Spatio-Temporal Partitioning for Improving Aerosol Prediction Accuracy*

Vladan Radosavljevic Slobodan Vucetic Zoran Obradovic†

Abstract

In supervised learning, on data collected over space and time, different relationships can be found over different spatio-temporal regions. In such situations, an appropriate spatio-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, partitions are typically decided based on prior knowledge. As an alternative to domain-based partitioning, we propose a method that automatically discovers a spatio-temporal partitioning through the competition of regression models. The method is evaluated on a challenging problem using satellite observations to predict Aerosol Optical Depth (AOD), which represents the amount of depletion that a beam of radiation undergoes as it passes through the atmosphere. Our experiments used more than 20,000 labeled data points collected during 3 years from more than 100 sites worldwide. Our partitioning-based approach was compared to the recently developed operational AOD prediction algorithm, called C5, which uses domain knowledge for spatio-temporal partitioning of the Earth and implements a region-specific deterministic predictor that utilizes forward simulations from the postulated physical models. Data partitioning used in C5 divides the world into three spatio-temporal regions that differ based on the location and the time of the year as decided by domain experts. The results showed that a neural network predictor trained on all the data has accuracy comparable to C5. When specialized neural network predictors were learned on C5-based partitions, the overall prediction accuracy was not improved. On the other hand, our competition-based spatio-temporal data partitioning approach resulted in large accuracy improvements. The most accurate results were obtained when (1) the data from each of the sites were split into two temporal subsets, one for winter-spring months and another for summer-fall months; and (2) two neural network predictors were competing for each of the identified spatio-temporal subsets.

*This work is funded in part by NSF grant IIS-0612149.

†Center for Information Science and Technology, Temple University, Philadelphia, PA 19122, USA, {vladan, vucetic, zoran}@ist.temple.edu

1 Introduction.

Data collected at discrete points in space over time are referred to as spatio-temporal data. In different spatio-temporal regions, relations among the attributes can be different. Therefore, a single global predictor constructed using the entire dataset could be biased toward the dominant distribution while being less accurate on data points that do not follow the dominant distribution. If both space and time are partitioned in such a way that the observed attributes in each spatio-temporal subset have the same distribution, training specialized predictors on the identified subsets can be beneficial. Applying those local predictors on the corresponding spatio-temporal partitions would increase overall prediction accuracy as compared to the accuracy achieved by applying a single predictor on the entire dataset.

When the data generating process changes as a function of time and location, the same values of observed attributes could result in very different target values at various spatio-temporal regions. Therefore, proximity in attribute space does not necessarily mean the data points should belong to the same spatio-temporal partition. In this situation, unsupervised clustering algorithms are not suitable for discovering spatio-temporal partitions.

The method for spatio-temporal partitioning explored in this study is inspired by our competition-based algorithm for learning from spatial data generated by a mixture of distributions [1]. In this approach multiple regression models are learned on disjoint spatial partitions, followed by repartitioning based on competition between models where a data point is assigned to the model, which has the highest prediction accuracy. The competition process is iterated as long as it leads to accuracy gains. The method was successfully applied for discovering homogeneous regions in heterogeneous spatial data. A similar idea was successfully exploited for improving accuracy of nonstationary time series prediction through competition based on time partitioning [2]. The novel challenge addressed in the current study is how to deal with the distributional change that occurs over both spatial and temporal dimension.

In Section 2 we introduce the problem of remote sensing of aerosols and prediction of Aerosol Optical

Depth (AOD). Our new method for learning from heterogeneous spatio-temporal data is described in Section 3. Evaluation of the approach on a large-scale problem of remote sensing of aerosols is described in Section 4.

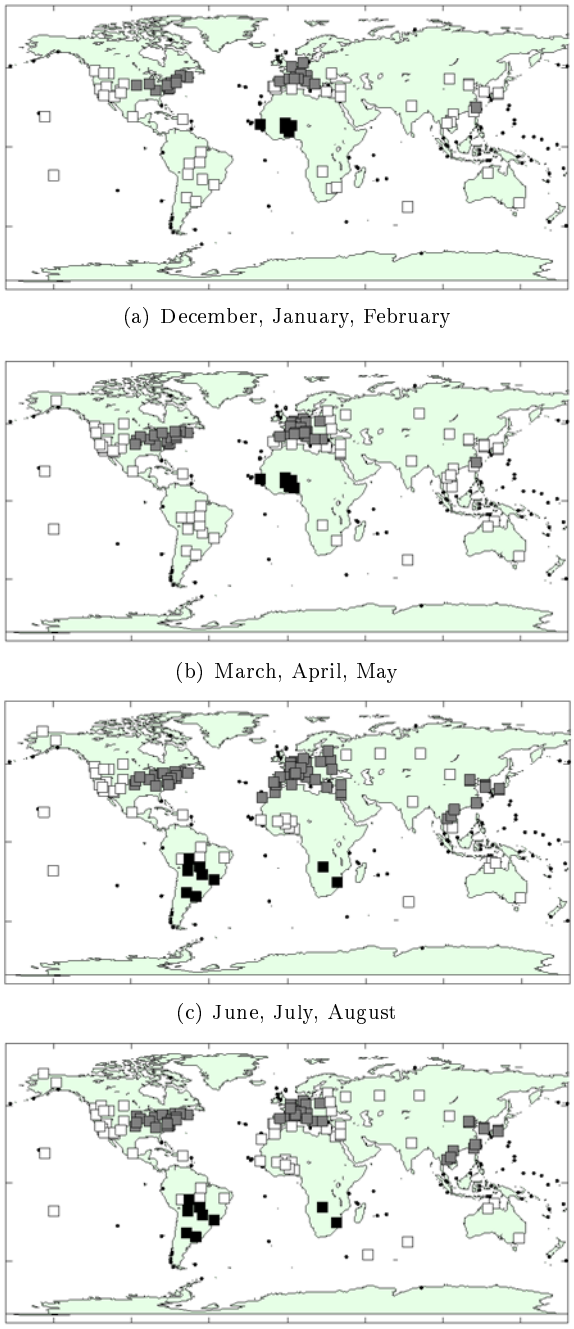
2 Background.

2.1 Ground-based and satellite observations of aerosols. Our study is motivated by one of the main challenges of current climate research consisting of using satellite observations to estimate Aerosol Optical Depth (AOD). AOD is a dimensionless quantity that represents the amount of depletion that a beam of radiation undergoes as it passes through the atmosphere. Large error in a global scale AOD prediction is one of the major limiting factors influencing simulation-based climate change studies [3].

The AOD can be predicted using ground [3] or satellite [4] based observations. Ground-based observations are mostly obtained by AEROsol robotic NETwork (AERONET), which is the global remote sensing network of about 540 radiometers 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 [5]. Satellite-based observations considered in our study are obtained by the MODerate resolution Imaging Spectrometer (MODIS), aboard NASAs Terra and Aqua satellites. This instrument observes reflected solar radiation from the Earth over a large spectral range with a high spatial resolution and has almost daily coverage of the entire planet. Designing accurate AOD predictors from satellite observations is a very challenging problem. In the following we outline two major approaches for AOD prediction.

2.2 Knowledge-driven AOD predictors. Operational algorithms used to predict AOD from MODIS observations are based on matching the atmospheric component of the observed reflected radiation to the simulated values stored in lookup tables. The atmospheric component is obtained by removing the effect of the surface and is dominantly influenced by aerosol optical properties.

Since aerosol properties and abundance change through time and over space, using a single model would not be able to fully describe the aerosol optical properties over a global scale. A recently developed operational AOD prediction algorithm, called C5, utilizes domain knowledge for spatio-temporal partitioning of the Earth. For each spatio-temporal partition, C5 consults the look-up table constructed by forward simulations of the physical model of aerosol optical properties.



(a) December, January, February

(b) March, April, May

(c) June, July, August

(d) September, October, November

Figure 1: Aeronet sites assigned to the spatio-temporal models of operational C5 AOD prediction algorithm. Three models are represented by white, gray and black colors.

C5 defines four aerosol models corresponding to prevalent atmospheric conditions over several characteristic spatio-temporal regions of Earth [6]. The partitioning was obtained by studying observations from AERONET ground-based instruments and combining this information with the climatology domain knowledge. One of those four models is invariant through time and can be applied globally while the other three models, used to adjust the global model, depend on the location and time. When defining aerosol models as a function of location and time, the assumption was that aerosol properties would not change a lot during a three-month season. For each AERONET site and each season, the percentage of data points best described by each of three models was determined. This was used to assign the dominant aerosol type to each AERONET site during each season. The resulting data partitioning used in C5 divides the world into three spatio-temporal regions that differ based on the location and the time of the year as summarized at Figure 1.

2.3 Data-driven AOD predictors. An alternative and a complement to knowledge-driven AOD predictors developed by aerosol experts, such as C5, is a data-driven statistical approach based on learning a regression model on fusion satellite- and ground-based data [7, 8, 9, 10]. In our previous study [11], neural networks trained to predict AERONET AOD over continental U.S. using attributes derived from MODIS observations were significantly more accurate than when using the C4 AOD prediction algorithm (C4 algorithm was the operational MODIS retrieval algorithm until 2007 when the C5 algorithm became operational.) This study suggested that the ability of a single predictor to explain the complex aerosol spatio-temporal variability was limited while an integration of global and local data-driven aerosol predictors was less subject to these limitations. In the related work [12], improved AOD prediction results were obtained by combining a global neural network trained to predict AERONET AOD over the continental U.S. with region-specific neural networks.

In our recent work [13], we developed an ensemble of neural networks to predict AOD over a global scale. We observed that Mean Squared Error (MSE) and other standard cost functions are not completely suitable for AOD prediction because (1) error variance increases with AOD, (2) the distribution of AOD is skewed toward small values and, (3) there are usually many outliers due to measurement imperfections. To reduce the influence of larger prediction errors at large AOD values, we employed the Mean Squared Relative Error instead of the MSE as the optimization criterion for training of AOD predictor.

The hypothesis investigated in our current study is that, due to the variability of aerosol, AOD predictors specialized to specific spatio-temporal regions should result in the increased AOD prediction accuracy. We argue that the existing domain-based spatio-temporal partitioning used in the C5 algorithm is not necessarily the best choice because of the fundamental differences in the nature of data-driven and knowledge-based AOD prediction approaches. For example, while the knowledge-based algorithms such as C5 eliminate surface effects from the satellite observations as a preprocessing step, data-driven algorithms use the observations directly as the input attributes. The goal of this study is to develop a method that automatically discovers a successful spatio-temporal partitioning for AOD prediction through the competition of regression models as an alternative to the domain based partitioning of space and time.

3 Methodology.

3.1 Statistical foundation. When the data generating process changes over time and space, prediction accuracy can be significantly improved by learning a number of regression models specialized for certain spatio-temporal partitions as compared to a single (global) model learned on whole dataset. Let us assume that a spatio-temporal dataset is a union of K disjoint partitions P_i, $i = 1 \dots K$, where the number of partitions and their spatio-temporal locations are not known in advance. The data generating process for P_i can be represented as:

$$y_{st} = f_i(\mathbf{x}_{st}) + e_{st}, e_{st} \sim N(0, \sigma^2), st \in P_i \tag{3.1}$$

where fi is the regression function of partition Pi, xst and est are the attribute vector and the error term of observation at location s and time t. Domains of the observed attributes at different partitions generally overlap, which means that the same vector x can produce quite different outputs at different partitions.

Without any prior knowledge about the spatio-temporal partitions, learning a global prediction model over the entire dataset would result in learning the global regression function defined as:

$$h^*(\mathbf{x}) = \underbrace{argmin}_{h(x)} E_{Y|x}[(Y - h(\mathbf{x})^2] \tag{3.2}$$

for any given $\mathbf{x}$. The MSE of the global prediction model h^* on the data from partition P_i, mse_i, can be expressed as [1]:

$$\begin{aligned} mse_i &= noise_i + bias_i \\ noise_i &= E_{D_i}[e^2] \\ bias_i &= E_{D_i}[(h^*(\mathbf{x}) - f_i(\mathbf{x}))^2] \end{aligned} \tag{3.3}$$

over the domain D_i that corresponds to the partition P_i. The term $noise_i$ corresponds to an unavoidable error which would be obtained by a local predictor specialized for partition P_i and the term $bias_i$ corresponds to the bias of the global prediction model on the data from partition P_i. If spatio-temporal partitions were already known, the $bias_i$ from the previous equation would be eliminated by learning a local model on each partition.

3.2 Spatio-temporal partitioning method. Since we know that introducing local prediction models can improve prediction accuracy when data generating process is heterogeneous, we propose a method that discovers the appropriate spatio-temporal partitions.

We first describe an algorithm by ignoring information about location and time of data points. The algorithm relies on the competition among specialized predictors for each point in the dataset S. It starts by randomly dividing the entire dataset into K disjoint subsets S_i, $i = 1 \ldots K$, where K is the number of the specialized prediction models. A specialized predictor M_i is trained on each subset S_i, $i = 1 \ldots K$. The resulting predictors are competing for points from dataset S such that all the points that are best predicted with predictor M_i are assigned to subset S_i. The competition procedure is repeated until convergence.

The described competition algorithm is noise-sensitive because it can easily lead to assignment of points near in space and time to different subsets. This is clearly an undesirable behavior and we need a mechanism that prevents this from happening.

Our solution is to group the data into spatio-temporal cells that contain multiple data points close in space and time and to run the competition procedure over the cells instead of the individual points. Similar to the original algorithm, the cell is assigned to the prediction model that achieves the smallest average prediction error over the data points in the cell. After reassigning all cells, competition procedure is iteratively repeated until there is no improvement in prediction accuracy.

The choice of the cell size is important because the small cells are sensitive to noise while the large cells could be heterogeneous. In the first case, the partitioning procedure would be unstable and the resulting specialized predictors would be just the artifacts of the procedure. In the second case, the partitioning would be too constrained and would result in highly similar specialized predictors. To achieve the best possible partitioning several choices for time period length should be considered. An outline of the partitioning method is presented in the Algorithm 1.

This is a simple example of how the proposed method operates. In the Figure 2, in the top picture, all spatio-temporal data points from three locations are assigned to one of the two models gray or white. In our example the entire time interval is divided into three parts. Therefore, there are nine spatio-temporal cells. Models are trained and predictions for all data points are obtained. Data points in a cell are assigned to the model achieving better accuracy. This partitioning results in a temporally more homogeneous partitioning. The competition procedure iterates until stable solution is found.

Algorithm 1 Spatio-temporal partitioning method

Input: $\mathbf{S}$ - set of spatio-temporal data points
Output: $\mathbf{S_1}, \mathbf{S_2}, \ldots, \mathbf{S_M}$ - spatio-temporal partitions of $\mathbf{S}$

Set number of models $\mathbf{M} = 2$
Set length of time period $\mathbf{T}$ = maximal possible value
repeat
  repeat
    Divide $\mathbf{S}$ randomly into $\mathbf{M}$ equal sized subsets $\mathbf{S_1}, \mathbf{S_2}, \ldots, \mathbf{S_M}$
    repeat
      Train model $\mathbf{M_i}$ on the subset $\mathbf{S_i}$
      Obtain predictions of all $\mathbf{M_i}$
      Divide time into consecutive parts of length $\mathbf{T}$ and group the points into spatio-temporal cells
      for each spatio-temporal cell **do**
        Select dominant model based on the smallest prediction error
        Assign all points from the cell to the dominant model
      end for
    until there is no further increase in the accuracy
    Decrease $\mathbf{T}$
  until there is no further increase in the accuracy
  Increase $\mathbf{M}$
until there is no further increase in the accuracy
return the best partition $\mathbf{S_1}, \mathbf{S_2}, \ldots, \mathbf{S_M}$

4 Spatio-Temporal Partitioning of Aerosol Remote Sensing Data.

4.1 Accuracy measures for AOD prediction. Regardless of the approach used for AOD prediction, the resulting predictor has to be evaluated and its accuracy adequately quantified. Considering AOD prediction as a regression problem, there are many possible measures that could be used to assess predictor performance. Given a target vector t of AERONET AOD measurements and vector y of the AOD predictions based on the satellite observations, the appropri-

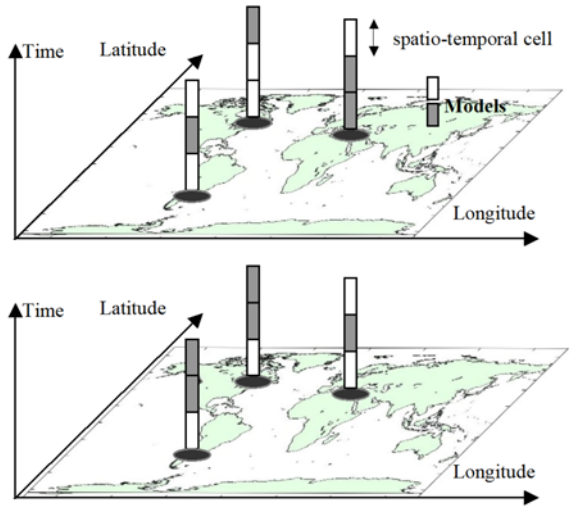


Figure 2: The simple example of partitioning procedure. Top picture three spatial locations, data points are assigned to two models gray and white. Bottom picture reassigned models.

ate measure of prediction accuracy could be *coefficient of determination* (R^2) defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - t_i)^2}{\sum_{i=1}^{N}(t_i - \overline{t})^2} \tag{4.4}$$

where $\overline{t}$ represents the mean value of vector $\mathbf{t}$, and summations are over all N points. In the regression analysis, R^2 is preferred to simple quadratic distance measure mean square error (MSE) defined as:

$$MSE = \frac{\sum_{i=1}^{N}(y_i - t_i)^2}{N} \tag{4.5}$$

because it takes into account variance in the target data. The fraction of the variance that a predictor successfully models is described by R^2 value. The highest R^2 accuracy is 1, while R^2 accuracy of the predictor that simply predicts the mean of the population is 0.

Another measure that is often used is *correlation coefficient* (CORR):

$$CORR = \frac{\sum_{i=1}^{N}(y_i - t_i)(t_i - \overline{t})}{\sqrt{\sum_{i=1}^{N}(y_i - \overline{y})^2}\sqrt{\sum_{i=1}^{N}(t_i - \overline{t})^2}} \tag{4.6}$$

where $\overline{y}$ represents the mean value of prediction vector $\mathbf{y}$, while other parameters were defined previously. CORR measure is insensitive to the prediction bias that is easily correctable [11].

We also consider several domain-specific measures of AOD prediction accuracy. Due to the inherent measurement errors of the MODIS instrument, the acceptable boundaries for AOD prediction error were proposed in [5]. Boundaries were defined as a linear function of the ground truth AOD value t_i measured as:

$$|y_i - t_i| \leq 0.05 + 0.15 t_i \tag{4.7}$$

Relation (4.7) directly implies that errors in AOD prediction are much more tolerable at large AOD than at small AOD values. Consequently, the AOD predictor should be much more accurate in predicting small AOD. In this sense, a new accuracy measure can be defined as *mean squared relative error* (MSRE):

$$MSRE = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{y_i - t_i}{0.05 + 0.15 t_i}\right)^2 \tag{4.8}$$

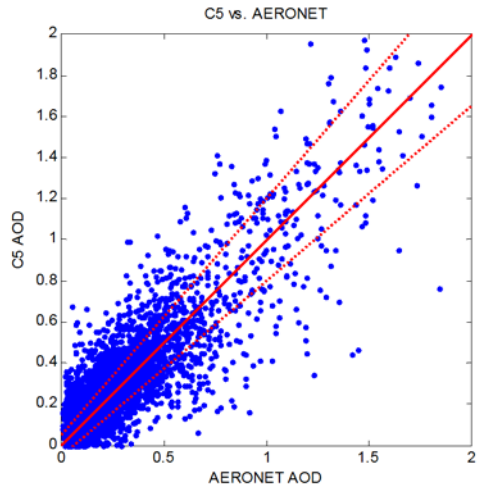
where the sum is over all N data points. If MSRE is closer to 0 the AOD predictor has better performance. However, keeping in mind the MODIS instrument uncertainty, it can be said that predictor performance is acceptable if MSRE is not much larger than 1. In addition, we measure FRAC defined as the fraction (FRAC) of predictions that are between the domain-expected boundaries:

$$FRAC = \frac{I}{N} \times 100\% \tag{4.9}$$

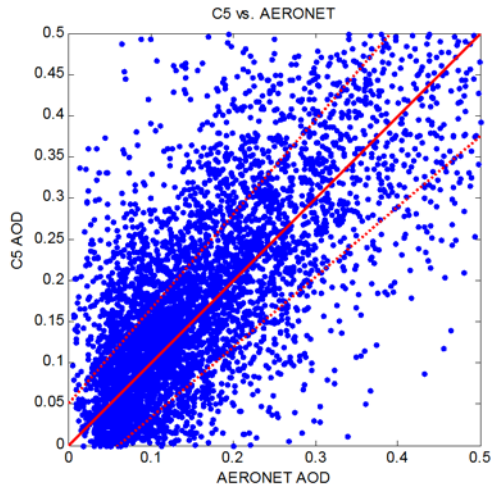
where I is the number of predictions that satisfy relation (4.7) and N is the total number of predictions.

In order to demonstrate the need for using various kinds of measures for AOD predictor evaluation, let us analyze the accuracy of C5 AOD prediction. A scatter plot of AOD prediction vs. true AOD during the 2005 over whole globe is depicted in Figure 3 while the values of the proposed accuracy measures are shown in Table 1. In Figure 3, a solid line represents an ideal, desirable AOD prediction, while dashed lines represent boundaries of an area containing predictions whose quality is deemed acceptable by domain scientists. In Figure 3a, the whole range of AOD values is plotted, while the zoomed-in portion of Figure 3a for small values of AOD (defined as AOD<0.5 [5]) is presented in Figure 3b.

From Table 1, we can conclude that C5 AOD predictor has a good performance based on the CORR accuracy. However, R^2 accuracy tells us that there is a significant portion of variance that the C5 algorithm was unable to model. MSE accuracy is difficult to judge when the accuracy of some simpler competing predictors is not available. Furthermore, the domain specific MSRE accuracy is higher than 1, which indicates lower than expected accuracy. Finally, the FRAC measure shows that almost 40% of predictions are of insufficient accuracy.



(a) Whole range of AOD



(b) Small AOD

Figure 3: Scatter plot of operational C5 satellite based AOD vs. AERONET AOD.

Table 1: Operational AOD vs. AERONET AOD accuracy for Jan-Aug 2005

# points	Deterministic AOD prediction				
	MSE	R^2	CORR	MSRE	FRAC
3224	0.020	0.66	0.87	2.02	63%

4.2 Relative error as a neural network cost function. Since AERONET AOD predictions are considered highly accurate [2], they can be used as target values in data-driven approaches for AOD prediction. Construction of the neural network AERONET AOD predictor based on the MODIS observations will be explored here.

The standard approach in building neural networks uses the MSE minimization as the optimization objective. This kind of cost function treats all errors equally regardless of the level of target value. Based on the domain knowledge, we know that large values of AOD can often be considered outliers. Hence, in this application, using an MSE function as the cost function for the neural network training is not the most appropriate.

To address this issue, we use an alternative neural network cost function defined as MSRE in (4.8). When using MSRE as the optimization criterion, the backpropagation algorithm for neural network training should be modified by replacing

$$\frac{\partial MSE}{\partial y_i} = \frac{2}{N}(y_i - t_i) \tag{4.10}$$

with

$$\frac{\partial MSRE}{\partial y_i} = \frac{2}{N}\frac{(y_i - t_i)}{(0.05 + 0.15t_i)^2} \tag{4.11}$$

By analyzing equation (4.11), we can conclude that the influence of prediction error on the backpropagation is large when ti is small and vice versa. The MSRE criterion also provides the solution for the problem with potential outliers in the training set, as influence of the errors made on large AOD during the training procedure weakens.

4.3 Fusion of satellite and ground based observations. Although MODIS instrument has high spatial resolution (one pixel is as small as $250 \times 250m^2$ at nadir), operational MODIS algorithms do not predict AOD for single pixels due to the low signal to noise ratio [5]. Instead, single pixels are aggregated to larger areas. Based on the assumption that AOD has small spatial variability, the operational C5 algorithm predicts AOD in $10 \times 10km^2$ blocks. After discarding cloud, snow, ice and water pixels along with 20% of the darkest and 50% of the brightest ones, the average of remaining pixels is taken as representative of the corresponding $10 \times 10km^2$ block.

Validation of satellite based AOD predictions is usually performed using AERONET AOD predictions as ground truth [4]. Whereas MODIS achieves an almost complete global coverage daily, AERONET predictions are provided many times every day but only over selected locations. Validation studies showed that

it would be inappropriate to compare AOD from a single MODIS block directly to an AERONET point measurement [14]. Hence, the method of fusion of the AERONET and the MODIS data has been proposed [14] (Figure 4). Essentially, this method involves aggregating initial MODIS blocks of $10 \times 10km^2$ size into blocks of size $50 \times 50km^2$ around each AERONET site, called spatial merging.

Because MODIS and AERONET AOD predictions may occur at different times, temporal data merging is necessary. AERONET AOD data are acquired on average at intervals of 15 min. Assuming slow AOD variation over short time periods, the MODIS AOD predictions are said to be temporally collocated with the corresponding AERONET AOD predictions if there is a valid AERONET AOD prediction within 30 minutes of the satellite flyover. The data collocated in this way can be obtained from the official MODIS website of NASA [5]. Each collocated data point is represented with time, date, average AERONET AOD, average MODIS observations, and ancillary attributes. In our experiments, we attached a collocated data point to the data set if we have at least one valid out of possible 25 MODIS AOD predictions in $50 \times 50km^2$ spatial block and at least one valid AERONET AOD prediction within the 30 minutes from the satellite overpass.

4.4 Data Description. There are several types of AERONET AOD ground-truth data that differ in amount and quality [2] and several versions of MODIS AOD predictions [5]. Although AERONET Level 1.5 data were cleaned and cloud contaminated measurements were removed, outliers are still present. To avoid a potential problem with outliers in ground-truth data, AERONET Level 2.0 observations were considered since they are both cloud screened and manually verified. The primary benchmark AOD predictor for comparison with our approach is the most recent version of the MODIS operational algorithm, C5, as validation studies show that version C5 is significantly more accurate than the previous version C4 [5]. However, at the time of this study, both AERONET Level 2.0 measurements and C5 predictions were available only for the first eight months of 2005. In order to give a more complete evaluation of our partitioning algorithm, we also compared with C4 predictions that were available during the longer period, between April 2003 and November 2005. The C4 data set consisted of 23,903 data points containing MODIS observations, C4 AOD prediction, and collocated AERONET Level 1.5 AOD measurements from 129 AERONET sites over the globe (Figure 5).

To compare data-driven AOD prediction to both C4 and C5 algorithms, we also collected 3,234 data points

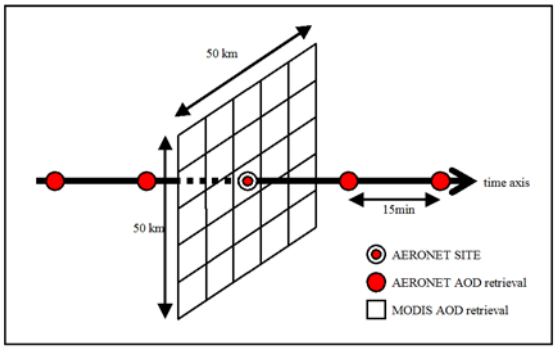


Figure 4: Spatio-temporal fusion of MODIS observations and AERONET AOD predictions.

Table 2: List of attributes collected from the data

Attribute index	Description
1-4	Mean of reflected radiation in $50 \times 50km^2$ blocks at three wavelengths
5-8	Std. deviation of reflected radiation in $50 \times 50km^2$ blocks at three wavelengths
9-12	Solar zentih, Solar azimuth, Sensor zenith, Sensor azimuth angles
13	Surface elevation

with MODIS observations where both C4 and C5 AOD predictions were available together with AERONET Level 2.0 AOD measurements. These data corresponded to the first eight months of 2005 collected at 110 AERONET sites.

As shown in Figure 5, AERONET sites are not uniformly distributed. The highest density is within the U.S. On the other hand, continental Asia, Africa, and Australia are poorly covered. Hence, as a cautionary note, global applicability of data-driven AOD prediction methods are somewhat limited. During the studied time period, measurements from the 129 AERONET sites were not uniformly distributed. For example, the data were undersampled during rainy seasons.

In order to make fair comparison between the deterministic and data-driven methods, we extracted only the satellite-based attributes that were used as inputs to both C4 and C5 algorithms. The attributes used are listed in Table 2. The four wavelengths were taken from the MODIS range between 440nm - 2100nm, as these are sufficient to describe aerosol properties [5].

By convention, AOD is reported at the 550nm wavelength. As AERONET instruments do not provide AOD at that wavelength, based on domain knowledge, we performed log-linear interpolation of AERONET

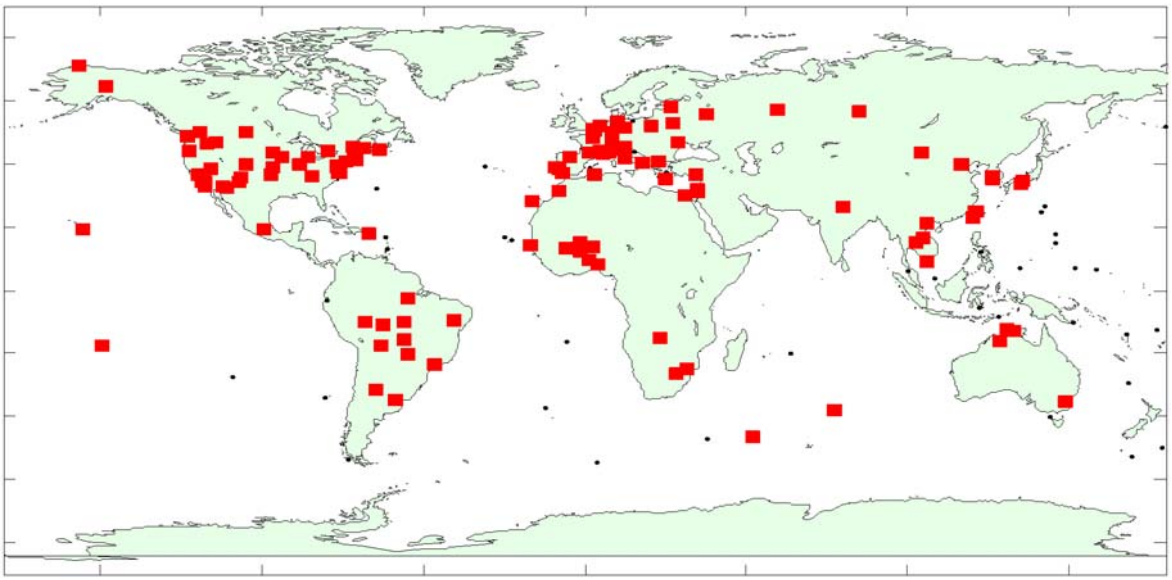


Figure 5: Location of 129 AERONET sites used in competition procedure.

AOD measurements at 440nm and 670nm to estimate AOD at 550nm [5].

4.5 AOD predictions by a single neural network predictor. We evaluated the performance of neural networks as AOD predictors from the MODIS attributes listed in Table 2. Neural networks were trained on 16,328 collocated data points, collected between April 2003 and November 2004. The remaining 7,575 data points, from December 2004 to November 2005, were used as test set for accuracy estimation. It should be noted that the test data covered the consecutive four seasons that were not seen during training. By dividing dataset in this way, we avoided the problem of memorizing training data, which would have occurred if the training-test split was performed randomly. The memorization would occur because of the temporal correlation in AOD values that can remain significant over periods of up to a few weeks.

Neural networks with one hidden layer and one neuron in the output layer were used throughout all experiments. Inputs to neural networks are listed in Table 2. The sigmoid activation function has been used for all hidden neurons while the linear activation function was used for the output neuron. The neural networks were trained using MSE function as a cost function. We note that the MSRE function (4.8) should be used if we wanted to avoid the problem with outliers and we are especially interested in high accuracy at low AOD values.

The results obtained by using a different number of neurons in the hidden layer are presented in Table 3. The values in the table represent averages from 10 repeated experiments. Compared to the results obtained by C4 algorithm on the same dataset presented in Table 4, we can conclude that neural networks were more accurate in predicting AOD than the C4 algorithm for all 5 accuracy measures. It should be noted that a simple linear predictor achieved AOD prediction accuracy comparable to the accuracy of C4 algorithm.

Table 3: Single neural network vs. AERONET AOD accuracy

# neurons in hidden layer	Neural Network AOD prediction				
	MSE	R^2	CORR	MSRE	FRAC
0(linear output)	0.034	0.46	0.69	3.08	49%
5	0.024	0.62	0.79	2.27	56%
10	0.021	0.67	0.82	1.95	60%
20	0.020	0.68	0.82	1.90	61%

Since there was no significant difference between neural networks with 10 and 20 hidden nodes, we used neural networks with 10 hidden neurons in all of the remaining experiments.

4.6 Experiments using predictors specialized for C5 spatio-temporal partitions. To test whether spatio-temporal partitions defined in operational C5 algorithm could be used in data-driven AOD prediction

Table 4: C4 AOD vs. AERONET AOD accuracy for Dec 2004 - Nov 2005

# points	Deterministic AOD prediction				
	MSE	R^2	CORR	MSRE	FRAC
7575	0.034	0.46	0.79	5.36	52%

Table 5: Neural networks specialized for C5 partitions vs. AERONET AOD accuracy for Dec 2004 - Nov 2005

# points	Neural Networks AOD prediction				
	MSE	R^2	CORR	MSRE	FRAC
7575	0.023	0.63	0.80	2.40	61%

Table 6: Neural networks specialized for partitions discovered by competition vs. AERONET AOD accuracy

# models	T (months)	Neural Networks AOD prediction				
		MSE	R^2	CORR	MSRE	FRAC
2	12	0.017	0.73	0.85	1.80	65%
	6	0.015	0.75	0.87	1.55	68%
	3	0.019	0.70	0.84	1.76	64%
	1	0.017	0.73	0.85	1.82	65%
3	12	0.019	0.70	0.84	1.83	65%
	6	0.018	0.72	0.85	1.77	67%
	3	0.018	0.72	0.85	1.70	66%
	1	0.022	0.65	0.82	2.23	67%

approach, we trained neural networks specialized for the three regions presented in Figure 1. Each of the three neural networks was trained on data belonging to one of the partitions (white, gray, or black) depicted in Figure 1. Similar to training-test partition in Section 4.5, we used data between April 2003 and November 2004 for training, while the test set was taken between December 2004 and November 2005. The results are presented in Table 5.

While the achieved accuracy was better than the accuracy of C4 algorithm (Table 4), considering all measures, the specialized predictor was worse than using a single predictor (Table 3).

4.7 Experiments using predictors specialized for spatio-temporal partitions discovered by competition. The results in Table 5 indicate that domain-based partitioning is not suitable for learning specialized AOD predictors. Instead, we applied the proposed competition method from Section 3 to find spatio-temporal partitions.

To run the procedure, we have to choose an appropriate size of spatio-temporal cells. Each cell is defined as a time interval for a specific AERONET site. We evaluated several cell size choices, empirically. The largest temporal size we considered was $T = 12$ months as the aerosol concentration is periodic with yearly cycles. In addition, we considered smaller temporal sizes of $T = 6$, $T = 4$ and $T = 1$ months. Cell temporal size is fixed during the competition procedure. The competition starts with $K = 2$ models. In the first step, the entire dataset is divided randomly into K equal sized subsets. Next, neural network predictors were trained on each of the K subsets. Data from each AERONET site were partitioned into the consecutive, disjoint temporal cells of size T. Given predictions of the competing predictors on all examples within a cell, the cell is assigned to the model achieving the smallest prediction error. The competition iterated until a stable solution was found. The experiment was repeated for various parameter values K and T. Finally, all possible partitionings were evaluated on the independent test set and the best one was chosen as the final solution.

There were several additional issues that had to be addressed. First, we wanted to avoid making evaluations of the competing predictors on the training data. Instead of training competing neural networks on the complete training data set, we applied 4-cross-validation procedure. Data from each month were partitioned into 4 weekly intervals; one week was used for validation, while the remaining three weeks were merged and used for training.

Second, the cost function for training the neural networks in the competition procedure had to be determined. Due to an abundance of outliers in the training data, the standard MSE function was not the most appropriate choice because the training procedure would be dominated by the outliers and it would be difficult to find a stable solution. To overcome this problem we used MSRE as a cost function. As discussed in Section 3.2, this function is less sensitive to the outliers.

Third, prediction models compete for the cells based on the prediction error, which has to be defined. To avoid the possibility that outliers could dominate the competition procedure, we used average MSRE error over the cell to determine the winning model. The model that achieves minimal MSRE was considered the winner. Finally, the neural network predictors were built on the discovered spatio-temporal partitions. Those networks had to be trained using the standard MSE cost function, since networks trained with MSRE as a cost function tend to underestimate large AOD values.

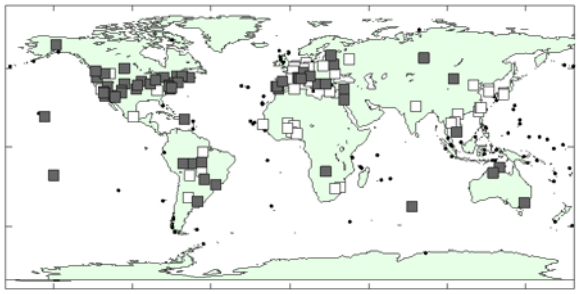
The competition procedure was applied on the explained training data from between April 2003 and November 2004. The learned spatio-temporal partitions were evaluated on the test data between December 2004 and November 2005. The results for the different K and T values are presented in Table 6.

Based on the results from Table 6, we can conclude that the proposed competition-based spatio-temporal data partitioning approach resulted in large accuracy improvements. The best results were obtained for cell size of six months ($T = 6$), where one interval covered winter-spring months and another summer-fall months, and for spatio-temporal partitioning that results in two specialized predictors ($K = 2$).

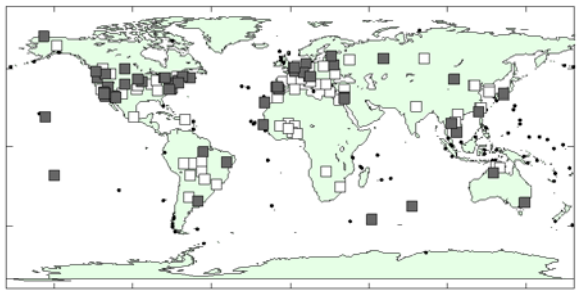
The resulting spatio-temporal partitions are shown in Figure 6. From Figure 6, we can see that during the winter-spring months the whole U.S. was assigned to the same partition, while during the summer-fall months some U.S. sites moved to the other partition. Also, AERONET sites in Africa did not change their assignment during the year. It is interesting to mention that average AOD in the gray partition is 0.13 with standard deviation 0.12 while average AOD in white partition is 0.29 with standard variation 0.35. Although it might appear that the competition procedure discovered partitions based on the average AOD values, the standard deviation suggests that the underlying process is more complicated.

4.8 Comparison of C4, C5 and data-driven predictors. We compared our specialized predictors to the recently developed C5 algorithm. As it is described in Section 3.4, data from the first eight months of 2005 were extracted because during that time period both C4 and C5 predictions are available. We tested our best predictor that consisted of two neural networks specialized for two spatio-temporal subsets from Figure 6. The results are presented in Table 7. C5 significantly outperformed C4 on all accuracy measures. The specialized neural networks were more accurate in predicting AOD than the operational C5 algorithm for all 5 of the accuracy measures.

The improvement in the AOD prediction can be seen in Figure 7, where comparative scatter plots of C5 vs. AERONET AOD and AOD predicted by specialized neural networks vs. AEROENT AOD are presented. By inspecting these plots, we can conclude that the specialized neural networks were equally successful for both small and large AOD. The higher accuracy of the proposed method in predicting small AOD can be seen in the zoomed-in plots in Figure 7. The bias in predicting small AOD values was significantly reduced using the proposed method as compared to C5.



(a) two partitions for winter-spring months



(b) two partitions for summer-fall months

Figure 6: Spatio-temporal partitions obtained by competition procedure.

Table 7: Predictors vs. AERONET AOD accuracy

Predictor	AOD prediction accuracy				
	MSE	R^2	CORR	MSRE	FRAC
C4	0.031	0.55	0.84	4.52	48%
C5	0.020	0.70	0.88	1.93	63%
Neural Network local predictors	0.014	0.81	0.90	1.50	68%

5 Conclusion.

Our previously reported results provide strong evidence that it is possible to develop a global data-driven AOD predictor which is as accurate as the operational knowledge-based AOD predictor [13]. In an effort to further improve accuracy of data-driven AOD prediction we explored benefits of learning an ensemble consisting of multiple regression models specialized for specific spatio-temporal data subsets.

In our experiments on three years of integrated AERONET and MODIS data, AOD prediction accuracy has not improved when training specialized data-driven models on knowledge-based partitions used by C5-algorithm. However, the competition-based spatio-temporal data partitioning proposed in this study resulted in large accuracy improvements. The resulting

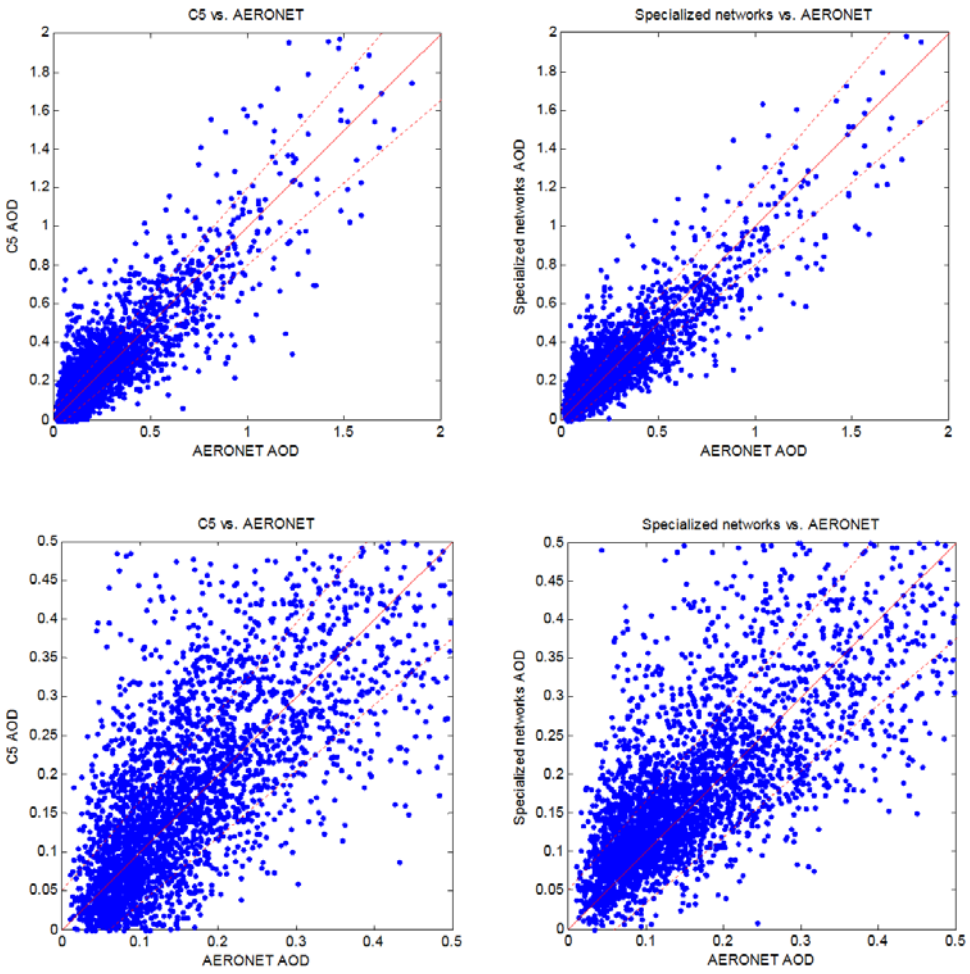


Figure 7: Scatter plots of C5 vs. AERONET (left panels) and specialized neural networks AOD predictions vs. AERONET (right panels); bottom panels zoomed scatter plots; solid line ideal prediction, dashed lines boundaries of acceptable error.

ensemble of specialized AOD predictors was a lot more accurate than both the global data-driven predictor and an ensemble of data-driven predictors trained on C5 partitions. Accuracy improvements were observed for a range of spatio-temporal partitioning parameter choices considered and for various accuracy measures.

The most accurate results were obtained when (1) the data from each AERONET location were first grouped in disjoint temporal cells of length equal to six months. In this manner, half of the cells from each location corresponded to winter-spring months and another half to summer-fall months; and (2) two specialized neural network predictors were competing for the cells using the proposed procedure.

References

[1] S. Vucetic and Z. Obradovic, *Discovering homogeneous regions in spatial data through competition*, in Machine learning: Proceedings of the 17th International Conference, (2000), pp. 1091–1098.

[2] S. Vucetic, Z. Obradovic, and K. Tomsovic, *Price-load relationships in California's electricity market*, IEEE Trans. Power Syst., 16 (2001), pp. 280–286.

[3] M. D. King, Y. J. Kaufman, P. W. Menzel, and D. Tanreacute, *Remote sensing of cloud, aerosol, and water vapor properties from the Moderate Resolution Imaging Spectrometer (MODIS)*, IEEE Trans. on Geosci. and Remote Sens., 30 (1992), pp. 2–27.

[4] R. C. Levy, L. A. Remer, et al., *Evaluation of the MODIS aerosol retrievals over ocean and land during CLAMS*, J. Atmos. Sci., 62(4) (2005), pp. 974-992.

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

[6] F. Aires, C. Prigent, W. B. Rossow, and M. Rothstein, *A new neural net-work approach including first guess for retrieval of atmospheric watervapor, cloud liquid water path, surface temperature, and emissivities overland from satellite microwave observations*, J. Geophys. Res., 106 (2001), pp. 14,887–14,908.

[7] V. M. Krasnopolsky, *Neural networks for standard and variational satelliteretrievals*, OMB Contrib. 148, NOAA Environ. Model. Cent., Ocean Model. Branch, Washington, D.C., 1997.

[8] F. Chevallier, F. Chruy, N. A. Scott, and A. Chdin, *A neural networkapproach for a fast and accurate computation of a longwave radiativebudget*, J. Appl. Meteorol., 37 (1998), pp. 1385–1397.

[9] M. D. Muller, A. K. Kaifel, M. Weber, S. Tellmann, J. P. Burrows and D. Loyola, *Ozone profileretrieval from Global Ozone Monitoring Experiment (GOME) data using a neural network approach (neural network ozone retrieval system (NNORSY))*, J. Geophys. Res., 108(D16),4497, 2003.

[10] C. Jamet, C. Moulin, and S. Thiria, *Monitoring aerosol opticalproperties over the Mediterranean from SeaWiFS images usinga neural network inversion*, Geophys. Res. Lett., 31, 2004.

[11] B. Han, Z. Obradovic, Z. Li and S. Vucetic, *Data Mining Support for Improvement of MODIS Aerosol Retrievals*, Proc. IEEE Intl Geoscience and Remote Sensing Symp., Denver, CO, Aug. 2006.

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

[13] V. Radosavljevic, S. Vucetic and Z. Obradovic, *Aerosol Optical Depth Retrieval by Neural Networks Ensemble with Adaptive Cost Function*, Proc. 10th Intl Conf. Engineering Applications of Neural Networks, Thessaloniki, Greece, Aug. (2007), pp. 266-275.

[14] C. Ichoku, D. A. Chu, et al., *A spatio-temporal approach for global validation and analysis of MODIS aerosol products*, Geophys. Res. Lett. 29(12): no.616, 2002.