Integration of Deterministic and Statistical Algorithms for Aerosol Retrieval*

Bo Han[1], Slobodan Vucetic[1], Amy Braverman[2], Zoran Obradovic[1]

[1]Center for Information Science and Technology, Temple University, 1805 N. Broad St, Philadelphia, PA 19122, USA

[2]Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr, Pasadena, CA 91109, USA

Keywords: geophysical retrievals, aerosols, regression, spatial interpolation, ensemble models.

Abstract – Aerosol optical thickness (AOT) is typically estimated from satellite radiance observations through computationally demanding deterministic retrievals based on manually constructed physical models. A statistical alternative to this deterministic method is to train regression models for AOT prediction from radiance data. This approach provides fast retrievals albeit with somewhat reduced accuracy. In this paper, we explore an integrative approach that combines statistical and deterministic algorithms to provide both inexpensive and accurate retrievals. Given a limited set of locations with AOT produced by the deterministic algorithm, and a full set of radiance data, we retrieve AOT at the remaining locations using several distinct statistical algorithms: (1) inverse distance spatial interpolation, (2) global neural networks learned on data from the entire domain, (3) region-specific neural networks, and (4) optimally weighted averaging ensembles of the first three algorithms. The integrated retrieval algorithms are evaluated using AOT and radiances observed by the Multi-angle Imaging SpectroRadiometer (MISR) instrument onboard NASA's Terra satellite during two 16-day periods in 2002 over the continental US. Results show that integration of the deterministic and statistical algorithms provide a range of options for selection of the best trade-off between accuracy and complexity. Moreover, on the statistical side, the best tradeoff between retrieval speed and accuracy was obtained through weighted averaging of global neural networks and spatial interpolation.

1 Introduction

Aerosols are small particles produced by natural and man-made sources that both reflect and absorb incoming solar radiation. Aerosol concentration and chemical properties are important parameters in climate change modeling, in studies of regional radiation balances and the hydrological cycle [11]. Using radiance observations from satellite instruments, it is possible to estimate the attenuation of solar energy as it passes through a column of atmosphere due to particulates, a quantity commonly known as Aerosol Optical Thickness (AOT). Since radiance intensity depends on AOT, deterministic forward simulation algorithms are used to "retrieve" AOT [2, 13]. These algorithms predict radiances for candidate aerosol types and amounts and select the types and amounts that most closely match the observed radiances. While deterministic algorithms provide accurate retrievals, they are computationally demanding, and this limits achievable spatial resolution and the ability to provide timely updates.

To address these issues we are exploring a complementary statistical approach based on supervised learning. In preliminary work, we used artificial neural networks (ANNs) to construct global aerosol predictors by learning from all available labeled aerosol data [10]. While the results were encouraging, the heterogeneous

*Contact: Zoran Obradovic, Professor and Director, Center for Information Science and Technology, Temple University. Tel: 1-215-2046265, Fax: 1-215-2045082, Email: zoran@ist.temple.edu

spatial-temporal nature of aerosol makes it unlikely that a single aerosol predictor could fully exploit distinct properties of specific spatial regions. An alternative is to construct a number of local predictors, each specific to a given spatial area. While development of local models addresses the data heterogeneity problem, the scarcity of region-specific data could raise issues related to the choice of model complexity and overfitting control in supervised learning. Therefore, integration of local and global models is an attractive alternative. In a more recent study, we proposed an integration approach where global and local neural network models were appropriately combined resulting in more accurate retrievals than provided by the component models in isolation [3]. It is worth noting that region-specific approach has been used to a certain extent in the existing deterministic aerosol retrieval algorithms [8, 9]. For example, until recently, the operational aerosol retrieval algorithm for the MISR instrument differentiated between water, dense, dark, vegetated land, and heterogeneous land. Region-specific approaches to retrieval of aerosol properties are also considered in [4].

In this paper, we report preliminary results on integration of statistical and deterministic retrieval algorithms. The objective is to combine the strengths of the two approaches – the higher accuracy of deterministic retrievals, and significantly lower computational cost of statistical retrievals. The idea is as follows: use deterministic algorithms over a training set of locations and retrieve aerosols at the remaining locations by applying the statistical algorithm learned from the training set. The goal is to develop an integration model that provides the best trade-off between retrieval speed and accuracy. Two basic statistical approaches explored here are based on spatial interpolation and construction of neural network AOT predictors. To combine predictions obtained from component neural networks and spatial interpolation models, we used weighted averages with weights optimized to minimize the mean squared prediction error over a specific region.

We evaluate our method using MISR data over the continental US during two, 16-day periods during 2002: 07/01/2002 – 07/16/2002 and 10/01/2002 – 10/16/2002. Experimental results suggest that the integration approach may provide a promising mechanism for producing simultaneously fast and accurate retrievals.

2 Methodology

2.1 Preliminaries

Given a set $\{x_i\}$ of satellite-based radiance observations, each data point x_i is represented as tuple $x_i = [t_i, lat_i, lon_i, x_{i1} \ldots x_{iM}]$, where t_i is the time of the observation, lat_i and lon_i denote the spatial location, and x_{ij}, $j = 1 \ldots M$, are derived from the observed radiances and the corresponding geometric parameters which describe satellite camera view angles, and sun angle at time t_i . The aerosol retrieval task can be stated as prediction of target variable y_i (i.e. AOT) based on the corresponding values of x_i.

Deterministic retrieval algorithms are based on the physical models of the relationship between AOT, aerosol type, and radiances. Those models are developed manually by the domain experts. Alternatively, given a labeled set D consisting of N pairs $d_i = (x_i, y_i)$, $i = 1 \ldots N$, statistical retrieval algorithms are based on learning an accurate regression model $f(x,b)$ that minimizes Mean Squared Error (MSE) defined as $E[(y - f(x,b))^2]$, where b is a set of model parameters.

While learning regression models can be time consuming, the prediction (i.e. evaluation of $f(x,b)$ for a given x) is rather inexpensive (not so for deterministic retrievals). It is important to emphasize that the quality of statistical algorithms depends on the appropriate attribute selection and, on the quality of labeled data. In practice, the true values of the target variable y are not known, and estimates based on the deterministic algorithms represent their best available proxy. Consequently, statistical algorithms could be only as accurate as the corresponding deterministic algorithm used to construct the labeled data. On the other hand, the accuracy of statistical algorithms is expected to increase with the amount of labeled data.

2.2 Problem Definition

Based on the properties of deterministic and statistical retrieval algorithms described in 2.1, it is evident that the two approaches are quite complementary – deterministic retrieval algorithms are accurate and computationally expensive, while statistical retrieval algorithms may be less accurate and are significantly less expensive. The goal of this paper is to explore integration of the two approaches that allows optimization of

the trade-off between accuracy and computational complexity. The approach here is as follows. Instead of applying the deterministic algorithm at high spatial resolution over a large region, apply it over a significantly reduced subset of spatial locations. Then use the statistical algorithm to provide retrievals at the remaining locations. It is evident that with the increase of the retrieval resolution by the deterministic algorithms, both retrieval accuracy and retrieval time are increased. The objective here is to explore the potential of the approach that integrates deterministic and statistical algorithms to significantly reduce retrieval time while allowing only a slight decrease in the retrieval accuracy. In sections 2.3-2.6 we outline four different statistical retrieval algorithms to be integrated with the deterministic algorithm used by MISR.

2.3 Construction of a Global Neural Network Model

Given a set of K data points with observation attributes $\{(t_i, lat_j, lon_j, x_{j1}, \ldots x_{jM}, y_j), j = 1\ldots K\}$ obtained from the deterministic algorithm over a period of time. Use them to construct a set of labeled points for training a regression model. ANNs have been successfully used in many applications and we use them here. Specifically, we use a feedforward neural network with a single hidden layer of sigmoidal units to predict AOTs. The design objectives are feature selection, identification of an appropriate ANN structure, as well as choice of training algorithm to maximize prediction accuracy on out-of-the-sample test data. Prediction accuracy is measured by the coefficient of determination $R^2 = 1 - MSE/Var(y)$, where $Var(y)$ is the variance of the AOT target variable.

MISR data provide a large number of informative attributes (t_i, lat_j, lon_j, $x_{j1}\ldots x_{jM}$) (M=111 in our case). This results in a neural network with a large number of weights, and requires significant training time. Under these circumstances, overfitting problems can result when learning from limited training data. Therefore, we used Principal Component Analysis (PCA) to reduce data dimensionality to $k < M$ attributes, and used the largest k principal components as inputs to our ANN. We determined the appropriate number of hidden units in our ANN, and the best training algorithm experimentally. We considered 5, 10, 15 and 20 units, and the following training methods: Bayesian regularization, Powell-Beale conjugate gradient, Polak-Ribiere conjugate gradient, scaled conjugate gradient, Levenberg-Marquardt and quasi-Newton backpropagation. We chose those resulting in minimal prediction errors. Following similar strategies, we also developed global ANN models trained only on non-spatial attributes to determine to what extent AOT prediction accuracy by a global neural network is independent of spatial location. If reasonably accurate, such a property would allow efficient retrievals in future time periods without the position-memorization concerns arising from training a statistical model on historic data.

2.4 Spatial Interpolation Using the Inverse Distance Method

Given a set of K data points $\{(lat_j, lon_j, y_j), j = 1\ldots K\}$ obtained from the deterministic algorithm, we applied inverse distance interpolation to predict AOT at unlabeled locations. The AOT value y_i at unknown location (lat_i, lon_i) is computed as

$$y_i = \sum_{j=1}^{K} y_j / d_{ij}^{p} \Big/ \sum_{j=1}^{m} 1 / d_{ij}^{p} \qquad (1)$$

where d_{ij} is the distance between points i and j, and p is a parameter determining the extent of weighting distance towards giving higher influence to near neighbors [12]. Larger p values assign larger weights to the near neighbors where $p = 0$ corresponds to an overall average estimator and $p = \infty$ is equivalent to the nearest neighbor predictor. We explored several choices of p over multiple local regions and used the best fitting for all interpolations.

2.5 Spatial Interpolation Using Region-Specific Neural Networks

More complex interpolation methods are possible. In this work, we have constructed locally-specific neural network interpolators. For each spatial region, called an orbit, a local neural network is constructed using a subset of labeled data corresponding to the region. In addition to using all available attributes for training local neural network, we also consider using only the non-spatial attributes. Since we only use labeled data from a spatially-constrained region, we expect that the neural network will memorize the specific properties of the region. This is desirable for this type of application and serves as an alternative type of spatial

interpolation. The input dimensionality for local neural networks is determined by performing PCA reduction on local data aimed at retaining about the same variance as with PCA on global data (as in section 2.2). We repeat this data reduction process over multiple local regions and the average reduction size over these experiments is used as local PCA choice.

2.6 Weighted Averaging Ensemble Model

To combine predictions from global ANN model and spatial interpolation models, we construct an integration predictor g_k for each local region r, $r = 1\ldots R$, by weighting the component predictors as

$$g_k(x) = \alpha_k f_G(x,\beta_G) + (1-\alpha_k) f_k(x) \ . \qquad (2)$$

Here $f_k(x)$ represents the AOT prediction from spatial interpolation, and a_k are the weight parameters constrained to lie in the interval [0, 1].

The weight parameter a_k in equation (2) is optimized to minimize mean squared prediction error of weighted averaging model g_k. It can be shown that the optimal choice of α_k is

$$a_k = \sum_{i=1}^{N_k} (y_i - f_k(x_i)) \cdot (f_G(x_i, b_G) - f_k(x_i)) \Big/ \sum_{i=1}^{N_k} (f_G(x_i, b_G) - f_k(x_i))^2 \qquad (3)$$

where N_k represents the number of labeled data points within the k-th region. If $a_k < 0$, its value is set to zero; and if $a_k > 1$, its value is set to one.

The value of the weight parameter can provide important information about the nature of a local region. If a_k is near 1, the properties of the local area do not differ significantly from the global properties, while if a_k is near 0, the local area is different from the overall data distribution.

3 Experimental Results

3.1 Data Set

MISR data used in our experiments were obtained from NASA's Langley Atmospheric Sciences Data Center. MISR measures reflected solar radiation from nine view angles along the direction of flight, and in four spectral bands at each angle [1, 6]. On each orbit, MISR sweeps out a 360 km wide swath of data from north to south at 1.1 km spatial resolution while in daylight. Since MISR does not collect data at night, consecutive swaths are separated geographically resulting in 14 or 15 evenly spaced half-orbits per day. MISR's ground footprint repeats in 16 days cycles, which is the time it takes the Terra satellite to fly 233 distinct orbital paths. We obtained about 12GB of MISR Level 1B2 radiance data covering two 16-day repeat cycles over the entire continental US (cycle 1 period July 1-16, 2002 and cycle 2 period October 1-16, 2002). There are 30 cycle 1 orbits and 29 cycle 2 orbits used in this study.

Table 1. Driving attributes constructed at 17.6k×17.6km resolution from 1.1 km pixels

Attribute Index	Name and Explanation
1,2,3,4	Time, Latitude, Longitude, Orbit number
5	Block number (8*32 aggregate of 17.6x17.6km regions)
6-41	36 mean values of radiance measurements
42-77	36 stand derivations of radiance measurements
78	Solar zenith angle
79-87	View zenith angle (9 cameras)
88-96	Relative view-Sun azimuth angle (9 cameras)
97-105	Scattering angle (9 cameras)
106-114	Glitter angle (9 cameras)

The learning target is AOT retrieved at 17.6k×17.6km spatial resolution [5]. The attributes listed at Table 1 are derived from 114 radiance variables, and their corresponding geometric parameters (e.g. sun angle, view angle, etc.) measured at 1.1 km spatial resolution. In the preprocessing stage, we average each of 36 radiance attributes over 17.6k×17.6km regions represented by a single AOT target value. Before averaging radiances,

the MISR quality flag was used to remove 1.1km pixels with non-valid radiance information. After data cleaning, we have 12,167 cloud-free examples for cycle 1 and 8,753 cloud-free examples for cycle 2.

3.2 Optimization of Statistical Retrieval Models

Data in each orbit in cycles 1 and 2 were independently and randomly divided into three, equal size, disjoint subsets used for training, validation and testing respectively. We study the influence of data reduction and choices of modeling parameters for global, local and weighted averaging models. Results reported at this section were averaged over 3 cross-validation experiments.

3.2.1 Optimization of a Global Neural Network Model

We first developed ANNs that use attributes of various dimension, all with one hidden layer of 15 neurons and trained using Bayesian regularization backpropagation. For cycle 1, any reduction by PCA starting from 114 attributes listed at Table 1 to more than 10 attributes results in similar accuracy while the best accuracy for cycle 2 is obtained when projecting to 25 dimensions. Therefore, in the remaining studies of both cycles we fixed PCA reduction to 25 dimensions and call the projected data PCA25 (retained variance 97% and 96% for cycle 1 and 2 respectively). For ANN with PCA25 trained using Bayesian regularization backpropagation we explored several choices for hidden nodes. The best accuracy was achieved with 15 and 10 hidden nodes for cycle 1 and 2 data respectively (Table 2). These choices were used in the remaining experiments.

Table 2. Hidden layer optimization for global ANNs with PCA25 and Bayesian regularization

Number of hidden nodes	5	10	15	20
R^2 (2002.7.1 – 2002.7.16)	0.867	0.871	**0.873**	0.871
R^2 (2002.10.1– 2002.10.16)	0.806	**0.825**	0.806	0.824

We also explored the influence of training algorithm choice for global ANNs with 15 and 10 hidden nodes (cycle 1 and 2 respectively) using PCA25 attributes. Table 3 reports accuracy, measured by the overall coefficient of determination R^2 (1−MSE/Var(AOT over all orbits in a cycle)). Bayesian regularization and Levenberg-Marquardt backpropagation are the best choices for cycles 1 and 2 respectively. These algorithms are fixed for the remaining global ANN experiments.

Table 3. Prediction accuracy of global ANNs with PCA25 achieved by different training algorithms

Training algorithm	Bayesian regularization	Powell-Beale conjugate gradient	Polak-Ribiere conjugate gradient	Scaled conjugate gradient	Levenberg-Marquardt	BFGS quasi-Newton
R^2 (2002.7.1 – 2002.7.16)	**0.873**	0.871	0.869	0.864	0.864	0.867
R^2 (2002.10.1– 2002.10.16)	0.786	0.767	0.767	0.786	**0.825**	0.806

Excluding latitude, longitude and other spatial information from the set of attributes, we repeated the experiments using PCA25 and ANN with 15 and 10 hidden nodes, and trained by Bayesian regularization and Levenberg-Marquardt backpropagation. A global model without spatial information achieves just slightly lower prediction accuracy (R^2 equals 0.8681and 0.8165 for cycle 1 and 2) than global model using all 114 attributes. This is appealing from a computational standpoint since such a global ANN model can be used on the next cycle data without retraining.

3.2.2 Optimization of Local Spatial Interpolation Models

To explore the influence of various ways of weighting spatial interpolation towards near neighbors we ran experiments for different *p*-parameter values in inverse distance model (equation (1)). Results listed at Table 4 suggest using p=2.5 as the best interpolation choice.

Table 4. Accuracy of inverse distance interpolation using different p value

P	0.5	1	1.5	2	2.5	3	3.5	4
R^2 (2002.7.1 – 2002.7.16)	0.692	0.811	0.865	0.882	**0.886**	0.884	0.880	0.879
R^2 (2002.10.1– 2002.10.16)	0.570	0.726	0.804	0.823	**0.844**	0.843	0.823	0.833

3.2.3 Optimization of Local Neural Network Models

To optimize settings in local ANN models, we applied PCA on each orbit to seek the minimal dimension for which retained variance is larger than 97% in 1 cycle and 96% in cycle 2. These are the proportions of variance retained by PCA25 on the entire cycle 1 and 2 datasets respectively (see section 3.2.1). The average reduced dimension over multiple orbits, retaining sufficient variance, was 19 and 17 for cycle 1 and 2 data. These choices were used for designing local ANN models with a single layer of hidden nodes. Furthermore, results shown in Table 5 suggests that, like the global ANN for local ANN models, it is beneficial to use all attributes instead of relying only on non-spatial radiance information (36 attributes). However, the difference in accuracy is fairly small, and training non-spatial local ANNs can be nevertheless be beneficial when resources are limited.

Table 5. Accuracy of local ANN reduced from 114 and 36 attributes with different number of hidden nodes

Number of hidden nodes	5		10		15		20	
Number of attributes	36	114	36	114	36	114	36	114
R^2 (2002.7.1 – 2002.7.16)	0.804	0.847	0.853	0.879	0.876	0.899	0.881	**0.905**
R^2 (2002.10.1– 2002.10.16)	0.746	0.759	0.830	0.746	0.836	0.860	0.842	**0.867**

3.3 Comparison of Statistical Retrievals using 2/3 of Deterministic AOT Retrievals for Training

Results of comparing the overall accuracies of global ANN (GANN), local ANNs (LANN36 and LAN114 using 36 and 114 attributes respectively), local inverse distance interpolation models (LIDI) and the corresponding weighted averaging models are listed in Table 6. Here, weighted averaging model $WA_{GANN+LIDI}$ integrates GANN with LIDI models, while $WA_{GANN+LANN36}$ and $WA_{GANN+LANN114}$ are weighted averaging models obtained by integrating GANN with LANN36 and LANN114 respectively. The weighted averaging models significantly outperformed local and global models. The accuracies also varied with time period and region (orbit) as data dependences changed in space as shown in Figure 1. The weighted averaging of global ANN and local models was clearly beneficial at all orbits while the integration was such that discovered relationships were mostly global at some orbits, mostly local at other orbits and exhibited both types of dependencies elsewhere. In Figure1 these results are shown as a plot of weight parameter over individual orbits.

Table 6. Comparison of accuracies achieved by different models

Model Name	GANN	LIDI	LANN36	LANN114	$WA_{GANN+LIDI}$	$WA_{GANN+LANN36}$	$WA_{GANN+LANN114}$
R^2 (2002.7.1 – 2002.7.16)	0.874	0.883	0.881	0.905	0.901	0.886	**0.911**
R^2 (2002.10.1– 2002.10.16)	0.817	0.842	0.842	0.867	0.870	0.859	**0.885**

The results suggest that AOT for orbits covering the western US are more difficult to predict than those in the east. This is consistent with previously published observations [7] comparing MISR AOT with AOT retrieved from ground-based instruments. By comparing weight parameter for the same path in different cycles, we see that they also change with time, suggesting global or local dominance changes with seasons.

3.4 Statistical Retrievals using Different Fractions of Deterministic AOT Retrievals for Training

Experiments described at Section 3.3 are performed using 2/3 of the deterministic AOT retrievals for training and validation. We also repeated the exercise using various other fractions of available deterministic AOT retrievals for training and validation. In nine sets of additional experiments, we varied fraction of training data from 10% to 90% of total data in increments of 10%. When learning from training data of lower spatial density accuracy was increasing nearly linear while improvement was less evident when learning from high fraction of deterministic retrievals (Figure 2). We view this as computational support of geoscientist's decisions to provide deterministic AOT retrievals AOT at 17.6k×17.6km spatial resolution rather than doing this at a lower resolution. From Figure 2 it can be observed that over a large range of training data densities, statistical models obtained by weighted averaging of global and local predictors were superior in accuracy to global or local statistical models by themselves.

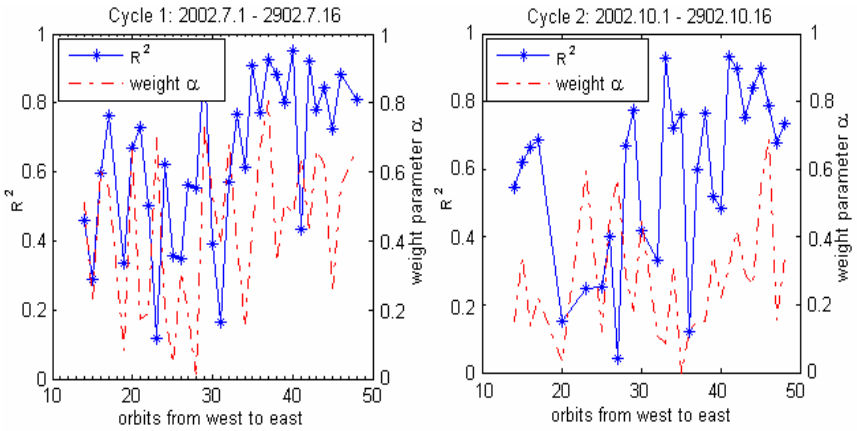

Figure 1: Accuracy R^2 and weight parameter of weighted averaging model $WA_{GANN+LIDI}$ for individual orbits

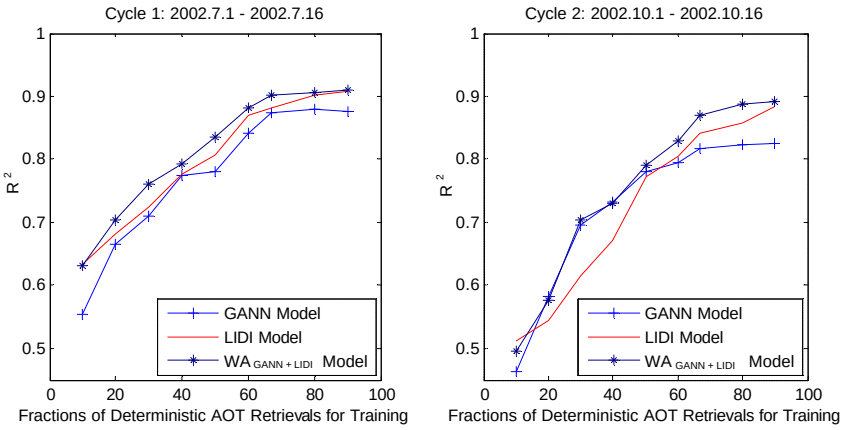

Figure 2: Overall accuracy of global model GANN, local inverse distance interpolation model LIDI and weighted averaging model $WA_{GANN+LIDI}$ when using 10% to 90% of available deterministically retrieved AOT for training statistical models

Finally, plots shown in Figure 2 can be used for analyzing consequences of trading retrieval accuracy for speed since time requirements of deterministic modeling are much higher than statistical. For example, if we retrieve AOT at 20% of the locations using the deterministic algorithm, and then use the results to train a statistical retrieval model, computational efforts would be reduced nearly five times assuming computation cost of deterministic retrievals is only linear with the number of retrieved AOT examples. However, in such a case the coefficient of determination on 80% AOT retrievals obtained statistically versus using deterministic retrievals would be about 0.6.

4 Conclusions and Next Steps

In this study, we proposed integration of deterministic AOT retrievals with less accurate but computationally more efficient statistical approaches. We considered global neural networks trained using data from the entire domain, local neural networks, and local spatial interpolation models developed using data from a limited region (a single orbit in our experiments). We also proposed a weighted average ensemble of global and local statistical model outputs to improve accuracy of geospatial predictors. Averaging takes advantage of large global data sets, but also exploits more specific spatial properties at local sites. The

methodology is evaluated in the context of AOT prediction using MISR data.

Our experimental results provide some evidence that statistical AOT retrievals can serve as a practically useful complement to traditional deterministic retrieval methods. We found that statistical prediction of AOT for orbits in the western US is more difficult than such prediction of AOT in the east. Changes in dependencies with seasons were also evident (Summer vs. Fall data of cycles 1 and 2). Therefore, further research is needed to fully understand where in space and time such integration is likely to be the most beneficial.

The results also showed that the global-local weighted average ensamble approach achieved higher overall accuracy than either local or global models did alone in our test datasets. Although the best overall results were obtained through weighted averaging of global and local ANNs, replacing local ANN with spatial interpolation models achieved comparable accuracy, and has the additional benefit of running faster. The benefits of weighted averaging statistical models were particularly clear when larger fraction of deterministic AOT retrievals are used for training (2/3 or more). This suggests that integrating statistical and deterministic AOT retrievals may be useful for obtaining high quality AOT retrievals at higher resolutions without significant additional computational burden. Exploring this topic in more details is a subject of our next steps research.

Acknowledgments: We thank Ralph A. Kahn and John Martonchik, MISR Science Team members at the Jet Propulsion Laboratory, for their generous help with understanding of the domain. We also thank Atmospheric Sciences Data Center at NASA Langley Research Center for their support in collection of MISR data and Qifang Xu at Temple University for data management and preprocessing help.

References

[1] Bothwell, G.W., Hansen, E.G., Vargo, R.E., and Miller K.C., *The Multi-angle Imaging SpectroRadiometer science data system, its products, tools, and performance*, IEEE Trans. Geosci. Remote Sens., 40 (7), 1467-1476, 2002.

[2] Diner, D.J., et al, *Level 2 Aerosol Retrieval Algorithm Theoretical Basis*. Jet Propulsion Lab., Pasadena, CA, JPL Tech. Doc. D-11400, 2001.

[3] Han, B., Braverman, A., Vucetic, S., Obradovic, Z., *Construction of an Accurate Geospatial Predictor by Fusion of Global and Local Models*, 8th Int'l Conf. Information Fusion, Philadelphia, PA, 2005, accepted.

[4] Hsu, N.C., Tsay, S.C., King, M.D. and Herman, J.R., *Aerosol Properties over Bright-Reflecting Source Regions*, IEEE Trans. Geosci. Remote Sensing, 42 (3), 2004.

[5] http://delenn.gsfc.nasa.gov/~imswww/pub/imswelcome/.

[6] Jet Propulsion Laboratory, California Institute of Technology, *Multi-angle Imaging Spectro-Radiometer Data Product Specifications*, April 2003.

[7] Liu, Y., et al., *Validation of Multiangle Imaging Spectroradiometer (MISR) Aerosol Optical Thickness Measurements using Aerosol Robotic Network (AERONET) Observations over the Contiguous United States*, J. Geophysical Research, 109, D06205, doi: 10.1029/2003JD003981, 2004.

[8] Martonchik, J.V., Diner, D.J., Crean, K. A. and Bull, M.A., *Regional Aerosol Retrieval Results From MISR*, IEEE Trans. Geosci. Remote Sensing, 40 (7), 2002.

[9] Martonchik, J.V., Diner, D.J., Kahn, R.A., Ackerman, T.P., Verstraete, M.M., Pinty, B., and Gordon, H., *Techniques for the Retrieval of Aerosol properties Over Land and Ocean Using Multiangle Imaging*, IEEE Trans. Geosci. Remote Sens., 36 (4), 1212-1227, 1998.

[10] Peng, K., Obradovic, Z., Vucetic, S.**,** *Towards Efficient Learning of Neural Network Ensembles from Arbitrarily Large Datasets*, 16th European Conference on Artificial Intelligence, Valencia, Spain, 623-627, 2004.

[11] Ramanathan, V., Crutzen, P.J., Kiehl, J.T., Rosenfeld, D., *Atmosphere – 'Aerosols, climate, and the hydrological cycle'*, Science, 294 (5549), 2119-2124, 2002.

[12] Vucetic, S., Fiez, T. and Obradovic, Z., *Examination of the Influence of Data Aggregation and Sampling Density on Spatial Estimation*, Water Resources Research, 36 (12), 3721-3731, 2000.

[13] Wen, G., Tsay, S.C., Cahalan, R.F. and Oreopoulos, L., *Path RadianceTtechnique for Retrieving Aerosol Optical Thickness over Land*, J. Geophys. Res., 104, 31321-31332, 1999.