Uncertainty Analysis of Neural-Network-Based Aerosol Retrieval

Kosta Ristovski, Slobodan Vucetic, and Zoran Obradovic

***Abstract*—Neural networks have the ability to represent and learn complex regression functions and are very suitable for retrieval of geophysical parameters from remotely sensed data. Neural networks trained to minimize the mean square error are able to estimate the conditional expectation of target variables. In many remote sensing applications, it is also critical to provide estimates of prediction uncertainty. In this paper, we evaluate an approach that, in addition to training a neural network for retrievals, also trains a neural-network-based estimator of retrieval uncertainty. The uncertainty estimator is built under the assumption that uncertainty is a function of input variables. The methodology was evaluated on aerosol-optical-depth retrieval. The data set consists of 38 238 collocated Moderate Resolution Imaging Spectrometer (MODIS) satellite instrument and Aerosol Robotic Network ground-based instrument measurements collected over the entire Earth during two years (in 2005–2006). The results indicate that a neural network ensemble is more accurate than the operational MODIS retrieval algorithm called Collection 5 and that the retrieval uncertainty of the ensemble can be estimated with satisfactory accuracy.**

***Index Terms*—Regression, remote sensing, uncertainty.**

I. Introduction

AEROSOLS, small particles emanating from natural and man-made sources, have been recognized as the largest source of uncertainty for understanding the Earth's radiative budget [1]. Aerosol retrieval is very important for climate research, weather forecasting, environmental monitoring, and understanding the impact of pollution on human health [2]. Aerosol optical depth (AOD), which measures the amount of depletion that a beam of solar radiation undergoes as it passes through the atmosphere, is one of the most important properties of atmospheric aerosols.

Multiple satellite and ground-based sensors have been deployed for remote sensing of aerosols. In this paper, we are proposing a machine learning procedure that exploits multisource observations from satellite and ground-based sensors. We consider aerosol-related data collected by highly accurate ground-based Aerosol Robotic Network (AERONET) instruments and by Moderate Resolution Imaging Spectrometer (MODIS) instruments aboard Terra and Aqua satellites. MODIS retrievals made by the operational algorithm called Collection 5 (C005) are reported together with quality assurance (QA) confidence (QAC) flags ranging from three (high confidence) to zero (low or no confidence). QAC is a useful qualitative measure of retrieval uncertainty, which can be a significant limitation in some applications [3].

A statistical retrieval approach can be treated as regression, where the goal is to learn a functional relationship between MODIS observations and true AOD that is assumed to equal AERONET retrieval. Experimental results provide strong evidence that neural-network-based AOD retrieval is more accurate than the physically based C005 operational algorithm [4]. However, neural networks, in their basic form, could not estimate the retrieval uncertainty. An alternative Gaussian process [5] approach has the ability to provide the conditional distribution and, thus, the uncertainty estimation. However, the use of this approach is limited for AOD retrieval due to high computational costs stemming from inversions of very large matrices.

To enhance the neural-network-based AOD retrieval, an appropriate uncertainty estimation model is needed. A method for neural network uncertainty assessment with application to remote sensing was studied in [6] and [7]. It was based on the assumption that targets are corrupted by Gaussian noise with zero mean and constant variance. In AOD retrieval, the noise variance is heteroscedastic (not constant) and depends on surface properties, aerosol microphysics and distribution, and viewing geometry. In this paper, we model retrieval error by Gaussian distribution with zero mean and input-dependent variance. It is worth noting that a Bayesian method for regression learning with input-dependent noise was proposed in [8]. However, this method requires calculating large Hessian matrices and their inverses during training of neural networks and is therefore prohibitively time consuming for large-scale applications in remote sensing. As an alternative, a bootstrap-based technique that is tractable for large data sets was proposed in [9]. In this paper, we will apply this approach to provide uncertainty estimation of neural-network-based retrievals. In the experimental section, we will study how correlated the uncertainty estimates are with the actual prediction errors. We will also study the relationship between QAC values from MODIS C005 retrieval algorithm and neural network retrieval accuracy and uncertainty estimation.

II. Methodology

In remote sensing and many other data-rich domains, data sets frequently contain observations of differing quality.

Manuscript received December 2, 2010; revised May 4, 2011; accepted July 10, 2011. Date of publication October 3, 2011; date of current version January 20, 2012. This work was supported in part by the National Science Foundation under Grants IIS-0612149 and IIS-1117433.

The authors are with the Center for Data Analytics and Biomedical Informatics, Temple University, Philadelphia, PA 19122 USA (e-mails: kosta.ristovski@temple.edu; vucetic@temple.edu; zoran.obradovic@temple.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2011.2166120

Recently, there has been an increasing interest [10], [11] in methods that assess data quality and then learn a prediction model based on the assessment. In aerosol retrieval, information about the quality of MODIS observations is available in advance through QAC flags. Our goal is to understand how data quality affects neural-network-based retrieval and to analyze uncertainty estimations of such models.

A. Problem Setup and Preliminaries

Let us assume that we are given a data set $D = \{(x_i, y_i), i = 1, 2, \ldots, N\}$, where x_i is a vector of input variables derived from MODIS observations and y_i denotes the AOD values retrieved by collocated AERONET retrievals. A standard regression model trained from D assumes that the target y is related to input vector x as

$$y(x) = f(x) + \varepsilon(x) \tag{1}$$

where $\varepsilon(x)$ is a random variation of y around regression function $f(x)$. The noise is typically assumed to be Gaussian with zero mean and constant variance σ_n^2. As discussed in the introduction, the constant variance assumption is not appropriate for aerosol retrieval. Instead, the noise variance is modeled as a function of inputs $\sigma_n^2(x)$. Given the Gaussian noise assumption, we can rewrite (1) as

$$P\left(y(x)|f(x)\right) \sim N\left(f(x), \sigma_n^2(x)\right). \tag{2}$$

The noise variance $\sigma_n^2(x)$ is unknown, and it has to be learned from data. Another unknown quantity is the regression function $f(x)$. Let us denote by $m(x)$ the estimate of $f(x)$ learned from the data. If the learning algorithm is a universal approximator, such as neural networks, a common assumption is that $m(x)$ is an unbiased estimate of $f(x)$ and that we can represent $P(f(x)|m(x))$ as the Gaussian distribution

$$P\left(f(x)|m(x)\right) \sim N\left(m(x), \sigma_m^2(x)\right) \tag{3}$$

where $\sigma_m^2(x)$ is the model variance that also has to be learned from the data. Given distributions (2) and (3) and assuming that the noise components $f(x) - m(x)$ and $\varepsilon(x)$ are independent, we can represent target distribution as

$$\begin{aligned} P\left(y(x)|m(x)\right) &= P\left(y(x)|f(x)\right) \cdot P\left(f(x)|m(x)\right) \\ P\left(y(x)|m(x)\right) &\sim N\left(m(x), \sigma^2(x)\right) \\ \sigma^2(x) &= \sigma_m^2(x) + \sigma_n^2(x) \end{aligned} \tag{4}$$

where $\sigma^2(x)$ is the target variance given the prediction $m(x)$ and is a sum of noise variance and model variance. Distribution (4) allows us to both provide the prediction in the form of $m(x)$ and estimate the prediction uncertainty in the form of input-dependent target variance $\sigma^2(x)$. In summary, to be able to provide prediction and prediction uncertainty, one should learn $m(x)$, $\sigma_n^2(x)$, and $\sigma_m^2(x)$ from data set D. In the following section, we will describe a previously proposed [9] robust learning procedure that achieves this objective.

B. Bootstrap Approach for Regression and Uncertainty Estimation

In [9], an approach was proposed to first train an ensemble of neural network predictors and use it to provide prediction function $m(x)$ and estimate model variance $\sigma_m^2(x)$. Then, a separate neural network is trained to estimate heteroscedastic noise variance $\sigma_n^2(x)$. The details are as follows.

The neural network ensemble consists of b neural networks, each trained on a different bootstrapped sample from the training data set D. The ith neural network $m_i(x)$ is trained from data set D_i which has N examples sampled with replacement from the original training set D. We note that D_i contains in average only 63% of the original examples which are included in the bootstrapped data set.

The b neural networks are averaged to provide the prediction function $m(x)$ as

$$m(x) = (1/b) \sum_{i=1}^{b} m_i(x). \tag{5}$$

Owing to the availability of an ensemble of predictors, the model variance can be estimated as

$$\sigma_m^2(x) = (1/(b-1)) \sum_{i=1}^{b} \left(m_i(x) - m(x)\right)^2. \tag{6}$$

Estimating noise variance $\sigma_n^2(x)$ is a nontrivial problem. It has been solved in [9] by introducing a separate neural network $m_n(x)$ trained to estimate

$$l^2(x) = \max\left(0, r^2(x) - \sigma_m^2(x)\right) \tag{7}$$

where $r(x)$ is the residual of the bootstrap committee defined as $r(x) = y(x) - m(x)$. Value $l(x)$ serves as a proxy for noise $\varepsilon(x)$ whose variance needs to be estimated. Neural network $m_n(x)$ is trained using data set $D_n = \{(x_i, l^2(x_i)), i = 1, 2, \ldots, N\}$, obtained from the original training data set D.

As can be seen from (7), the model variance $\sigma_m^2(x)$ is needed to obtain $l^2(x)$ value. To provide unbiased estimates of noise variance, out-of-sample examples have to be used. An example from the training set is an out-of-sample example for a particular neural network if it has not been used for its training. This is a consequence of sampling with replacement applied for making bootstrap replicates. Assuming that example x_a did not appear in k out of the m replicates, the corresponding k neural networks in the committee can use that data point as a test example. Taking into account outputs of these k neural networks, estimates of the mean and model uncertainty for out-of-sample example x_a can be calculated in similar way as in (5) and (6) and used for estimation in (7). As the noise is modeled with Gaussian distribution, the additional neural network for noise variance estimation is trained to minimize negative log likelihood L using

$$L = -\sum_{i=1}^{N} \log\left(\frac{1}{\sqrt{2\pi m_n^2(x_i)}} \exp\left(-\frac{l^2(x_i)}{2m_n^2(x_i)}\right)\right). \tag{8}$$

C. Accuracy Measures

To compare accuracies of $C005$ versus the bootstrap committee, we use the standard accuracy measures: the coefficient of determination (R^2), correlation $(CORR)$, and the root mean square error (mse) $(rmse)$. In addition, we also use a domain-specific accuracy measure called fraction of successful predictions $(FRAC)$ [12] and the average negative log-predictive density $(NLPD)$ used previously to estimate success of uncertainty prediction [13]. The $FRAC$ and $NLPD$ accuracy measures are defined in this section.

1) Fraction of Successful Predictions ($FRAC$): AOD predictions are considered sufficiently good [14] if they fall within the region specified by

$$|y_i - m(x_i)| \leq 0.05 + 0.15 y_i. \tag{9}$$

$FRAC$ is defined as the percentage of successful predictions

$$FRAC = (I/N) \cdot 100\% \tag{10}$$

where N is the total number of points and I is the number of points that satisfy (9).

2) Average NLPD: The average NLPD [13] of the true targets on n data points is a measure calculated as

$$NLPD = (1/n) \sum_{i=1}^{n} \left[\log \sigma(x_i) + (y_i - m(x_i))^2 \,/\, \left(2\sigma^2(x_i)\right) \right] \tag{11}$$

where $m(x_i)$ and $\sigma^2(x_i)$ are the mean and variance of the target distribution for point x_i. $NLPD$ is sensitive to the quality of both prediction and uncertainty estimation. Smaller values of $NLPD$ correspond to better quality of the estimates.

III. Data Set

MODIS sensors aboard Terra and Aqua satellites observe reflected solar radiation through multiple spectral bands [15] and provide almost daily global coverage with high spatial resolution [16]. The operational MODIS retrieval algorithm called C005 is an inverse operator derived from a forward-simulation model according to the domain knowledge of aerosol physical properties. During the retrieval process, the C005 algorithm tests how well the observed data meet certain criteria specified by domain scientists, using the QA plan [17], [18]. The result is reported as the QA flag. QA flags are designed not only to report success or failure of criteria being used in retrieval but also to estimate the data quality. The QAC flag, used in this study, is derived from QA values. It has four possible values, $QAC \in \{0, 1, 2, 3\}$, with $QAC = 3$ indicating high-quality retrievals and $QAC = 0$ indicating very low quality retrievals.

AERONET sites are located at fixed locations over the globe and acquire data every 15 min. Level 2.0 AERONET AOD retrievals have estimated uncertainties of around ± 0.01 [19], and thus, they are often considered as ground truth for validation of MODIS AOD retrievals. Alternatively, as is done in this paper, AERONET retrievals can be used as target variables during training of statistically based retrieval algorithms.

TABLE I
LIST OF ATTRIBUTES COLLECTED FROM MODIS

Attribute index	Description
1-4	Mean reflectance in 50x50km^2 blocks at four wavelengths
5-8	Standard deviation of reflectance
9-13	Solar Zenith, Solar Azimuth, Sensor Zenith, Sensor Azimuth, Scattering Angle
14	AERONET site elevation

TABLE II
NUMBER OF DATA POINTS ON YEARLY BASIS COLLECTED FROM MODIS AND GROUPED BY QAC FLAGS

	Year	
	2005	2006
QAC = 0	6052	9040
QAC = 1	2292	2575
QAC = 2	1554	1725
QAC = 3	10504	4496

TABLE III
DISTRIBUTION OF DATA POINTS OVER CONTINENTS GROUPED BY QAC FLAG

	North A.	South A.	Europe	Africa	Asia&Aust.
QAC = 0	4693	1015	5267	1970	2147
QAC = 1	1288	338	1878	671	692
QAC = 2	784	256	1375	468	396
QAC = 3	3748	1209	5820	2084	2139

TABLE IV
DISTRIBUTION OF DATA POINTS WITH DIFFERENT QACs AT DIFFERENT MAGNITUDES OF AOD

AOD	Points with QAC=3	Points with QAC<3
0-0.05	2022	3373
0.05-0.1	3739	5769
0.1-0.2	4345	6794
> 0.2	4986	7210

The data set has been obtained after a spatiotemporal collocation of the two sources of data [20]. Spatial collocation is achieved creating a grid of 5×5 MODIS retrievals with an AERONET site placed in the middle. The data are said to be temporally collocated if AERONET AOD observation is obtained within 30 min of the satellite overpass. Our resulting data set contains 38 238 observations from MODIS collocated with 93 AERONET sites over the whole globe during the period of two years (in 2005–2006). Each of the 93 sites has more than 15 high-quality $(QAC = 3)$ points per year. Fourteen satellite-based attributes listed in Table I have been used together with QAC flags for AOD retrieval and uncertainty estimation. To get a better insight into the distribution of QAC flags, in Table II, we summarize its temporal distribution, in Table III its spatial distribution, and in Table IV its distribution with respect to AERONET AOD values.

IV. Experimental Results

A. Experimental Design

To build an ensemble of neural networks for AOD prediction, we used hundred feedforward neural networks with a single hidden layer of ten neurons. Overfitting was avoided by

TABLE V
ACCURACIES FOR MODELS TRAINED AND TESTED ON THE DATA WITH DIFFERENT QUALITIES

Dataset	R^2	FRAC	CORR	RMSE	NLPD
Training (QAC=0,1,2,3) Test (QAC = 3)	0.8060±0.0001	76.10±0.03%	0.9020±0.0001	0.1030±0.0001	-1.96±0.01
Training (QAC=3) Test (QAC = 3)	0.8189±0.0001	76.07±0.02%	0.9050±0.0001	0.0995±0.0002	-2.02±0.02
C005 Test (QAC=3)	0.786	71.2%	0.898	0.108	N/A
Training (QAC= 0,1,2,3) Test (QAC =0,1,2)	0.7632±0.0004	68.93±0.01%	0.8743±0.0002	0.1180±0.0001	-1.82±0.02
Training (QAC= 3) Test (QAC =0,1,2)	0.6929±0.0001	63.61±0.01%	0.8444±0.0001	0.1344±0.0001	-1.70±0.01
C005 Test(QAC=0,1,2)	0.685	63.8	0.86	0.136	N/A

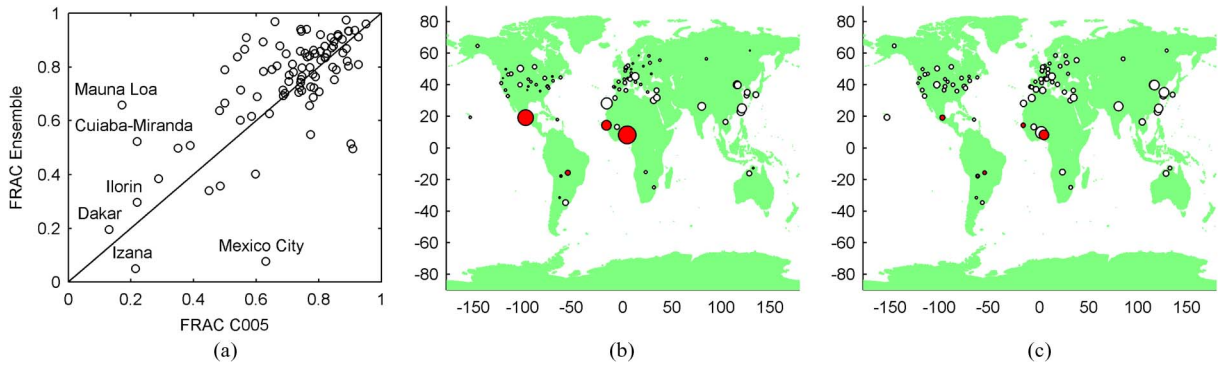

Fig. 1. (a) Scatter plot of site-specific fractions of successful predictions: Ensemble versus C005 (b) averaged prediction errors of a neural network ensemble and (c) averaged retrieval uncertainties by our method.

monitoring the mse on the separate validation data. To prevent overly long training time, the training process was stopped if the number of training epochs reached 300. The additional neural network for uncertainty estimation had five hidden neurons, and it was trained to minimize the negative log likelihood from (8).

The accuracy was estimated using one-site-out validation. Specifically, data from all AERONET sites except one were used as a training set while the predictor was tested on the remaining site. This procedure was repeated 93 times, such that each of the AERONET sites was used once as the test set. Test predictions for all 93 rounds were pooled together and used to calculate accuracy. To be able to establish statistical significance of obtained accuracies, we repeated each of the one-site-out-validation experiments ten times. In this way, we obtained ten accuracies, and we report their mean and standard deviation.

B. Analysis of Retrieval Accuracy

We evaluated two types of neural network ensembles for AOD prediction: One trained using all available training data, and another trained using only high-quality training data with $QAC = 3$. In Table V, we summarize the accuracies of the two ensembles and C005 retrieval algorithm on high-quality $(QAC = 3)$ and low-quality $(QAC < 3)$ test data. For statistical comparison of accuracies, we used the one-tailed t-test with 5% significance level. As expected, all three predictors have significantly higher accuracy on higher quality data. Both neural network ensembles are significantly more accurate than C005 on both high- and low-quality test data and over all studied accuracy measures. It is worth noting that the ensemble trained including low-quality data is significantly more accurate on low-quality test data, while the one trained on high-quality data is significantly more accurate on high-quality test data. This result shows that QAC quality flag is indeed a very useful qualitative measure of retrieval uncertainty.

To illustrate the difference between neural network ensembles and C005, in Fig. 1(a), we compare average $FRAC$ accuracies on each of the 93 AERONET sites. Neural networks were more accurate on 64 of the 93 sites. It could also be seen that $FRAC$ accuracy on majority of AERONET sites is relatively high, while there are several sites where the $FRAC$ accuracy is significantly lower. For further insight, in Fig. 1(b), we show the spatial distribution of rmse accuracies at 93 AERONET sites (rmse is proportional to the circle size). It can be seen that the four sites with the lowest accuracy (the largest circles) are in regions poorly covered by ground-based stations. We observe that the largest discrepancies between neural network ensembles and C005 algorithm are in Mexico City (C005 is more accurate) and Mauna Loa in Hawaii (C005 is less accurate). Both sites will be discussed in the next section.

C. Analysis of Uncertainty Estimation

In this section, we will present retrieval uncertainty analysis on high-quality data $(QAC = 3)$, showing that there are sizeable differences in retrieval uncertainty among such data. This is an important result because such differences are not

TABLE VI
SITES AT WHICH THERE IS UNDERESTIMATION OF UNCERTAINTY

AERONET site	Number of data points	NLPD at site
Mexico City	130	8.52
Ilorin	209	0.59
Dakar	46	0.24
Cuiaba-Miranda	218	-0.32

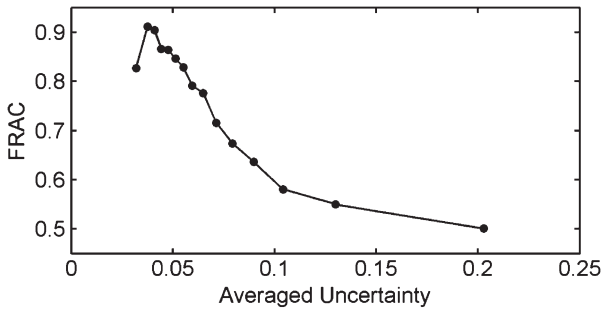

Fig. 2. Prediction accuracy measured as the fraction score for equal-width bins of 1000 points sorted from lower to higher uncertainty.

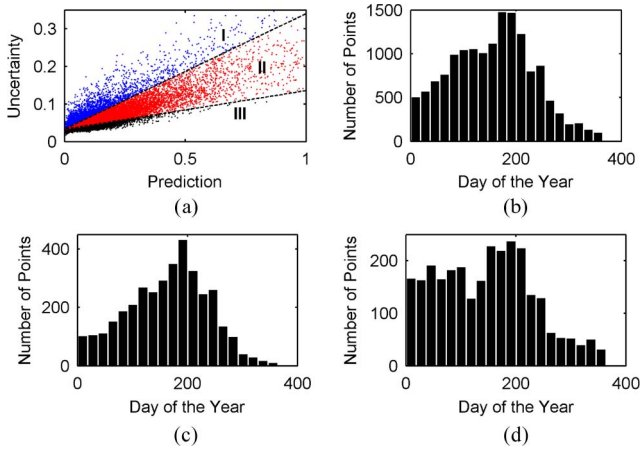

Fig. 3. (a) Scatter plot of prediction versus uncertainty. (b) Temporal distribution of all data points with $QAC = 3$. (c) Temporal distribution of points from the black region in (a). (d) Temporal distribution of points from the red region in (a).

detectable using QAC flag. The results presented here correspond to a neural network ensemble trained on high-quality $(QAC = 3)$ data. In Fig. 1(c), we show the site-averaged uncertainties at 93 sites, where the smaller circle means smaller average uncertainty. Comparing rmse and uncertainties from Fig. 1(b) and (c), we can see that they follow very similar pattern. The correlation of site-averaged estimated retrieval uncertainties and rmse is impressive, i.e., 0.74, which indicates the success of retrieval uncertainty estimation.

However, it is also worth noting that, at several sites, there is a large difference between rmse and estimated uncertainty. These sites are presented as red circles in Fig. 1(b) and (c) and are listed in Table VI. Interestingly, all of these sites are also among the six sites stated in Fig. 1(a), where accuracy is very low.

To get a further insight into the proposed uncertainty estimation method, we analyzed AOD retrievals of high-quality data versus the corresponding $FRAC$ scores. It is expected that AOD retrievals with small uncertainty estimates have higher $FRAC$ than AOD retrievals that have larger uncertainty estimates. Therefore, we sorted AOD retrievals according to uncertainty estimates in ascending order and then split them in equal-width bins (groups) of 1000 points. For each bin, $FRAC$ was measured and shown in Fig. 2. It can be seen that $FRAC$ had a decreasing trend with uncertainty except within the first bin. The first (leftmost) bin in Fig. 2 had a lower fraction value because it had a large fraction of data points from Mexico City site.

We compare AOD prediction and uncertainty estimate on all 38 238 data points from our data set in Fig. 3(a). It can be observed that prediction uncertainty grows with the value of AOD prediction. We defined three regions based on the relationship between uncertainty and prediction: the middle region denoted as II in Fig. 3(a), the small region denoted as III at the bottom, and the small region denoted as I at the top. Region I contained points for which uncertainty estimate was unusually high and had 2825 points. Region III contained points for which uncertainty estimate was unusually low and had 3656 points.

Our analysis revealed two sites where uncertainty estimates were unusually large. Those were Mauna Loa and Izana, which had almost all points within the red region in Fig. 3(a). Further analysis found that uncertainty at those sites is dominated by the model uncertainty, which was at the highest level among all 93 sites. As performed experiments were conducted by leave-one-site-out cross-validation protocol, the high model uncertainty suggests that those two sites are different from the other sites. Indeed, after investigating literature about these two sites, we found that both are located on high elevations, on islands (Mauna Loa in Hawaii; Izana-Tenerife in Spain), and that they serve for calibration of AERONET instruments.

Temporal distributions of high-quality data over the studied two-year period are shown in Fig. 3(b)–(d). Days of the year are labeled horizontally as 1–365 and correspond to the January 01 to December 31 period, and bins represent the number of high-quality data points $(QAC = 3)$ observed within the period covered by the bin. Temporal distribution of all high-quality data points is shown in Fig. 3(b). Points corresponding to the bottom part of Fig. 3(a) [denoted as III in Fig. 3(a)] belong to different sites around the world, but most of them are observed in summer. However, we found that points corresponding to the upper region of Fig. 3(a) [denoted as I in Fig. 3(a)] are evenly distributed over seasons, as shown in Fig. 3(d).

NLPD values at most sites vary in the range from -1.8 to -2.9, depending mostly on AOD retrieval errors. By NLPD analysis, we found that Mexico City is an outlier $(NLPD = 8.52)$. This is a very high elevation site (2268 m), and elevation is an important input attribute of our AOD retrieval model. We observed that neural network ensemble tends to predict small AOD at high elevation sites. Other sites with high elevations are BSRN BAO Boulder (1604 m), Mauna Loa (3397 m), and Izana (2268 m). The average AOD values at these sites are very small (less than 0.1). That is why Mexico City is an outlier, and so, it is a difficult site for neural network AOD retrieval as sites with similar AOD properties were not seen in the training set.

V. Conclusion

We have shown that data quality is very important not only for neural-network-based AOD retrieval but also for good uncertainty estimation of such retrievals. Leave-one-site-out experiments using two years of collocated satellite and ground-based observations were performed over the entire Earth in such a way that retrievals are made on locations not seen during training. This allowed us to evaluate how accurate neural network ensembles are in AOD retrieval, to compare their accuracy with that of the operational MODIS C005 retrieval algorithm, and to determine the quality of retrieval uncertainty. Committees of neural networks for AOD retrieval were more accurate than the operational C005 algorithm. The obtained uncertainty estimates were found to be reasonably accurate. This very promising result shows that it is possible to complement AOD retrievals with accurate quantitative estimates of retrieval uncertainty. The uncertainty analysis presented in this paper provides multiple opportunities to further improve our understanding of global properties of aerosols and to allow more informed use of AOD retrievals in the subsequent climatology studies. We also reported some shortcomings of the statistically based retrieval algorithms, as evident from AERONET site in Mexico City. Similarly, we found that data points collected at AERONET sites Mauna Loa and Izana are very different from other AERONET sites. Potential solutions for further improvements that we are currently investigating are aimed at an appropriate integration of the physically and statistically based AOD retrieval methods.

References

[1] J. Hansen, M. Sato, and R. Ruedy, "Radiative forcing and climate response," *J. Geophys. Res.*, vol. 102, no. D6, pp. 6831–6864, 1997.

[2] Z. Hu, "Spatial analysis of MODIS aerosol optical depth, PM2.5, and chronic coronary heart disease," *Int. J. Health Geograph.*, vol. 8, p. 27, May 2009.

[3] R. C. Levy, L. A. Remer, R. G. Kleidman, S. Mattoo, C. Ichoku, R. Kahn, and T. F. Eck, "Global evaluation of the collection 5 MODIS dark-target aerosol products over land," *Atmos. Chem. Phys.*, vol. 10, no. 21, pp. 10 399–10 420, Nov. 2010.

[4] V. Radosavljevic, S. Vucetic, and Z. Obradovic, "Aerosol optical depth retrieval by neural networks ensemble with adaptive cost function," in *Proc. 10th Int. Conf. Eng. Appl. Neural Netw.*, Thessaloniki, Greece, 2007, pp. 266–275.

[5] C. E. Rasmussen and C. K. I. Williams, *Gaussian Proccesses for Machine Learning*. Cambridge, MA: MIT Press, 2006.

[6] F. Aires, C. Prigent, and W. B. Rossow, "Neural network uncertainty assessment using Bayesian statistics: A remote sensing application," *Neural Comput.*, vol. 16, no. 11, pp. 2415–2458, Nov. 2004.

[7] C. M. Bishop, *Neural Networks for Pattern Recognition*. London, U.K.: Oxford University Press, 1995.

[8] C. M. Bishop and C. Qazaz, "Regression with input-dependent noise: A Bayesian treatment," in *Advances in Neural Information Processing Systems*, vol. 9, M. Mozer, M. Jordan, and T. Petsche, Eds. Cambridge, MA: MIT Press, 1997, pp. 347–353.

[9] T. Heskes, "Practical confidence and prediction intervals," in *Advances in Neural Information Processing Systems*, vol. 9, M. Mozer, M. Jordan, and T. Petsche, Eds. Cambridge, MA: MIT Press, 1997, pp. 176–182.

[10] V. C. Raykar, S. Yu, L. H. Zhao, G. H. Valadez, C. Florin, L. Bogoni, and L. Moy, "Learning from crowds," *J. Mach. Learn. Res.*, vol. 11, pp. 1297–1322, Mar. 2010.

[11] J. Ting, A. D'Souza, and S. Schaal, "Bayesian regression with input noise for high-dimensional data," in *Proc. Int. Conf. Mach. Learn.*, 2006, pp. 937–944.

[12] V. Radosavljevic, S. Vucetic, and Z. Obradovic, "A data mining technique for aerosol retrieval across multiple accuracy measures," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 2, pp. 411–415, Apr. 2010.

[13] J. Quinonero-Candela, C. E. Rasmussen, F. Sinz, O. Bousquet, and B. Scholkopf, *Evaluating Predictive Uncertainty Challenge, Machine Learning Challenges*, vol. 3944. Heidelberg, Germany: Springer-Verlag, 2006, pp. 1–27.

[14] L. A. Remer, D. Tanre, and Y. J. Kaufman, "Algorithm for Remote Sensing Tropospheric Aerosol from MODIS: Collection 5," 2006. [Online]. Available: http://modis-atmos.gsfc.nasa.gov/_docs/ATBD_MOD04_C005_rev2.pdf

[15] M. King, W. P. Menzel, Y. J. Kaufman, D. Tanre, B. Gao, S. Platnick, S. A. Ackerman, L. A. Remer, R. Pincus, and P. A. Hubanks, "Cloud and aerosol properties, precipitable water, and profiles of temperature and water vapor from MODIS," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 2, pp. 442–458, Feb. 2003.

[16] L. A. Remer, Y. J. Kaufman, D. Tanre, S. Mattoo, D. A. Chu, J. V. Martins, R.-R. Li, C. Ichoku, R. C. Levy, R. G. Kleidman, T. F. Eck, E. Vermote, and B. N. Holben, "The MODIS aerosol algorithm, products and validation," *J. Atmos. Sci.*, vol. 62, no. 4, pp. 947–973, Apr. 2005.

[17] R. C. Levy, L. A. Remer, D. Tanre, S. Mattoo, Y. Kaufman, "Algorithm for Remote Sensing of Tropospheric Aerosol Over Dark Targets From MODIS: Collections 005 and 051: Revision 2," MODIS Theoretical Basis Document, 2009

[18] P. Hubanks, MODIS Atmosphere QA Plan for Collection 005. Deep Blue Update Version, 2007. [Online]. Available: http://modis-atmos.gsfc.nasa.gov/reference_atbd.html

[19] T. F. Eck, B. N. Holben, J. S. Reid, O. Dubovik, A. Smirnov, N. T. O'Neill, I. Slutsker, and S. Kinne, "Wavelength dependence of the optical depth of biomass burning, urban, and desert dust aerosols," *J. Geophys. Res. Atmos.*, vol. 104, no. D24, pp. 31 333–31 349, Dec. 1999.

[20] C. Ichoku, D. A. Chu, S. Mattoo, Y. J. Kaufman, L. A. Remer, D. Tanre, I. Slutsker, and B. N. Holben, "A spatio-temporal approach for global validation and analysis of MODIS aerosol products," *Geophys. Res. Lett.*, vol. 29, no. 12, pp. MOD121–MOD124, 2002.

Kosta Ristovski was born in Valjevo, Serbia, in 1981. He received the B.S. degree in electrical engineering from the University of Belgrade, Belgrade, Serbia, in 2000. He is currently working toward the Ph.D. degree in computer science at Temple University, Philadelphia, PA.

His research interests include spatiotemporal data mining and machine learning.

Slobodan Vucetic received the B.S. and M.S. degrees in electrical engineering from the University of Novi Sad, Novi Sad, Serbia, in 1994 and 1997, respectively, and the Ph.D. degree in electrical engineering from Washington State University, Pullman, in 2001.

He is currently an Associate Professor with the Department of Computer and Information Sciences, Temple University, Philadelphia, PA. His research interests are in data mining and machine learning.

Zoran Obradovic received the B.S. and M.S. degrees from the University of Belgrade, Belgrade, Serbia, in 1985 and 1987, respectively, and the Ph.D. degree in computer science from the Pennsylvania State University, University Park, in 1991.

He is currently a Professor of computer and information sciences and the Director of the Center for Information Science and Technology, Temple University, Philadelphia, PA. He has published about 240 articles addressing data mining challenges in health informatics, the social sciences, environmental management, and other domains. His research focuses on improving predictive modeling and decision support through data-driven discovery and modeling of hidden patterns in large data sets.