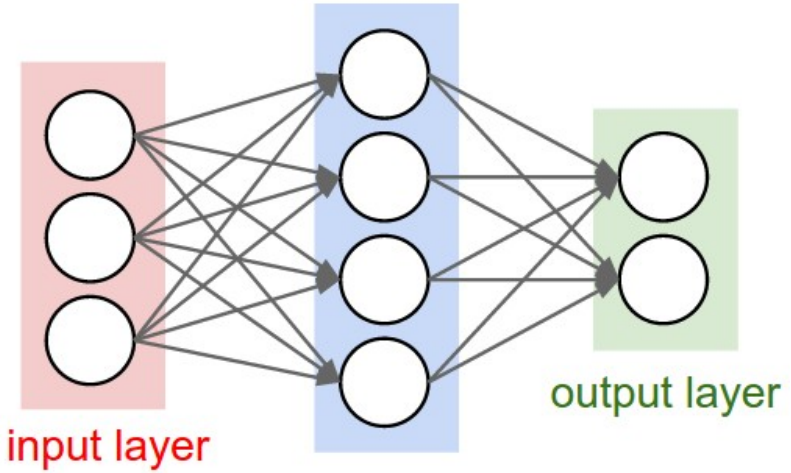
- A set of nodes connected by numerical weights.
- The weights are tuned based on a specific training dataset containing input-output data pairs.
- The superposition of many nonlinear transfer functions (nodes) allow NN's to approximate extremely nonlinear functions

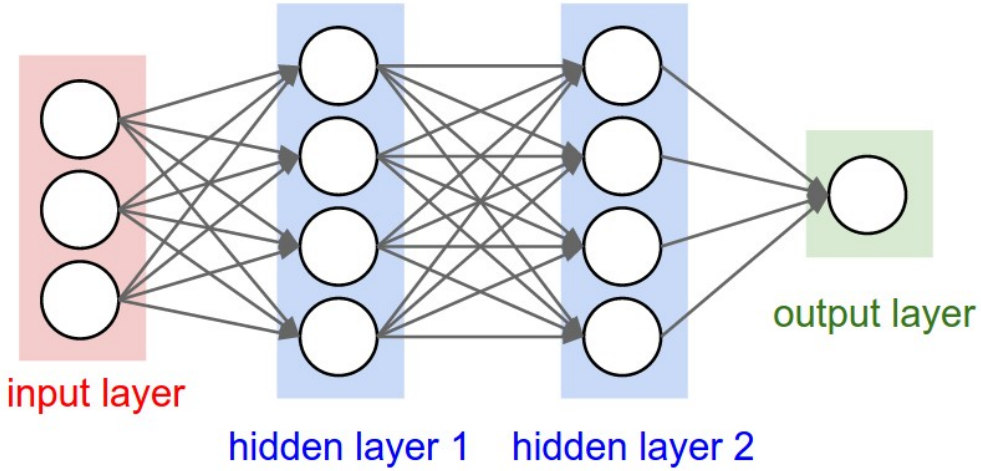


$$\sigma(\boldsymbol{w}, \boldsymbol{x}) = \frac{1}{1 + \exp\left(-\boldsymbol{w} \cdot \boldsymbol{x} - c\right)}$$



Feed-forward Neural Network





L1 Reflectance

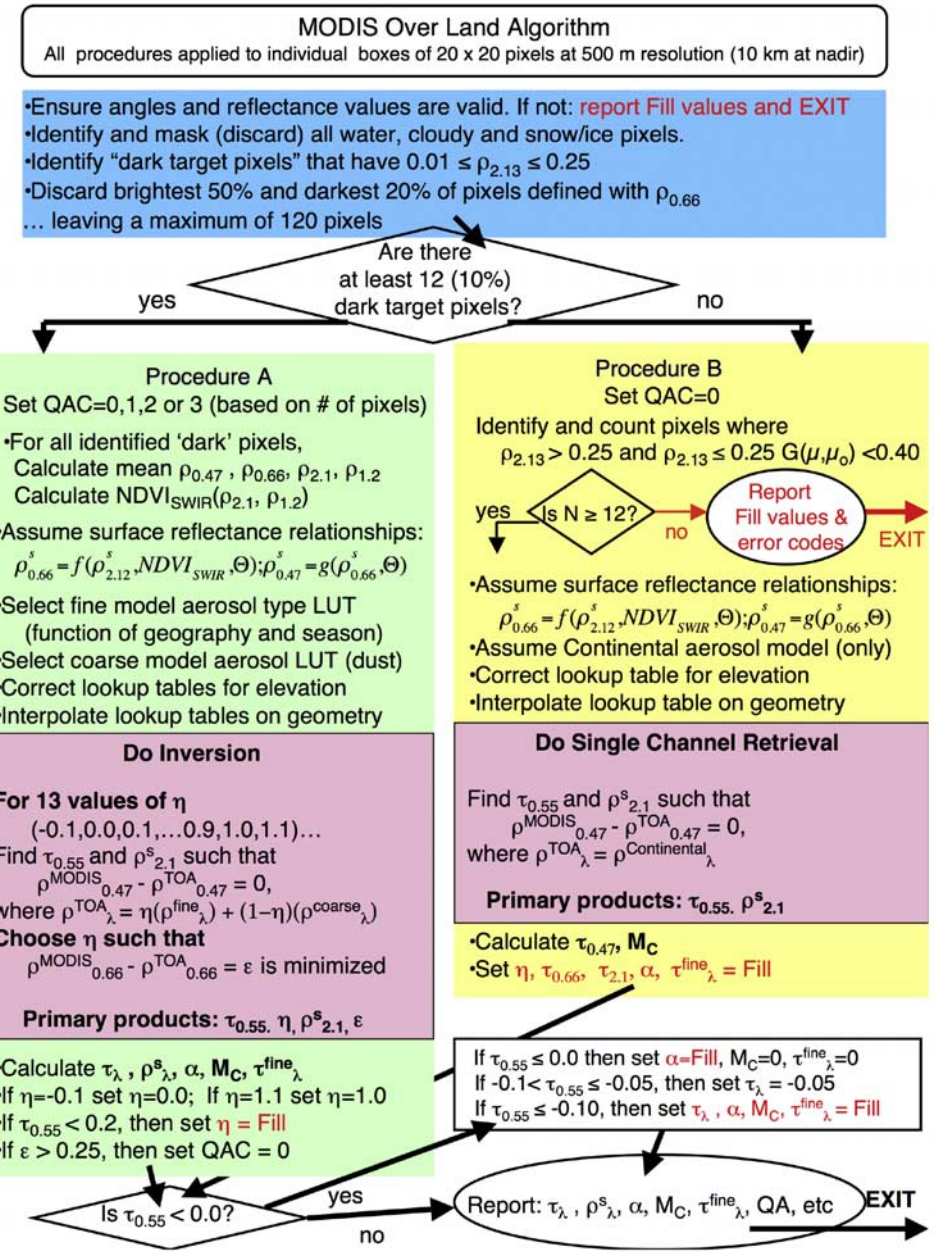
Generate L2 Reflectance
- ❑ Cloud mask
- ❑ Quality control

Collect Relevant Auxiliary Data
- ❑ Surface reflectance
- ❑ Aerosol type – optical properties

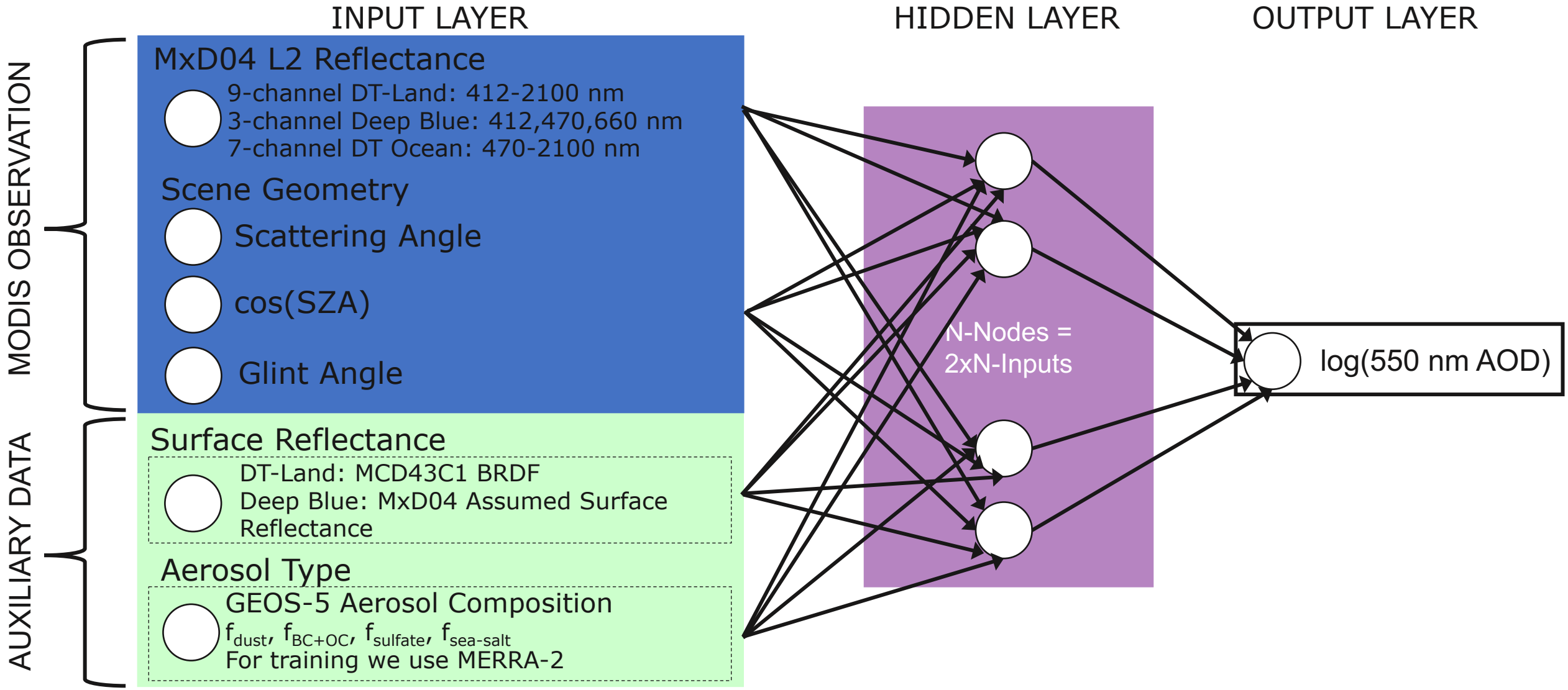
Neural Network trained with MODIS-AERONET observations

Do Inversion

550 nm AOD

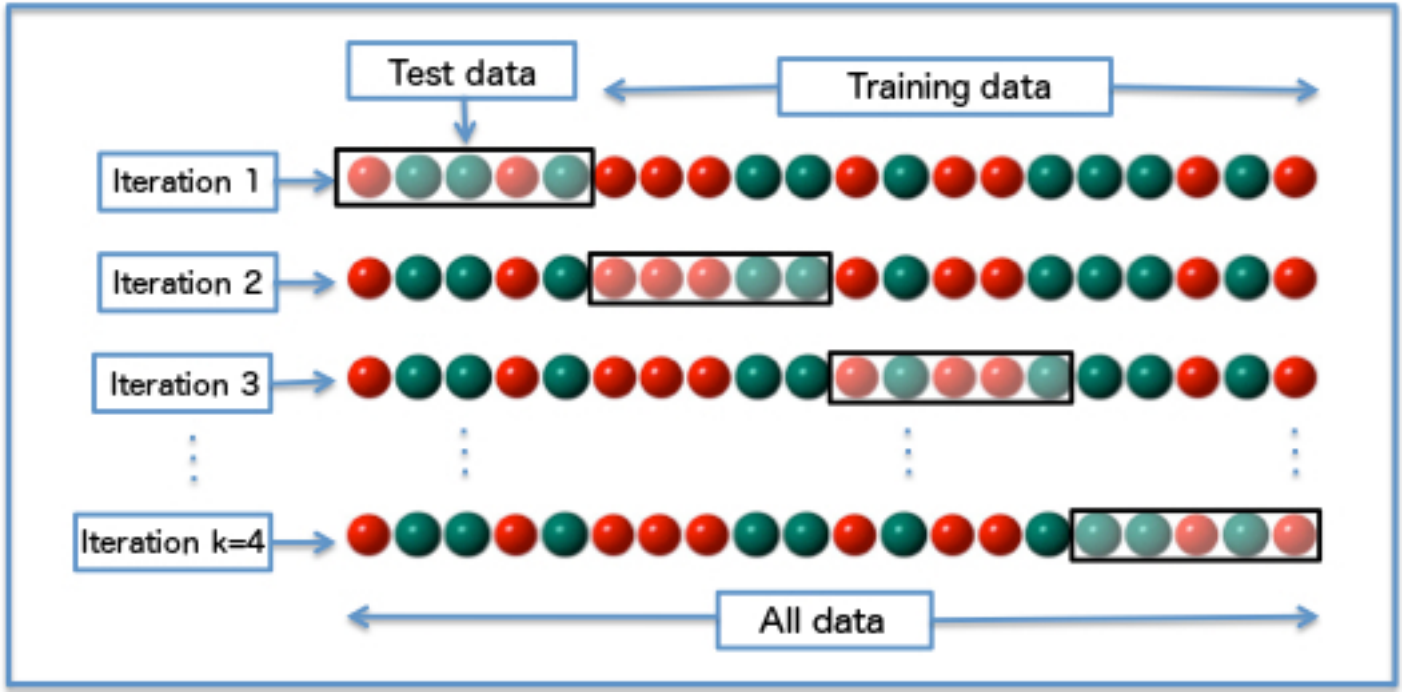


GEOS-5 NNR for AOD

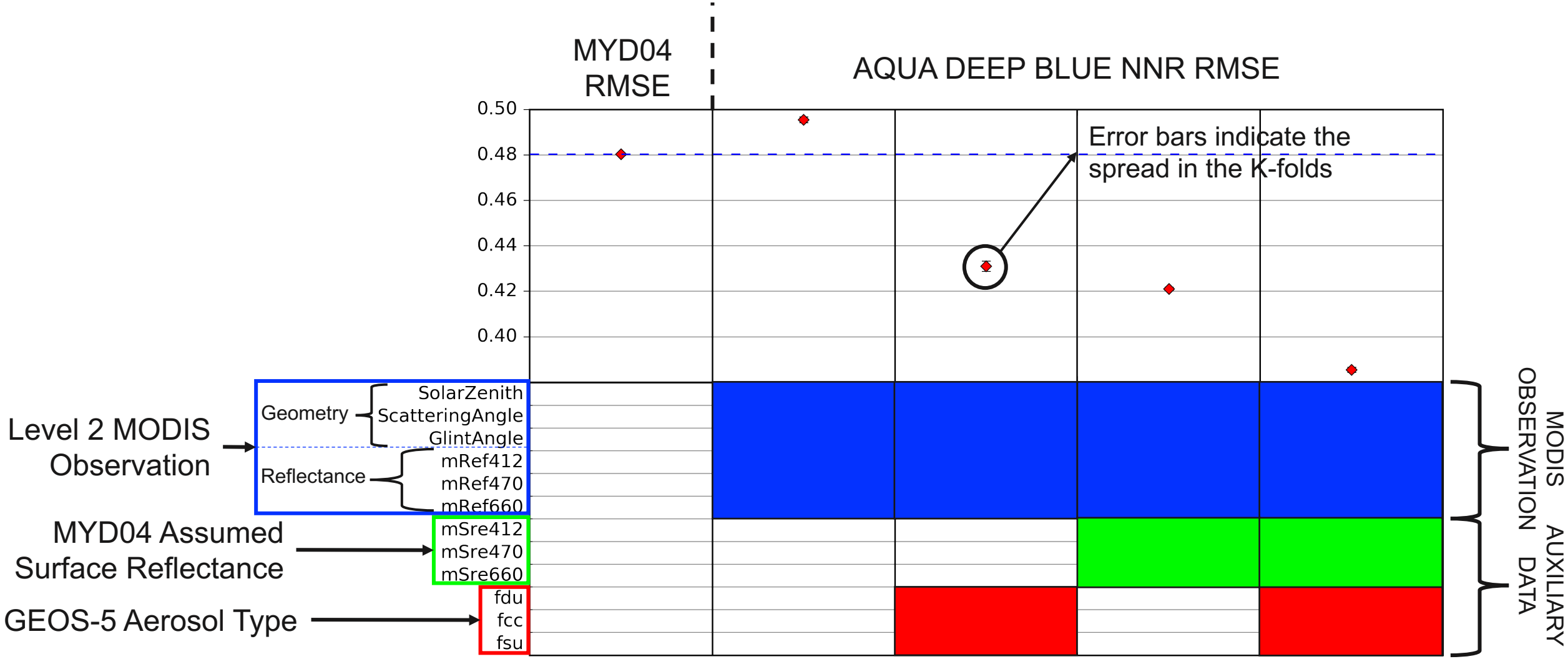


NNR Training, Testing, Validation

- Train & Test
  - Iteratively train and test adjusting input variables, architecture, etc. to optimize neural network
  - Cross Validation
    - K-folding: create K subsets of data, using K-1 for training, and 1 for testing. Iterate K times.
- Validate
  - Use a separate dataset, not used to train/test
  - Observations after 2015



NNR Training & Testing



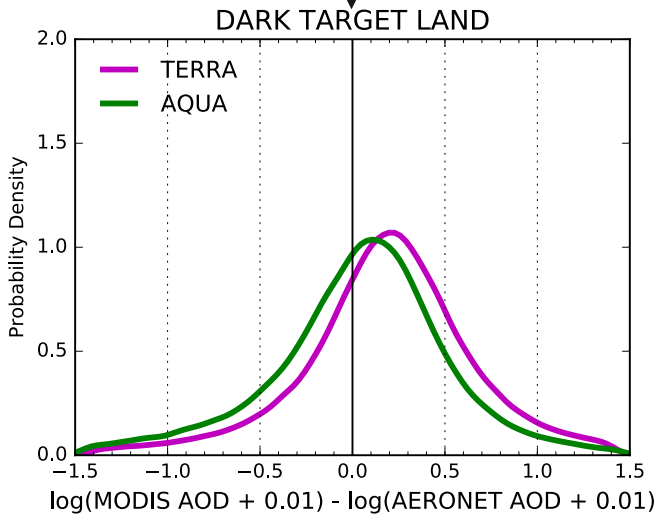
NNR Testing

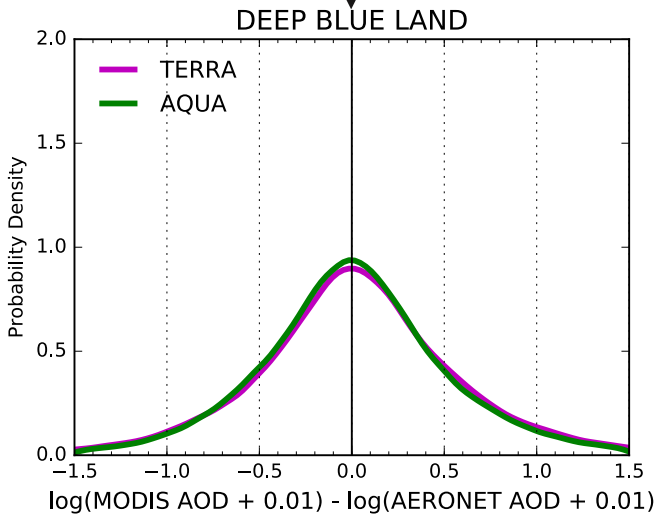
LAND - Dark Surface

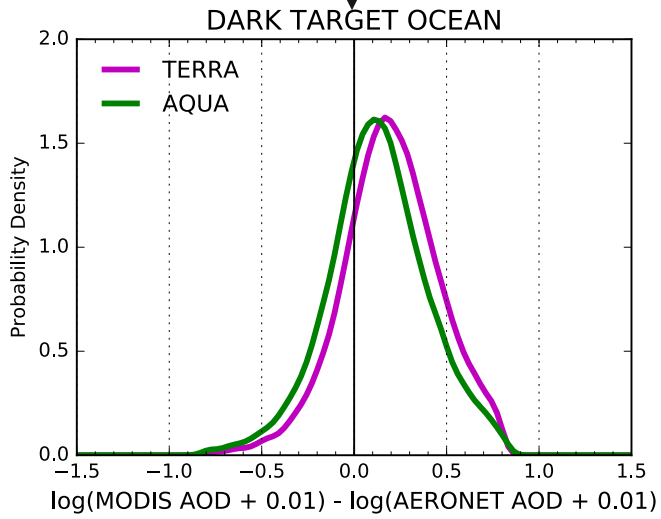
LAND - Bright Surface

OCEAN

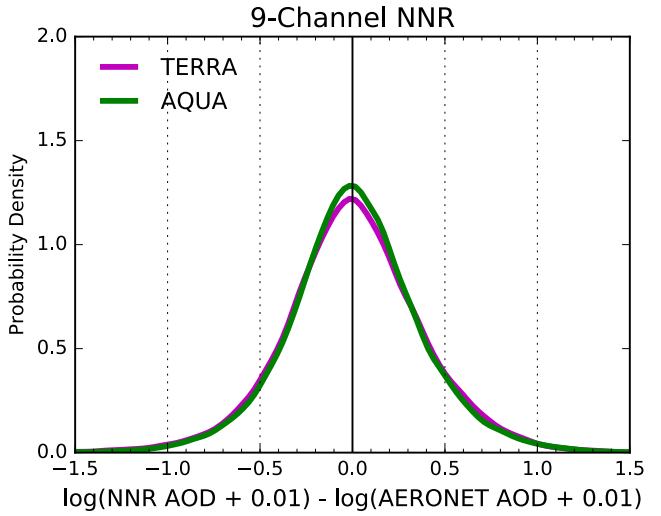
MODIS Standard Retrievals →

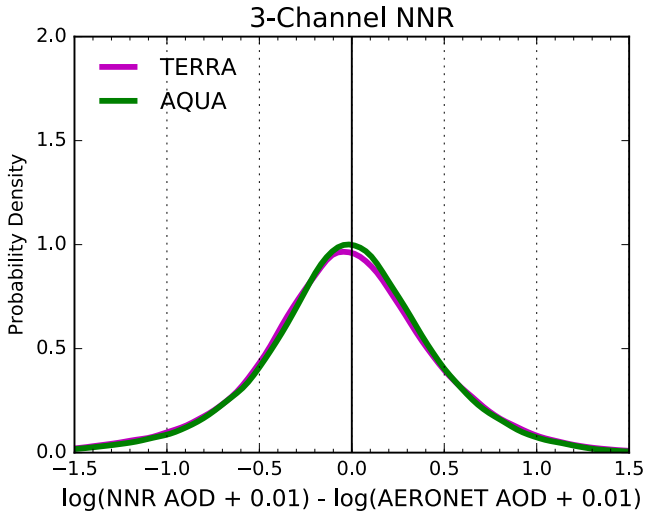


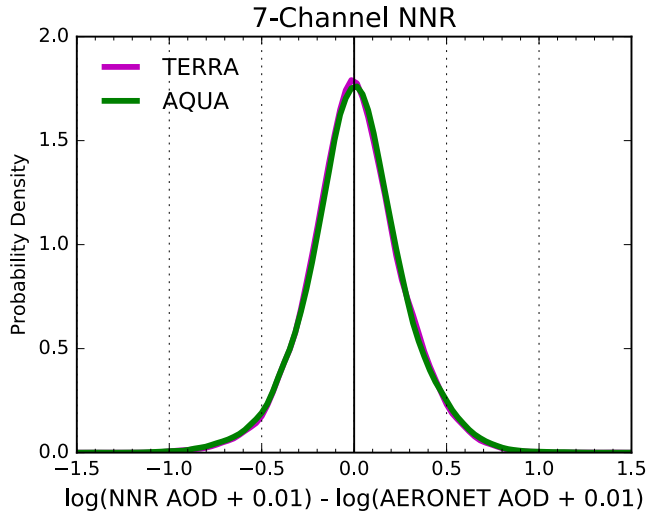




Neural Network Retrievals →

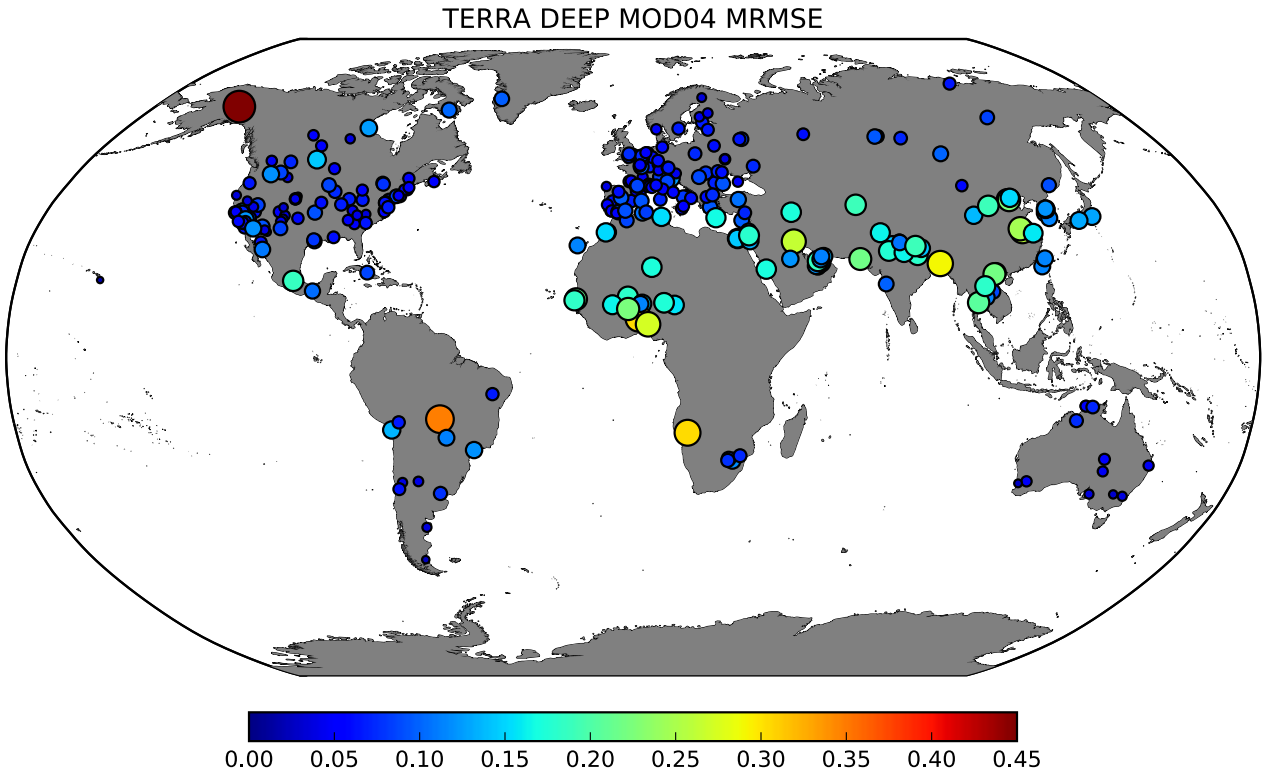


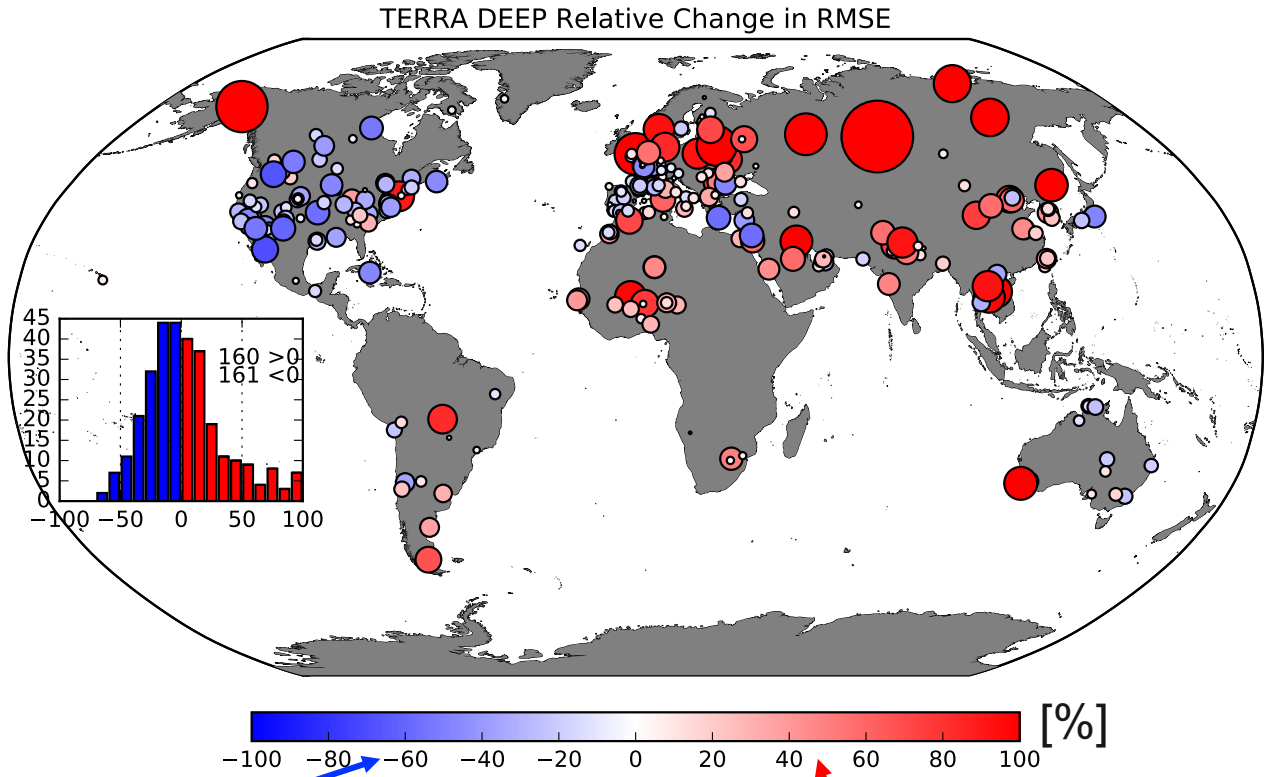




LAND - Bright Surface

NNR Testing



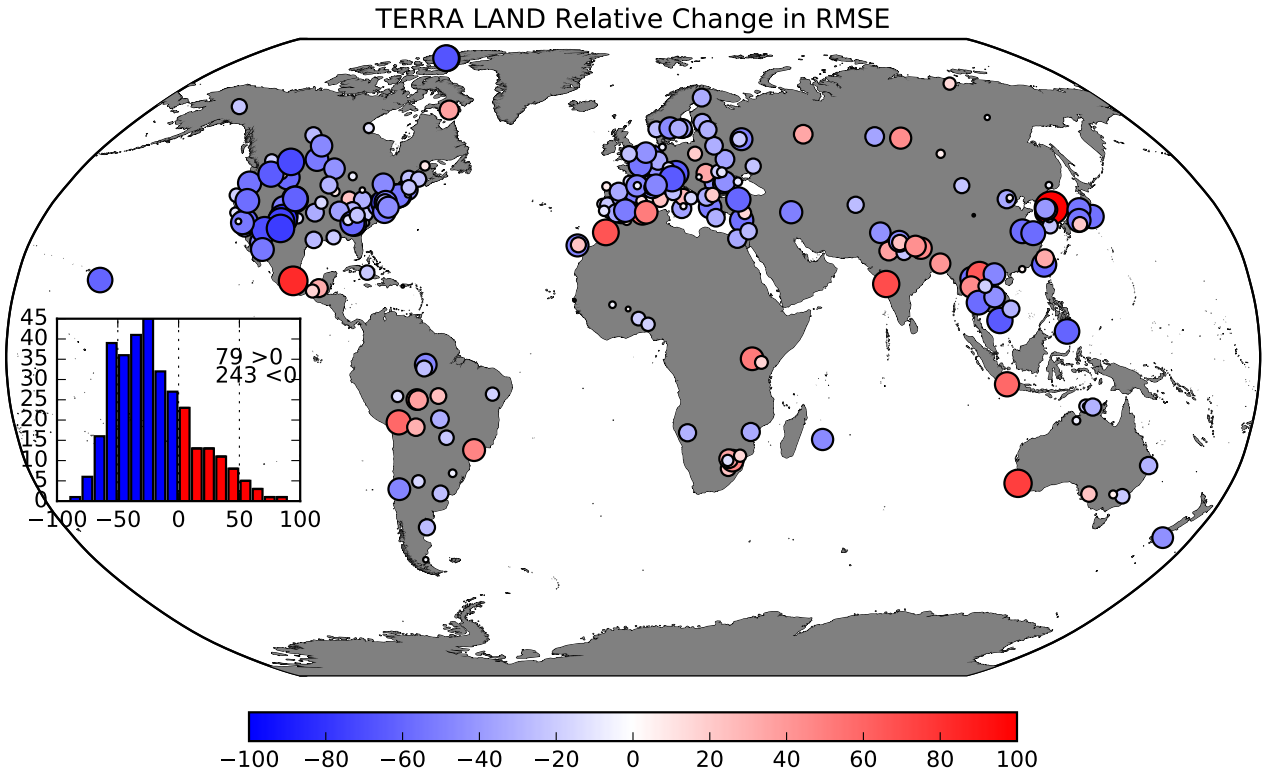


NNR RMSE < MODIS Standard Retrieval

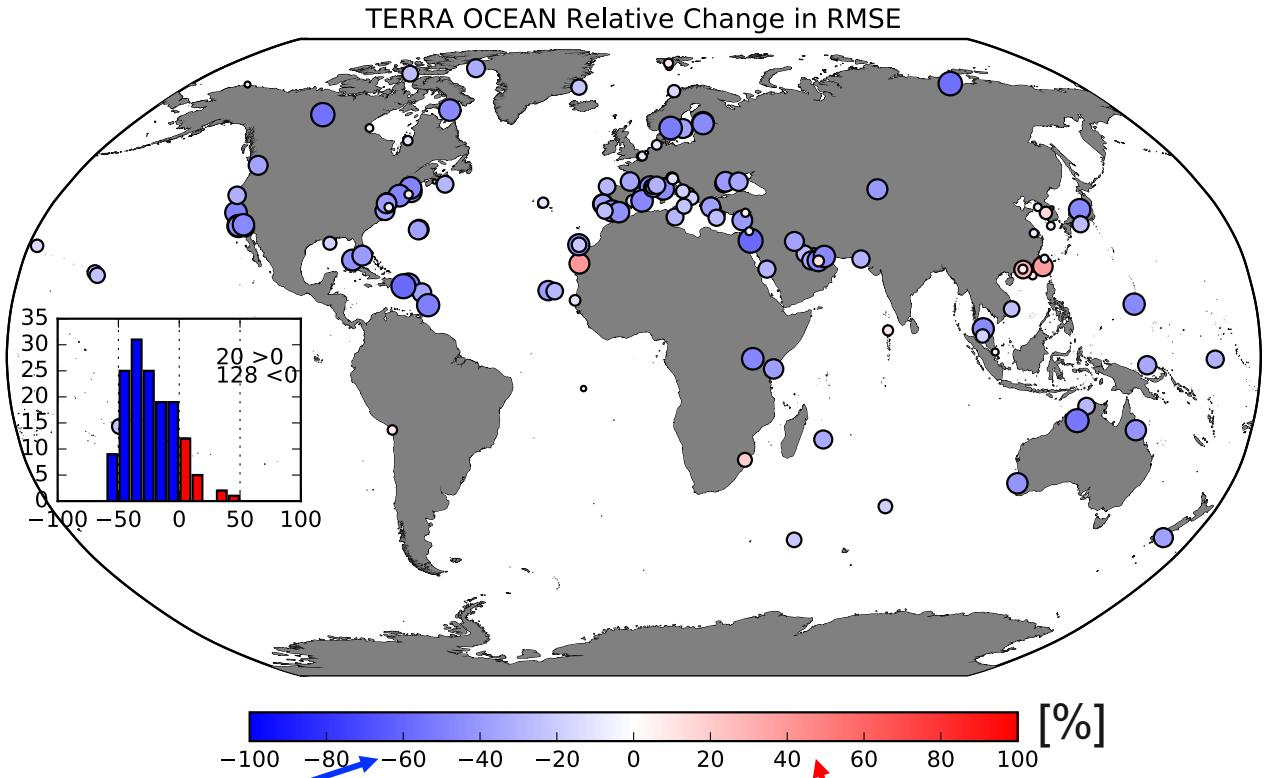
NNR RMSE > MODIS Standard Retrieval

NNR Testing

LAND - Dark Surface



OCEAN

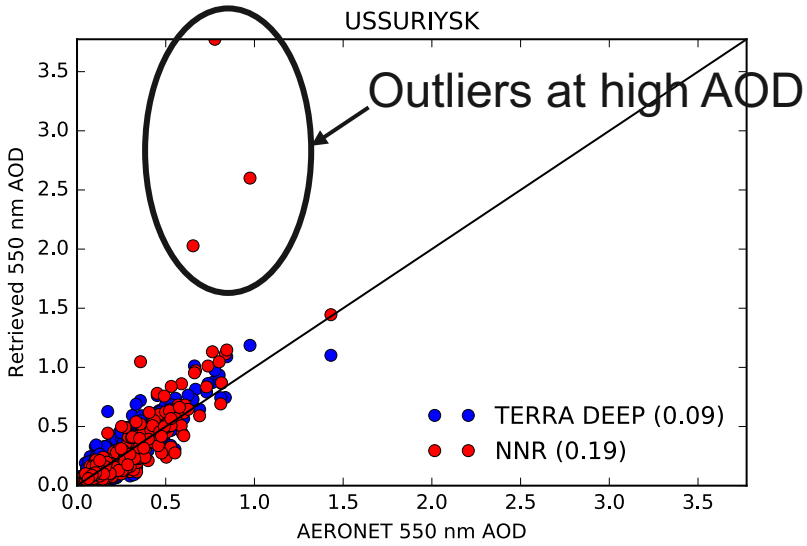


NNR RMSE < MODIS Standard Retrieval

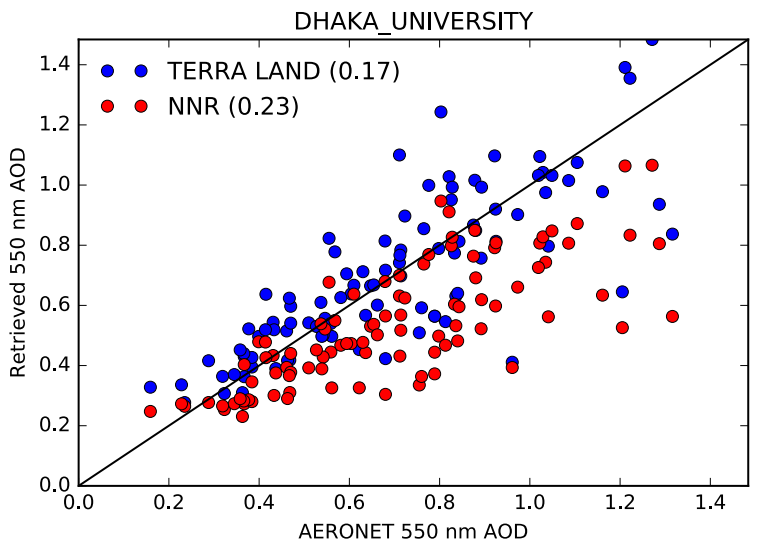
NNR RMSE > MODIS Standard Retrieval

NNR Testing at Some Individual Sites

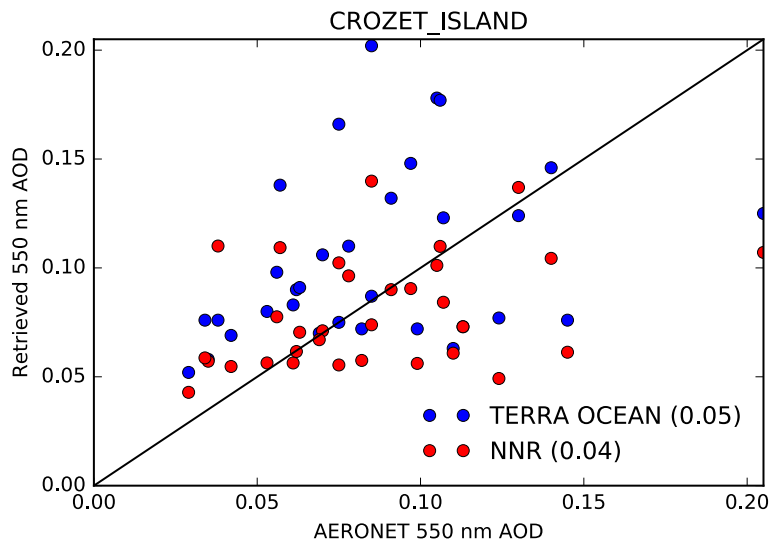
LAND - Bright Surface

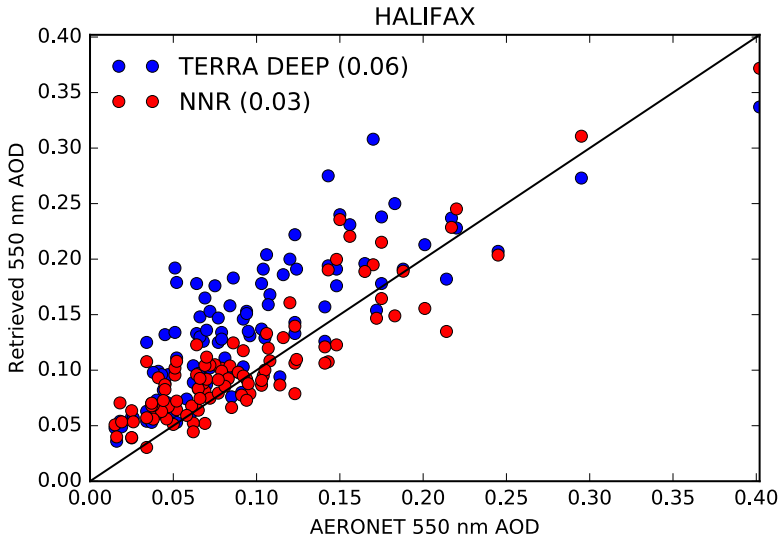


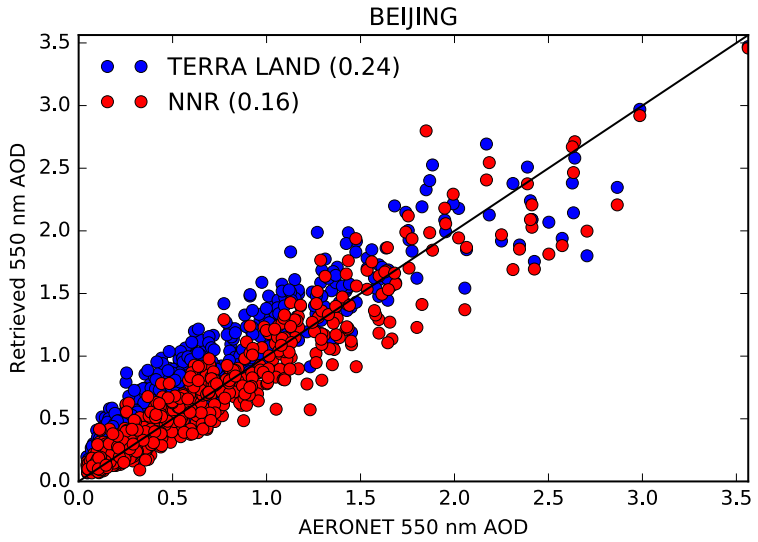
LAND – Dark Surface

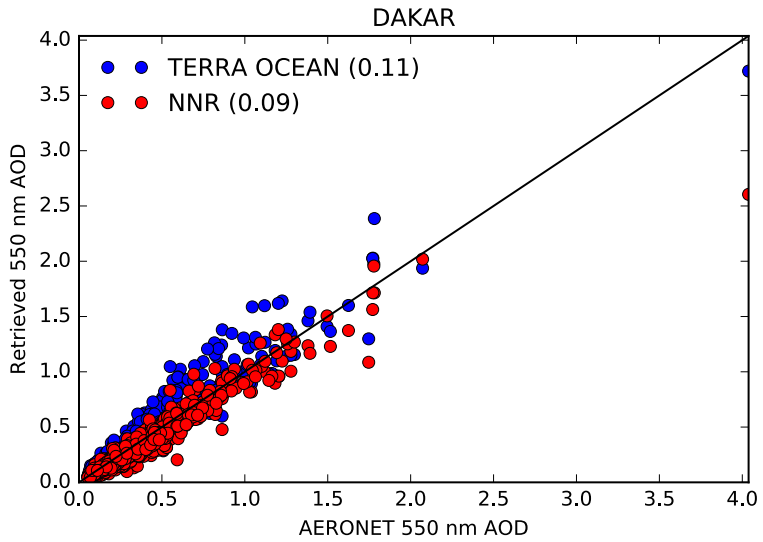


OCEAN









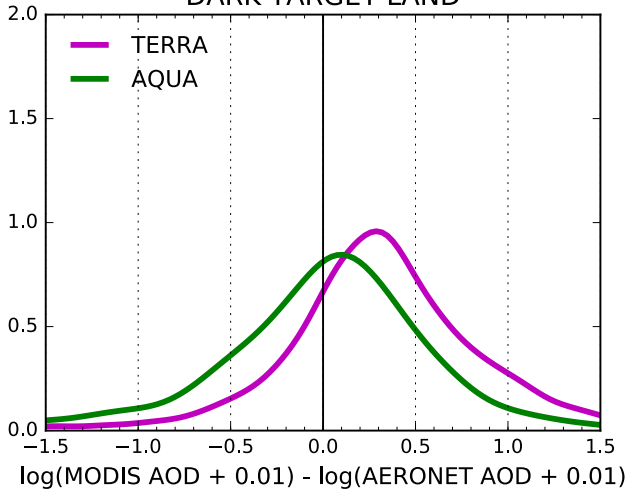
NNR Validation (Observations after 2015)

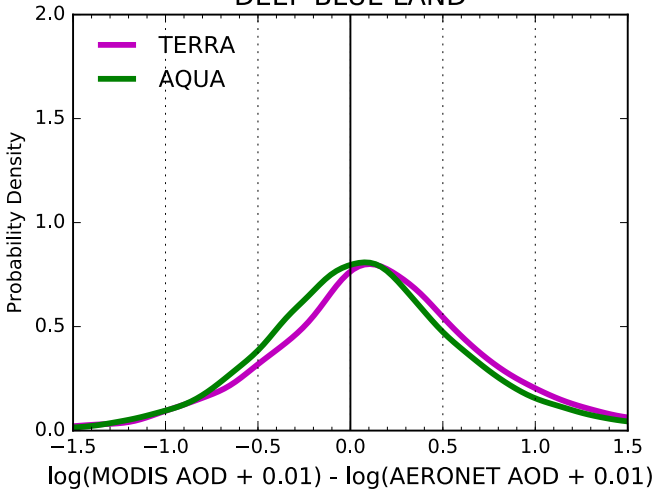
LAND - Dark Surface

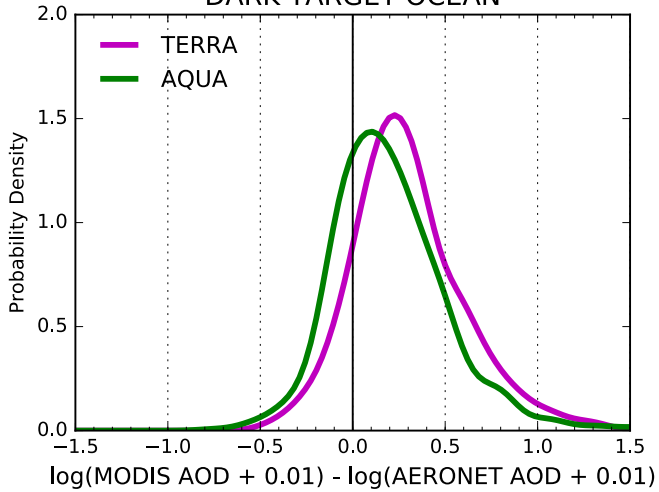
LAND - Bright Surface

OCEAN

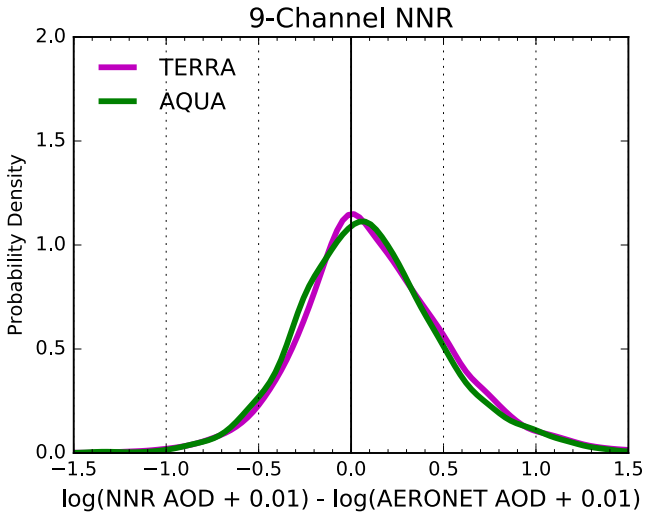
MODIS Standard Retrievals →

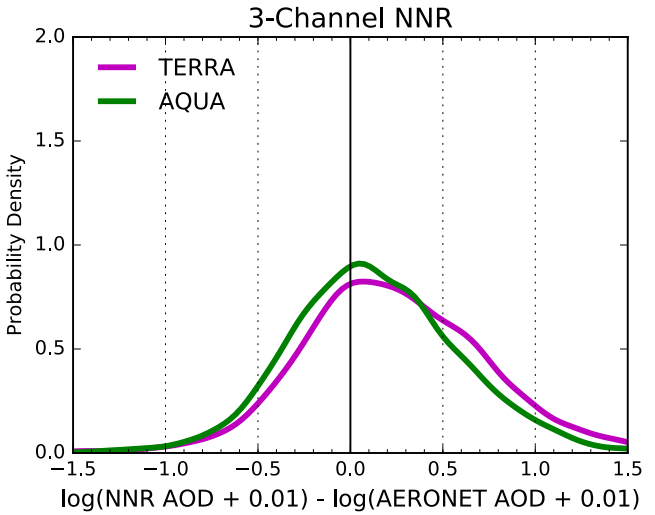


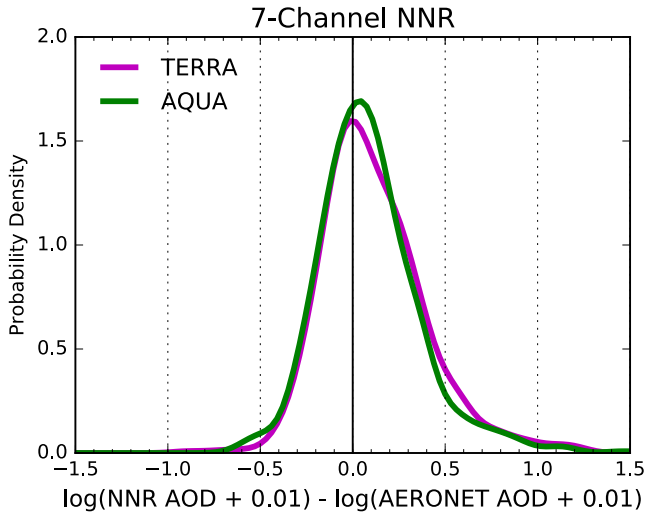




Neural Network Retrievals →

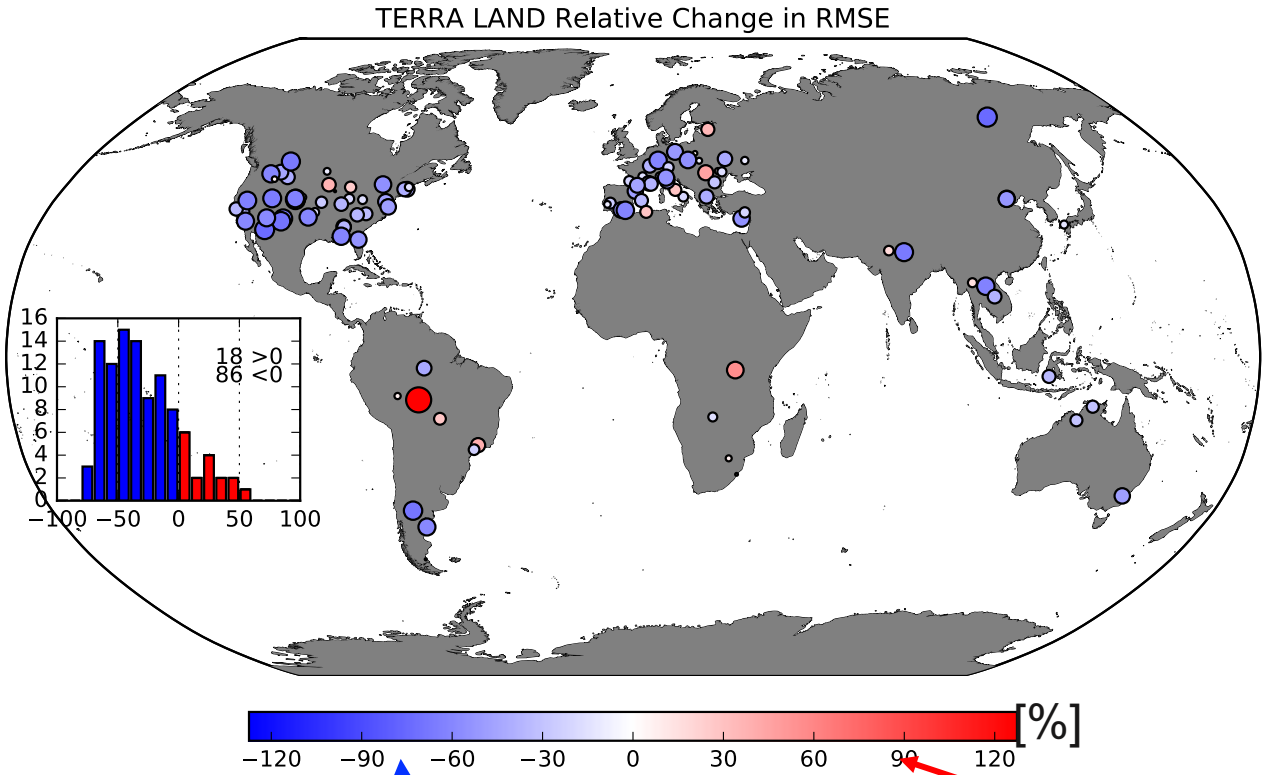




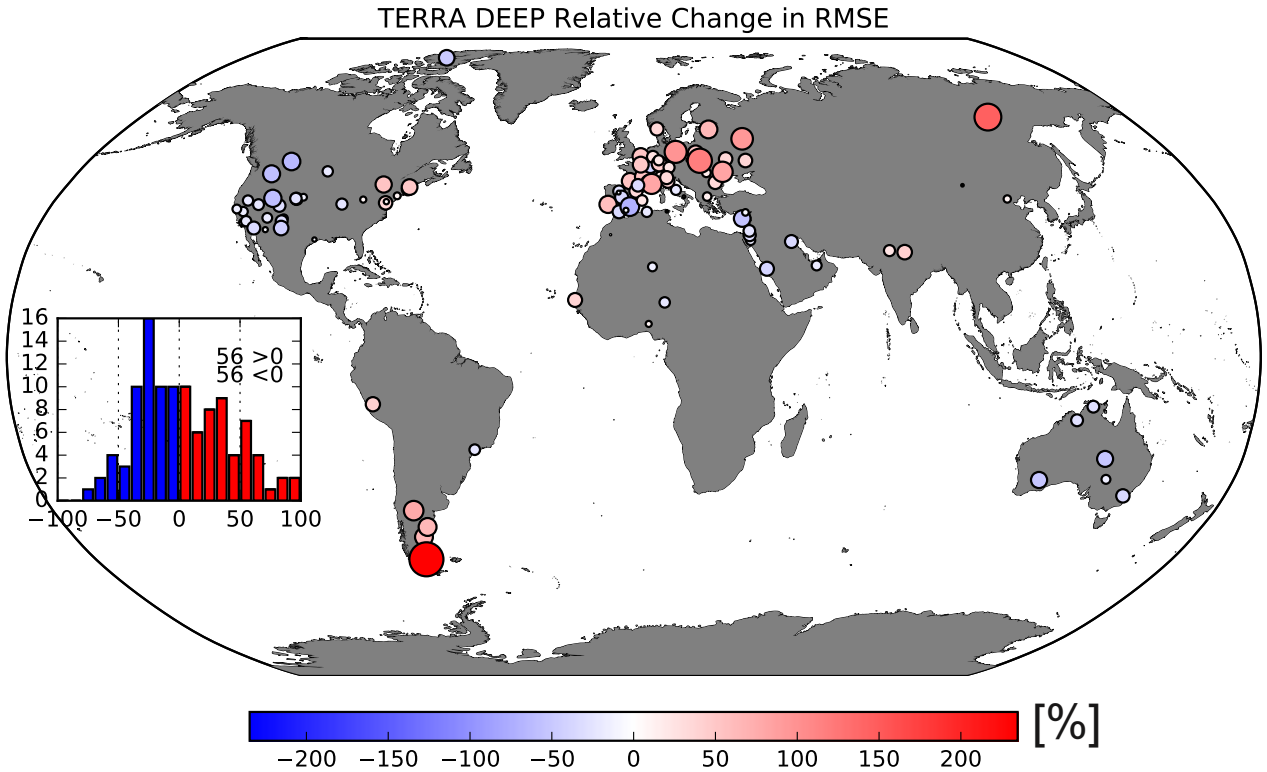


NNR Validation

LAND - Dark Surface



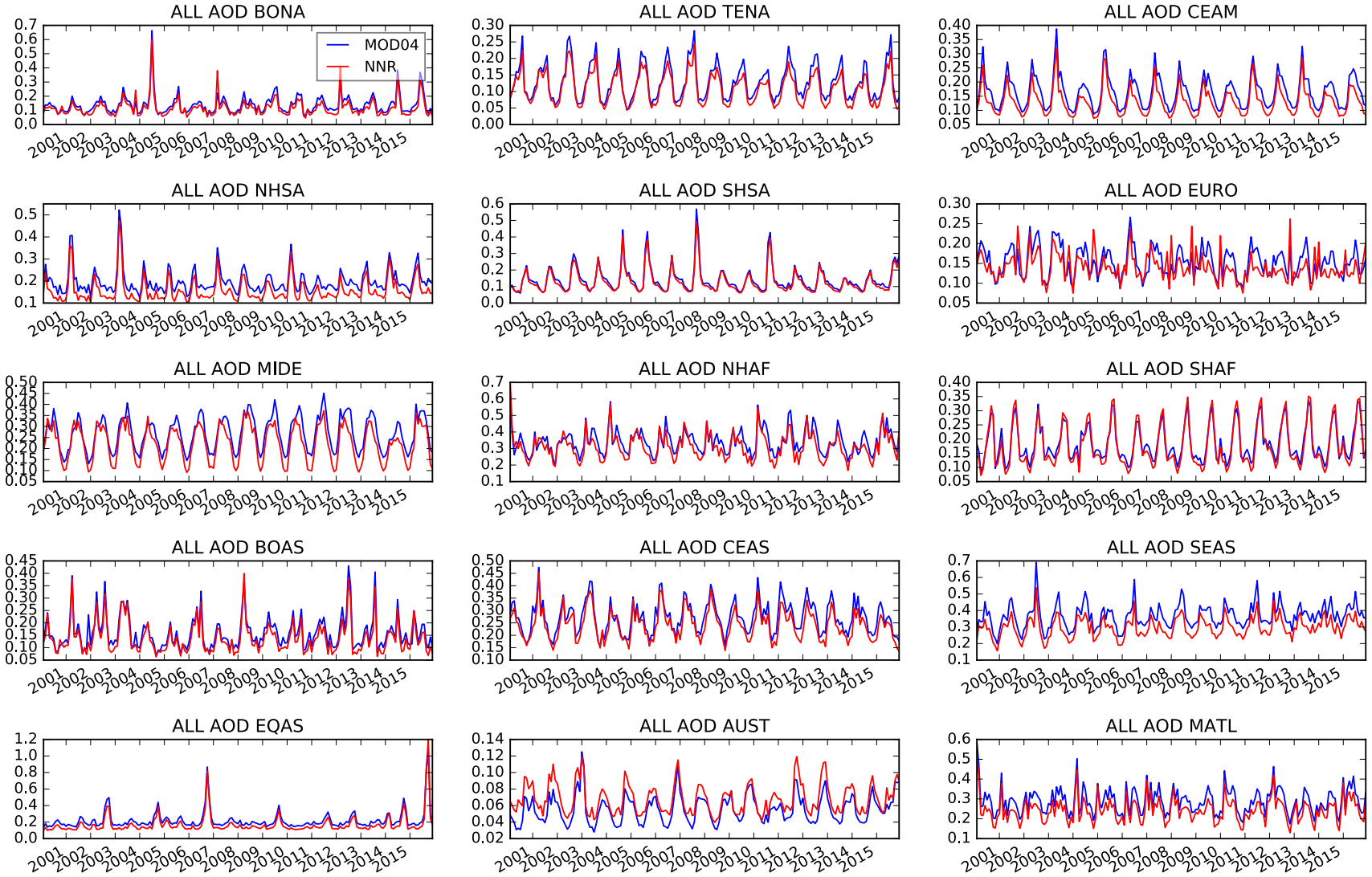
LAND - Bright Surface



NNR RMSE < MODIS Standard Retrieval

NNR RMSE > MODIS Standard Retrieval

NNR Applied to Entire MODIS Time Series: TERRA

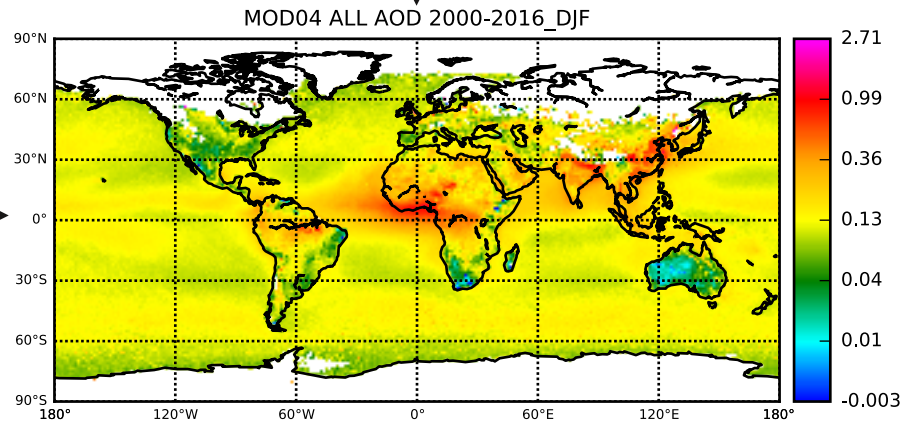


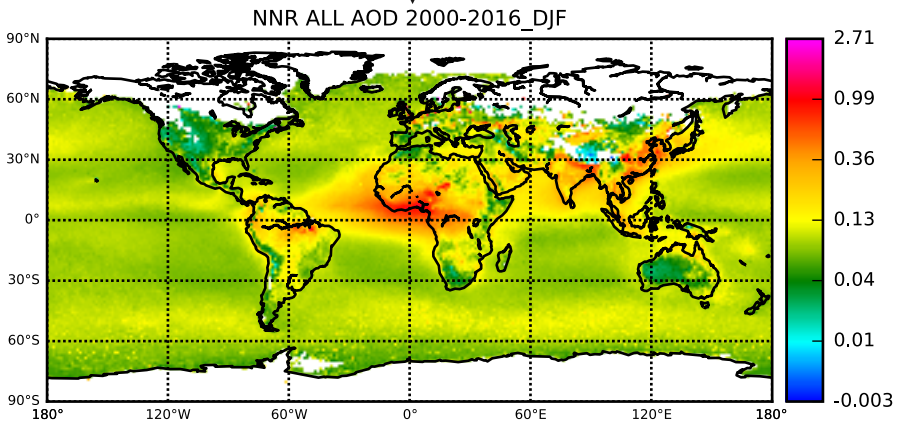
NNR Applied to Entire MODIS Time Series: TERRA Climatology

MODIS Standard Retrieval

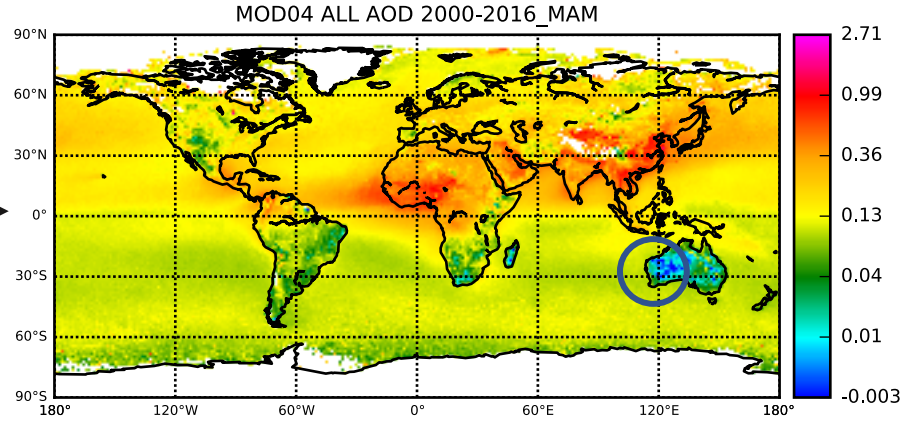
Neural Network Retrieval

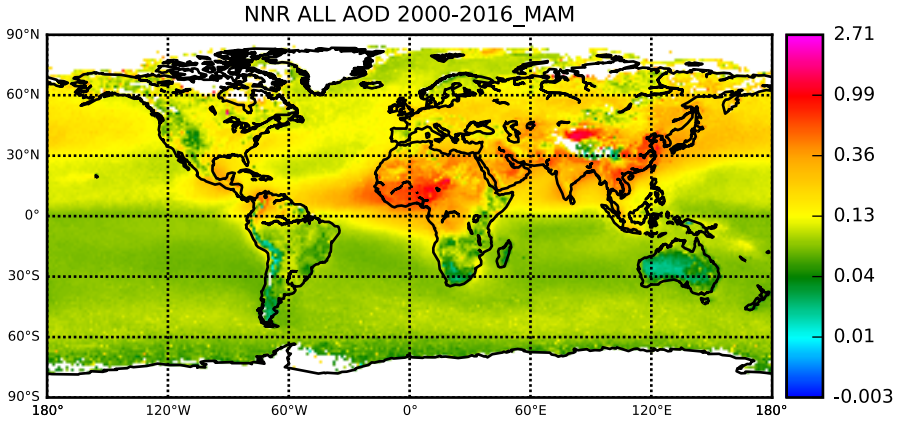
DJF





MAM



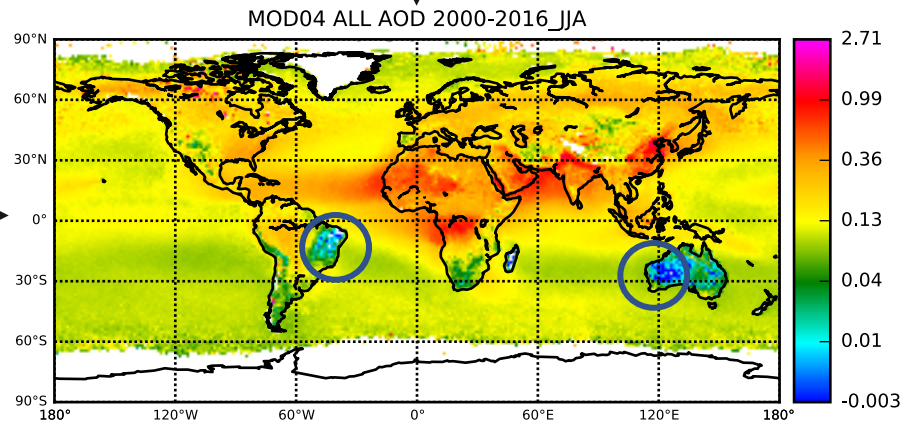


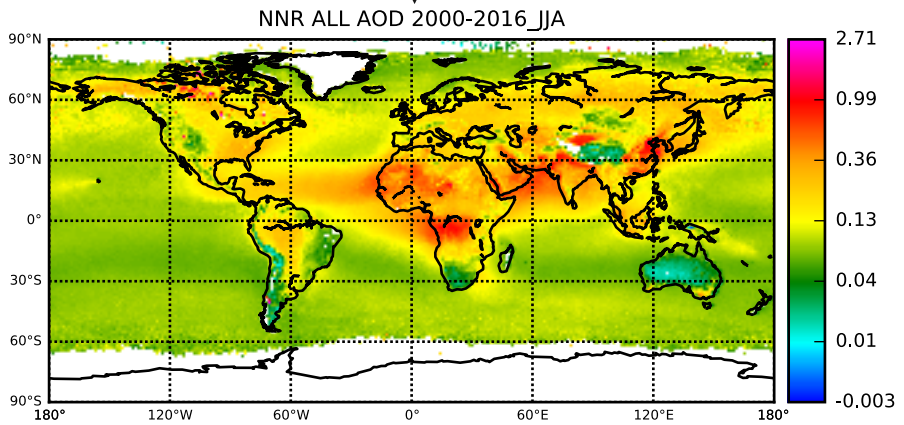
NNR Applied to Entire MODIS Time Series: TERRA

MODIS Standard Retrieval

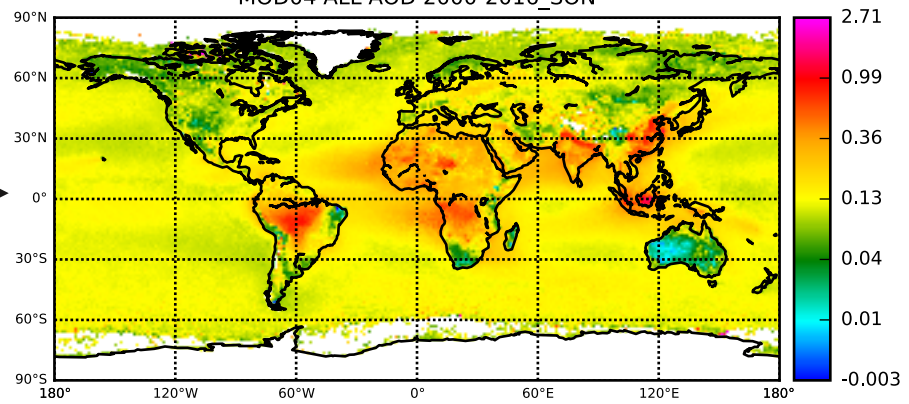
Neural Network Retrieval

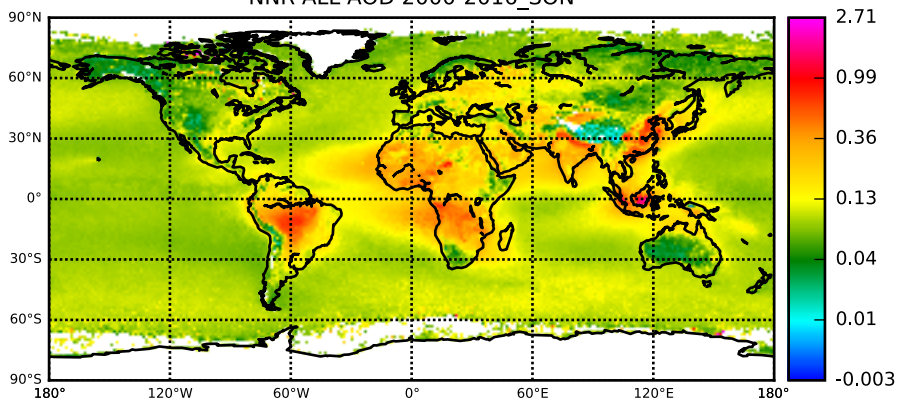
JJA





SON





Summary & Outlook

- The NNR provides a way to homogenize the AOD observing system for data assimilation
- Validation with independent data indicates that the error characteristics of the NNR are stable [for the most part]

- Future Work
  - Validate over open ocean with Marine Aerosol Network (MAN) Observations
  - Periodic retraining for near real-time application
    - Do calibration drifts require multiple networks?
  - Improve training data set
    - More QA filtering, balancing
  - Multiple targets
    - Multi-channel AOD
    - SSA
    - Angstrom Exponent
  - Geostationary observations