

Information system for molecular spectroscopy.

1. Structure of information resources

A.D. Bykov, B.A. Voronin, A.V. Kozodoev, N.A. Lavrent'ev,
O.B. Rodimova, and A.Z. Fazliev

*Institute of Atmospheric Optics,
Siberian Branch of the Russian Academy of Sciences, Tomsk*

Received September 15, 2004

The data model used in the information system for molecular spectroscopy (<http://saga.atmos.iao.ru>) is described. The principal parts of the data model are fundamental molecular characteristics, spectral line parameters, and spectral functions. The data input system, allowing users to create their own data sets, is described. The relations between the parts of the data model are demonstrated by calculation of the absorption coefficient as an example.

Introduction

Molecular spectroscopy refers to key fundamental sciences whose advances have a large number of applications. For atmospheric sciences, most important are spectral data about tens of molecules. It should be mentioned that only the data on the water molecule consist of information about hundreds millions of spectral lines. The operation with such large data arrays necessitates the use of automated information computation systems (ICS).

There are a sufficient number of information systems for molecular spectroscopy, and most of them use spectral line parameters as data. The HITRAN¹ and GEISA² data banks accumulate the information about almost forty molecules. These data banks are accessible via the Internet and CD. As we know, the first attempt to organize automatic ICS with direct Internet access for scanning spectral line parameters and calculating spectral functions was made by Golovko and Babikov with co-workers.³⁻⁵ The base of SPECTRA ICS (<http://spectra.iao.ru>) developed by them were spectral line parameters from the data banks^{1,2} and original data obtained at IAO SB RAS (<ftp://ftp.iao.ru/pub/CDS-296> and <ftp://ftp.iao.ru/pub/CDS-1000>).

In the Internet-accessible systems, which involve sharing of information resources, of great importance is understanding of the data model employed in ICS. In this paper, we describe the data model used in our ICS for molecular spectroscopy (<http://saga.atmos.iao.ru>). A feature of this ICS is the data model augmentability and the possibility for users to input private data in the framework of the used data model.

By their origin, spectroscopic data are of two types: experimental and calculated, and their descriptions differ principally. In the description of experimental data, the key information, beyond the experimental conditions, are parameters of the experimental setup. In description of the calculated data, the key information includes the used computational methods and the used data models. In

the current practice of ICSs development, these descriptions are not employed in calculations with the use of spectral line parameters. This is mainly caused by the fact that computer processing of such descriptions is still not accessible for most spectroscopy specialists, on the one hand, because of their misunderstanding of capabilities of modern information technologies and, on the other hand, because of the increasing cost of such work.

Below we describe the structure of the information resources, as well as the data input and data display procedures in the molecular spectroscopy information model. The information model is materialized in the form of information resources,⁶ which are divided into the data and metadata.⁷ The user-accessible level of the information service is demonstrated by the calculation of the absorption coefficient as an example.

Structure of information resources

The operation cycle involves several stages. The basic stages are data acquisition, storage, processing, and delivery to the user. Data acquisition and input in the ICS are performed with the aid of html forms, accessible to both the administrator and the user. Information is stored in databases. Data processing is the most laborious part in the program realization of this cycle, because this stage requires a detailed knowledge of data semantics. Data are delivered through the Internet.

Three qualitatively different structures can be distinguished in the data model used by us:

1. Fundamental molecular characteristics.
2. Spectral line parameters.
3. Spectral functions.

The order of the structures in this list is important in preparing metadata. The molecular structure data are used in calculating spectral line parameters, which, in turn, are used in calculating spectral functions. When calculating, for example, the absorption coefficient and constructing the metadata hierarchy,

the hierarchy's lower level is represented by metadata for fundamental characteristics, the next level – metadata for spectral line parameters and, finally, the last level contains metadata for spectral functions.

Parameters of the equilibrium configuration, potential energy function, and dipole moment or polarizability function are usually considered as fundamental molecular characteristics. In our data model, fundamental molecular characteristics are parameters, which allow the calculation of spectral line parameters, namely, line positions, intensities, halfwidths, pressure-induced shifts, or cross-relaxations. In other words, they are characteristics determining the energy of a molecule. Depending on the description method, they can be represented by the parameters of the total molecular Hamiltonian (potential energy, dipole moment) or the effective Hamiltonian (rotational, centrifugal distortion, and resonance constants, parameters of effective dipole moment). They, certainly, should be complemented with quadrupole, octupole molecular moments and other parameters characterizing the intermolecular interaction in gases.

The use of spectral line parameters, whose lists differ in different data banks, is based on the assumption that the hypothesis of isolated line is applicable in calculations. These data are oriented at the use in calculations of spectral functions (absorption coefficients, transmission functions, and others).

Systematization of experimental values of spectral functions is needed, in the first turn, for comparison with the calculated ones. When compiling the data bank, we paid a particular attention to experiments involving the study into wings of spectral lines.

As was mentioned above, the information resources have two components: data and metadata. The widely used extended definition of metadata is that they are data about other data. This definition includes neither the criterion of completeness of metadata for the data of some subject field, nor indication, from the position of what scientific discipline the metadata are considered. By this reason, the metadata taken by us for data description are defined below by enumeration. Metadata, which are, in our opinion, of primary importance in molecular spectroscopy, must point to the source of information for all three data structures, as well as contain parameters determining the domains of applicability, the error magnitude, experimental conditions, methods of calculation, and others. They can be used for logical justification of applicability of the results obtained.

The data also differ in the degree of user's authority for data access. In the ICS, there are two access levels: administrative and user. The data input interface is the same for the both access levels. The only difference is that the data inputted by the administrator are accessible to all users, while the data inputted by some user are accessible only to this user. The administrative level allows for both data input and data integrity control, while the user level authorizes only data input. At present, it is possible to input data of two structures: spectral line parameters

(the minimal set includes line positions and intensities) and spectral functions (absorption coefficient, transmission functions, etc.).

There are own data processing methods for every data structure. The methods of operation with spectral line parameters and spectral functions are described below.

The stage of data processing includes some technical operations (database inquiring, graphic and tabulated displaying) and subject processing. The technical stage is described only for graphical displaying. The subject aspect is associated with calculation of absorption coefficients of a gas mixture.

Note that, when operating in the SPECTRA ICS, users cannot input private data on spectral line parameters and perform calculations with them. The only authorized data input option for users in the SPECTRA ICS is input of private data in the form of a two-dimensional array in order to compare them graphically with the calculated results.

Operations with information resources

In this section, we describe the data processing, namely, procedures of data input into the information system and data display for the end user. Consider two structures: spectral line parameters and spectral functions.

According to the physical meaning, spectral line parameters for every molecule are divided into several groups, namely

- parameters of an isolated spectral line (line intensity, position, lower-level energy, statistical weight of the lower and upper levels, transition moment, etc.);
- parameters of line assignment (vibrational and rotational-vibrational assignment);
- parameters caused by collisions (line halfwidth, pressure induced shift, temperature dependence of halfwidth, etc.).

Some physical characteristics have attributes, such as the indicator of data origin (experiment, calculation), type of a scale (absolute, relative), and units of physical characteristics.

Experts usually fill spectral data banks taking into account the multiple values of line parameters. The frequency of data bank updating varies from one to several years, which poses some difficulties in updating calculated spectral functions. The solution of the problem lies on the way of augmenting ICS services, allowing the user, on the one hand, to create private information resources and, on the other hand, to use the resources available in the ICS. The important procedure in the first case is the data input, and in the second case – the compilation of resource collections.

The common point in the input of different data structures is the creation of the data source. The description of the data source includes abbreviation, authors, the title of the paper (monograph), the title of the journal, its volume, the number, the year,

pages, URL address, and comments. Before inputting data, the user describes the data source or makes linking between the source already described in the ICS and the data to be inputted.

The data input procedure uses the private data in the form of a file including data columns. The order of the columns is arbitrary, but synchronized with the list of physical characteristics, compiled by the user during the data input. At the first stage of input, the user creates the data source. At the second stage, the user selects a molecule and interactively forms the list of characteristics in accordance with the order of columns in the file. Columns containing positions and intensities of spectral lines are compulsory. At the last stage, the user defines the attributes of physical parameters.

When inputting data on the absorption coefficient, most parameters are indispensable, in particular, spectral parameters (resolution, path length), thermodynamic parameters (temperature and pressure), interacting substances (absorbing or buffer gases), wavenumber and absorption coefficients (units), and the method of obtaining the absorption coefficient (experiment, calculation).

Looking the diagram of spectral line intensities for some selected molecule is conducted in the menu Molecule. There are two stages of the looking. First, the molecule or its isotopic modification is selected from the available list, and the spectral range is defined, along with the thermodynamic conditions

and the highest or lowest value of the line intensity, which is used in the data access. At the second stage, the user selects a needed source from the displayed list of data sources. Note that for each source the list presents the number of lines, falling within the selected spectral range, and the number of lines, not falling within the range, but corresponding to the bands falling within this range.

Data are visualized with the aid of the graphic application in the form of applet. This application allows the user to scale the picture, to select different data sources for displaying, and to compare plots by pairs as mirror images. Figure 1 shows the comparison of two data sets. The detailed description of display options can be found in <http://saga.atmos.iao.ru>.

Calculation of the absorption coefficient

Spectral functions are calculated using the parameters of an isolated line and parameters conditioned by collisions. To calculate them, the ICS includes two physically different situations: pure gas and a mixture of gases. The data needed for calculation of the absorption coefficient for a pure gas (homogeneous in composition) are described below. The extension to the case of some gas mixture is rather simple.

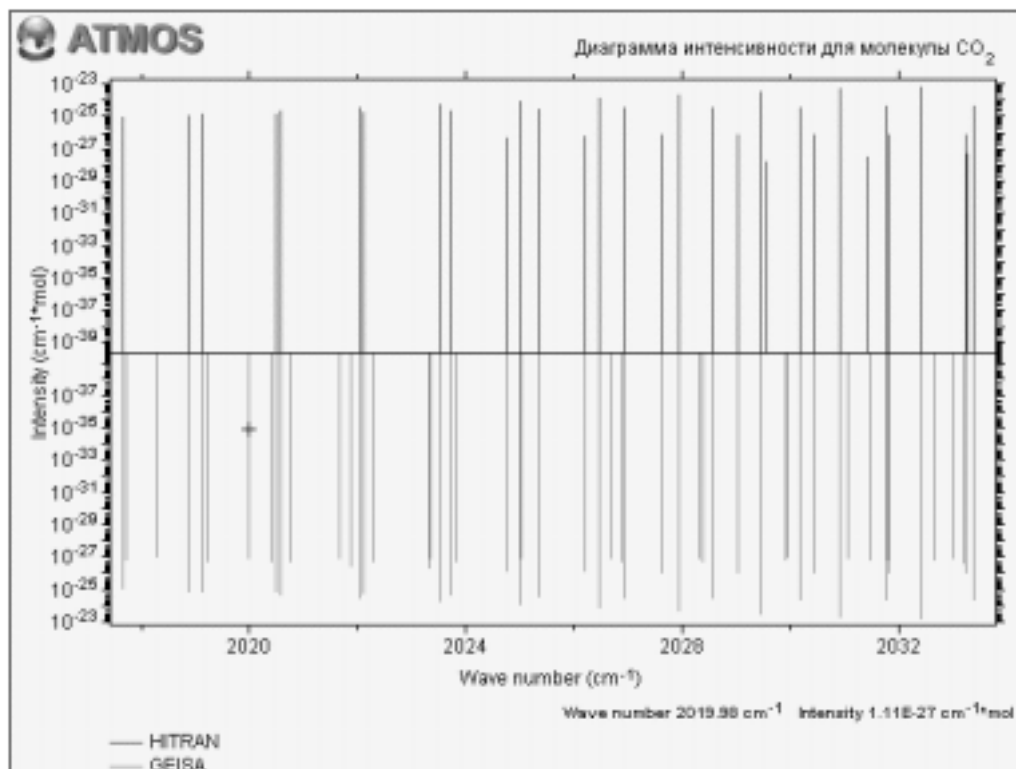


Fig. 1. Comparison of spectral line parameters for the CO₂ molecule in the spectral range 2018–2034 cm⁻¹ from two information sources: HITRAN (upper panel) and GEISA (lower panel). The cross indicates the handler position, and parameters of the chosen spectral line are displayed at the bottom.

As is well known, the absorption coefficient of a pure gas s at the frequency ω motivated by the absorption by all lines in the range $[\omega - \Delta\omega/2, \omega + \Delta\omega/2]$ can be written as

$$\kappa^{(s)}(\omega, T, p^{(s)}) = \sum_j \frac{\omega}{\omega_j} \frac{(1 - e^{-h\omega/kT})}{(1 - e^{-h\omega_j/kT})} S_j^{(s)}(T) G^{(s)}(\omega, \omega_j, T, p^{(s)}), \quad (1)$$

where the summation is performed over all lines j in the selected range $[\omega - \Delta\omega/2, \omega + \Delta\omega/2]$; ω is the frequency, at which the absorption coefficient is calculated, ω_j is the central frequency of the j th line; $S_j^{(s)}$ is the intensity of the j th line; $G^{(s)}(\omega, \omega_j, T, p^{(s)})$ is the profile of the j th line in the case of self-broadening, T is the considered temperature, in K, T_0 is the temperature, at which the line parameters are defined in the source of information about spectral line parameters; $p^{(s)}$ is the pressure (in atm.) of the considered gas. The range $\Delta\omega$ is specified from heuristic assumptions and can depend on the gas type and thermodynamic conditions.

The intensity of the j th line at the arbitrary T is

$$S_j(T) = S_j(T_0) \frac{(1 - e^{-C_2\omega_j/T})}{(1 - e^{-C_2\omega_j/T_0})} \times e^{-C_2E_j \frac{(T_0 - T)}{T_0 T}} \frac{Q_V(T_0)}{Q_V(T)} \frac{Q_R(T_0)}{Q_R(T)}, \quad (2)$$

where Q_V and Q_R are the vibrational and rotational statistical sums, respectively; E_j is the energy of the lower level of the transition, in cm^{-1} .

The tables of the ratios

$$\frac{Q(T_0)}{Q(T)} = \frac{Q_R(T_0)}{Q_R(T)} \frac{Q_V(T_0)}{Q_V(T)}$$

can be found, for example, in HITRAN-92.

The first stage prior to the calculation involves selection of the substance, spectral range, thermodynamic conditions, and the threshold value of the isolated line intensity, below which the data about the line are neglected. At the second stage, the user obtains the list of the possible sources of information about spectral line parameters. Once one source is selected, the user has to specify the type of the spectral line profile, interval of the line profile $\Delta\omega$, and points for calculation. The points can be defined in one of three ways: by enumeration (in the file), by indication in the html form for input of the number of uniform divisions of the selected spectral range (no less than the number of lines) or the division factor (in this case, the number of divisions is equal to the product of the division factor by the number of spectral lines falling within the selected spectral range).

To calculate the absorption coefficient, the following data set is used: positions and intensities of spectral lines, statistical sums, the coefficient of temperature dependence of the halfwidth, the line halfwidth and shift.

On the user's request, the calculated result can be saved at the server. Then it can be compared with the experimental results or basic calculations, included in the data bank. Figure 2 demonstrates the comparison of the calculated results with the experiment.⁸

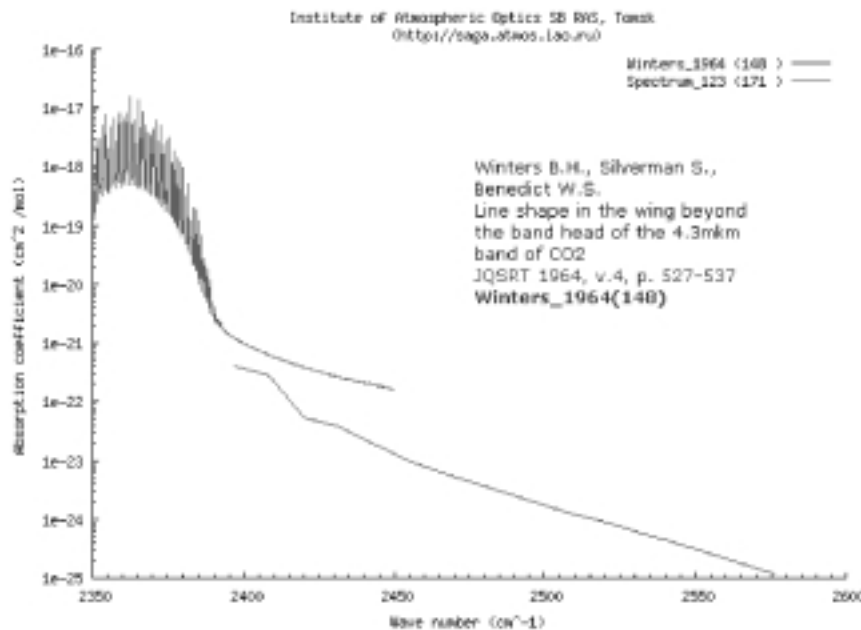


Fig. 2. Comparison of experimental data for CO₂ (Winters_1964, lower curve) with calculation (Spectrum_123).

Annotation		DublinCore XML	
Substance		Thermodynamical Conditions	
Absorbing gas	CO2	Temperature (°K)	296
Broadening gas	self	Pressure (atm)	0.25
Data array		Broadening gas pressure (atm)	
Wave number (number of values, unit) (cm ⁻¹)	15	Spectral parameters	
Absorption coefficient (number of values, unit) (Am ⁻² cm ⁻¹ , exp)	15	Spectral resolution (cm ⁻¹)	0.6
Errors	Yes	Path length (m)	10
Reference		Frequency range (cm ⁻¹)	
Authors	Winters B.H., Silverman S., Benedict W.S.		
Title	Line shape in the wing beyond the band head of the 4.3μm band of CO2		
Journal	JQSRT 1964, v.4, p. 527-537		
Commentary			

Fig. 3. Metadata for Winters_148 information source.

The system provides for browsing of metadata for experimental results stored in the database (Winters_1964). They can be browsed from the same page. Figure 3 exemplifies the representation of the metadata.

Conclusions

In this paper, we have described the structure of the information resources used in the Atmospheric Spectroscopy information computation system. The particular attention was paid to description of the structured data model. It should be noted that most molecular spectroscopy data can be presented now within the framework of semistructured or unstructured data models, and they are included in the system under consideration. The description of operations with such data is beyond the scope of this paper.

Further development of the data model, first of all, fundamental molecular characteristics, will be carried out in a new version of the system. The first steps in this direction are already made.⁹

Acknowledgments

The authors are grateful to Prof. S.D. Tvorogov, Corresponding Member of RAS (Grant "Optical

molecular spectroscopy and radiative processes in the atmosphere" NSh No. 373.2003.5) and Russian Foundation for Basic Research (Grant No. 02-07-90139).

References

1. <http://www.hitran.com>
2. <http://ara.ltd.polytechnique.fr>
3. Yu.L. Babikov, A. Barbe, V.F. Golovko, S.N. Mikhailenko, and V.I.G. Tyuterev, Proc. RCDL, 183-187 (2001).
4. Yu.L. Babikov, A.S. Fasliev, V.F. Golovko, and O.B. Rodimova, Proc. SPIE **4341**, 604-615 (2000).
5. S.N. Mikhailenko, Yu.L. Babikov, V.I.G. Tyuterev, and A. Barbe, Vychisl. Tekhnol. **7**, 64-70 (2002).
6. M.R. Kogalovskii, *Promising Technologies of Information Systems* (DMK Press, Moscow, 2003), 288 pp.
7. M.R. Kogalovskii, in: *Proc. of the First All-Russian Sci. Conf. Electronic Libraries: Promising Methods and Technologies, Electronic Collections*, (St. Petersburg, 1999), pp. 16-31.
8. B.H. Winters, S. Silverman, and W.S. Benedict, J. Quant. Spectrosc. Radiat. Transfer **4**, No. 4, 527-537 (1964).
9. A.V. Kozodoev, R.K. Azizov, A.D. Bykov, A.M. Pshenichnikov, and A.Z. Fazliev, in: *Abstracts of Reports Presented at Int. Symp. on Atmospheric Radiation* (St. Petersburg, 2004), pp. 92-93.