SPATIO-TEMPORAL CHARACTERIZATION OF AEROSOLS THROUGH ACTIVE USE OF DATA FROM MULTIPLE SENSORS

Z. Obradovic[a,] *, D. Das[a], V.Radosavljevic[a], K.Ristovski[a], S.Vucetic[a]

[a] Center for Information Science and Technology, Temple University, Philadelphia, PA, USA

Technical Commission VII Symposium 2010

KEY WORDS: Atmosphere, Environment, Analysis, Data Mining, Retrieval, Algorithms, Spatial, Temporal

ABSTRACT:

One of the main challenges of current climate research is providing Earth-wide characterization of Aerosol Optical Depth (AOD), which indicates the amount of depletion that a beam of radiation undergoes as it passes through the atmosphere. Here, a comprehensive overview will be presented of our ongoing data mining based study aimed at better understanding of spatio-temporal distribution of AOD by taking advantage of measurements collected from multiple ground and satellite-based sensors. In contrast to domain-driven methods for AOD retrieval (prediction from satellite observations), our approach is completely data-driven. This statistical method consists of training a nonlinear regression model to predict AOD using the satellite observations as inputs where the targets are obtained from a network of unevenly distributed ground-based sites over the world. Challenges and our proposed solutions discussed here in context of global scale AOD estimation include (i) AOD regression from mixed-distribution spatio-temporal data; (ii) training such a statistical predictor for robust performance across multiple accuracy measures; (iii) uncertainty analysis of AOD estimation, (iv) active selection of sites for ground based observations, (v) discovery of major sources of correctable errors in deterministic models, and (vi) using conditional random fields to combine nonlinear regression models and a variety of correlated knowledge sources in a unified and more accurate AOD prediction model. The proposed methods is illustrated on experiments conducted using three years of global observations obtained by merging satellite data of high spatial resolution (MODIS Level 2 data from NASA's Terra and Aqua satellites) with ground-based observations of high temporal resolution (a remote-sensing network of radiometers called AERONET network). The experiments revealed that the proposed methods result in more accurate AOD retrieval than the baseline statistical and domain-based predictors.

1. INTRODUCTION

The global impact of environment change to climate is monitored largely by use of remote sensing instruments that measure radiances emitted or reflected from Earth. The observed radiances are used to estimate underlying geophysical characteristics through the predictive process called *retrieval*. The retrieved parameters are then used in various applications ranging from monitoring change of atmospheric temperature, the extent of snow, ice or vegetation cover, cloud and aerosol properties to the development of general circulation models for climate studies. Accurate and timely retrievals of geophysical parameters are therefore critical for the success of many climate change related studies.

In recent years remote sensing instruments of various properties have been employed in an attempt to better characterize important geophysical phenomena. The technology of new generation sensors has improved dramatically, but the collected data still contain large uncertainties due to high noise and a large fraction of missing values. As a consequence, retrieval from such high dimensional spatio-temporal observations is a very challenging problem (Jeong *et al*, 2005).

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 models, in studies of regional radiation balances, and studies of the hydrological cycle (Ramanathan *et al*, 2002). Using radiance observations, it is possible to estimate the attenuation of solar energy as it passes through a column of atmosphere, a quantity commonly known as aerosol optical depth (AOD).

The AOD can be retrieved using ground (Levy *et al*, 2005) or satellite (Remer *et al*, 2006) based observations. Ground-based observations are mostly obtained by the AEROsol robotic NETwork (AERONET) which is the global remote sensing network of about 250 radiometers (spatial distribution shown at Fig. 1) that measure AOD several times an hour at specific locations. AERONET AOD prediction is considered very accurate and is often taken as the ground truth for validation of various satellite-based AOD prediction algorithms aimed at providing global coverage .

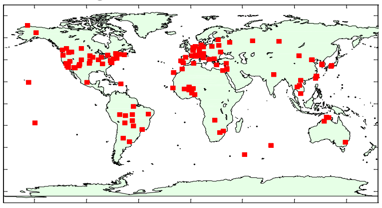

Figure 1: Spatial distribution of AERONET sites.

* Corresponding author (e-mail zoran.obradovic@temple.edu).

The main limitation of ground based observations is their low spatial coverage. Therefore, satellite-based aerosol related observations from several new instruments of high spatial coverage are also considered. The MODerate resolution Imaging Spectrometer (MODIS), aboard NASA's Terra and Aqua satellites launched in 1999 and 2002 with a single camera observes reflected solar radiation from the Earth over a large spectral range in 7 bands. It has a repeating cycle of 16 days and high spatial resolution with almost daily coverage of the entire planet. In comparison, MISR, also aboard Terra satellite, is a nine camera instrument with four bands per camera that provides global coverage every 9 days. MISR collects raw data at 1.1 km resolution, but retrieves aerosol properties at 17.6 km resolution for twenty-four postulated aerosol types. MODIS on the other hand collects data at 1km resolution and its retrievals are provided at 10km resolution.

Designing accurate AOD predictors from satellite observations is a very challenging task due to various problems including reflectance superposition from multiple sources (effects of clouds and surface reflectance are illustrated at Fig. 2). Therefore, satellite based retrievals are less accurate than ground based retrievals. However, they provide high spatial coverage and so are very important for climate studies.

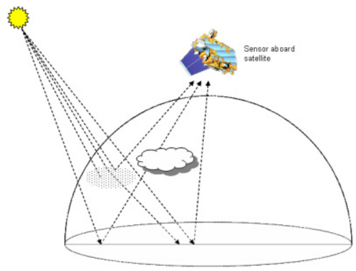

Figure 2: Physics of satellite-based retrieval.

Most operational aerosol retrieval algorithms are constructed as inverse operators of high-dimensional non-linear functions derived from forward-simulation models according to the domain knowledge of aerosol physical properties [35,36,67,68]. For example, MISR uses 24 compositional aerosol models in the Aerosol Climatology Product. These aerosol models, such as mineral dust, biomass burning particles and urban soot, are considered to be representative of the types found over the globe. They are mixtures of individual component aerosols, where each component is defined by a size distribution, particle shape, spectral index of refraction and vertical distribution within the atmosphere. Up to three components can comprise an aerosol mixture, and the fractional optical depths of the components making up a given mixture are pre-specified. For each component aerosol, the corresponding radiative properties are computed using wavelength, illumination, and view geometry information. The results are recorded in a look-up table. By using a modified linear mixing theory, the radiative properties of a mixture are calculated during the retrieval process. These simulated data are then compared to actual observations for the appropriate scene type (land or ocean). According to a set of goodness-of-fit criteria based on the domain knowledge, the matched aerosol model in the look-up table is used for AOD computation.

Drawbacks of deterministic retrieval methods include (1) high computational cost due to inversion of nonlinear forward models; (2) slow development due to manual construction of the postulated physical models; (3) suboptimality due to difficulties in capturing complex radiance-aerosol relationships in all realistic scenarios; and (4) significant retrieval inaccuracies that are due both to the instrument limitations and imperfections in the retrieval algorithms.

Our team has demonstrated that more accurate retrieval is achievable by a completely data-driven approach using spatio-temporally collocated satellite and ground based observations as shown at Fig 3 (Han *et al*, 2005a; 2005b; 2006a; 2006b; Das *et al*, 2008; Obradovic *et al*, 2006; Xu *et al*, 2005; Zhuang *et al* 2008). This statistical method consists of training a nonlinear regression model using the satellite observations as inputs and ground based AOD measurements as target. Some of our recent related activities and findings are discussed in this article (Das *et al*, 2009; Radosavljevic *et al* 2008; 2009; 2010a; 2010b; Ristovski *et al*, 2009; Vucetic *et al*, 2008).

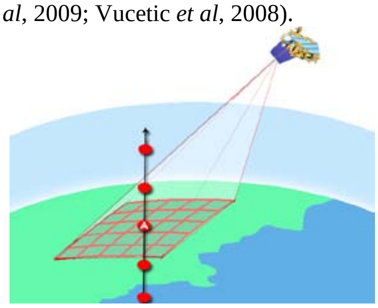

Figure 3: Spatio-temporal collocation of MODIS and AERONET data. A is an AERONET site with AOD retrieved within a short time before and after the satellite overpass (circle dots). The square regions are MODIS observations in a proximity of AERONET site A at the satellite overpass time.

2. METHODOLOGY

The overall objective of our study is to facilitate aerosol retrieval algorithms development, application and modification by developing data mining methodology that utilizes satellite data together with ground-based measurements. Our specific aims are do determine if data mining can

- provide accurate statistical AOD retrievals;
- help discovering the major sources of correctable retrieval errors of deterministic retrievals;
- improve understanding spatio-temporal properties of deterministic retrievals.

An overview of our recently proposed approaches towards achieving these aims is provided in this section.

2.1 Spatio-Temporal Data Partitioning for AOD Retrieval

In principle, AOD retrievals of high spatial resolution can be obtained from satellite observations by training a regression model on a dataset that consists of the satellite observations as inputs and more accurate ground based AOD measurements as output. However, challenges of such supervised learning on aerosol data collected over space and time include existence of different relationships among observations and AOD over various spatio-temporal regions. In such situations an appropriate spatial-temporal data partitioning followed by building specialized predictors could often achieve higher overall prediction accuracy than when learning a single predictor on all the data. In practice, such partitions are typically decided based on prior knowledge.

As an alternative to the domain-based partitioning, we have proposed a method that automatically discovers a soft spatio-temporal partitioning of Earth AOD through the competition of gating regression models (Radosavljevic *et al*, 2008). To address the spatio-temporal dependence the algorithm takes information about location and time of data points as inputs for gating function and performs competition among specialized predictors for each point in the dataset. It starts by randomly dividing the dataset into two disjoint subsets. A specialized predictor is then trained on each subset. Iteratively data are reassigned with some weight to each predictor. Weight is determined based on gating output and accuracy of regression models. Predictors and gating network are then retrained taking into account new assignment.

2.2 AOD Retrieval across Multiple Accuracy Measures

Well known accuracy measures such as Mean Squared Error (MSE) are often not informative enough because (1) retrieval error increases with AOD, (2) distribution of AOD is skewed towards small values, and (3) there are many outliers. Instead, domain scientists use an array of accuracy measures to gain better insight into the retrieval accuracy. For example, the Mean Squared Relative Error (MSRE) makes larger absolute errors more tolerable when predicting large AOD than when predicting small AOD. Ideally, one would like to have a retrieval algorithm that provides good accuracy with respect to these alternative accuracy measures.

To address this issue we considered training of neural networks that minimize MSRE instead of MSE. In order to construct a predictor that is also accurate with respect to MSE and several other accuracy measures, we proposed an approach that builds an ensemble of neural networks, each trained with slightly different MSRE measure (Radosavljevic *et al*, 2010a). The outputs of the ensemble are then used as inputs to a meta-level neural network that produces the actual AOD predictions.

2.3 Uncertainty Analysis of AOD Retrieval

In this task our objective was to explore if neural networks can provide estimates about retrieval uncertainty in addition to providing accurate retrievals. Uncertainty estimation for the confidence of retrieval requires modeling of the whole conditional distribution of the target variable. A standard approaches for neural network uncertainty estimation assume constant noise variance. However, this assumption is not valid for AOD retrieval where noise is heteroscedastic (variance of noise is input-dependent). This is why we explored the Bayesian approach for uncertainty estimation, based on the previous work by Bishop and Quazaz. We also considered alternatives based on the bootstrap technique that are more tractable for large data sets.

A neural network-based regression assumes that target y is related to input vector $\boldsymbol{x}$ by stochastic and deterministic components. The stochastic component is a random variation of target values around its mean caused by heteroscedastic noise with zero-mean Gaussian distribution and input-dependent variance. The deterministic component determines a functional relationship between attributes and prediction. Our goal was to estimate both the stochastic and deterministic component as good as possible.

In (Ristovski *et al*, 2009) we have evaluated three approaches for estimating the stochastic component. The first was based on training a neural network to predict squared error from attributes. We used a standard Mean Squared Error (MSE) criterion to train this network. The second approach assumed heteroscedastic noise and defined the conditional target distribution. The uncertainty estimation neural network is obtained by maximizing the corresponding log-likelihood. The second method assumes that the conditional mean is exactly estimated by the bootstrap committee. Since this is only an estimate, in the third approach we also considered the model uncertainty. In this approach error occurs due to both uncertainty in the model and noise in target.

2.4 Selection of Sites for Ground Based Observations

Ground based AOD stations are often located without a rigorous statistical design. Decisions are typically based on practical circumstances (e.g. overrepresentation in urban regions and industrialized nations) and according to domain experts' assumptions about the importance of specific sites. Given these circumstances, our aim was to evaluate performance of the current design of AERONET sensor network and to apply data mining techniques to assist in future modifications of the sensor network.

In (Radosavljevic *et al*, 2009) we assumed that there is a pending budget cut for maintenance of the existing AERONET sites. The objective was to remove a fraction of the AERONET sites while making sure that the utility of the remaining sites is as high as possible. We made a simplifying assumption that operational costs for each AERONET site around the globe are equal. Common to most selection techniques originating from the spatial statistics is a tendency to overlook the time dimension of data collected by the sensor network. Therefore, we considered series of observations and proposed to optimize AERONET sensor selection based on the concept of retrieval accuracy. Each AERONET site provides a time series which we used for training a regression model to retrieve future AOD. Sites that can be removed are those whose observations are best predicted by the model built on data from the remaining sites.

In (Das *et al*, 2009) our objective was to determine appropriate locations for the next set of ground-based data collection sites as to maximize accuracy of AOD prediction. Ideally, a new site should capture the most significant unseen aerosol patterns and should be least correlated with the previously observed patterns. We proposed achieving this aim by selecting the locations on which the existing prediction model is most uncertain. Several criteria were considered for site selection, including uncertainty, spatial diversity, temporal similarity, and their combination.

Spatial diversity selects sites that are farthest away from the existing sites. The traditional approach in active learning is to label the most uncertain data points. In our application, instead of selecting an individual data point, we select a site. To address this, we defined uncertainty of a site as the average uncertainty over all its observations. For this purpose, we trained a number of neural networks on data obtained from the existing AERONET sites using the bootstrap method. Then we used these neural networks to predict the value of AOD at all satellite observations over potential AERONET sites. We measured the variance among the network predictions and considered this variance as the uncertainty of prediction at the individual data-points. The selected sites are those with the highest measured uncertainty. One drawback of the site uncertainty selection is that a global measure like average uncertainty might fail to compare the similarity in temporal variation of the uncertainty among sites. Each of the potential sites can be regarded as a time-series of uncertainty values over a year. It has been observed that the shapes of uncertainty time-series closely match with that of AOD time-series because the uncertainty values are highly correlated with AOD. So, we used uncertainty estimates as a proxy for the actual AOD labels that cannot be observed over candidate sites. A potential drawback of the previous three site selection algorithms is that selection by one metric is not guaranteed to be the same as that selected by another metric. Therefore, we modified these algorithms to combine uncertainty, spatial and temporal correlation criteria in a single measure. Our objective was to evaluate which approach is the most appropriate for AERONET site selection

2.5 Discovering Correctable AOD Retrieval Error

We analyzed performance of the operational MODIS aerosol retrieval algorithm. Overall, the main sources of MODIS aerosol retrieval errors are the separation of surface and atmospheric components of the observed radiances, the inaccuracies in the forward-simulation model, and inversion errors. Some sources of retrieval uncertainties, such as bright surfaces or cloud-contaminated scenes, are due to the limitations of the MODIS instrument and cannot be corrected, while others, such as imperfections in the retrieval algorithm, are correctable. Aerosol scientists' major goal is to understand the primary sources of correctable retrieval errors and to use such knowledge to improve the retrieval algorithms. The goal of this study was to explore if data mining could facilitate this process.

Our approach consisted of the three main components: 1) use collocated AERONET and MODIS data to train neural networks for the retrieval of AOD; 2) compare the accuracy of neural networks and the MODIS operational algorithm, and 3) understand the present conditions in instances when the neural network is more accurate than MODIS retrievals. A neural network trained in the first step is a completely data-driven retrieval algorithm, distinct from the model-driven MODIS operational algorithm. The drawback of neural network retrieval is that its high accuracy is not guaranteed for the conditions unlike those at the AERONET sites. As such, neural networks are not a completely viable alternative to model-driven retrieval algorithms. However, if neural networks can achieve higher retrieval accuracy over the AERONET locations, then it is clear that the accuracy of a model-driven algorithm can be further improved.

2.6 Unifying Multiple Retrievals by Structural Regression

The aerosol data are characterized by strong spatial and temporal dependencies. To exploit these dependencies we have recently developed Continuous Conditional Random Fields (CRF) for AOD retrieval that are able to exploit by defining interactions among outputs using feature functions (Radosavljevic *et al*, 2010b). The use of features to define the CRF models allowed us also to include arbitrary properties of input-output pairs into the compatibility measure. Our CRF probabilistic model for structured regression uses multiple non-structured predictors as its features. Features were constructed as squared prediction errors of deterministic and statistical models and we showed that this results in multivariate Gaussian conditional $P(\mathbf{y}|\mathbf{x})$ distribution. Consequently, in the proposed approach learning is a convex optimization problem with a global solution for a set of parameters and inference is conveniently conducted through matrix computation.

3. RESULTS

3.1 Spatio-Temporal Data Partitioning for AOD Retrieval

Following methodology summarize in Section 2.1 we performed large scale experiment using 2 years of data from more than 200 ground based AERONET sites located at six continents spatio-temporally collected with data from MODIS instrument aboard NASA's Earth observing Terra and Aqua satellites. The obtained soft partitioning results (illustrated at Fig. 4) were compared to the data partitioning used in the MODIS operational algorithm that divides the world into three spatial-temporal regions based on domain knowledge. The experiments showed that the new soft partitioning of Earth results in significant AOD retrieval accuracy improvements (Radosavljevic *et al*, 2008).

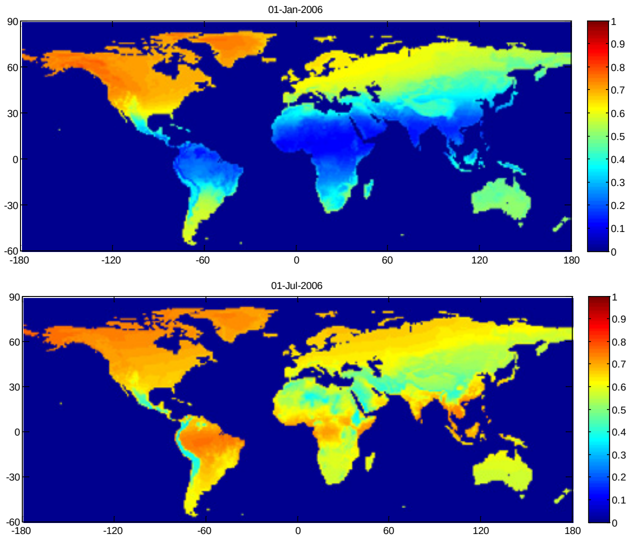

Figure 4: Spatio-temporal partitioning of Earth discovered by competition of two AOD prediction models. Pixel color corresponds to weight w assigned to one AOD predictor in a mixture. The other predictor has weight 1-w. Top panel: winter partitions, Bottom panel: summer partitions.

3.2 AOD Retrieval across Multiple Accuracy Measures

Neural networks from the ensemble described in Section 2.2 were trained using collocated data points whose attributes were derived from MODIS instrument satellite observations and whose target AOD variable was obtained from the ground-based AERONET instruments. Instead of relying on MSE minimization criterion for neural network training, we used the relative error REL, which can be considered as generalization of MSE.

We observed that REL criterion allowed us to achieve increased accuracy over certain ranges of AOD values. To provide a predictor that is accurate over the whole range of AOD values for each of the 5 commonly used accuracy measures, we developed an ensemble of neural networks with adaptive cost functions. Some networks in the ensemble were specialized in predicting small AOD while others were specialized in predicting large AOD. The experiments showed that the proposed ensemble outperformed an ensemble that used standard MSE optimization; it managed to achieve as high MSE, R^2 and CORR accuracies while it significantly improved MSRE and FRAC accuracies. In addition, AOD prediction accuracy of the proposed ensemble was compared to the recently developed operational MODIS Collection 005 retrieval algorithm. Results obtained over the entire globe during the first six months of year 2005 showed that the proposed ensemble of neural networks was significantly more accurate for all the considered accuracy measures (Radosavljevic *et al*, 2010a).

3.3 Uncertainty Analysis of AOD Retrieval

Three years of MODIS data collocated with 201 AERONET sites over whole the globe have been used for uncertainty analysis of AOD retrieval. The average negative log-predictive density (NLPD) of the true targets was used as a measure of the quality of uncertainty estimation. Committees of 30 neural networks were trained on the subset of sites which have data in both training and testing years. The obtained results allowed analysis of uncertainty of AOD retrieval at a given site over time and also uncertainty comparison at multiple sites. As an example, we compared uncertainty of AOD retrieval at Beijing site in China vs. Muana Loa site in Hawaii to conclude that properties of aerosols are much more stable at Muana Loa than in Beijing. By further investigation we found that this discovery is consistent with domain experts' expectations as this site serves for calibration statistics of AERONET instruments.

Our analysis of seasonal uncertainty levels over three years also revealed existence of different interesting patterns. For example, we compared sites with the highest and the lowest average uncertainty over the seasons. The highest uncertainty levels occur in Asia over all seasons, in Africa during the winter and fall, and in the central part of South America during the summer. These levels reach extreme values in summer while for other seasons are almost equal. On the other hand, the lowest levels of uncertainty appear in North America and Europe during winter, summer, and fall, and in South America during the spring (Ristovski *et al*, 2009).

3.4 Selection of Sites for Ground Based Observations

Data used in sites reduction experiments were distributed over entire globe at 217 AERONET sites during years 2005 and 2006. We performed training on 2005 data and used 2006 data for testing. We considered a scenario when current operational AERONET sites have to be reduced by 33% or 66%. In all experiments, we started from a set of 30 AERONET sites and applied the proposed method and the two alternatives to identify a subset of 20 or 10 AERONET sites to be retained. The two alternatives included a random selection of sites as well as an approach based on spatial distance among the sites was also considered. Sites were selected such that their spatial coverage was maximized. In our experiments, the proposed accuracy-based selection achieved consistently better results than the alternatives. Also, accuracy of the proposed site reduction method did not change much even after removing 20 of the 30 AERONET sites. Interestingly, on average, the spatial selection strategy performed slightly worse than the random selection strategy. According to presented results we concluded that the proposed accuracy-based sites reduction method is superior to spatially-based and random selection alternatives (Radosavljevic *et al*, 2009).

Our extensive experiments on globally distributed data over 90 AERONET sites from the years 2005 and 2006 provide strong evidence that sites selected using the algorithm proposed in Section 2.5 improve the overall AOD prediction accuracy at a faster rate than those selected randomly or based on spatial diversity among sites (Das *et al*, 2009). The evaluation data spanned the entire world, ranging from January of 2005 to December of 2006. A committee of 20 neural networks was used to estimate the model uncertainty, each having 10 hidden nodes. Initial training set was created with 700 training points from 10 randomly selected AERONET sites. We assumed that the AOD values corresponding to the remaining training points are unknown. Then we proceeded to select t = {1, 2, … 20} sites in twenty independent experiments using the proposed site selection algorithms. The R-squared accuracy was computed on the test data before and after the selection of prospective sites.
Our results show that uncertainty-based site selection gives significantly higher accuracy over random and spatial distance-based selection (especially when only a few sites are to be selected) and marginally higher accuracy than temporal distance-based method. Performance improves further if correlation among the sites is taken into consideration. A comparison between accuracies uncertainty-based and all three other methods which selects site based on combination of uncertainty, spatial and temporal distance shows that there is some improvement in performance due to inclusion of spatial and temporal correlation metrics over purely uncertainty-based selection. Although the improvement in accuracy is small, it is not negligible, keeping in mind the huge variability of AOD over the entire earth.

At the continent scale, we observed that different continents favor different selection algorithms. For Europe, random selection works quite well because a large number of unlabeled sites are from Europe and therefore random selection favors sites from Europe. In North America, despite the large number of candidate sites, the random selection was not very successful. The accuracies for North America generally followed the accuracies obtained on overall test set. In South America and Africa, spatial selection performed better than other methods. Especially in South America, it was able to attain a significantly higher accuracy (Das *et al*, 2009)

3.5 Discovering Correctable AOD Retrieval Error

Our approach described in Section 2.5 was applied on 3,646 collocated MODIS and AERONET observations within the continental United States. The results showed that neural networks are more accurate than the operational MODIS algorithm over the observed locations. A study of differences between neural networks and the MODIS algorithm revealed interesting findings. For example, NN are more accurate than MODIS when the retrieval is contaminated by clouds, snow, or water (i.e. NDVI <= 0). These and other discoveries have been found to be mostly consistent with expert knowledge and revealed some new insights into the MODIS algorithm performance (Vucetic *et al*, 2008).

3.6 Unifying Multiple Retrievals by Structural Regression

Experiments were conducted on MODIS data at 50x50 km^2 resolution collocated with observations at 217 AERONET sites during years 2005 and 2006. In nested spatio-temporal 5-cross-validation experiments we determined parameters corresponding to the operational *C005* and neural network (*NN*) retrievals over five continents (Asia and Australia were treated as a single region). In our experiments CRF achieved significantly better accuracy than either *NN* or *C005* alone.

In follow-up experiments when using indicator functions for 5 continents (Asia and Australia were treated as one continent) CRF achieved better accuracy than either *NN* or *C005* alone over almost all continents. Values of obtained parameters suggest that we should trust *NN* more in the Eastern US, an area well covered by AERONET sites where SR is expected to work well, while in Africa, an area poorly covered by AERONET, we should trust *C005* more (Radosavljevic *et al*, 2010b).

4. CONCLUSION

The reported results provide strong evidence that accurate statistical AOD retrieval is possible by problem partitioning through competition of local spatio-temporal models (Radosavljevic *et al*, 2008) and developing an ensemble statistical model with components specialized for small and for large AOD prediction (Radosavljevic *et al*, 2010a). It was also found in our study how to analyze uncertainly of statistical AOD retrievals (Ristovski *et al*, 2009) and how to use this tool for optimized placement of ground-based observation sensors (Das *et al*, 2009). A data mining analysis can also reveal the conditions when deterministic AOD retrievals can be significantly enhanced (Vucetic *et al*, 2008). Our work in progress on structural regression by continuous conditional random fields suggests that further major improvements of retrieval accuracy can be achieved by unifying multi-source retrievals (Radosavljevic *et al*, 2010b).

References

Das, D., Obradovic, Z., Vucetic, S. 2009. Active Selection of Sensor Sites in Remote Sensing Applications, *Proc. IEEE Int'l Conf. Data Mining*, December, Miami, FL.

Das, D., Radosavljevic, V., Vucetic, S., Obradovic, Z. 2008. Reducing Need for Collocated Ground and Satellite based Observations in Statistical Aerosol Optical Depth Estimation, *IEEE Int'l Geoscience and Remote Sensing Symposium,* Boston, MA.

Han, B., Vucetic, S., Braverman, A., Obradovic, Z. 2006a. A Statistical Complement to Deterministic Algorithms for Retrieving Aerosol Optical Thickness from Radiance Data, *Engineering Applications of Artificial Intelligence*, vol. 19, no. 7, pp. 787-795.

Han, B., Obradovic, Z, Li, Z., Vucetic, S., 2006b. Data Mining Support for Improvement of MODIS Aerosol Retrievals, *Proc. IEEE Int'l Geoscience and Remote Sensing Symp*., Denver, CO.

Han, B., Vucetic, S., Braverman, A., Obradovic, Z. 2005a. Integration of Deterministic and Statistical Algorithms for Aerosol Retrieval, *Proc. Int'l Conf. Novel Applications of Neural Networks in Engineering*, Lillie, France, pp. 85-92.

Han, B., Vucetic, S., Braverman, A., Obradovic, Z. 2005b. Construction of an Accurate Geospatial Predictor by Fusion of Global and Local Models, *Proc. IEEE 8th Int'l Conf. Information Fusion*, B.11.2 pp. 1-8, Philadelphia, PA.

Jeong, M.-J., Li, Z., Chu, D. A., Tsay, S.-C. 2005. Quality and Compatibility Analyses of Global Aerosol Products Derived from the Advanced Very High Resolution Radiometer and Moderate Resolution Imaging Spectroradiometer, *J. Geophys. Res.*, 110, D10S09.

Levy, R. C., Remer, L. A., Martins, L.V., Kaufman, Y.J., Plana-Fattori A., Redemann, J., Wenny, B. 2005. Evaluation of the MODIS Aerosol Retrievals over Ocean and Land during CLAMS, *J. Atmos. Sci*., vol. 62(4), pp. 974-992.

Obradovic, Z, Han, B., Xu, Q., Li, Y., Braverman, A., Li, Z. Vucetic, S. 2006. Data Mining Support for Aerosol Retrieval and Analysis – Project Summary, *NASA Data Mining Workshop*, Pasadena, CA.

Radosavljevic, V., Vucetic, S., Obradovic, Z. 2010a. A Data Mining Technique for Aerosol Retrieval Across Multiple Accuracy Measures, *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no.2, pp. 411-415.

Radosavljevic, V., Obradovic, Z., Vucetic, S. 2010b. Continuous Conditional Random Fields for Regression in Remote Sensing, *Proc. 19th European Conf. on Artificial Intelligence*, Lisbon, Portugal.

Radosavljevic, V., Vucetic, S., Obradovic, Z. 2009. Reduction of Ground-Based Sensor Sites for Spatio-Temporal Analysis of Aerosols, *Proc. 3rd International Workshop on Knowledge Discovery from Sensor Data at the 15th ACM SIGKDD Conf. Knowledge Discovery and Data Mining*, Paris, France.

Radosavljevic, V., Vucetic, S., Obradovic, Z. 2008. Spatio-Temporal Partitioning for Improving Aerosol Prediction Accuracy, *Proc. Eight SIAM Int'l Conf. Data Mining,* Atlanta, GA.

Ramanathan, V., Crutzen, P.J., Kiehl, J.T., Rosenfeld, D. 2002. Atmosphere - Aerosols, climate, and the hydrological cycle, Science, vol. 294, pp. 2119-2124.

Remer, L. A., Tanré, D. , Kaufman, Y. J., 2006. Algorithm for Remote Sensing of Tropospheric Aerosol from MODIS for Collection 005, http://modis-atmos.gsfc.nasa.gov/atbd02.pdf.

Ristovski, K., Vucetic, S., Obradovic, Z. 2009. Evaluation of a Neural Networks based Approach for Aerosol Optical Depth Retrieval and Uncertainty Estimation, *Proc. Int'l Conf. on Space Technology*, Thessaloniki, Greece.

Vucetic, S., Han, B., Mi, W., Li. Z., Obradovic, Z. 2008. A Data Mining Approach for the Validation of Aerosol Retrievals, *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 1, pp. 113-117.

Xu, Q., Han, B., Li, Y., Braverman, A., Obradovic, Z., Vucetic, S. 2005. Improving Aerosol Retrieval Performance by Integrating AERONET, MISR, and MODIS Data Products, *Proc. IEEE 8th Int'l Conf. Information Fusion*, B.11.3 pp. 1-8, Philadelphia, PA.

Zhuang, W., Radosavljevic, Han, B., Obradovic, Z., Vucetic, S. 2008. Aerosol Optical Depth Prediction from Satellite Observations by Multiple Instance Regression, *Proc. Eight SIAM Int'l Conf. Data Mining,*, Atlanta, GA.

Acknowledgement

This work was supported in part by the National Science Foundation under Grant IIS-0612149.