PAPER • OPEN ACCESS

On estimation of cloudiness characteristics and parameters of DOAS retrieval from spectral measurements using a neural network

To cite this article: O V Postylyakov et al 2020 IOP Conf. Ser.: Earth Environ. Sci. 489 012031

View the article online for updates and enhancements.

You may also like

- <u>Comparison of measured and simulated</u> <u>NO2 integral content in the lower</u> <u>troposphere in Moscow region</u> O Postylyakov, A Borovski, A Kirsanov et al.
- <u>Vertical profile of aerosol extinction based</u> on the measurement of O₂ of multielevation angles with MAX-DOAS Fusheng Mou, , Jing Luo et al.
- <u>Development of a DOAS System for</u> <u>ToTAL-DOAS Applications with</u> <u>Temperature Control</u> Javier A Ramos and Erna Frins





DISCOVER how sustainability intersects with electrochemistry & solid state science research



This content was downloaded from IP address 18.217.189.213 on 16/05/2024 at 20:49

IOP Publishing

On estimation of cloudiness characteristics and parameters of **DOAS** retrieval from spectral measurements using a neural network

O V Postylyakov^{1*}, S V Nikitin², A I Chulichkov² and A N Borovski¹

¹A.M. Obukhov Institute of Atmospheric Physics, Russian Academy of Science, Pyzhevsky per.3, Moscow, 119017, Russia ²M.V. Lomonosov Moscow State University, Leninskiye Gory 1, Moscow, 119991,

Russia

*Email: oleg.postylyakov@gmail.com

Abstract. Light scattering by clouds significantly affects the values associated with the content of NO₂, H₂CO and other small gases in the lower troposphere, which are obtained by the differential optical absorption spectroscopy (DOAS) technique. Since there are a large databases of optical observations of trace gases by DOAS technique that are not accompanied by other measurements of clouds, the development of approaches to the refinement of scattering characteristics and coefficients linking the DOAS slant column depth with the gas vertical content directly from spectral measurements remains an important task. The paper considers the tasks of determining the coefficient F used for transformation of the DOAS slant column depth of a gas to its vertical column from quantitates obtained from ZDOAS measurements (the O_4 slant column, the color index, the absolute intensity, etc.). It was shown in numerical experiments that an algorithm based on a neural network can estimate the coefficient F in cloudy conditions. It looks like the better approach that two step estimation of this parameter using a neural network for estimation of cloud characteristics in the first step with the following radiative transfer simulation at the second step.

Keywords: determination of air mass factor of DOAS measurements, neural networks, DOAS technique, cloud and aerosol scattering in the atmosphere.

1. Introduction

To determine the vertical distribution and/or the integral content of trace gases and aerosol in the troposphere, the differential optical absorption spectroscopy (DOAS) technique is widely used [1][2]. Multi AXis (MAX-DOAS) and Zenith (ZDOAS) versions of the technique use measurements of the scattered solar radiation from several elevation directions or from the zenith, respectively. Due to such measurement geometries, estimations of gas and aerosol tropospheric contents, based on the DOAS technique, are under the strong influence of clouds.

Quantitatively, this dependence can be described using the following formalism. A value called the slant column density (SCD) of a trace gas, which is initially determined from DOAS spectrum, is a weighted average

$$S = \int m(h) \cdot n(h) dh \tag{1}$$

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI. Published under licence by IOP Publishing Ltd 1

107 (2020) 012031 u01.10.1088/1755-1515/489/1/012

IOP Publishing

of the gas number density n(h) along height h. The weighting coefficient m(h) of integral (1) depends on light scattering in the atmosphere and, hence, on cloud properties. The coefficient m(h) coincides with the efficient layer (box) air mass factor (AMF) [3][4]. The gas vertical column density (VCD) V, the value in which researchers are interested in, is determined as

$$V = \int n(h)dh .$$
⁽²⁾

Hence, VCD V may be calculated using equivalence

$$V = S \cdot F , \qquad (3)$$

where coefficient F

$$F = \frac{1}{\int k(h) \cdot m(h)dh} \tag{4}$$

depends on the layer AMFs m(h) and the shape of the gas vertical profile k(h). We determine the shape k(h) as the normalized gas number density: $\int k(h)dh = 1$. k(h) can be set using some model assumptions in case of ZDOAS measurement or determined from the measurements itself in case of MAX-DOAS observations. The layer AMFs m(h) can be estimated using a linearized radiative transfer model (LRTM) if we know the scattering characteristics and the content of the strong absorbers of the atmosphere in time of the observation [5][3][4]. The effects of possible inaccuracies in description of cloudy conditions during DOAS observations on m(h) were briefly discussed, for example, in papers [6][7].

As it follows from equality (3), the determination of the coefficient F with the greatest possible accuracy is an important task of retrieval the vertical distribution and/or the integral content of trace gas by the DOAS method [8]. There are possible a few approaches to treat cloud-contaminated DOAS measurements.

The first approach is to select and process only DOAS measurements obtained in clear sky conditions. A few algorithms for the cloud detection basing on several quantities derived from DOAS measurements have been proposed and used [9]-[12].

The second approach is to obtain cloud properties from some measurements and calculate AMFs m(h) and coefficient *F* using RTM. We previously developed an algorithm for the nitrogen dioxide [7] and formaldehyde [8] retrieval for continuous cloud conditions when height of the cloud base is greater than height of the atmospheric mixture layer. Statistical comparison of data on the formaldehyde content near Moscow obtained in overcast with its content for cloudless days showed the similarity of values for the background air masses and differences in observations in the pollution conditions [13]. Since accuracy of gas and aerosol retrieval in the cloudy conditions strongly depend on knowledge of cloud characteristics [6][8], to determine them, ceilometers, lidars [9] and other instruments [14]-[19] may be used.

The significant advances were achieved in investigations of the absolute calibration of DOAS spectrometers [20][21], what can provide additional information on clouds (see, for example [22]) from DOAS measurements itself. We showed the early results on development an algorithm for determination of the height of the cloud base and the cloud optical depth based on some quantities derived from absolutely calibrated DOAS spectra in paper [23]. As a basis for the algorithm a neural network trained using a sample set of DOAS quantities obtained using a LRTM.

In this paper we present the early results of the third approach. We train an artificial neural network to obtain coefficients F (4) directly from absolutely calibrated DOAS spectra and some standard DOAS product without LRTM calculation at this step. A sample set of DOAS quantities and coefficients F for different cloud and aerosol scenarios were calculated by a LRTM.

IOP Publishing

2. Mathematical model of DOAS measurement

For the solution of a problem on estimating any atmospheric characteristics, one has to describe mathematically the relationship of the parameters to be estimated with the quantities measured in the experiment.

For calculations of the spectral radiance measured in DOAS experiment and the derived from its quantities we use the linearized radiative transfer model (LRTM) MCC++ [3][4][24]. The model MCC++ is a combination of the Monte Carlo method to simulate multiple scattering with the direct integration procedure to simulate single scattering. This combination makes optimum use of the computing resources. To simulate multiple scattering, the method of conjugate walks (in other words "backward simulation") was applied. An approximation of the spherically symmetrical atmosphere is applied to further shorten the time of the simulation. The vector (with polarization) version of MCC++ code was used with taking into account Lambertian surface albedo, aerosol scattering and absorption. The model has participated in several international comparisons (see overview in [24]), and been used repeatedly in remote sensing investigations related to the atmosphere [25]-[27]. The model is linearized, thus it may efficiently calculate both the radiance and the layer air mass factors. The layer air mass factor is a coefficient which connects the vertical distribution of any gas with the slant column of this gas obtained by the DOAS technique.

We performed radiative transfer calculations under different scenarios varying the parameters of the atmosphere and the underlying surface within the following limits:

- the zenith angle of the sun from 30° to 84°;
- albedo of the underlying surface 5% (for summer) and 34% (for winter);
- the height of the cloud bottom from 0.4 km to 3.2 km;
- the aerosol optical depth at 550 nm from 0.1 to 1.1;
- the cloud geometrical thickness from 0.1 km to 3.2 km;
- the cloud particle number concentration from 0.25 to 4 of the concentration of the cloud model C1 from [28], and the clear sky.

The microphysical parameters of the cloud particles were taken from [28], the cloud model C1 was used. The combinations of geometrical thickness and particle number concentrations give the cloud optical depth at 550 nm within 0.4 and 51. A model of the urban aerosol [29] was used for the calculation of the tropospheric aerosol phase function [30].

We call these parameters the input RTM parameters.

The calculated by LRTM parameters are:

- the coefficient F for 475 nm (for the retrieval of NO₂ integral content in the troposphere);
- the absolute intensities at 4 selected wavelengths (345nm, 374nm, 430nm, 480nm);
- the O4 slant columns at 8 selected wavelengths (340nm, 350nm, 369nm, 379nm, 425nm, 435nm, 475nm, 485nm);
- the color indices in the UV (345nm/374nm) and visible (430nm/480nm) ranges;
- the coefficient at the linear term of DOAS analysis for the UV (345nm/374nm) and visible (430nm/480nm) ranges.

We used two scenarios of the normalized gas vertical profile k(h) to simulate coefficient F: winter and summer (similarly to [8]). In both scenarios, 90% of NO₂ content is located in atmospheric boundary layer (ABL) and has constant volume mixing ratio. NO₂ content above ABL decreases exponentially. In winter scenario, we set the height of ABL to 400 m, in summer – 1000 m.

We will call them the out parameters of LRTM. The total number of parameters involved in the LRTM calculation is 21, of which 6 are input and 16 are output. In total, we performed radiative transfer simulations for 5400 scenarios. These calculations were used further for estimating the coefficient F from DOAS data.

3. The problem of estimating the coefficient F from DOAS data

We would like to study the possibility of estimating the coefficient F from measurements of spectra of the scattered solar radiation.

To simplify the problem, we will further consider not the measured spectra of scattered radiation as a whole, but only some quantities given in Section 2 which may be derived from radiance scattered in the zenith by DOAS technique. Information on the relationship of the coefficient F with the registration results is contained in the data set described in Section 2. The following quantities were considered as known parameters:

- the absolute radiation intensity at 4 wavelengths;
- the O4 slant columns at 8 wavelengths;
- 2 color indices;
- 2 coefficient at the linear term of DOAS analysis;
- zenith angle of the sun;
- albedo of the underlying surface;
- cloud particle number concentration.

The total number of known quantities is 18, and we will consider them as the coordinates of the vector $\mathbf{g} = (g_1, g_2, ..., g_{18}) \in \mathbb{R}^{18}$. The task is to find the relationship between the studied coefficient F and the measured quantities \mathbf{g} in the form of a function $A(\cdot): \mathbb{R}^{18} \to \mathbb{R}^1$, acting from \mathbb{R}^{18} to \mathbb{R}^1 :

$$F = A(\mathbf{g}) . \tag{5}$$

To find this function, the data set obtained by calculations of LRTM is used.

Formally, the task is to find a function $A(\cdot): \mathbb{R}^{18} \to \mathbb{R}^1$ from a certain class of functions A for which a minimum of the error functional is achieved [31]

$$\Phi(A) = \inf_{A' \in \mathbf{A}} \left\{ \sum_{i=1}^{N} \left| F_i - A'(\mathbf{g}_i) \right|^2 \right\}.$$
(6)

Here F_i and \mathbf{g}_i are the values of the estimated atmospheric parameters and the values of the specified parameters, respectively, for the *i*-th sample of the data set described in Section 2. Solution of minimization problem (6) is carried out by training a neural network [32] (See also other applications of neural networks for estimation of cloud properties [33]-[35],[23]).

4. Neural Network Architecture and Training

A fully connected neural network [32] is used, consisting of three layers - the input (18 neurons), the output (1 neuron) and the hidden one, consisting of 64 neurons with the activation function RELU. The number of settings (neural network weights) is 1603. A standard technique of neural networks was used in which the set of 5400 examples was divided into three parts. The first part is used for training, i.e. to select a function $A(\cdot) \in A$. The second part is used for validation in the learning process; the purpose of this part is to avoid setting the network to the training sample (retraining) too accurately. The third part is a test one; it is used to determine the accuracy of the estimates. The training set was 64% of the total number of examples, the validation set was 16% of the total set, and the test set was 20%. Iterative teaching method was used, the learning process was stopped if the value of the minimized functional changed slightly during 50 iterations; the total number of iterations (epochs) during training did not exceed 1000.

To prevent overfitting of the neuron network, the regularization was used in the form of a restriction on the weighted sum of a combination of the Euclidean and uniform norms of the weight vector. I.e. instead of minimizing functional (6), the regularized functional was minimized

$$\Psi(A) = \inf_{A' \in A} \left\{ \sum_{i=1}^{N} |F_i - A'(\mathbf{g}_i)|^2 \right\} + \alpha_1 \sum_{j=1}^{M} c_j^2 + \alpha_2 \max_{j=1,\dots,M} |c_j|, \qquad (7)$$

in which a linear combination of the Euclidean and uniform norms of the vector was used as stabilizer, the coordinates of which are the weights of the neural network.

5. Neural network learning outcomes

The learning outcomes of the neural network are shown in Figures 1 and 2.

Figures 1(a) and 2(a) show histograms of the distribution of the estimation errors when estimating the coefficient F for summer and for winter. The corresponding scatter plots are shown in Figures 1(b) and 2(b); the horizontal axis shows the true value of the parameter, and the vertical axis shows the estimates of these parameter obtained by the neural network.

The r.m.s. retrieval error of the coefficient *F* is 0.046 for both summer and winter albedo.

Note that the retrieval errors of the cloud parameters, which were estimated in work [23], were rather large and amounted to 1.9 for the cloud optical depth and 0.81 km for the cloud base height. It is significantly better results that may be expected in two step estimation of this parameter using a neural network for estimation of cloud characteristics with the following radiative transfer simulation at the second step.



Figure 1. Estimating the coefficient F for summer: (a) - histogram of the estimation errors; (b) the scatter plots.



Figure 2. Estimating the coefficient F for winter: (a) - histogram of the estimation errors; (b) the scatter plots.

6. Conclusion

It was shown in numerical experiments that an algorithm based on a neural network can estimate the coefficient F related to the air mass factor of the DOAS technique in cloudy conditions. We obtained

IOP Publishing

preliminary the accuracy of determining the coefficient F at 475 nm as 0.046 for both summer and winter albedo. It is better results that may be expected in two step estimation of this parameter using a neural network for estimation of cloud characteristics in the first step with the following radiative transfer simulation at the second step.

Acknowledgements

The study was carried out with the financial support of the Russian Science Foundation by project 16-17-10275 (remote sensing method in clouds) and the Russian Foundation for Basic Research by projects 19-01-00790 (neural network investigation) and 18-35-00682 (slant column retrieval).

References

- [1] Platt U 1994 Differential optical absorption spectroscopy (DOAS) Chem. Anal. Series 127 27-83
- [2] Platt U, Stutz J 2008 Differential Optical Absorption Spectroscopy, Principles and Applications (Berlin: Springer)
- [3] Postylyakov O V 2004 Radiative transfer model MCC++ with evaluation of weighting functions in spherical atmosphere for use in retrieval algorithms Adv. Space Res., 34(4) 721-726 doi:10.1016/j.asr.2003.07.070.
- [4] Postylyakov O V 2004 Linearized vector radiative transfer model MCC++ for a spherical atmosphere J. Quant. Spectrosc. Radiat. Transfer 88(1-3) 297-317 doi:10.1016/j.jqsrt.2004.01.009
- [5] Postylyakov O V, Belikov I B, Elansky N F, Elohov A S 2006 Observations of the ozone and nitrogen dioxide profiles in the TROICA-4 experiment Adv. Space Res. 37(12) 2231-2237 doi:10.1016/j.asr.2005.07.023
- [6] Postylyakov O V, Borovski A N 2014 Measurements of formaldehyde total content using DOAS technique: a new retrieval method for overcast *Proc. SPIE* **9259** 925918 doi:10.1117/12.2069595
- [7] Ivanov V A, Postylyakov O V 2010 Estimation of integral NO₂ content in atmospheric boundary layer using observations of scattered in zenith radiation *Optika Atmosfery i Okeana (Atmos. and Ocean. Optics)* 23(6) 471-475 (in Russian)
- [8] Postylyakov O, Borovski A, Ivanov V 2015 On determination of formaldehyde content in atmospheric boundary layer for overcast using DOAS technique *Proc. SPIE* 9680 96804O doi: 10.1117/12.2205925
- [9] Wagner T, Apituley A, Beirle S, Dörner S, Friess U, Remmers J, Shaiganfar R 2014 Cloud detection and classification based on MAX-DOAS observations *Atmos. Meas. Tech.* 7 1289-1320 doi:10.5194/amt-7-1289-2014
- [10] Gielen C, Van Roozendael M, Hendrick F, Pinardi G, Vlemmix T, De Bock V, De Backer H, Fayt C, Hermans C, Gillotay D, Wang P 2014 A simple and versatile cloud-screening method for MAX-DOAS retrievals Atmos. Meas. Tech. 7 3509-3527 doi:10.5194/amt-7-3509-2014
- [11] Wang Y, Penning de Vries M, Xie P H, Beirle S, Dörner S, Remmers J, Li A, Wagner T 2015 Cloud and aerosol classification for 2.5 years of MAX-DOAS observations in Wuxi (China) and comparison to independent data sets *Atmos. Meas. Tech.* 8 5133-5156 doi:10.5194/amt-8-5133-2015
- [12] Postylyakov O V, Borovski A N 2016 Measurement of formaldehyde total content in troposphere using DOAS technique: improvements in version 1.3a of IAP retrieval algorithm *Proc. SPIE* 9876 98761N doi:10.1117/12.2229231.
- [13] Ivanov V, Borovski A, Postylyakov O 2017 First comparison of formaldehyde integral contents in ABL retrieved during clear-sky and overcast conditions by ZDOAS technique *Proc. SPIE* 10424 1042400 doi: 10.1117/12.2278235
- [14] Chulichkov A I, Andreev M S, Golitsyn G S, Elansky N F, Medvedev A P, Postylyakov O V 2017 On cloud bottom boundary determination by digital stereo photography from the Earth's surface Atmospheric and Oceanic Optics 30(2) 184–190 doi: 10.1134/S1024856017020075
- [15] Andreev M, Chulichkov A I, Medvedev A P, Postylyakov O V 2014 Estimation of cloud base height using ground-based stereo photography: Method and first results *Proc. SPIE* 9242 924219 doi: 10.1117/12.2069824
- [16] Chulichkov A I, Nikitin S V, Medvedev A P, Postylyakov O V 2016 Stereoscopic ground-based determination of the cloud base height: camera position adjusting with account for lens distortion *Proc.* SPIE 10035 100353B doi:10.1117/12.2248633

- [17] Chulichkov A I, Nikitin S V, Emilenko A S, Medvedev A P, Postylyakov O V 2017 Selection of optical model of stereophotography experiment for determination the cloud base height as a problem of testing of statistical hypotheses *Proc. SPIE* 10424 1042411 doi: 10.1117/12.2279553
- [18] Zuev S V, Krasnenko N P 2006 Passive method for cloud base height detection Proc. SPIE 6522 65221S doi:10.1117/12.723217
- [19] Zuev S V, Krasnenko N P 2008 Determination of cloud base height using the passive monostatic method Proc. SPIE 6936 69360N doi: 10.1117/12.78315
- [20] Wagner T, Beirle S, Dörner S, Penning de Vries M, Remmers J, Rozanov A, Shaiganfar R 2015 A new method for the absolute radiance calibration for UV-vis measurements of scattered sunlight *Atmos. Meas. Tech.* 8 4265-4280 doi:10.5194/amt-8-4265-2015
- [21] Wagner T, Beirle S, Remmers J, Shaiganfar R, Wang Y 2016 Absolute calibration of the colour index and O4 absorption derived from Multi AXis (MAX-)DOAS measurements and their application to a standardised cloud classification algorithm *Atmos. Meas. Tech.* **9** 4803–4823 doi:10.5194/amt-9-4803-2016
- [22] Khatri P, Iwabuchi H, Hayasaka T, Irie H, Takamura T, Yamazaki A, Damiani A, Letu H, Kai Q 2019 Retrieval of cloud properties from sky radiometer observed spectral zenith radiances *Atmos. Meas. Tech. Discuss.* doi:10.5194/amt-2019-273
- [23] Nikitin S V, Chulichkov A I, Borovski A N, Postylyakov O V. 2019 Estimation of cloudiness and aerosol characteristics in the atmosphere from spectral measurements of scattered solar radiation using a neural network *Proc. SPIE* 11152 111521H doi:10.1117/12.2535490
- [24] Postylyakov O V 2004 Spherical radiative transfer model with computation of layer air mass factors, and some of its applications *Izvestiya Atmospheric and Oceanic Physics* **40(3)** 276-290.
- [25] Borovski A N, Dzhola A V, Elokhov A S, Grechko E I, Postylyakov O V, Kanaya Y 2014 First measurements of formaldehyde integral content at Zvenigorod Scientific Station Int. J. of Remote Sensing 35(15) 5609-5627 doi:10.1080/01431161.2014.945011
- [26] Elansky N F, Mitin I V, Postylyakov O V 1999 Maximum accuracy of Umkehr measurements of vertical ozone profiles *Izvestiya - Atmospheric and Oceanic Physics* 35(1) 65-77.
- [27] Elansky N F, Grechko G M, Plotkin M E, Postylyakov O V 1991 The OZAFS experiment: observing the fine structure of the ozone and aerosol distributions in the atmosphere from the Salyut 7 orbiter. Part III: Experimental results J. Geophys. Res. 96(D10) 18661-18670 doi: 10.1029/91JD01394
- [28] Deirmendjian D 1969 Electromagnetic scattering on spherical polydisprsions (New York: American Elsevier Publishing company Inc.)
- [29] WMO. World Climate Research Program. 1986 A preliminary cloudless standard atmosphere for radiation computation *Technical report WCP-112*, *WMO-TD N24* (Geneva:World Meteorological Organization) 60 p.
- [30] Wiscombe W L 1980 Improved Mie scattering algorithms Applied Optics 19(9) 1505-1509
- [31] Pyt'ev Yu P 2010 Measurement-Computation Converter as a Measurement Facility Automation and Remote Control 71(2) 303–319 doi:10.1134/S0005117910020116
- [32] Goodfellow I, Bengio Y, Courville A 2016 Deep Learning (MIT Press)
- [33] Strandgren J, Bugliaro L, Sehnke F, Schröder L 2017 Cirrus cloud retrieval with MSG/SEVIRI using artificial neural networks *Atmos. Meas. Tech.* **10** 3547–3573 doi:10.5194/amt-10-3547-2017
- [34] Strandgren J, Fricker J, Bugliaro L 2017 Characterisation of the artificial neural network CiPS for cirrus cloud remote sensing with MSG/SEVIRI Atmos. Meas. Tech. 10 4317–4339 doi:10.5194/amt-10-4317-2017
- [35] Hakansson N, Adok C, Thoss A, Scheirer R, Hörnquist S 2018 Neural network cloud top pressure and height for MODIS Atmos. Meas. Tech. 11 3177–3196 doi:10.5194/amt-11-3177-2018
- [36] <u>https://www.tensorflow.org/</u>
- [37] https://keras.io/