Evaluation of a Neural Network-Based Approach for Aerosol Optical Depth Retrieval and Uncertainty Estimation

Kosta Ristovski, Slobodan Vucetic, and Zoran Obradovic

***Abstract*— In many applications of the neural networks, predicting the conditional average of the target variable is not sufficient. Often, real life problems also require estimation of the uncertainty. In this study, uncertainty analysis is applied on a remote sensing problem of Aerosol Optical Depth (AOD) estimation. AOD is one of the most important properties of the atmosphere that indicates the amount of depletion that a beam of radiation undergoes as it passes through the atmosphere. To predict AOD, we used a data-driven approach based on training neural networks. Several techniques for uncertainty estimation which are tractable for large amounts of high-dimensional remote sensing data are considered. Under the assumption that the noise in targets is input-dependent, the uncertainty of AOD predictions is computed as the variance of the conditional distribution of targets given the input data. Several methods for uncertainty estimation were applied to a real data set with 67,907 observations collected over the whole Earth during three years (2005-2007) with the attributes derived from MODIS satellite instrument and with the targets obtained from ground-based AERONET instruments. Knowledge discovered from the uncertainty analysis of this data set can potentially be very useful for better understanding of aerosol properties.**

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

I. INTRODUCTION

Aerosols, small particles emanating from natural and man-made sources, along with green house gases have been recognized as very important factors in ongoing climate changes [1]. The accurate study of aerosol composition and their concentration is one of the main challenges in current climate research. Aerosol optical depth (AOD) is a dimensionless measure of the degree to which aerosols prevent the transmission of light. The process of predicting AOD using ground or satellite based observations is known as AOD retrieval.

We will consider aerosol-related data collected by the ground based Aerosol Robotic Network (AERONET) instruments and also by the Moderate Resolution Imaging Spectrometer (MODIS) instrument aboard both Terra and Aqua satellites. AERONET is a global remote sensing network that provides aerosol information from the ground with a relatively high accuracy but with low spatial coverage. The MODIS instrument observes reflected solar radiation through multiple spectral bands with high spatial resolution and provides almost daily global coverage.

To retrieve AOD the operational MODIS retrieval algorithm, called C005 (for Collection 5), relies on domain knowledge of aerosol properties to construct tables of expected AOD for various aerosol compositions [2]. In C005, observations provided by MODIS are matched to the values stored in limited-size lookup tables. However, this algorithm is not designed to estimate uncertainty of predicted AOD.

More recently, statistical models based on artificial neural networks emerged as successful tools for nonlinear regression modeling of various remote sensing problems. In a neural network based approach for AOD retrieval, satellite observations are viewed as the inputs to a regression model while ground-based AOD data provided by AERONET instruments act as the corresponding target values. Such an approach [3] provided strong evidence that an application of neural networks for retrieving AOD could be more accurate than alternatives.

Uncertainty estimation for the confidence of retrieval requires the modeling of the whole conditional distribution of the target variable. Previous studies [4,5] assume constant noise variance which is not valid for many remote sensing situations where noise is heteroscedastic (variance of noise is input-dependent). To overcome possible bias in the maximum likelihood approach, Bishop and Quazaz introduced a Bayesian method [6]. However, this method requires calculating Hessian matrices and their inverses during a training process of neural networks and is therefore prohibitively time-consuming for large scale applications like ours. So, in this paper we will consider several methods for uncertainty estimation based on the bootstrap technique that are tractable for large data sets [7].

Several approaches for uncertainty estimation that are considered in our study are explained in Section II. In Section III, the MODIS and AERONET data used in our experiments

This work was supported in part by U.S. National Science Foundation grant IIS-0612149.

Kosta Ristovski, Slobodan Vucetic and Zoran Obradovic are with the Center of Information Science and Technology, Temple University, (e-mails:kosta@ist.temple.edu, vucetic@ist.temple.edu, zoran@ist.temple.edu).

Corresponding address: Z. Obradovic, 1805 N. Broad Street, 303 Wachman Hall, is with the Center of Information Science and Technology, Computer and Information Science Department, Temple University, Philadelphia PA 19122-6094 USA (phone: 215-204-6265; fax: 215-204-5082).

is described. A quality measure for comparing uncertainty estimation by various methods is defined in section IV, followed by a summary of the obtained results in section V. Finally, section VI contains the discussion and conclusions.

II. Methodology

We are given a set of data pairs $\{\boldsymbol{x_i}, y_i\}$ $i = 1,2,\dots,N$ where $\boldsymbol{x_i}$ and $\boldsymbol{y_i}$ represent input attribute vectors and target variables, respectively. A neural network-based regression assumes that target y is related to input vector $\boldsymbol{x}$ by stochastic and deterministic components [8]

$$y = \mu(x) + \varepsilon(x)\,. \tag{1}$$

The stochastic component is a random variation of y around its mean $\mu(\boldsymbol{x})$ caused by heteroscedastic noise $\varepsilon(\boldsymbol{x})$ with zero-mean Gaussian distribution and input-dependent variance $\sigma_n^2(\boldsymbol{x})$. The deterministic component determines a functional relationship between $\mu(\boldsymbol{x})$ and $\boldsymbol{x}$**.** Our goal is to estimate both the stochastic and deterministic component as good as possible.

Suppose we have chosen n training points out of N and want to train b neural networks in a bootstrap committee. First, we generate b bootstrap replicates B_i of the original training data set using sampling with replacement where each replicate contains n examples. We train a neural network $m_i(\boldsymbol{x})$ on each replicate B_i, $i = 1,2,\dots,b$. By averaging outputs of the b neural networks in the committee the deterministic component can be estimated as

$$m(\boldsymbol{x}) = (1/b)\sum_{i=1}^{b} m_i(\boldsymbol{x}). \tag{2}$$

The stochastic component is determined by training an additional neural network $s(\boldsymbol{x})$. AOD prediction on a test data point $\boldsymbol{x}$ is computed as $m(\boldsymbol{x})$ and uncertainty of that prediction is based on $s(\boldsymbol{x})$. In the following, we evaluate three different approaches for training $s(\boldsymbol{x})$ taking a special care to reduce bias in uncertainty estimation.

A. Prediction of squared error

A straightforward idea to estimate uncertainty $\sigma^2(\boldsymbol{x})$ is to train a neural network to predict squared error of $m(\boldsymbol{x})$ defined as

$$(y - m(\boldsymbol{x}))^2\,. \tag{3}$$

In our experiments a standard Mean Squared Error (MSE) criterion was used to train this network.

B. Maximization of log likelihood

Assuming heteroscedastic noise in (1), the conditional target distribution can be written as

$$P(y_i \mid \boldsymbol{x}_i) = (1/\sqrt{2\pi\sigma_n^2(\boldsymbol{x}_i)})\cdot \\ \cdot\exp[-(y_i - \mu(\boldsymbol{x}_i))^2/(2\sigma_n^2(\boldsymbol{x}_i))] \tag{4}$$

If $\mu(\boldsymbol{x})$ is estimated sufficiently well by $m(\boldsymbol{x})$, uncertainty $\sigma^2(\boldsymbol{x})$ can be estimated by a neural network $s(\boldsymbol{x})$ that maximizes the log-likelihood

$$\log L = \sum_{i=1}^{N}\log P(y_i \mid \boldsymbol{x}_i) = \\ = (1/2)\sum_{i=1}^{N}[-(y_i - m(\boldsymbol{x}_i))^2/2s(\boldsymbol{x}_i) + \log s(\boldsymbol{x}_\mathrm{i})] \tag{5}$$

Observe that $s(\boldsymbol{x})$ in (5) is replaced by $\sigma_n^2(\boldsymbol{x})$ in (4).

C. Model uncertainty and noise variance

In the negative log-likelihood approach it is assumed that the conditional mean $\mu(\boldsymbol{x})$ is exactly estimated by the bootstrap committee $m(\boldsymbol{x})$. Since $m(\boldsymbol{x})$ is only an estimate, the model uncertainty should be also considered. Model uncertainty for a particular pattern can be estimated as [9]

$$\sigma_m^2(\boldsymbol{x}_i) = [1/(b-1)]\sum_{j=1}^{b}[m_j(\boldsymbol{x}_i) - m(\boldsymbol{x}_i)]^2\,. \tag{6}$$

In this approach error occurs due to both uncertainty in the model and noise in target. Under the assumption that noise inherent to the data and model uncertainty are independent, squared error which comes from noise can be approximated with

$$r^2(\boldsymbol{x}_i) = \max([(y_i - m(\boldsymbol{x}_i))^2 - \sigma_m^2(\boldsymbol{x}_i)],0)\,. \tag{7}$$

Noise variance $\sigma_n^2(\boldsymbol{x})$ is estimated by training a neural network $s(\boldsymbol{x})$ to maximize log likelihood ([9])

$$(-1/2)\sum_{i=1}^{N}[r^2(\mathbf{x}_i)/2s(\mathbf{x}_i) + \log s(\mathbf{x}_i)]\,. \tag{8}$$

Notice that term $(y_i - m(\boldsymbol{x}_i))^2$ from (5) is replaced by $r^2(\boldsymbol{x}_i)$ to obtain (8). Uncertainty of the prediction is estimated as

$$\sigma^2(\boldsymbol{x}_i) = \sigma_m^2(\boldsymbol{x}_i) + s(\boldsymbol{x}_i)\,.$$

III. Data Set

A regression model was trained using satellite observations to predict targets that are AERONET AOD measurements. MODIS instruments abroad Terra and Aqua satellites have global daily coverage with spatial resolution 250x250m^2 while, on the other hand, AERONET sites are located at fixed locations on the globe and acquire data every 15 minutes. The data set has been obtained after a spatio-temporal collocation of the two sources of data where AERONET site-based AOD values obtained within a short time before and after the satellite overpass are integrated with satellite-based observations at the overpass time. The data set contains 67,907 observations from MODIS collocated with AERONET points at 201 AERONET sites over whole the globe in the period of three years (2005-2007). Fourteen satellite-based attributes listed in Table I have been used for uncertainty estimation and for AOD retrieval.

IV. NLPD

The average negative log-predictive density (*NLPD*) of the true targets is a measure which is used for estimation of the quality of uncertainty estimation. The main purpose of this measure is to compare different methods applied on the same data set. This loss function is defined as [10]

$$NLPD = -(1/n)\sum_{i=1}^{n}\log p(y_i \mid \boldsymbol{x}_i)\,. \quad (9)$$

It penalizes both over and under-confident predictions. For Gaussian predictive distribution *NLPD* is calculated as

$$(1/n)\sum_{i=1}^{n}\{\log\sigma(\boldsymbol{x}_i)+[y_i - m(\boldsymbol{x}_i)]^2/2\sigma^2(\boldsymbol{x}_i)+c\}\,. \quad (10)$$

The constant c can be ignored for comparison purposes. Since NLPD is calculated as the negative logarithm, smaller NLPD scores (larger absolute values) correspond to better quality uncertainty estimates.

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

V. EXPERIMENTAL RESULTS

The considered methods have been applied to the Aerosol data set described in Section III. In experiments based on year-out cross-validation, two years of data were used as a training set while the model was tested on the remaining third year. Among 201 AERONET sites some do not have data for a particular year. Models composed of 30 neural networks in the committee were trained on the subset of sites which have data in both training and testing years. In that way AOD retrievals and uncertainty estimation were obtained for all three years which is of utmost importance for the following analysis.

TABLE II
R^2 STATISTICS FOR BOOTSTRAP TECHNIQUE AND C005 OVER THREE YEARS

Algorithm	R^2		
	Year 2005	Year 2006	Year 2007
Bootstrap	0.797	0.782	0.809
C005	0.724	0.685	0.722

TABLE III
NLPD SCORES FOR CONSIDERED METHODS

Model for uncertainty estimation	NLPD
Squared error prediction	-0.69
Log-likelihood	-2.01
Model uncertainty and noise variance	-2.03

Accuracies in AOD prediction achieved by the bootstrap technique were the same for all three methods because of the same structure of the committee, but they were much higher in comparison to the accuracy of the C005 algorithm. R^2 accuracies are reported in Table II. Besides good prediction accuracy, our objective was to identify the method which gives the best uncertainty estimation. The difference between three methods was measured through *NLPD* score which was calculated by gathering estimates over all three years. *NLPD* scores are presented in Table III. In our experiments, taking into account model uncertainty and noise variance was slightly better than using a model trained on log-likelihood and was a lot better than the square error prediction. Further analysis of uncertainty reported in this section is based on the results achieved applying the uncertainty estimation method that accounts for model uncertainty and noise variance.

According to a popular domain-specific measure [2], AOD predictions are considered to be good if they fall in the region specified by

$$|\,y_i - m(\boldsymbol{x}_i)\,| \le 0.05 + 0.15\,y_i\,. \quad (11)$$

To use this measure we have sorted AOD retrievals according to uncertainty estimates in an ascending order and then we split them in equal-width bins of 3,000 points. In each bin we measured fraction of predictions defined as [3]

$$FRACTION = (I/N)\times 100\%\,, \quad (12)$$

where I is a number of data points in the bin which satisfy relation (11) and N is 3000. At Fig 1 each bin is represented by the average of standard deviations of uncertainty estimates for 3,000 points it contains (horizontal) and FRACTION score for these points (vertical). We can see that for bins that contain more certain data points, the obtained corresponding FRACTION score is higher, thus satisfying our expectation.

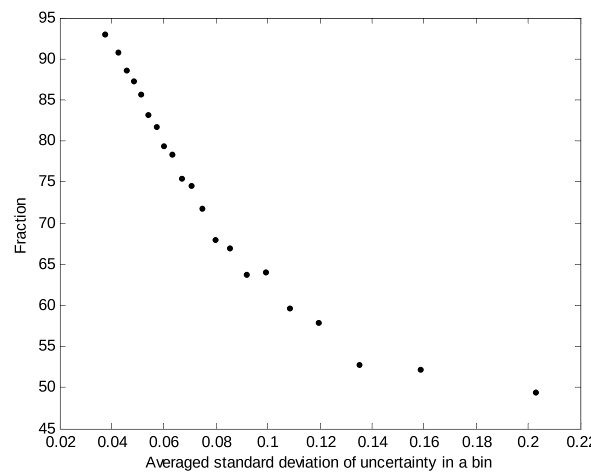

Figure 1. Prediction accuracy measured as the Fraction score for equal-width bins of 3,000 points sorted from lower to higher uncertainty.

The obtained results allow analysis of uncertainty of AOD retrieval at a given site over time and also uncertainty comparison at multiple sites. As an example, we compare uncertainty of AOD retrieval at Beijing site in China (Fig 2) vs. Muana Loa site in Hawaii (Fig 3) to conclude that properties of aerosols are much more stable at Muana Loa than in Beijing. It is easily noticeable that absolute errors for those sites have similar patterns as uncertainty estimates. Also, uncertainty level at Muana Loa site is very stable and extremely low. By further investigation we found that this discovery is consistent with domain experts' expectations as this site serves for calibration of AERONET instruments.

By an analysis of the NLPD scores for 450 site-year datasets (148, 169, and 133 sites in years 2005, 2006, and 2007) we found that 12 site-year patterns are outliers with high values which indicates failure of the algorithm. An example of such AERONET site in Taiwan in shown in Fig 4. Obviously, here absolute error values are much larger than a standard deviation of uncertainty estimates.

Average uncertainties at 148 sites in 2007 were calculated and they are presented in Fig 5 where the smaller circle means the smaller average uncertainty. We could see that annually averaged uncertainty has spatial patterns.

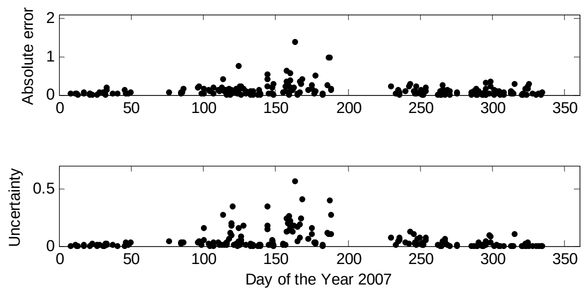

Figure 2. Absolute error of AOD prediction and estimated uncertainty (standard deviation of the total variance) for Beigin AERONET site in 2007.

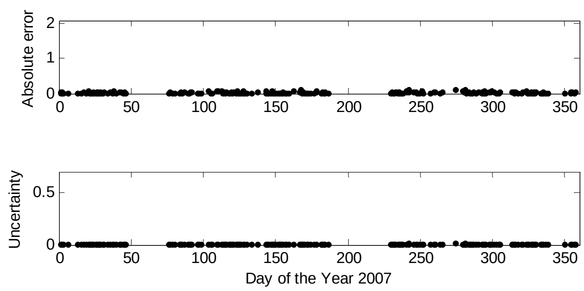

Figure 3. Absolute error of AOD prediction and estimated uncertainty for Muana Loa AERONET site in 2007.

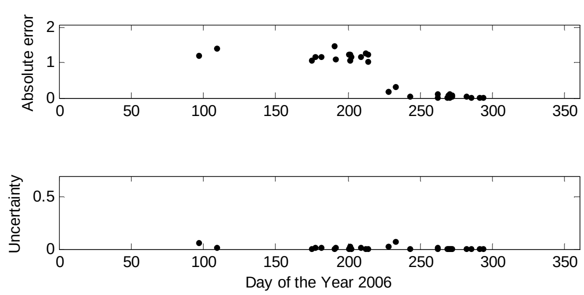

Figure 4. Absolute error of the prediction, and estimated uncertainty for Taiwan AERONET site in 2006 with NLPD=29.41.

Our analysis of seasonal uncertainty levels over three years also indicates existence of different interesting patterns (omitted from this report for lack of space). For example, we compared sites with the highest and the lowest average uncertainty over the seasons. For the most uncertain sites average seasonal variance was between 0.06 and 0.18. 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. Variances of these uncertainties are fairly constant over all seasons ranging from 0.0019 to 0.0024.

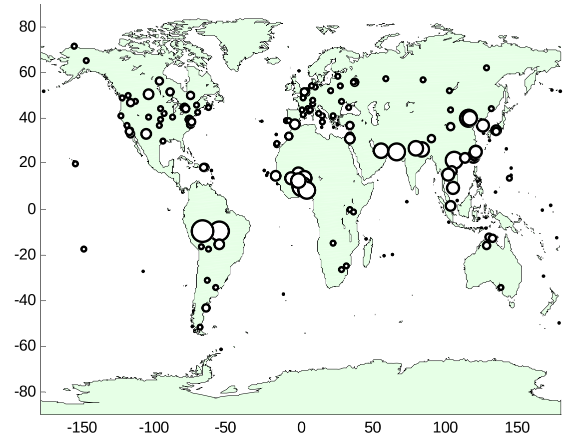

Figure 5. Average uncertainty levels at 148 sites in year 2007.

VI. Conclusion

Three methods for uncertainty estimation were presented and compared. Committees of neural networks for AOD prediction were more accurate than the operational algorithm C005. It was shown that the Bootstrap technique which takes into account model uncertainty and noise variance provides the best results among the studied alternatives. Spatial and temporal uncertainty analysis presented in this work provides multiple opportunities to improve our understanding of global properties of aerosol in a cost-effective way. Searching for new algorithms and more sophisticated analysis methods are the objectives of our future work.

References

[1] Hansen, J., Sato, M., and Ruedy, R., 1997. Radiative Forcing and Climate Response, *Journal of Geophysical Research*, 102, 6831-6864.

[2] Remer, L. A., Tanre, D., and Kaufman, Y. J., 2006. Algorithm for Remote Sensing Tropospheric Aerosol from MODIS: Collection 5, http://modisatmos.gsfc.nasa.gov/_docs/ATBD_MOD04_C005_rev2.pdf.

[3] Radosavljevic, V., Vucetic, S., and Obradovic, Z., 2007. Aerosol Optical Depth Retrieval by Neural Networks Ensemble with Adaptive Cost Function, *Proc. 10th Int'l Conf. on Engineering Applications of Neural Networks*, Thessaloniki, Greece, pp. 266-275.

[4] Aires, F., Prigent, C., and Rossow, W. B., 2004. Neural Network Uncertainty Assessment Using Bayesian Statistics: A Remote Sensing Application, *Neural Computation,* 16, 2415-2458.

[5] Bishop, C. M., 1995. *Neural Networks for Pattern Recognition*, Oxford University Press.

[6] Bishop, C. M., and Qazaz, C., 1997. Regression with Input-Dependent Noise: A Bayesian Treatment, in Mozer, M., Jordan, M., and Petsche, T., (Eds), *Advances in Neural Information Processing Systems*, 9, 347-353.

[7] Davidson, A. C., and Hinkley, D. V., 1997. *Bootstrap Methods and Their Application*, Cambridge University Press.

[8] Dybowski, R., and Roberts, S. J. 2001. Confidence Intervals and Prediction Intervals for Feed-Forward Neural Networks, *Clinical Applications of Artificial Neural Networks*, University Press, 298-326.

[9] Heskes, T., 1997. Practical confidence and prediction intervals, in Mozer, M., Jordan, M., and Petsche, T., Editors, *Advances in Neural Information Processing Systems*, 9, 176-182.

[10] Quinonero-Candela, J., Rasmussen, C. E., Sinz, F., Bousquet, O., and Scholkopf, B., 2006. Evaluating Predictive Uncertainty Challenge, *Machine Learning Challenges*, 3944, 1-27.