23

Decomposition of Absorption Spectra of Natural Gas Samples Using the Spectral Complexity Criterion

© Yu.V. Kistenev, D.A. Vrazhnov, A.V. Borisov

Tomsk State University,
Tomsk, Russia

e-mail: yuk@iao.ru

Received June 16, 2025
Revised July 28, 2025
Accepted November 25, 2025

The main challenge in analyzing absorption spectra of natural gas mixtures lies in the fact that their composition is only partially known. Existing methods for decomposing spectra of such gas mixtures are effective when the number of components is small. This paper proposes a method for decomposing absorption spectra of gas samples with an unknown composition and an arbitrary number of components. The method is based on reducing the „complexity" of the spectrum by accurately removing one of the components from the composite spectrum. The method is demonstrated through an example assessing the presence and restoring the concentration of trace gas impurities in atmospheric air.

Keywords: gas mixtures of natural origin, spectral analysis, decomposition.

DOI: 10.61011/EOS.2025.12.63178.45-25

Introduction

Component analysis of gas samples is essential for environmental applications and the study of volatile molecular markers in patient exhaled air. The absorption spectrum of gas samples $S^0_{\text{mix}}(\nu)$ is itself a superposition of the absorption spectra of its individual components $S_i(\nu)$ [1]:

$$S^0_{\text{mix}}(\nu) = \sum_{i=1}^{N} c_i S_i(\nu), \tag{1}$$

where c_i — concentration of i-th component, ν — frequency, N — number of components in the mixture. The objective of spectrum decomposition $S^0_{\text{mix}}(\nu)$ is to determine the unknown values c_i.

For the absorption spectra of gas samples of known composition (the so-called „white" system [2]) the efficient decomposition methods were developed, in particular, methods of multivariate curve resolution (MCR) [3–5], univariate calibration (UC) [6] in combination with the least-square method [7] and Levenberg-Marquardt extremal search method (LM) [8–10].

The main problem in analyzing absorption spectrum of gas mixtures of natural origin is that their composition is only partially known (the so-called „gray" system [2]). Component analysis of „gray" systems is significantly more complex, and methods for decomposing mixtures with only a few components have been developed for them.

For the decomposition of the spectra of two-component mixtures of unknown composition, Lawton and Sylvestre developed a self-modeling curve resolution (SMCR) method [11] based on transformation of the original spectra of a mixture into the space of hidden variables using the principal component analysis and the fact that the graph of the normalized spectra of two-component mixtures is a straight line in this space [12]. Rasmussen et al. [13] used SMCR to determine the number of chemical components in a mixture. Borgen and Kowalski presented a method that extends the SMCR procedure to three-component systems [14]. Rajko and Istvän [15] proposed a simplification of the Borgen and Kowalski approach based on computational geometry. In the study [16] by Ohta the expansion of SMCR method to the three-component mixture was also proposed. Meister [17] developed a SMCR-based algorithm that enabled to find the solutions for the three-component mixtures using a criterion of maximal spectral difference between the mixture components. Kawata et al. [18] developed an entropy-based optimization method for estimating the component concentrations of three-component mixtures. Vandeginste et al. [19] developed a similar approach, using the condition that the spectra of a single-component mixture should have a „simplest profile", defined mathematically as the curve with the smallest area for a given normalization. The uniqueness of the solution found using SMCR is analyzed in [20,21]. The method of searching for spectral regions where two-component mixtures can be decomposed was developed by Vosough et al. [22]. Expansion of Vosough method to three- and four-component mixtures is proposed by Golshan [23,24].

An alternative approach to solving „gray" systems is to determine the presence and concentration of a target component with a known spectrum, regardless of the composition and concentration of the remaining components in the mixture. This approach is realized in HAMAND method (hypothetical addition multivariate analysis with numerical

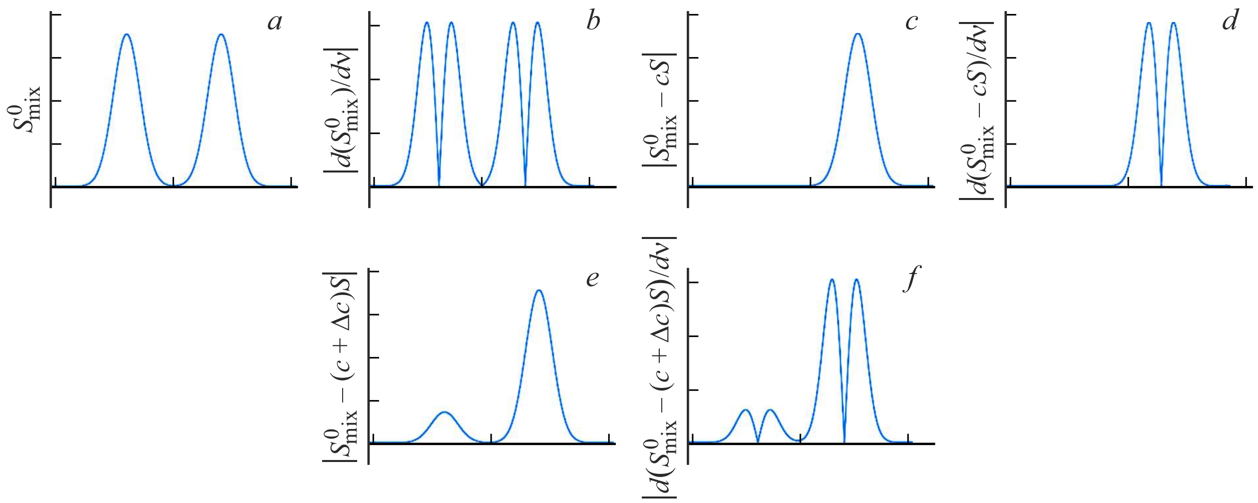

Figure 1. Illustration of RSC method: *a* — a model spectrum of mixture with two peaks corresponding to two components, *b* — modulus of the mixture spectrum derivative, *c* — modulus of the mixture spectrum after an exact subtraction of the spectrum of one of the components, *d* — modulus of the mixture spectrum derivative after an exact subtraction of the spectrum of one of the components, *e* — modulus of the mixture spectrum after subtraction with the spectrum error of one of the components (the estimated concentration differs from the true one), *f* — modulus of the mixture spectrum derivative after subtraction with the spectrum error of one of the components (the estimated concentration differs from the true one).

differentiation) [25], as well as in the method of reducing the spectrum complexity (RSC) [26–28].

The main idea of RSC method is that if the spectrum of a component is completely removed from the mixture spectrum (taking into account its concentration), then the complexity of the remaining spectrum should decrease [28]. The complexity criterion is the integral area of the module of the first derivative of the spectrum:

$$\delta f(\tilde{c}) = \int \left| \frac{d(S^0_{\mathrm{mix}}(\nu) - \tilde{c}S(\nu))}{d\nu} \right| d\nu, \tag{2}$$

where $\tilde{c}$ — unknown concentration which in the point of minimum of the functional $\delta f(\tilde{c})$ should be equal to the true concentration c. Here, $S(\nu)$ — spectrum of the searched component that shall be known a priori. The idea of the method is shown in Figure 1. The RSC method has proven to be an effective tool for determining the concentrations of components of interest in biological samples [27–29] and in atmospheric air [30]. If it is necessary to determine the presence and concentration of several components, the RSC method can be used sequentially. However, such approach increases the time of analysis multiply. Additionally, if the order of component concentration recovery is chosen suboptimal, the accuracy of decomposition may decrease significantly.

This paper considers the generalization of RSC method to the multidimensional case as an alternative to its sequential application.

Materials and methods

Let's consider the experimentally measured spectrum of a multi-component gas sample as follows:

$$S_{\mathrm{mix}}(\nu) = |S^0_{\mathrm{mix}}(\nu) + r(\nu)|, \tag{3}$$

where $r(\nu)$ — random additive noise. It is possible to search for multiple components by RSC method using the following approaches.

1. The form of the minimized functional is the same as in (2), but the minimization is performed by varying the concentrations of several components simultaneously:

$$\delta f(\mathbf{c}) = \int \left| \frac{d(S_{\mathrm{mix}}(\nu) - \Sigma_i \tilde{c}_i S_i(\nu))}{d\nu} \right| d\nu. \tag{4}$$

2. The minimized functional is the sum of functionals (2), each of which is associated with the assessment of the presence and search for the concentration of one of the analyzed components:

$$\delta f(\mathbf{c}) = \int \sum_{i=1}^{N} \left| \frac{d(S_{\mathrm{mix}}(\nu) - \tilde{c}_i S_i(\nu))}{d\nu} \right| d\nu. \tag{5}$$

3. A combination of approaches 1 and 2, when the minimized functional is expressed as

$$\delta f(\mathbf{c}) = \int \sum_{i=1}^{N} \left| \frac{d(\tilde{S}_i(\nu) - \tilde{c}_i S_i(\nu))}{d\nu} \right| d\nu, \tag{6}$$

$$\tilde{S}_i(\nu) = \tilde{S}_{i-1}(\nu) - \tilde{c}_i S_i(\nu), \ \ \tilde{S}_0(\nu) = S_{\mathrm{mix}}(\nu).$$

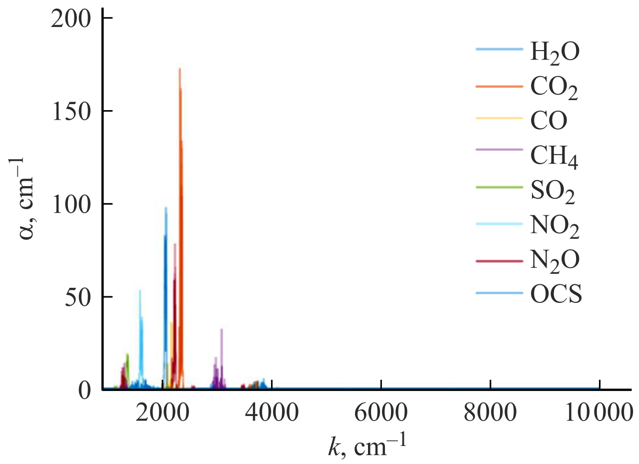

Figure 2. Absorption spectra for H_2O vapor, CO_2, CO, CH_4, SO_2, NO_2, N_2O, OCS in the range from 1000 to 3700 cm^{-1}.

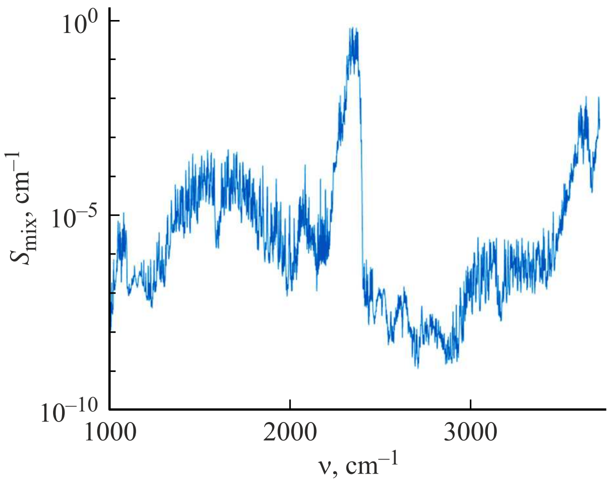

Figure 3. Example of calculated spectraS_{mix}.

Here $\mathbf{c} = (\tilde{c}_1, \tilde{c}_2, \ldots, \tilde{c}_N)$, $\tilde{c}_i$ — the sought-for concentrations of each of the components, N — number of the analyzed components.

Let's consider the application of the multidimensional RSC (mRSC) method using approaches (4)–(6) with an example of decomposing model spectra of atmospheric air, which includes the following main components: N_2 ($\sim 78.08\,\%$), O_2 ($\sim 20.95\,\%$), Ar ($\sim 0.93\,\%$), CO_2 ($\sim 0.04\,\%$), vapors of H_20 (content in the air — up to 4 %), as well as small components that are volatile molecular biomarkers and/or industrial air pollutants: CO, CH_4, SO_2, N_2O, NO_2, OCS. The examples also tested the hypothesis of the presence of NO_2 and OCS, which were actually absent in the model gas mixture.

Random noise was modelled by function

$$r(\nu) = R \cdot \max(S^0_{\mathrm{mix}}) rand(\nu), \tag{7}$$

where function $rand(\nu)$ takes random values from -0.5 to 0.5. At that, the spectrum $S_{\mathrm{mix}}(\nu)$ (3) is always not negative.

In the spectral range we are considering, the absorption of N_2, O_2, and Ar can be neglected, so these gases were not included in the model mixture. Fig. 2 shows the absorption spectra of the model mixture components, including small gas components (SGC). The spectra were calculated with a spectral resolution of 1 cm^{-1} for normal conditions using HITRAN 2020 spectral parameter database [31].

Next, a sample of model spectra of mixtures of the specified components S_{mix} was constructed. Concentration of H_2O changed in a random way from 0 to 4 %, and concentration of each of SGC changed as a normal distribution with parameters (average value, dispersion) mentioned in Table 1. Small gas components that are not shown in the table had a zero concentration. If the generator produced a negative concentration value, it was replaced with zero. Noise level varied from $R = 0$ to $R = 3 \cdot 10^{-4}$ with a step 10^{-6}. For each noise level value, 200 spectra S_{mix} were generated with different concentrations according to Table 1.

Table 1. SGC concentrations modeling parameters

Impurity	σ, ppm	Mean, ppm	Source
Carbon dioxide, CO_2	50	420	[32–35]
Carbon monoxide, CO	1.0	2.5	[34–37]
Methane, CH_4	1.0	1.9	[34,38,39]
Sulfur dioxide, SO_2	0.2	0.2	[40–42]
Nitrous oxide, N_2O	0.1	0.3	[34,36,43]

Results

Fig. 3 illustrates example of calculated spectra S_{mix}. By using mRSC method for spectra S_{mix}, let's find $\tilde{c}_i$ and relative error $\delta_i = |\frac{c_i - \tilde{c}_i}{c_i}|$ for the functionals (4)–(6). Table 2 gives an example of true (actual) c_i and calculated $\tilde{c}_i$ concentrations for one of the spectra S_{mix} at noise level $R = 0$.

It can be seen that the functional (6) provides a higher accuracy in reconstruction of SGC concentration compared to using the functionals (4) and (5). This is confirmed by the average relative error $\langle\delta\rangle$ and its mean-square error of SGC concentration reconstruction on the full sample of model spectra (Table 3). Gases with zero concentration were not considered.

Fig. 4 illustrates the calculated average relative error $\langle\delta\rangle$ of SGC concentrations reconstructed using functional (6) versus noise amplitude R and „signal/noise" ratio (SNR).

Conclusion

This paper proposes a method for decomposing the absorption spectra of gas samples of unknown composition with an arbitrary number of components. The method

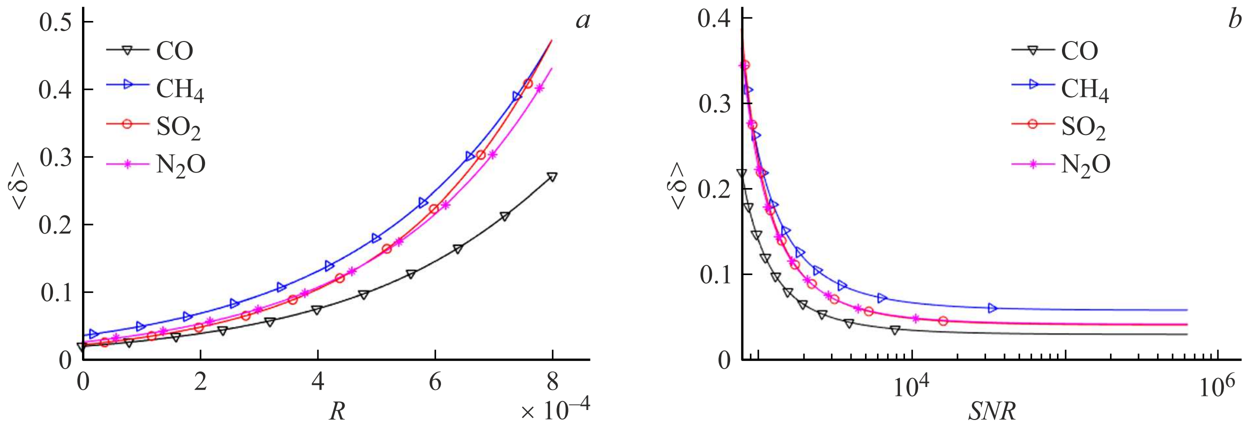

Figure 4. Calculated dependence of the average relative error $\langle\delta\rangle$ of SGC concentrations reconstructed using functional (6) on the noise level R (a) and „signal/noise" ratio SNR (b).

Table 2. Example of reconstructed $\tilde{c}_i$ and true c_i concentrations for one of the spectra S_{mix} at level of noise $R = 0$

Impurity	c_i, ppb	$\tilde{c}_i$, ppb		
		Functional (4)	Functional (5)	Functional (6)
CO	2500.0	2168.8	1755.0	2486.5
CH_4	10	9.0	9.1	9.8
SO_2	200.0	205.4	142.6	202.0
NO_2	0	0.1	−106.9	0.00003
N_2O	20.0	19.1	24.0	19.4
OCS	0	−0.78	154.6	0.1

Table 3. The average relative error $\langle\delta\rangle$ and the mean-square error of the reconstructed SGC concentrations on the full sample of model spectra at the noise level $R = 0$

Gas	$\langle\delta\rangle$		
	Functional (4)	Functional (5)	Functional (6)
CO	0.062 ± 0.050	0.122 ± 0.090	0.021 ± 0.008
CH_4	0.076 ± 0.034	0.136 ± 0.034	0.028 ± 0.002
SO_2	0.062 ± 0.049	0.122 ± 0.080	0.021 ± 0.006
N_2O	0.061 ± 0.031	0.131 ± 0.061	0.023 ± 0.006

is based on reducing the „complexity" of the spectrum by using a precise removal of one of the components from the total spectrum. The complexity criterion is the integral area of the module of the first derivative of the spectrum. Three variants of the minimized functional are considered, which are suitable for decomposing the absorption spectra of gas samples of unknown composition with an arbitrary number of components. The method is illustrated by assessing the presence and reconstruction of SGC concentrations in atmospheric air. The calculated decomposition error, including at different noise levels, showed the effectiveness of the proposed method in SGC concentration reconstruction up to the level of tens of ppb, despite the presence of the main components that significantly contribute to the absorption of air in the considered spectral range.

Funding

This study was funded by the Ministry of Science and Higher Education of the Russian Federation (grant № 075-15-2024-557 of 25.04.2024).

Conflict of interest

The authors declare that they have no conflict of interest.

References

[1] A. Trtyakov, D. Vrazhnov, A. Shkurinov, V. Zasedatel, Y. Kistenev. Appl. Sci., **14** (24), 11521 (2024). DOI: 10.3390/app142411521

[2] Y.Z. Liang, O.M. Kvalheim, R. Manne. Chemometr. Intell. Lab., **18** (3), 235 (1993). DOI: 10.1016/0169-7439(93)85001-W

[3] A. de Juan, R. Tauler. Anal. Chim. Acta, **1145**, 59 (2021). DOI: 10.1016/j.aca.2020.10.051

[4] S. Ishihara, Y. Hattori, M. Otsuka, T. Sasaki. Crystals, **10** (9), 760 (2020). DOI: 10.3390/cryst10090760

[5] O.E. Rodionova, A.L. Pomerantsev. Zhurn. analit. khimii, **71** (1), 58 (2016) (in Russian). DOI: 10.7868/S0044450216010126
[6] P. Kościelniak, M. Wieczorek. Anal. Chim. Acta, **944**, 14 (2016). DOI: 10.1016/j.aca.2016.09.024
[7] M.A. Merriman. Academy, **4**, 1 (1877).
[8] H.P. Gavin. Department of Civil and Environmental Engineering, Duke University, **3**, 1 (2019).
[9] K. Madsen, H.B. Nielsen, O. Tingleff. *Methods for non-linear least squares problems*. (Informatics and Mathematical Modelling Technical University of Denmark, 2004).
[10] K. Levenberg. Q. Appl. Math., **2** (2), 164 (1944).
[11] W.H. Lawton, E.A. Sylvestre. Technometrics, **13** (3), 617 (1971).
[12] S.K. Karimvand, M. Maeder, K. Bakhshi, H. Abdollahi. Anal. Chim. Acta, **1154**, 338320 (2021). DOI: 10.1016/j.aca.2021.338320
[13] G.T. Rasmussen, T.L. Isenhour, J.O. Lephardt. Anal. Chim. Acta, **103** (3), 213 (1978). DOI: 10.1016/S0003-2670(01)84040-X
[14] O.S. Borgen, B.R. Kowalski. Anal. Chim. Acta, **174**, 1 (1985). DOI: 10.1016/S0003-2670(00)84361-5
[15] R. Rajkó, K. István. J. Chemometr., **19** (8), 448 (2005). DOI: 10.1002/cem.947
[16] N. Ohta. Anal. Chem., **45** (3), 553 (1973). DOI: 10.1021/ac60325a010
[17] A. Meister. Anal. Chim. Acta, **161**, 149 (1984). DOI: 10.1016/S0003-2670(00)85786-4
[18] S. Kawata, H. Komeda, K. Saito, S. Minami. Appl. Spectrosc., **39** (4), 610 (1985).
[19] B.G.M. Vandeginste, W. Derks, G. Kateman. Anal. Chem., **57** (6), 971 (1985). DOI: 10.1021/ac00283a005
[20] P.J. Gemperline. Anal. Chem., **71** (23), 5398 (1999). DOI: 10.1021/ac990648y
[21] G. Ahmadi, H. Abdollahi. Chemometr. Intell. Lab., **120**, 59 (2013). DOI: 10.1016/j.chemolab.2012.11.007
[22] M. Vosough, C. Mason, R. Tauler, M. Jalali-Heravi, M. Maeder. J. Chemometr., **20** (6−7), 302 (2006). DOI: 10.1002/cem.1022
[23] A. Golshan, H. Abdollahi, M. Maeder. Anal. Chem., **83** (3), 836 (2011). DOI: 10.1021/ac102429q
[24] A. Golshan, M. Maeder, H. Abdollahi. Anal. Chim. Acta, **796**, 20 (2013). DOI: 10.1016/j.aca.2013.08.007
[25] M. Ando, I.K. Lednev, H. Hamaguchi. In: *Frontiers and Advances in Molecular Spectroscopy* (Elsevier, 2018). P. 369. DOI: 10.1016/B978-0-12-811220-5.00011-3
[26] S. Banerjee, D. Li. Appl. Spectrosc., **45** (6), 1047 (1991).
[27] A.V. Borisov, D.A. Vrazhnov, Yu.V. Kistenev, A.P. Shkurinov, V.V. Zasedatel, A.A. Karapuzikov. J. Breath Res., **15** (2), 027104 (2021). DOI: 10.1088/1752-7163/abebd4
[28] A.V. Borisov, M.S. Snegerev, S. Colón-Rodríguez, M.A. Fikiet, I.K. Lednev, Yu.V. Kistenev. Sci. Rep., **14** (1), 23070 (2024). DOI: 10.1038/s41598-024-73563-w
[29] Yu.V. Kistenev, A.V. Borisov, A.A. Samarinova, S. Colón-Rodríguez, A. Viner, O.P. Cherkasova, D.A. Vrazhnov, M.A. Fikiet, I.K. Lednev. Sci. Rep., **13** (1), 5384 (2023). DOI: 10.1038/s41598-023-31918-9
[30] E.Yu. Yerushin, N.Yu. Kostyukova, A.A. Boyko, I.B. Miroshnichenko. PTE, **3**, 67 (2024) (in Russian). DOI: 10.31857/S0032816224030082
[31] I.E. Gordon, L.S. Rothman, R.J. Hargreaves, R. Hashemi, E.V. Karlovets, F.M. Skinner, E.K. Conway, C. Hill, R.V. Kochanov, Y. Tan, P. Wcisło, A.A. Finenko, K. Nelson, P.F. Bernath, M. Birk, V. Boudon, A. Campargue, K.V. Chance, A. Coustenis, B.J. Drouin, J.-M. Flaud, R.R. Gamache, J.T. Hodges, D. Jacquemart, E.J. Mlawer, A.V. Nikitin, V.I. Perevalov, M. Rotger, J. Tennyson, G.C. Toon, H. Tran, V.G. Tyuterev, E.M. Adkins, A. Baker, A. Barbe, E. Cane, A.G. Császár, A. Dudaryonok, O. Egorov, A.J. Fleisher, H. Fleurbaey, A. Foltynowicz, T. Furtenbacher, J.J. Harrison, J.-M. Hartmann, V.-M. Horneman, X. Huang, T. Karman, J. Karns, S. Kassi, I. Kleiner, V. Kofman, F. Kwabia-Tchana, N.N. Lavrentieva, T.J. Lee, D.A. Long, A.A. Lukashevskaya, O.M. Lyulin, V.Yu. Makhnev, W. Matt, S.T. Massie, M. Melosso, S.N. Mikhailenko, D. Mondelain, H.S.P. Müller, O.V. Naumenko, A. Perrin, O.L. Polyansky, E. Raddaoui, P.L. Raston, Z.D. Reed, M. Rey, C. Richard, R. Tobias, I. Sadiek, D.W. Schwenke, E. Starikova, K. Sung, F. Tamassia, S.A. Tashkun, J. Vander Auwera, I.A. Vasilenko, A.A. Vigasin, G.L. Villanueva, B. Vispoel, G. Wagner, A. Yachmenev, S.N. Yurchenko. JQSRT, **277**, 107949 (2022). DOI: doi.org/10.1016/j.jqsrt.2021.107949
[32] S.M. Semenov Fundament. i prikl. klimatologiya, **2** (105), 2018 (in Russian). DOI: 10.2172/768563
[33] R. Dryden, M.G. Morgan, A. Bostrom, W. Bruine de Bruin. Risk Anal., **38** (3), 525 (2018). DOI: 10.1111/risa.12856
[34] M. Meinshausen, S.J. Smith, K. Calvin, J.S. Daniel, M.L.T. Kainuma, J.-F. Lamarque, K. Matsumoto, S.A. Montzka, S.C.B. Raper, K. Riahi, A. Thomson, G.J.M. Velders, D.P.P. van Vuuren. Clim. Change, **109**, 213 (2011). DOI: 10.1007/s10584-011-0156-z
[35] V.N. Arefyev, N.E. Chubarova, E.I. Grechko, A.V. Zharkov, and G.S. Rivkin. Izvestiya RAN. Fizika atmosfery i okeana, **50** (6), 655 (2014) (in Russian). DOI: 10.7868/S0002351514060030
[36] M. Shahgedanova, T.P. Burt, T.D. Davies. Water Air Soil Pollut., **112**, 107 (1999). DOI: 10.1023/A:1005043916123
[37] H.M. Worden, M.N. Deeter, D.P. Edwards, J.C. Gille, J.R. Drummond, P. Nédelec. Atmos. Chem. Phys., **13** (2), 837 (2013). DOI: 10.5194/acp-13-837-2013
[38] A. Van Amstel. J. Integr. Environ. Sci., **9** (S1), 5 (2012). DOI: 10.1080/1943815X.2012.694892
[39] V.I. Bogoyavlensky, G.M. Tretyakova, and V.Yu. Zhuravlev. Arktika: ekologia i ekonomica, **12** (3), 351 (2022). DOI: 10.25283/2223-4594-2020-3-304-319
[40] D.V. Melnikov, S.V. Ushakov. V sb.: *Geofizichesky monitoring i problemy seismicheskoy bezopasnosti Dal'nego Vostoka Rossii: trudi regionalnoy nauch.-tekhn. konferentsii*, 11−17 November 2007. (GS of RAS, Petropavlosk-Kamchatsky, 2008). V. 1. P. 101 (IN RUSSIAN).
[41] E. Robinson, R.C. Robbins. J. Air Pollut. Control Assoc., **20** (4), 233 (1970). DOI: 10.1080/00022470.1970.10469396
[42] A.V. Eliseev, I.I. Mokhov, and A.V. Timazhev. Izvestiya RAN. Fiz. Atmos. Okeana, **55** (1), 41 (2019) (in Russian). DOI: 10.31857/S0002-351553141-53
[43] H. Tian, R. Xu, J.G. Canadell, R.L. Thompson, W. Winiwarter, P. Suntharalingam, E.A. Davidson, P. Ciais, R.B. Jackson, G. Janssens-Maenhout, M.J. Prather, P. Regnier, N. Pan, S. Pan, G.P. Peters, H. Shi, F.N. Tubiello, S. Zaehle, F. Zhou, A. Arneth, G. Battaglia, S. Berthet, L. Bopp, A.F. Bouwman, E.T. Buitenhuis, J. Chang, M.P. Chipperfield, S.R. Cranborne, S. Dangal, E. Dlugokencky, J.W. Elkins, B.D. Eyre, B. Fu,

B. Hall, A. Ito, F. Joos, P.B. Krummel, A. Landolfi, G.G. Laruelle, R. Lauerwald, W. Li, S. Lienert, T. Maavara, M. MacLeod, D.B. Millet, S. Olin, P.K. Patra, R.G. Prinn, P.A. Raymond, D.J. Ruane, M.A. Saunois, J. Schroeder, R.J. Sindelar, K.M. Smith, R. Tohjima, F.N. Tubiello, G.R. van der Werf, N. Vuichard, J. Wang, R.F. Weiss, K.C. Wells, C. Wilson, J. Yang, Y. Yao. Nature, **586** (7828), 248 (2020). DOI: 10.1038/s41586-020-2780-0

Translated by J.Savelyeva