

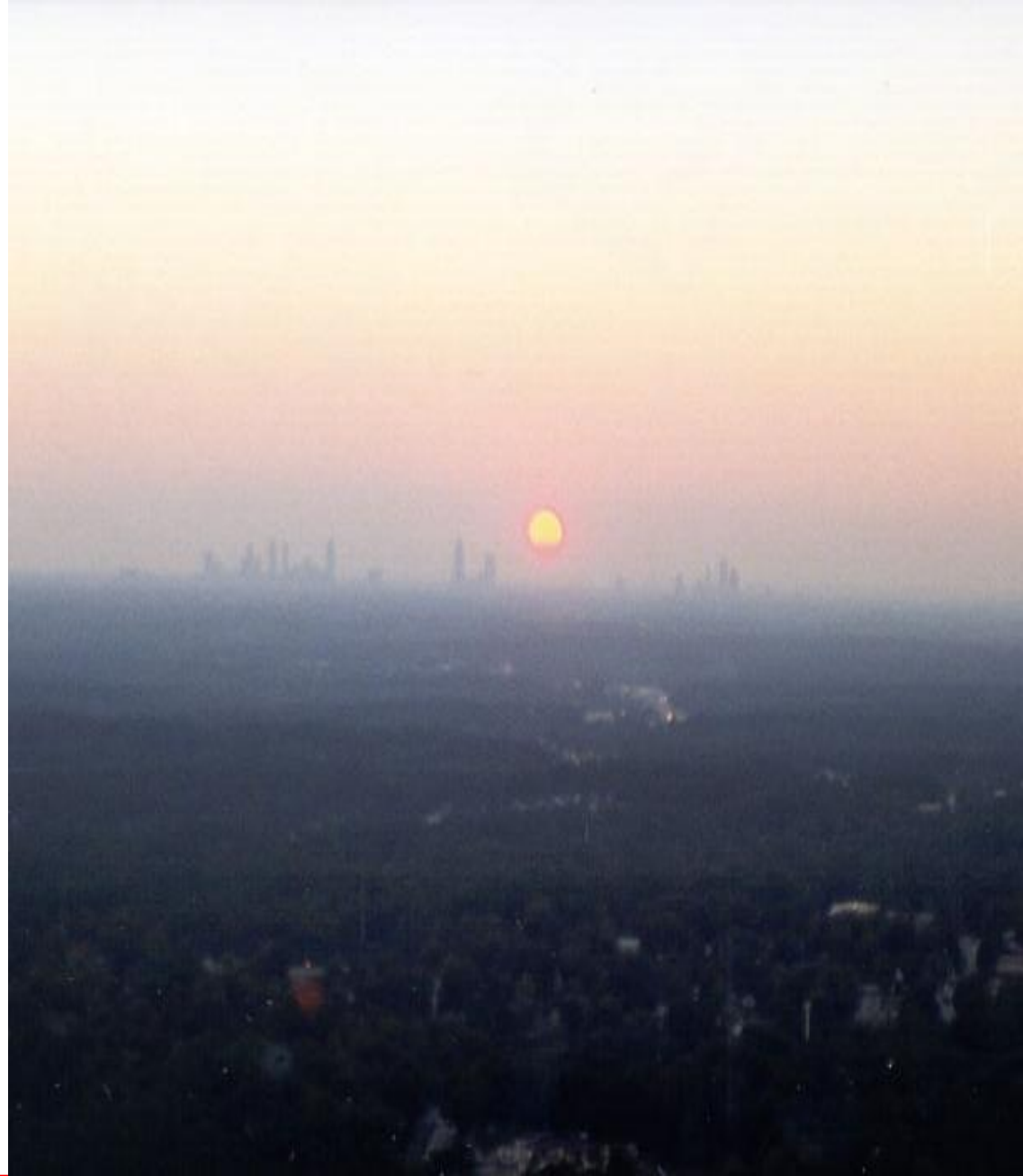
Predicting Aerosol Density for Climate Modeling Using Data Mining Techniques

***Vladan Radosavljevic, Debasish Das, Kosta Ristovski,
Qiang Lou, Haidong Shi***

***Advisors: Dr. Zoran Obradovic, Dr. Slobodan Vucetic,
Center for Information Science and Technology, Temple University***

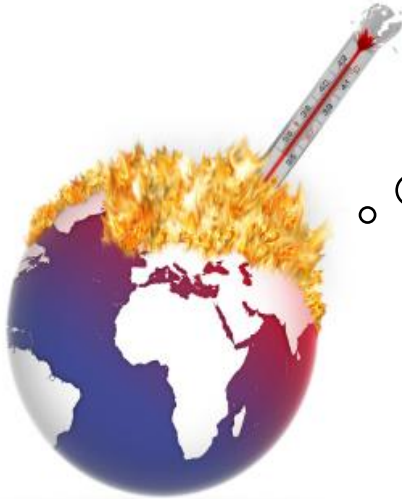
Aerosols

- Minute particles suspended in the atmosphere
- Scatter and absorb sunlight
- Reduce visibility (haze)
- Redden sunrises and sunsets



Are aerosols important?

Influence on
future
climates



Less sunlight
to vegetation



May be more
important than
ozone in causing
cancer problems

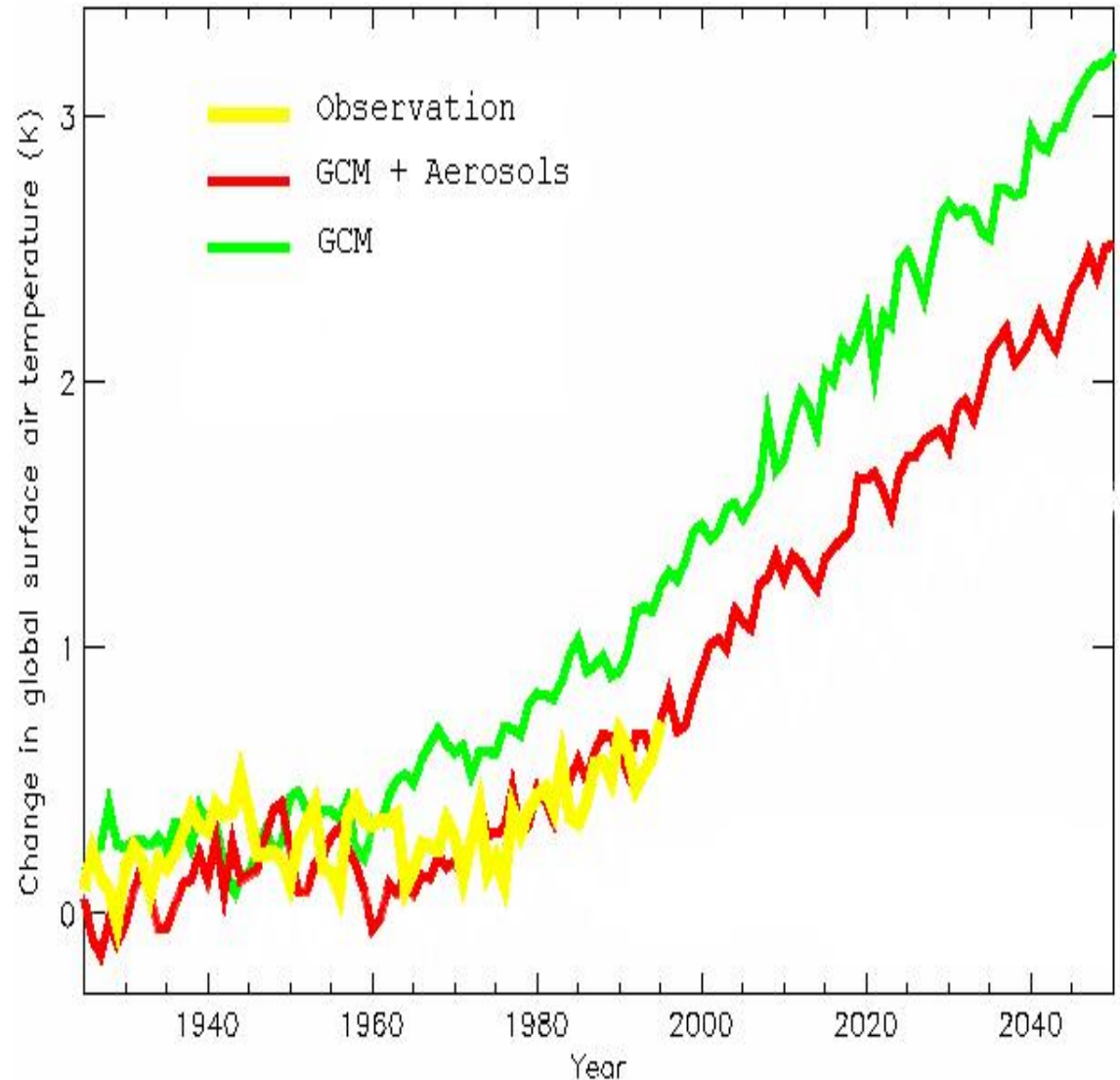


Less clouds
and
precipitation



Role of Aerosols in Climate Modeling

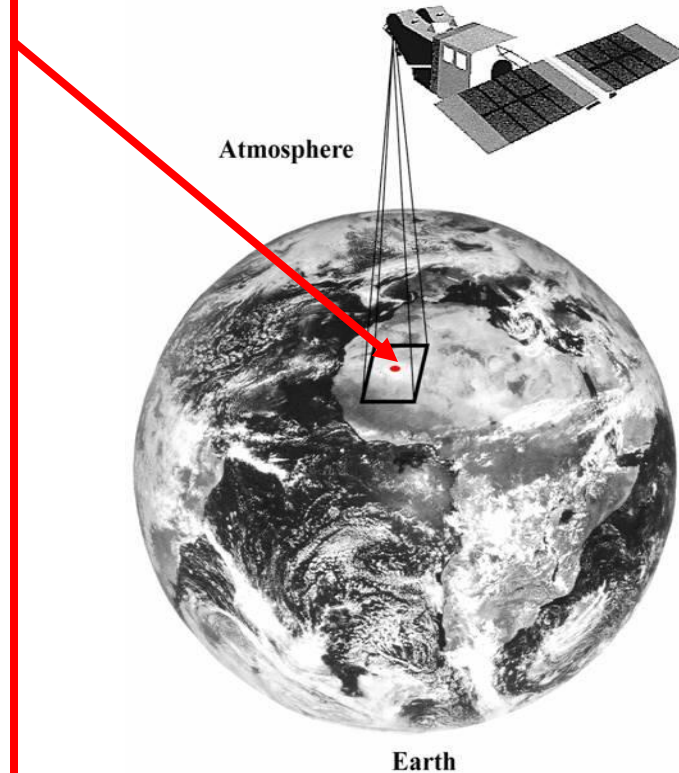
- The role of aerosols in climate was not widely recognized until 1990s
- Aerosols were identified as a central missing component in most general circulation models (GCMs) that simulated climate changes
- After introducing aerosols, model-simulated climate changes became more realistic



How to Measure Aerosols?

Two types of measurements

- Ground based - AERONET
- Satellite based – MODIS



	Ground-based	Satellite-based
Spatial Coverage	poor	high
Temporal Coverage	high	medium
Cost	high	low
Accuracy	high	poor

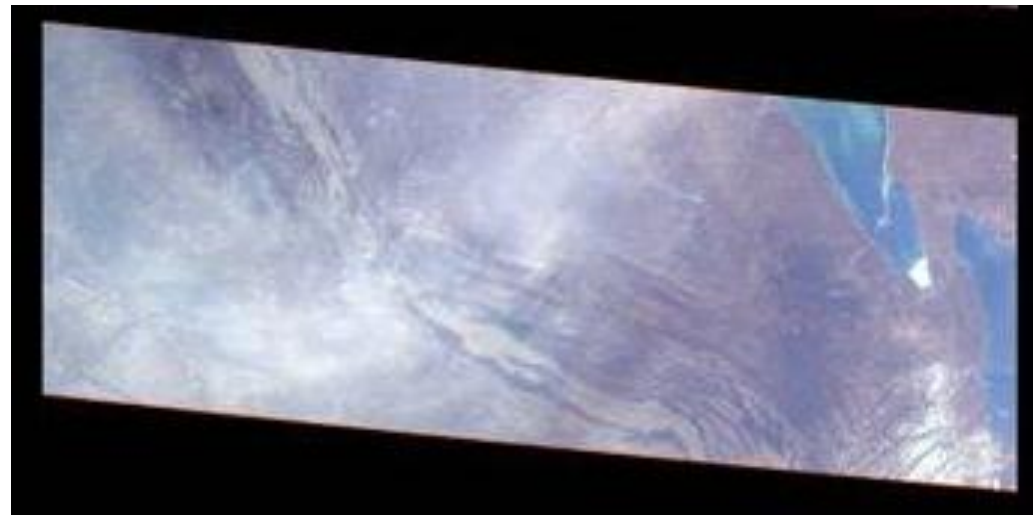
Satellite observations offer the potential of achieving the global coverage that is required to understand aerosols more completely

What to Measure About Aerosols?

- The most important characteristic of aerosols is Aerosol Optical Depth (AOD)
- AOD indicates the amount of depletion that light undergoes as it passes through the atmosphere
- Related to visibility and blurriness of satellite images



Small AOD ~ 0



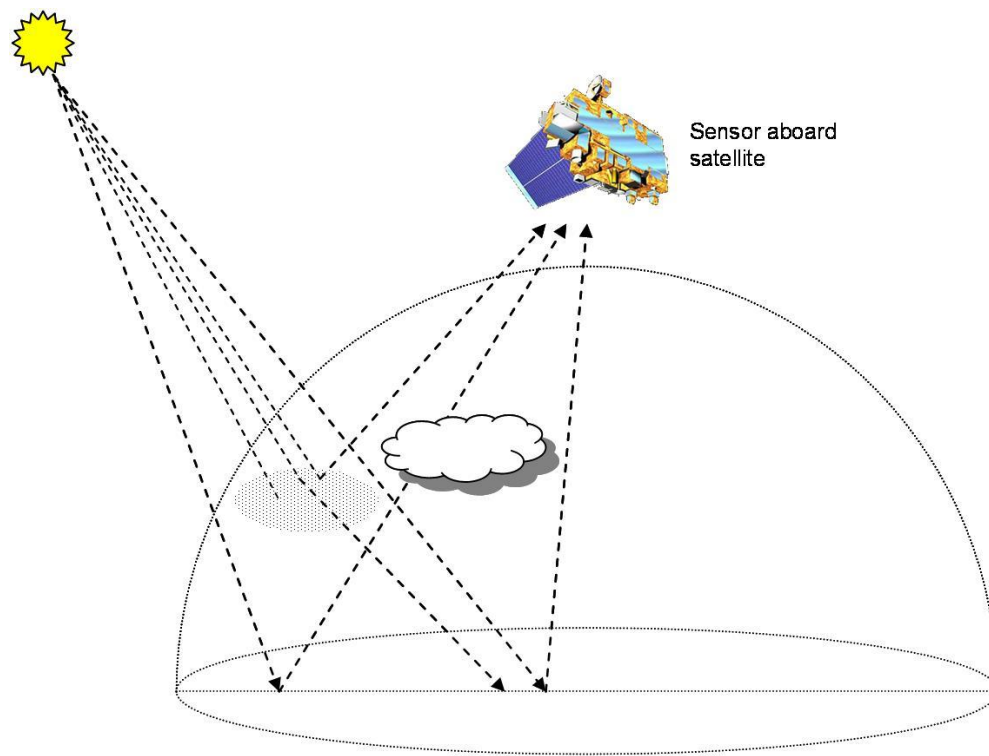
Large AOD > 1

The Satellite Signal Is Complicated!

- Noise caused by:

- scattering
- bright surfaces
- clouds
- reflections from neighboring area

complicate the AOD prediction process.



Domain Approach to Estimate AOD

- Based on physical principles
- Manually tuned
- Observed reflectance is matched to the simulated values stored in lookup tables
- Lookup tables are generated by the model that estimates reflectance given the aerosol type and amount
- The new algorithm, MODIS Collection 005 (C005), has been developed by including more realistic aerosol models [1]

$$\rho_{0.66}^s = f(\rho_{2.12}^s) = \rho_{2.12}^s * slope_{0.66/2.12} + yint_{0.66/2.12}$$

and

$$\rho_{0.47}^s = g(\rho_{0.66}^s) = \rho_{0.66}^s * slope_{0.47/0.66} + yint_{0.47/0.66}$$

$$slope_{0.66/2.12} = slope_{0.66/2.12}^{NDVI_{SWIR}} + 0.002\Theta - 0.27,$$

$$yint_{0.66/2.12} = 0.00025\Theta + 0.033,$$

$$slope_{0.47/0.66} = 0.49, \text{ and}$$

$$yint_{0.47/0.66} = 0.005$$

$$slope_{0.66/2.12}^{NDVI_{SWIR}} = 0.48; NDVI_{SWIR} < 0.25,$$

$$slope_{0.66/2.12}^{NDVI_{SWIR}} = 0.58; NDVI_{SWIR} > 0.75$$

$$slope_{0.66/2.12}^{NDVI_{SWIR}} = 0.48 + 0.2(NDVI_{SWIR} - 0.25); 0.25 \leq NDVI_{SWIR} \leq 0.75$$

Complicated?!

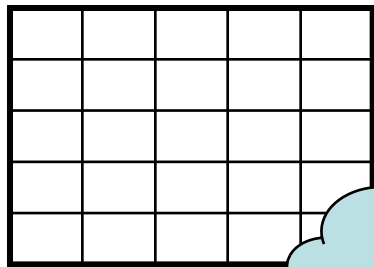
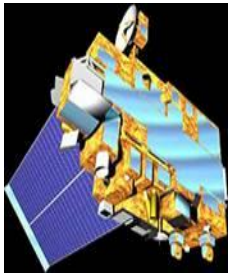
Too many assumptions!



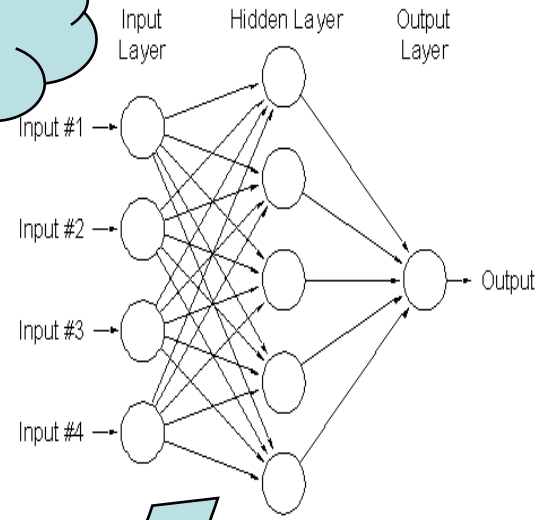
Our Approach

Data: satellite + ground based
matched in space and time

Satellite

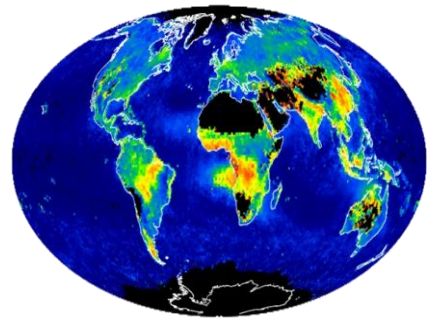


Data mining!



Neural Network

AOD
prediction



Domain-driven vs. Data Mining Approach

Prediction task

- given a set of tuples (\mathbf{x}, \mathbf{z}) where
 - \mathbf{x} are observed radiances at different wavelengths,
 - \mathbf{z} are ancillary attributes (angles, elevation, cloud cover mask)
- determine y where
 - y is underlying geophysical parameter (AOD)

Domain-driven approach

- develop forward-simulation model $\mathbf{x} = F(\mathbf{y}, \mathbf{z})$ based on the domain knowledge
- construct lookup table $LUT = \{(\mathbf{x}_i, \mathbf{z}_i, \mathbf{y}_i)\}$ representing the most common conditions
- use LUT to find the most similar entry to the observed pair (\mathbf{x}, \mathbf{z})

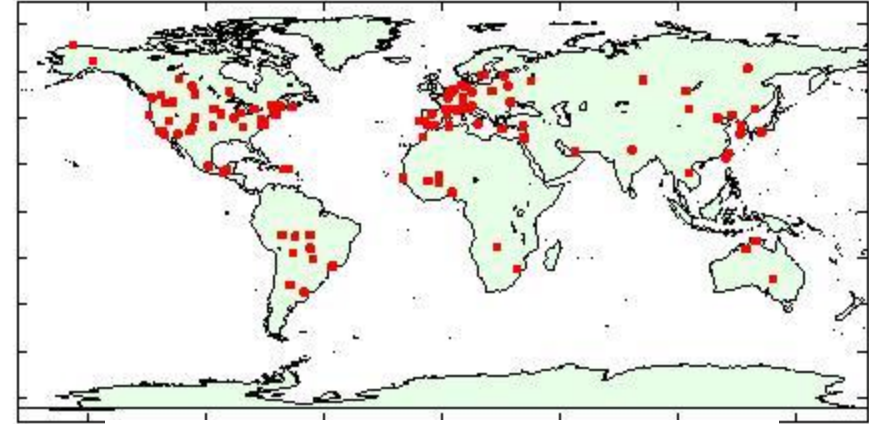
Data Mining approach

- merge MODIS data and AERONET AOD to build a data set $D = \{(\mathbf{x}_i, \mathbf{z}_i, \mathbf{y}_i)\}$
 - train a predictor (e.g. Neural Net NN) $\mathbf{y} = NN(\mathbf{x}, \mathbf{z})$ from D
- this is completely data-driven approach – no need for domain knowledge**

Dataset

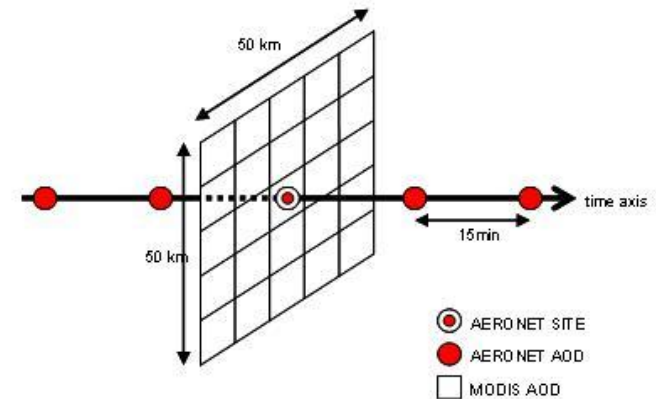
Data

- AERONET AOD in first eight months of 2005 from 125 sites distributed globally
- Collocated MODIS reflectance



Merging

- spatial overlap: 50x50km² around AERONET site
- temporal overlap: +/- 30 min of the satellite over flight



Data description

- 5,074 tuples are collected

Index	Description
1-7	Mean reflectance at 7 wavelengths
8-14	Std. dev. of reflect. at 7 wavelengths
15-19	Angles
20	Surface elevation

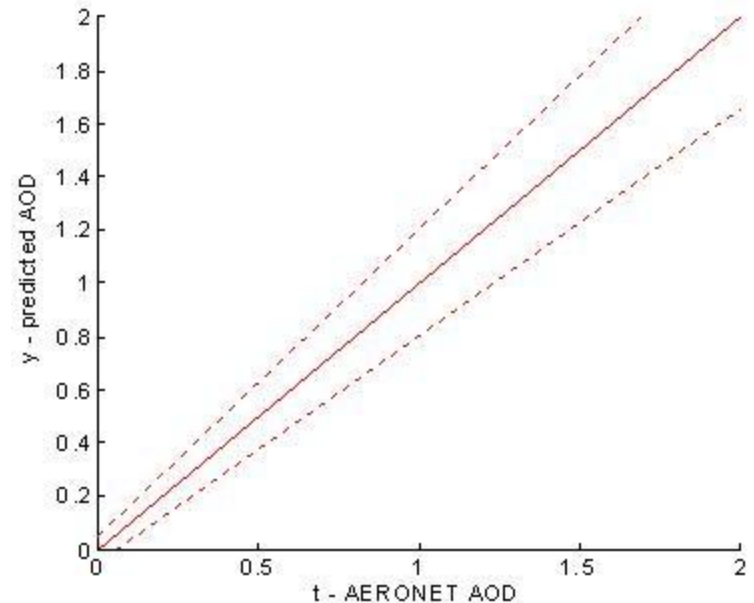
Our Results

Evaluation measures

- Domain-based boundaries for AOD prediction error rate are [1]:

$$|y-t| \leq 0.05+0.15t$$

t - AERONET AOD,
 y - AOD predicted using
satellite observations



- **Predictor should be much more accurate in predicting small AOD than in predicting large AOD.** In this sense, we defined a new accuracy measure as *mean squared relative error* (MSRE):

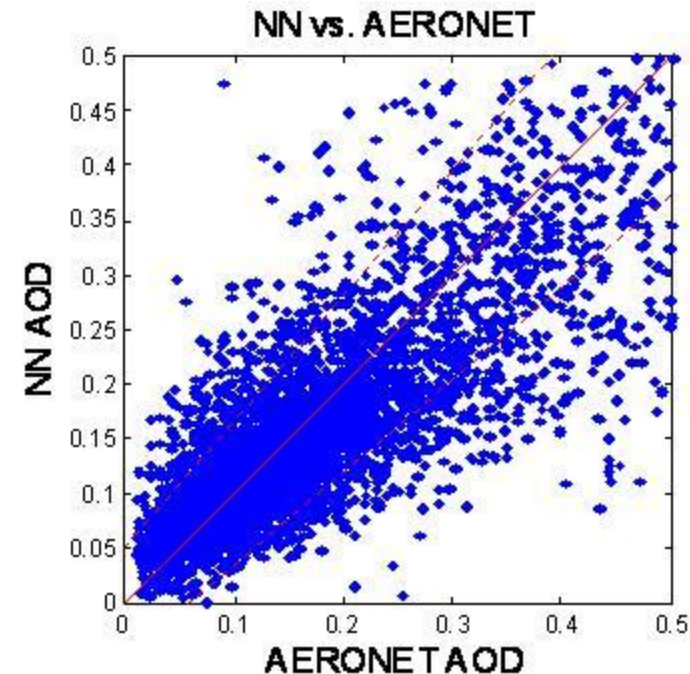
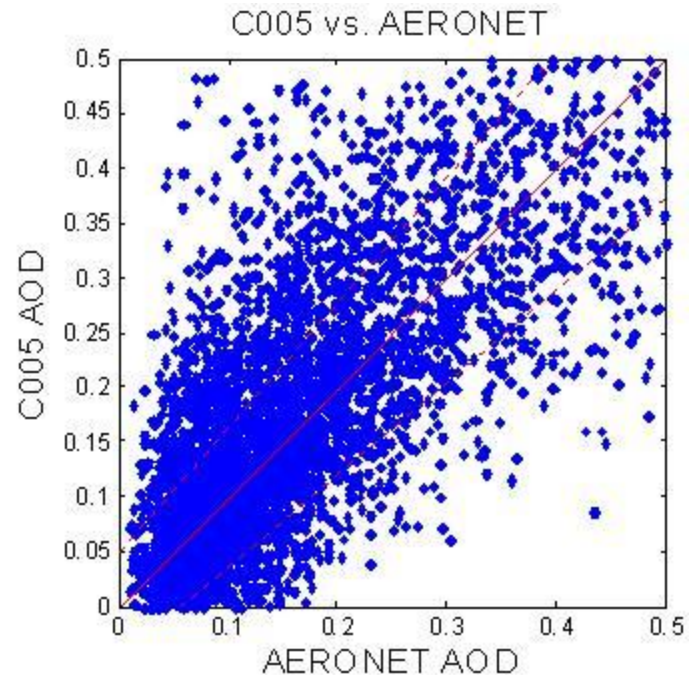
$$MSRE = \frac{1}{N} \sum_{i=1}^N \left[\frac{y_i - t_i}{0.05 + 0.15t_i} \right]^2, \text{ N is number of tuples}$$

- We also report *mean square error* (MSE), R^2 , *correlation* between t and y (CORR) and *fraction* of predictions within expected boundaries (FRAC)

Learning with Adaptive Cost Function *published*

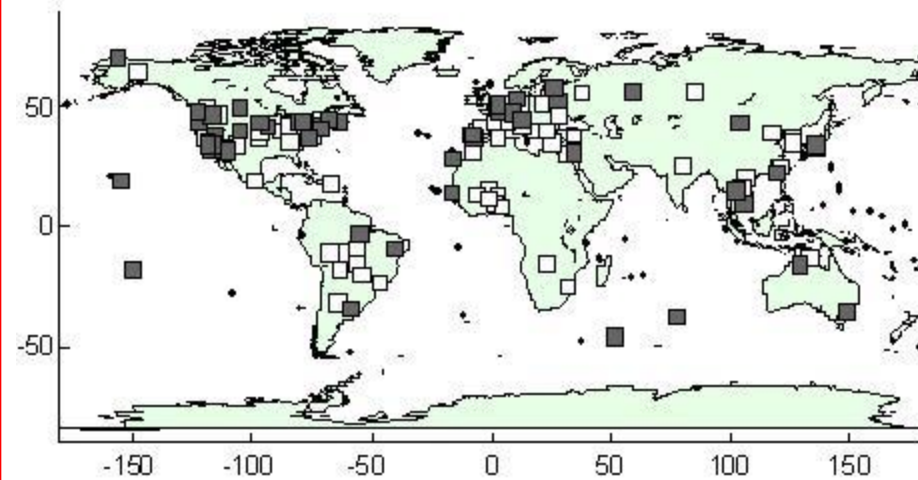
- We proposed a committee of Neural Networks (NN) trained to optimize MSRE which is more accurate in predicting AOD than C005 considering all measures [2]

	C005	NN
MSE	0.018	0.012
R ²	0.74	0.83
CORR	0.89	0.91
MSRE	1.80	0.88
FRAC	64%	77%

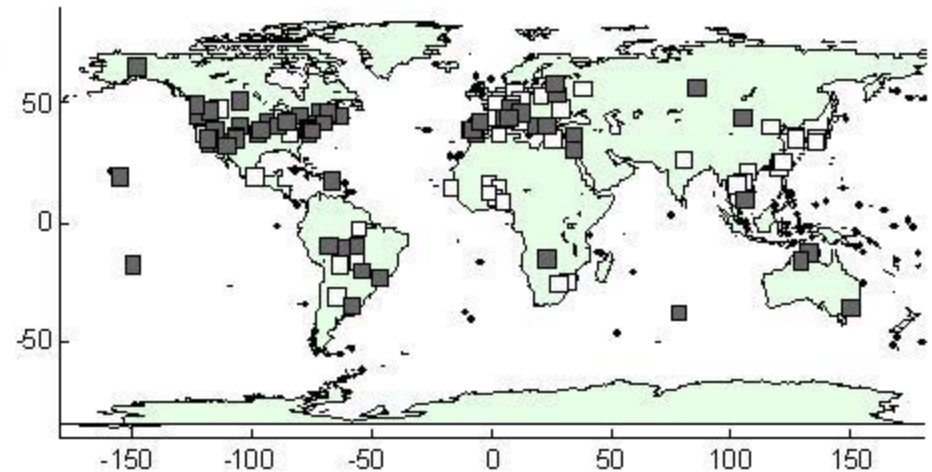


Spatio-temporal Partitioning *in press*

- We developed a competition based method that discovered spatio-temporal partitions of Earth that allowed improved AOD prediction as compared to the C005 algorithm from NASA [3]
- The most accurate results are obtained when the data from each site were split into two temporal subsets and two NNs were competing



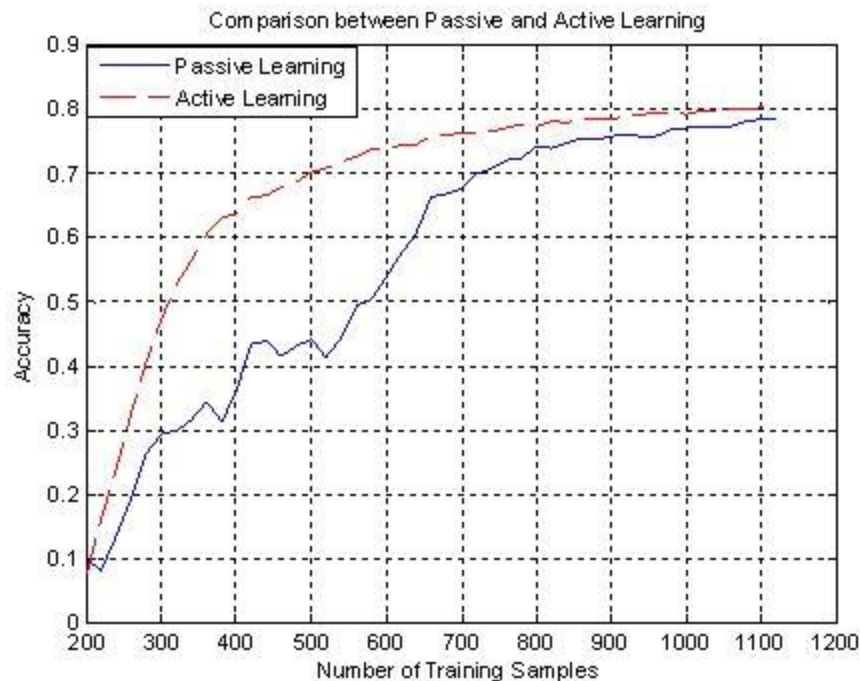
Two partitions for summer-fall months



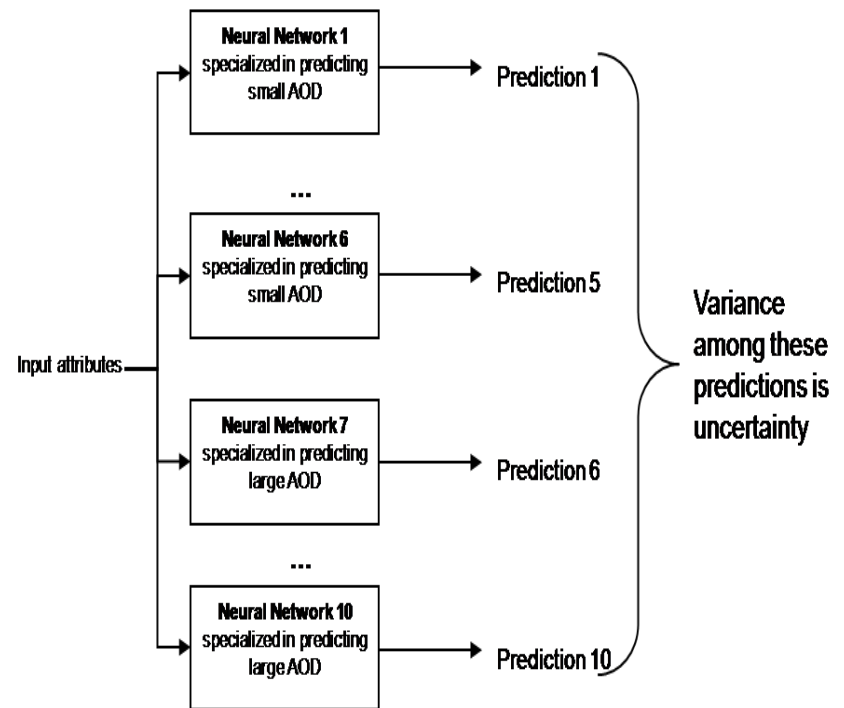
Two partitions for winter-spring months

Method - Query By Committee

- An ensemble of Neural Networks with different cost functions serves as a committee [4]
- A point is regarded as most informative if the committee has maximum disagreement in prediction on that point



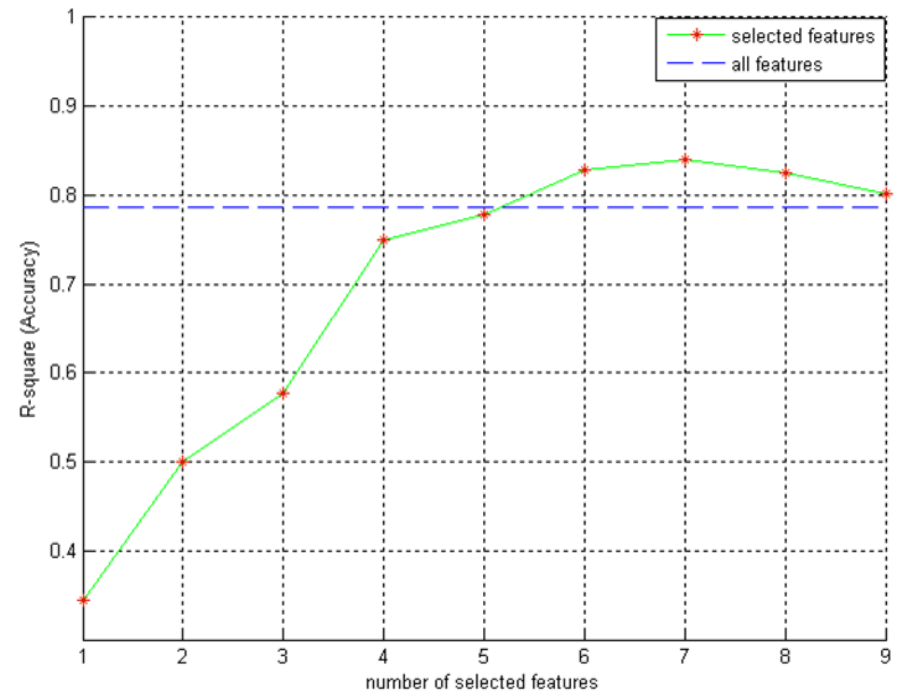
Comparison of learning curves for passive and active learning



Committee of Neural Networks

Informative Features *preliminary results*

- Irrelevant features might reduce prediction accuracy
- We applied 'forward feature selection' to determine the most informative subset of features
- Prediction accuracy improved when subset of 7 features are used instead of entire set of features
- Different seasons have different informative features

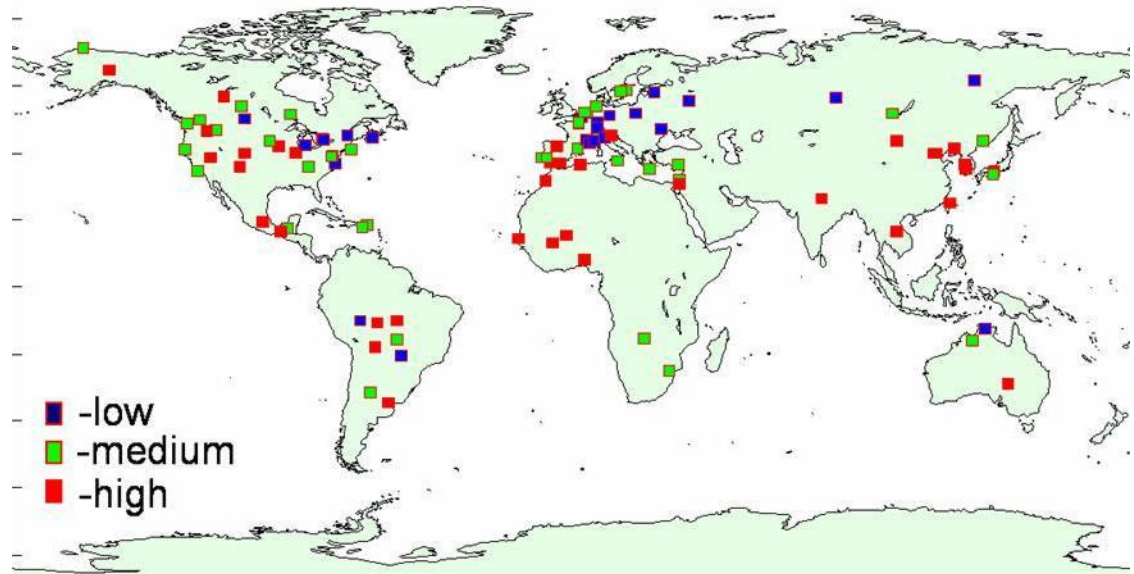


Informative features in specific seasons

spring	summer	winter
reflectance 470nm	reflectance 550nm	solar azimuth
sensor zenith	sd. dev. at 550nm	elevation
solar zenith	solar zenith	solar zenith
reflectance 840nm	sd. dev. at 660nm	sd. dev. at 660nm

Uncertainty in AOD Prediction *preliminary results*

- Uncertainty of Neural Network output is assessed using Bayesian statistics
- In general, low uncertainty is achieved in the East US and Europe which is in correlation with domain expectations as the same industrial type of aerosols is dominant there



Future Work

- Determining length of time intervals automatically in spatial-temporal competition procedure
- Increasing the efficiency of the active learning algorithm
- Feature selection based on different aerosol models
- Exploiting uncertainty information, improve the accuracy in regions where uncertainty is high

References

- [1] *Remer, L. A., Tanré, D., Kaufman, Y. J., “Algorithm for Remote Sensing of Tropospheric Aerosol from MODIS for Collection 005”, <http://modis-atmos.gsfc.nasa.gov/atbd02.pdf>, 2006.*
- [2] *Radosavljevic, V., Vucetic, S., Obradovic, Z., “Aerosol Optical Depth Retrieval by Neural Networks Ensemble with Adaptive Cost Function”, 10th Int'l Conf. on Engineering Applications of Neural Networks, Thessaloniki, Greece, 2007.*
- [3] *Radosavljevic, V., Vucetic, S., Obradovic, Z., “Spatio-Temporal Partitioning for Improving Aerosol Prediction Accuracy”, 2008 SIAM Conf. on Data Mining (SDM), Atlanta, GA, 2008.*
- [4] *Das, D., Radosavljevic, V., Vucetic, S., Obradovic, Z., “Reducing Need for Collocated Ground and Satellite based Observations in Statistical Aerosol Optical Depth Estimation”, Accepted into International Geoscience and Remote Sensing Symposium, Boston, MA, 2008.*