The current status of the W@DIS information system

A. Akhlyostin[1], Z. Apanovich[2], A. Fazliev[1], A.Kozodoev[1], N. Lavrentiev[1],
A. Privezentsev[1], O. Rodimova[1], S. Voronina[1], A.G. Császár[3], and J. Tennyson[4]

[1] *V. E. Zuev Institute of Atmospheric Optics SB RAS, Tomsk 634055, Russia,*
[2] *Institute of Informatics Systems SB RAS, Novosibirsk, Russia,*
[3] *MTA-ELTE Complex Chemical Systems Research Group, Budapest, Hungary,*
[4] *Department of Physics and Astronomy, University College London, London WC1E 6BT, United Kingdom*

ABSTRACT

The current status of the W@DIS information system used for the systematization of spectroscopic data, including rovibronic transitions and energy levels, and data sources is reviewed, where the abbreviation W@DIS stands for Water Internet @ccesible Distributed Information System. Functionalities of W@DIS are outlined. The primary emphasis of W@DIS is on properties of data sources characterizing data quality. Several examples describe the interfaces used to create molecular spectral line lists and representation of binary relations between data sources and typical individuals of the ontology of information resources, states and transitions. The discussion employs the water molecule as an example.

Keywords: systematization of spectral data, ontologies of resources, transitions, and rovibrational states.

1. INTRODUCTION

Research into a subject domain involves several steps: collection of facts, construction of subject domain models, comparison of the proposed models with those developed by other researchers, and, finally, publication of the models. This process raises several questions: (a) is the set of facts collected and generated by a researcher complete; (b) are these facts consistent with each other; (c) does the formal language of the model specification enable researchers to build adequate subject domain models which represent faithfully the collected facts; (d) are the proposed model**s** consistent with those developed by other researchers; and (e) how can access to the results be facilitated. These questions are of special relevance in the case of high-resolution molecular spectroscopy, one of the fields of physics and chemistry widely used in many applied research areas.

Data sets call for the systematization of information and the design of computer data processing software. Implementation of the software implies construction of subject domain models related to the data of interest as well as development of the required tools associated with a search for the relevant information resources. In high-resolution molecular spectroscopy, such resources are solutions of several related spectroscopic tasks. In quantitative spectroscopy, some of the expert data acquired by different research groups[1,2] proved to be inconsistent. The consistency among the data available and those obtained by different investigators involved in collecting expert data is one of the main tasks dealt with not only in spectroscopy,[3] but also in other subject domains.

In the mid-2000s, an IUPAC project was launched,[4–8] whereby an information system intended for collection of all currently available published experimental and theoretical rovibrational data on the water molecule and its main isotopologues was built. The information system contains tools for detecting inconsistency between transition wavenumbers. There are formal and informal criteria for identifying inconsistent spectroscopic data. The formal criteria include selection rules, root-mean-square deviations (RMSD) concerning the difference between the wavenumbers of identical transitions, duplicates, etc. The informal criteria imply expert assessment of data quality, like those used in the W@DIS information system discussed here. Ontology (logical theory) of information resources for the water molecule was built to describe the properties of published solutions to spectroscopic tasks.[3] At present two groups of ontologies

22nd International Symposium Atmospheric and Ocean Optics: Atmospheric Physics, edited by
Gennadii G. Matvienko, Oleg A. Romanovskii, Proc. of SPIE Vol. 10035, 100350D ·
· doi: 10.1117/12.2249235

developed in W@DIS provide descriptions of the state-of-the-art of published data and of transitions and states used in water spectroscopy. Note that these ontologies can be accessed via the Internet at http://wadis.saga.iao.ru. Goals of the present work within W@DIS are as follows: (a) collect publications on different molecules, specifically, water and its isotopologues; (b) provide an import facility of the spectral data describing the results obtained from computer processing of the measured and computed data; (c) develop software for the alignment of data relating to solutions of spectroscopic tasks;[3] (d) build information sources and a few ontologies for quantitative spectroscopy; and (e) assess the quality of the complete set of the relevant data.

Among the subjects of an IUPAC-sponsored activity[4–8] were the critical evaluation and validation of all the measured high-resolution spectra of nine water isotopologues. Utilizing the functionalities of the MARVEL algorithm,[9,10] where MARVEL stands for Measured Active Rotational-Vibrational Energy Levels, the measured transitions and the associated energy levels were made fully consistent. During this process a small part of the measured transitions had to be rejected and some of the published spectra had to be recalibrated. The results obtained and the lists of the validated and rejected transitions and energy values were imported into the W@DIS[11] and RESPECTH[12] information systems.

The development of the W@DIS information system was motivated in part by the publication of a large number of articles about the spectral parameters of the water molecule, including energy levels, transitions and transition intensities, and spectral lineshape parameters, every year. These publications contain new data about spectral parameters of the water molecule, including a list of more accurately measured transitions. In some instances, the newly published data are found to be inconsistent with those presented in previous publications.

W@DIS contains several tools which can be used for spectral data export and import, comparison of the spectral data associated with particular spectroscopic tasks, and representation of data properties. Most of the data properties are indicative of data quality, *i.e.*, of the validity of and trust in the available expert data. In this work, user interfaces are described and computer-generated reports on the quality of spectral data for all isotopologues of the water molecule are presented.

2. THE FUNCTIONALITIES OF W@DIS

In quantitative spectroscopy, expert spectroscopic data accessible to users are scattered in at least 20 databases. The VAMDC project developed a single interface connecting these databases.[13,14] Since each database is intended for solving a wide range of applied tasks, occasionally, inconsistent transitions, *e.g.*, different wavenumbers with identical quantum numbers, are made available to the users. The W@DIS information system considered in this work contains one of the databases available in the VAMDC project. W@DIS is designed to provide access to a complete set of published spectral data for a number of atmospheric molecules. W@DIS is of interest to experts and one of its main functionalities is the high-quality characterization of spectral data.

The information system presented in this work is based on a digital library of publications on quantitative high-resolution molecular spectroscopy, the associated data sources found in the published articles, and the information resources describing the properties of the data sources. The information resources of the system are intended for researchers involved in solving spectroscopic tasks by means of graphical user interfaces and ontologies oriented to programmable agents. The data manipulation software[13] comprises some twenty programs used for importing data sources, computing information sources for each of the data sources, searching and presenting the data available in W@DIS, assessing trust in expert data according to several criteria,[14,15] developing ontologies of information resources related to each of the molecules included in the information system and to several spectroscopic tasks,[3] and ontologies of molecular states and transitions, among which are ontologies of vibrational bands and states.

W@DIS was built in 2006[16] to accumulate, to provide access to, and to comment upon water transitions selected by experts.[4–7] Since 2009, W@DIS was supplemented by complete sets of spectral data for other atmospheric molecules, including information about twenty molecular species and their isotopologues. Notably, some isotopologues (for instance, water isotopologues HTO, TDO, T_2O, and $T_2{}^{18}O$) are not found in commonly employed expert datasets.

Table 1. The number of publications and data sources for the water molecule within W@DIS (VQN and NM stand for the vibrational quantum number and normal mode, respectively). Tasks T1-T7 denote different spectroscopic tasks[3] (for example, T5 is an inverse task whose solution results in line profile characteristics).

Molecule	T1		T2		T3		T5	T6	T7	Publications	
	VQN	NM	VQN	NM	VQN	NM	NM	NM	NM	T1,T2,T6,T7	T3,T5
H_2O	8	8	4	11	6	26	317	165	34	257	358
$H^{18}OH$	3	4	4	8	4	19	62	82	18	97	15
$H^{17}OH$	3	3	4	6	4	8	34	60	19	52	5
HOD	1	2	2	7	3	9	19	113	32	106	20
$H^{17}OD$			2	3	2	3	14	13	3	13	0
$H^{18}OD$	-	1	2	4	2	5	17	31	9	24	0
D_2O	-	2	3	4	3	5	9	41	19	89	5
$D_2{}^{18}O$	-	1	2	3	2	4	3	7	5	17	1
$D_2{}^{17}O$			2	2	2	3	2	2	1	5	0
HTO			-	1	-	1		2		9	1
Total	15	21	25	49	28	83	477	516	140	669	405

3. EXPORT AND IMPORT OF DATA SOURCES

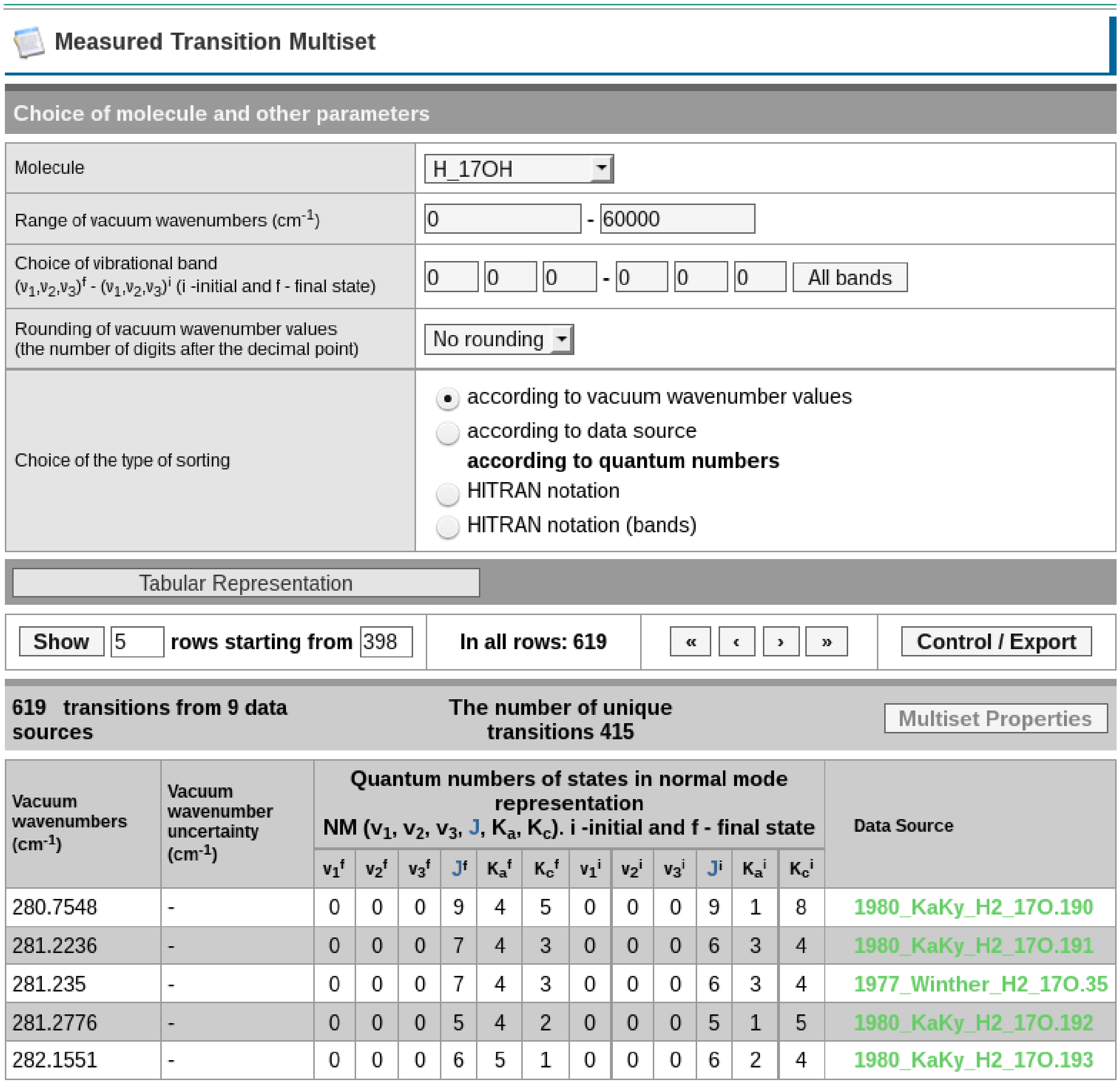
The work devoted to spectral data of the isotopologues of the water molecule was performed in three steps. First, publications produced between 1929–2015 containing observed spectroscopic data about the water molecule were collected and systematized, and a notation for quantum numbers was selected to build a W@DIS database. Second, the data were imported into W@DIS, and individual and relative criteria for assessing the validity of the data sources of interest were chosen. In addition, applications for generation of a list of characteristics of the measured transitions, collection of individual and relative properties of data sources were developed and made available to researchers (in addition, transitions that fail to satisfy selection rules or different types of binary relations between the data sources). Third, investigations were performed to create computer-generated ontologies for information resources, states, and transitions. Associated with the imported data sources are the properties of the data sources obtained with the use of a special metadata set. The metadata are intended to solve the task of search for information resources in accordance with a number of criteria. A key criterion is the validity of collected values of the physical quantities involved.

In W@DIS, data sources are parts of publications containing data about the solution of one of the above-mentioned spectroscopic tasks.[3] Data source import and export are detailed in a publication of Fazliev and co-workers.[17] The data distribution statistics for water isotopologues and spectroscopic tasks to be solved are presented in Table 1. More than 1300 data sources contain energy levels, transitions, and line profile data for the water isootpologues.

4. A LIST OF SPECTRAL WATER LINES

A special tool is used to create a list of rovibrational water lines containing measured values of transitions available in W@DIS. Figure 1 shows an interface to this tool, which allows for the generation of both a complete list of spectral lines and of its individual parts. For example, all transitions are listed which can be found in a given spectral range or in a certain vibrational band. The representation of the values of quantum numbers can be changed by specifying the number of digits in the menu item entitled “Rounding of vacuum wavenumbers”. It is also possible to choose the type of sorting of the rows in the list of lines according to the values of wavenumbers, name of the data source, and the values of the quantum numbers.

Further information about the number of transitions and data sources in the generated list of spectral lines is available at the top of the relevant table (see Fig.1). A sample list of the spectral lines for the fundamental band of the $H_2{}^{17}O$ molecule is given in Fig. 1. The fragment contains 619 transitions found in 9 data sources, 415 of which are unique. As to the list of spectral lines, consistency of the values of the physical quantities retrieved from different data sources is of great importance. A more detailed analysis of the inconsistency between the data sources is presented in Section 7.



Measured Transition Multiset

Choice of molecule and other parameters

Molecule	H_17OH
Range of vacuum wavenumbers (cm^{-1})	0 - 60000
Choice of vibrational band $(v_1,v_2,v_3)^f$ - $(v_1,v_2,v_3)^i$ (i -initial and f - final state)	0 0 0 - 0 0 0 All bands
Rounding of vacuum wavenumber values (the number of digits after the decimal point)	No rounding
Choice of the type of sorting	● according to vacuum wavenumber values ○ according to data source **according to quantum numbers** ○ HITRAN notation ○ HITRAN notation (bands)

Show 5 **rows starting from** 398 **In all rows: 619**

619 transitions from 9 data sources **The number of unique transitions 415**

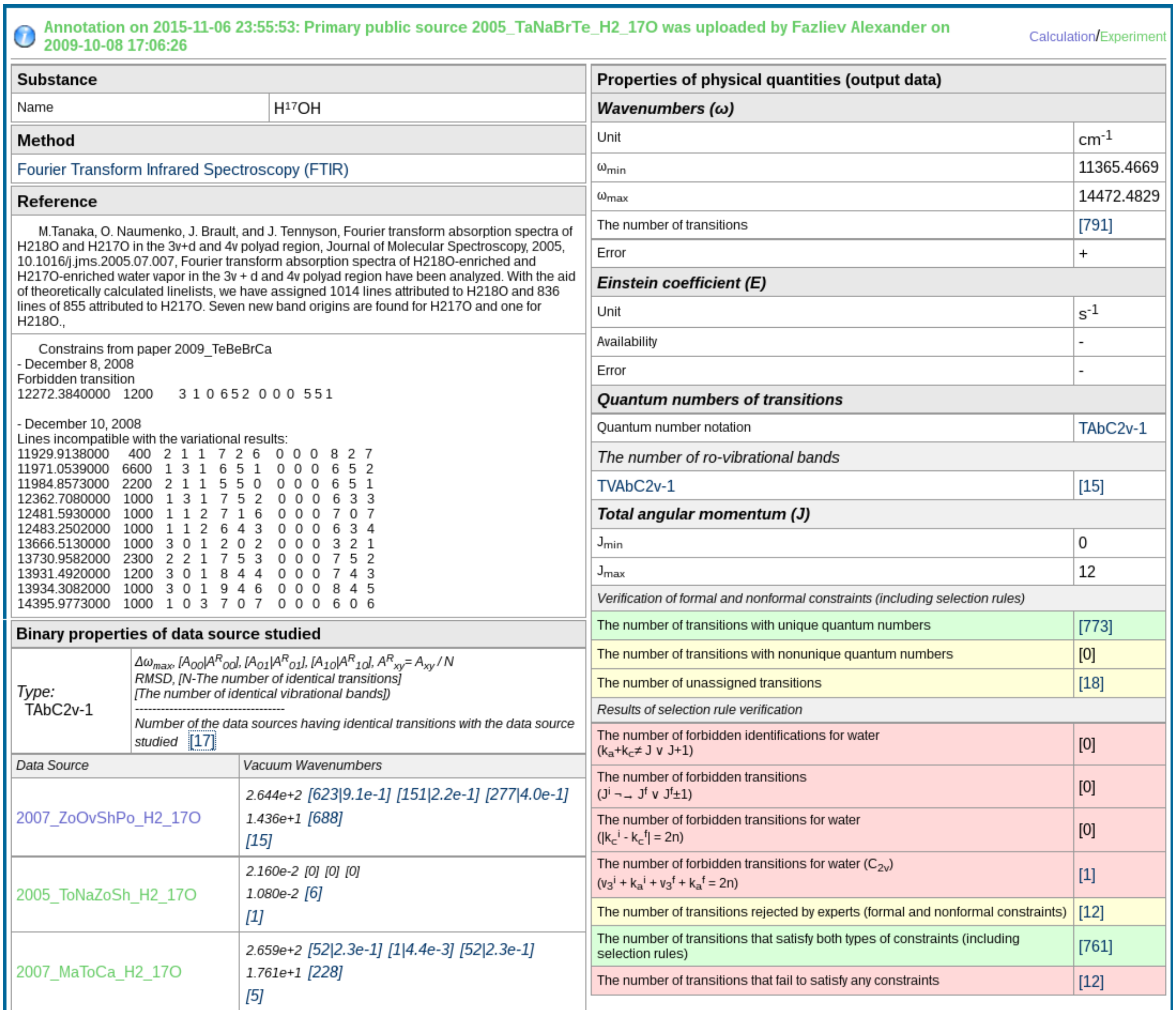
Vacuum wavenumbers (cm^{-1})	Vacuum wavenumber uncertainty (cm^{-1})	Quantum numbers of states in normal mode representation NM (v_1, v_2, v_3, J, K_a, K_c). i -initial and f - final state												Data Source
		v_1^f	v_2^f	v_3^f	J^f	K_a^f	K_c^f	v_1^i	v_2^i	v_3^i	J^i	K_a^i	K_c^i	
280.7548	-	0	0	0	9	4	5	0	0	0	9	1	8	1980_KaKy_H2_17O.190
281.2236	-	0	0	0	7	4	3	0	0	0	6	3	4	1980_KaKy_H2_17O.191
281.235	-	0	0	0	7	4	3	0	0	0	6	3	4	1977_Winther_H2_17O.35
281.2776	-	0	0	0	5	4	2	0	0	0	5	1	5	1980_KaKy_H2_17O.192
282.1551	-	0	0	0	6	5	1	0	0	0	6	2	4	1980_KaKy_H2_17O.193

Figure 1. Representation of a list of spectral lines, using $H_2{}^{17}O$ as an example.

5. A COMPUTED DESCRIPTION OF DATA SOURCES. AN INFORMATION SOURCE

During processing of computer data, the information stored in databases can be represented with different structures. In W@DIS, use is made of two groups of data representations. One group describes data sources as parts of publications and imported in the W@DIS database. The other group represents particular transitions, states, vibrational states and ro-vibrational bands of the molecule. Each data source of the second group is only associated with a set of identical transitions (states, …) (ones that have identical values of quantum numbers) and represented in W@DIS as individual of relevant OWL-ontology of transitions (see, http://wadis.saga.iao.ru/saga2/ontology/). The properties of a data source (resources, transitions, and states) are an annotation or an information source describing states and transitions. The properties are chosen so that their analysis allows for the assessment of the quality of the data source. In W@DIS, the values of the data properties are computed automatically. The data source and its properties, as well as information objects describing states and transitions and their properties, comprise the information source. For the end user, the properties retrieved from an information source are represented in W@DIS in tabular form (see Fig. 2).

An information source comprises two groups of properties of a data source, the notation used is explained elsewhere.[4] The individual properties include information about the molecule and the measurement technique, a reference to the pertinent publication (e.g., 2005_TaNaBrTe_H2_17O[20]), a list of transitions rejected by experts followed by a reference to the publication where the transitions are mentioned, the range of the wavenumbers and the total rotational angular momentum, the number of vibrational-rotational bands (and the reference to the list of the bands), a notation for quantum numbers, the number of non-unique transitions, the number of transitions without quantum numbers, the selection rules (see row $v_3{}^f + k_a{}^f + v_3{}^i + k_a{}^i = 2n$), etc. The relative properties are summarized in a subtable entitled "Binary properties of a data source". The quantities $\Delta\omega_{max}$, RMSD, and A_{00}, A_{10}, and A_{01} are described in Section 7.



Annotation on 2015-11-06 23:55:53: Primary public source 2005_TaNaBrTe_H2_17O was uploaded by Fazliev Alexander on 2009-10-08 17:06:26

Substance	
Name	$H^{17}OH$

Method

Fourier Transform Infrared Spectroscopy (FTIR)

Reference

M.Tanaka, O. Naumenko, J. Brault, and J. Tennyson, Fourier transform absorption spectra of H218O and H217O in the 3v+d and 4v polyad region, Journal of Molecular Spectroscopy, 2005, 10.1016/j.jms.2005.07.007, Fourier transform absorption spectra of H218O-enriched and H217O-enriched water vapor in the 3v + d and 4v polyad region have been analyzed. With the aid of theoretically calculated linelists, we have assigned 1014 lines attributed to H218O and 836 lines of 855 attributed to H217O. Seven new band origins are found for H217O and one for H218O.,

Constrains from paper 2009_TeBeBrCa
- December 8, 2008
Forbidden transition

```
12272.3840000   1200     3 1 0 6 5 2  0 0 0  5 5 1
```

- December 10, 2008
Lines incompatible with the variational results:

```
11929.9138000    400   2  1  1   7  2  6   0  0  0   8  2  7
11971.0539000   6600   1  3  1   6  5  1   0  0  0   6  5  2
11984.8573000   2200   2  1  1   5  5  0   0  0  0   6  5  1
12362.7080000   1000   1  3  1   7  5  2   0  0  0   6  3  3
12481.5930000   1000   1  1  2   7  1  6   0  0  0   7  0  7
12483.2502000   1000   1  1  2   6  4  3   0  0  0   6  3  4
13666.5130000   1000   3  0  1   2  0  2   0  0  0   3  2  1
13730.9582000   2300   2  2  1   7  5  3   0  0  0   7  5  2
13931.4920000   1200   3  0  1   8  4  4   0  0  0   7  4  3
13934.3082000   1000   3  0  1   9  4  6   0  0  0   8  4  5
14395.9773000   1000   1  0  3   7  0  7   0  0  0   6  0  6
```

Binary properties of data source studied

Type: TAbC2v-1	*$\Delta\omega_{max}$, $[A_{00}\|A^R{}_{00}]$, $[A_{01}\|A^R{}_{01}]$, $[A_{10}\|A^R{}_{10}]$, $A^R{}_{xy} = A_{xy} / N$* *RMSD, [N-The number of identical transitions]* *[The number of identical vibrational bands])* ----------------------------------- *Number of the data sources having identical transitions with the data source studied* [17]

Data Source	*Vacuum Wavenumbers*
2007_ZoOvShPo_H2_17O	*2.644e+2 [623|9.1e-1] [151|2.2e-1] [277|4.0e-1]* *1.436e+1 [688]* *[15]*
2005_ToNaZoSh_H2_17O	*2.160e-2 [0] [0] [0]* *1.080e-2 [6]* *[1]*
2007_MaToCa_H2_17O	*2.659e+2 [52|2.3e-1] [1|4.4e-3] [52|2.3e-1]* *1.761e+1 [228]* *[5]*

Properties of physical quantities (output data)	
Wavenumbers (ω)	
Unit	cm^{-1}
ω_{min}	11365.4669
ω_{max}	14472.4829
The number of transitions	[791]
Error	+
Einstein coefficient (E)	
Unit	s^{-1}
Availability	-
Error	-
Quantum numbers of transitions	
Quantum number notation	TAbC2v-1
The number of ro-vibrational bands	
TVAbC2v-1	[15]
Total angular momentum (J)	
J_{min}	0
J_{max}	12
Verification of formal and nonformal constraints (including selection rules)	
The number of transitions with unique quantum numbers	[773]
The number of transitions with nonunique quantum numbers	[0]
The number of unassigned transitions	[18]
Results of selection rule verification	
The number of forbidden identifications for water ($k_a+k_c \neq J$ v $J+1$)	[0]
The number of forbidden transitions (J^i ¬→ J^f v $J^f \pm 1$)	[0]
The number of forbidden transitions for water ($\|k_c{}^i - k_c{}^f\| = 2n$)	[0]
The number of forbidden transitions for water (C_{2v}) ($v_3{}^i + k_a{}^i + v_3{}^f + k_a{}^f = 2n$)	[1]
The number of transitions rejected by experts (formal and nonformal constraints)	[12]
The number of transitions that satisfy both types of constraints (including selection rules)	[761]
The number of transitions that fail to satisfy any constraints	[12]

Figure 2. Tabular representation of the properties of the primary data source 2005_TaNaBrTe.[18]

6. THE SYSTEMATIZATION OF INFORMATION SOURCES, MOLECULAR STATES, AND TRANSITIONS

Many research teams have been dealing with the compilation of spectral data. The amount of data in the collections is increasing at an alarming rate, to say nothing of the growing body of information about the properties of the data found in these collections. Users choose one or the other collection, or parts thereof, according to a set of properties of this collection. This is why the data collections must have an acceptable set of data properties that allow for an automatic comparison of collections, thus assisting in making an optimum choice of the collection by the end user. In W@DIS, the generation of reports is automated by means of ontologies. The sources of data about publications,[17] transitions, and states[21] that play a key role in the reports are presented. The structure of the sources of information about a data source representing data from publication[17] and a transition[21] are presented in Figs. 3 and 4.

The OWL-ontologies of information sources, molecular transitions, and states can be found at http://wadis.saga.iao.ru/saga2/ontology/.

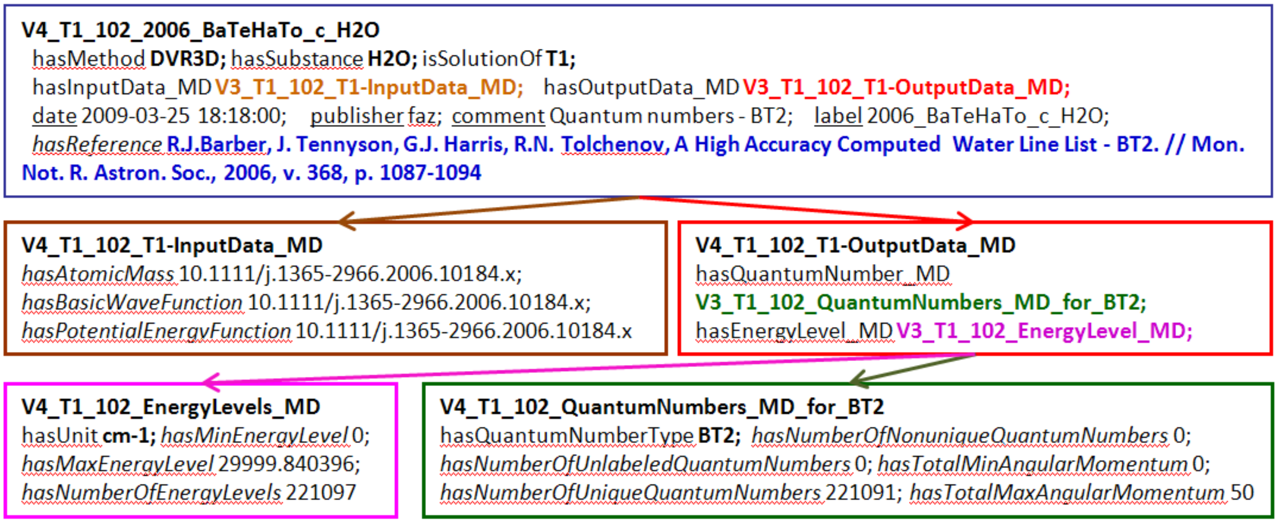


Figure 3. Structure of an information source describing a data source.

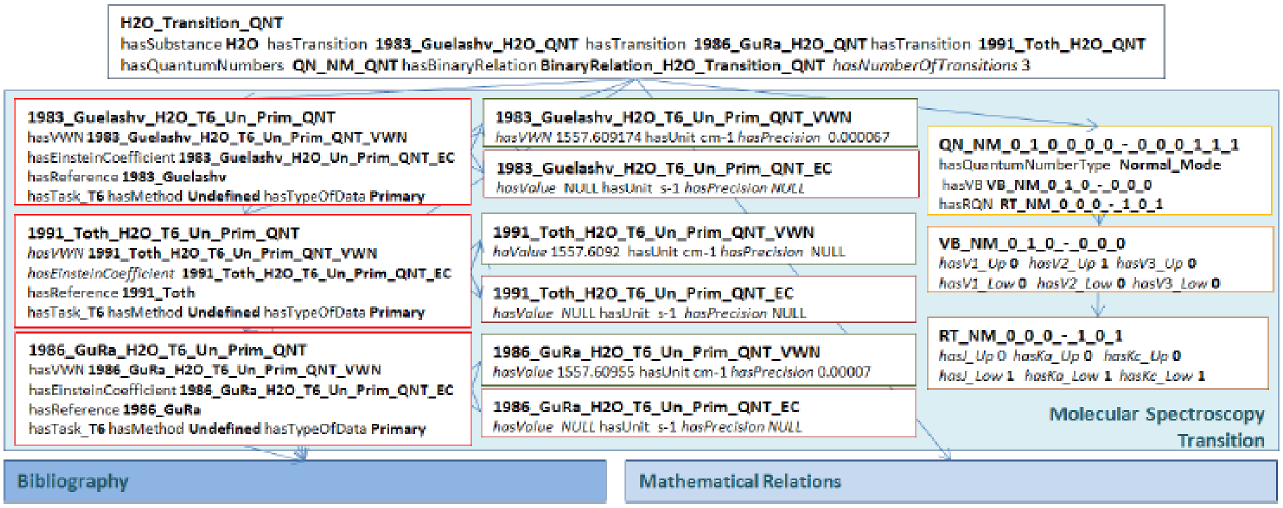


Figure 4. Structure of an information source describing a molecular transition.

7. AN ANALYSIS OF SPECTRAL DATA QUALITY

In the present work, the analysis of spectral data quality is focused on the validation and assessment of the trust in the data under consideration. This task is divided into four subtasks: (a) creation of lists of spectral lines; (b) provision of vacuum wavenumber consistency; (c) formulation of quantitative constraints on the publication criterion in different ranges of change of the physical quantities involved; and (d) decomposition of expert data.

In our context the validity of data means their compliance with formal constraints (criteria) obtained from molecular models and conditions for consistency of identical parts of data sources. Trust in expert data is assessed in the cases where it is possible to use informal criteria. We have formulated a publication criterion,[16,17] an informal criterion, for the qualitative analysis of trust in expert data, which enables untrusted data to be detected. The questionable data can be unpublished data or those exhibiting inconsistency with primary data. Investigators engaged in applied subject domains making use of expert spectral data have to check the untrusted data for compliance with criteria accepted in associated subject domains.

7.1. The validity of data

Solving the problem of fitting global data, computing reference energy levels, and constructing expert databases generates a need to select valid energy levels and wavenumbers from a list of measured spectral lines. Let us consider two groups of criteria that define the validity of spectral data. One group includes constraints obtained from a selection of mathematical and physical models for the water molecule. The other group of criteria is connected with constraints resulting from procedures used to provide consistency between the measured and calculated characteristics of identical molecular states and transitions.

Among the constraints originating from a mathematical molecular model are restrictions imposed on quantum numbers of states and transitions. The constraints are validated in W@DIS during data import and generation of an information source. Other constraints are requirements for the absence of duplicate states and transitions and of the states and transitions described by an incomplete set of quantum numbers.

7.1.1. Binary relations in quantitative spectroscopy

Another group of criteria describing the consistency of data sources includes the following:[22] (a) when a pairwise comparison of data sources is performed, the maximum difference between vacuum wavenumbers must be smaller than a certain value $\mathit{\Delta}$ assigned by the researcher, (b) the RMSDs of vacuum wavenumbers of two data sources must be smaller than a certain value δ specified by the researcher, and (c) the disordering factors A_{00}, A_{01}, and A_{10} must approach zero.

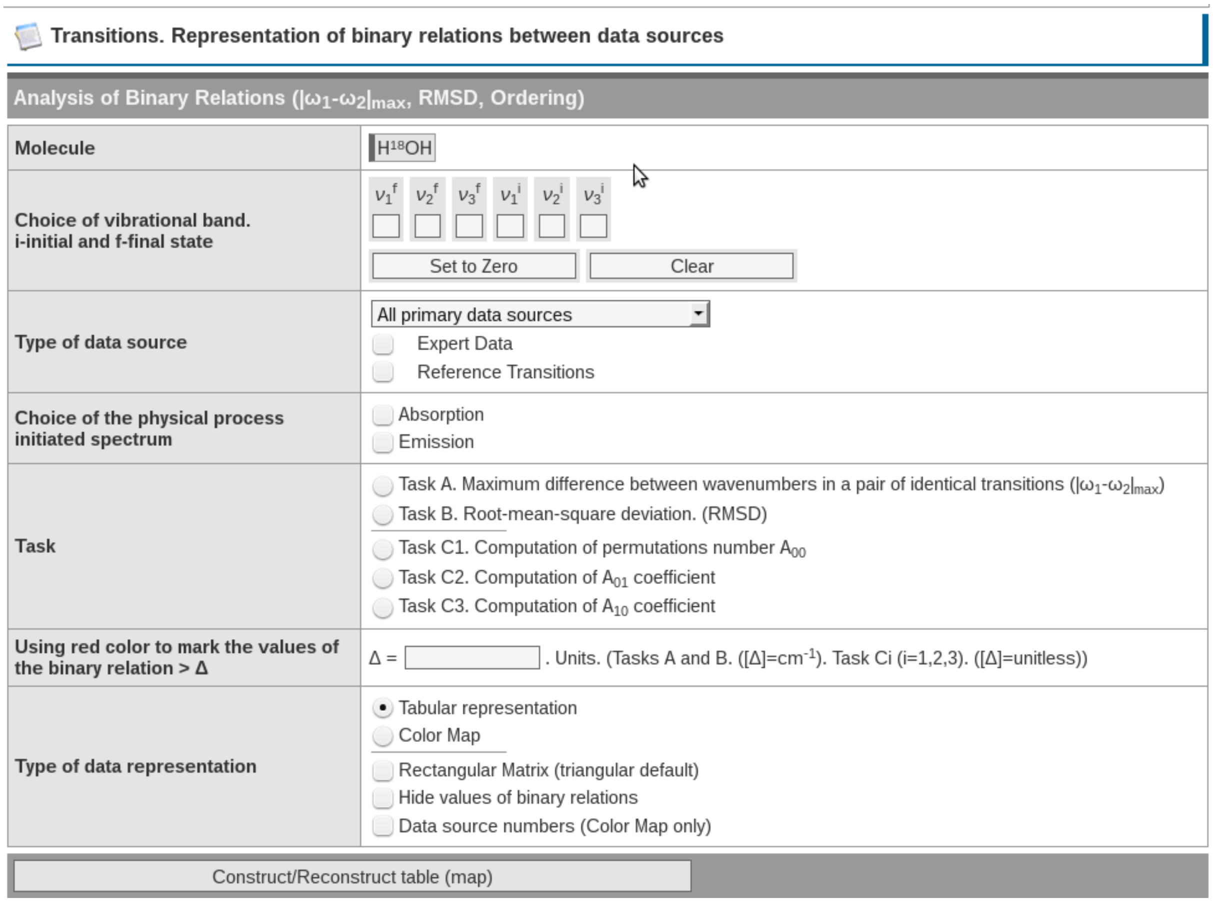
7.1.2. A comparison of ordered data sources

When analyzing the solutions of spectroscopic tasks, it is possible to derive the required smallness of the difference between the physical quantities involved in identical transitions and the desired RMS deviations. This measure is not sufficient to provide consistent solutions of the spectroscopic tasks. In a number of physical and chemical subject domains, physical quantities describe the states and transitions of complex systems. The set of values of each of the physical quantities characterizing a state (transition) can be ordered. In doing so, the sequence of values of a particular quantity can be assigned to a sequence of names of the states or transitions (in practice, the names are the quantum numbers). In the case where the measured values of a particular quantity for a pair of identical states are different, the sequence of states in the two sets of identical states may change. A situation such as this is referred to as the disordering effect (the difference between the sequence of transitions or states in a pair of the data sources being compared). The difference is indicative of the inconsistency of the data. For example, a comparison of two expert datasets for water molecule has revealed that the order is disrupted for more than 700 transitions in the water molecule.

A quantitative estimation of the degree of disorder of the quantum numbers in a pair of data sources can be performed in several different ways. Let there be two sets A_1 (a1i and qi) and A_2 (a2i and qi), containing values of the physical quantities ((ai), vacuum wavenumbers, spectral line intensities, etc.) and associated quantum numbers (qi). Furthermore, it is assumed that to the set of quantum numbers in set A_1 there corresponds an identical set of quantum numbers in set A_2. Let us arrange the sequences of values of the wavenumbers in an ascending order. The number of the each definite value of wavenumber in ordered sequence is assigned to the relevant quantum numbers. We have defined and subsequently calculated three disordering factors. Disordering factor A_{00} is the minimum number of permutations of transitions in one of the data sources required for the sequence of the quantum numbers in one source to be the same as that found in the other data source. Let us suppose that the quantum numbers in set A_1 are arranged in a proper order. Then disordering factor A_{01} gives the number of pairs (a2i, qi) to be deleted from set A_2 so that the remaining identical pairs in the two sets are arranged in the same order. Assuming that the quantum numbers in set A_2 are arranged in a proper order, disordering factor A_{10} determines the number of pairs **(**ai, qi) to be deleted from set A_1 so that the remaining identical pairs in the two sets are arranged in the same order.

7.1.3. Tabular representations of the binary relations between data sources

Figures 5–7 show interfaces for the representation of the results obtained from an analysis of the binary relations between data sources. An interface consists of three parts: (a) an interface for selecting the molecule of interest, type of data sources, form of representation, and task associated with the binary relation (Fig. 5); (b) a table containing the values of the binary relation for the task chosen and the number of identical transitions (Fig. 6) for all data sources chosen by the user, and (c) an interface for the representation of detailed information about the binary relations between a particular pair of data sources (Fig. 7).



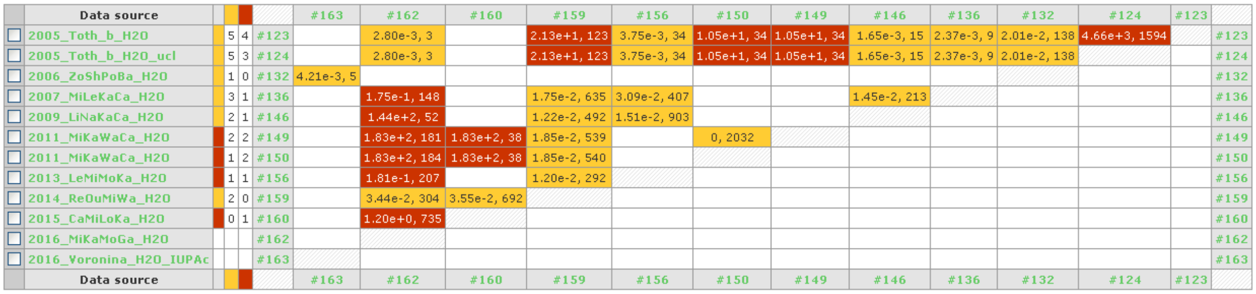
Transitions. Representation of binary relations between data sources

Analysis of Binary Relations ($|\omega_1-\omega_2|_{max}$, RMSD, Ordering)

Molecule	$H^{18}OH$
Choice of vibrational band. i-initial and f-final state	ν_1^f ☐ ν_2^f ☐ ν_3^f ☐ ν_1^i ☐ ν_2^i ☐ ν_3^i ☐ Set to Zero Clear
Type of data source	All primary data sources ☐ Expert Data ☐ Reference Transitions
Choice of the physical process initiated spectrum	☐ Absorption ☐ Emission
Task	○ Task A. Maximum difference between wavenumbers in a pair of identical transitions ($|\omega_1-\omega_2|_{max}$) ○ Task B. Root-mean-square deviation. (RMSD) ○ Task C1. Computation of permutations number A_{00} ○ Task C2. Computation of A_{01} coefficient ○ Task C3. Computation of A_{10} coefficient
Using red color to mark the values of the binary relation > Δ	Δ = ☐ . Units. (Tasks A and B. ([Δ]=cm^{-1}). Task Ci (i=1,2,3). ([Δ]=unitless))
Type of data representation	● Tabular representation ○ Color Map ☐ Rectangular Matrix (triangular default) ☐ Hide values of binary relations ☐ Data source numbers (Color Map only)

Construct/Reconstruct table (map)

Figure 5. Interface for selecting the molecule of interest and associated binary relations between a pair of data sources. A fragment of the table that lists the RMSDs for pairs of data sources is shown in Fig. 6.



	Data source		🟨	🟥	▨	#163	#162	#160	#159	#156	#150	#149	#146	#136	#132	#124	#123	▨
☐	2005_Toth_b_H2O	🟨	5	4	#123		2.80e-3, 3		2.13e+1, 123	3.75e-3, 34	1.05e+1, 34	1.05e+1, 34	1.65e-3, 15	2.37e-3, 9	2.01e-2, 138	4.66e+3, 1594	▨	#123
☐	2005_Toth_b_H2O_ucl	🟨	5	3	#124		2.80e-3, 3		2.13e+1, 123	3.75e-3, 34	1.05e+1, 34	1.05e+1, 34	1.65e-3, 15	2.37e-3, 9	2.01e-2, 138	▨		#124
☐	2006_ZoShPoBa_H2O	🟨	1	0	#132	4.21e-3, 5									▨			#132
☐	2007_MiLeKaCa_H2O	🟨	3	1	#136		1.75e-1, 148		1.75e-2, 635	3.09e-2, 407			1.45e-2, 213	▨				#136
☐	2009_LiNaKaCa_H2O	🟨	2	1	#146		1.44e+2, 52		1.22e-2, 492	1.51e-2, 903			▨					#146
☐	2011_MiKaWaCa_H2O	🟥	2	2	#149		1.83e+2, 181	1.83e+2, 38	1.85e-2, 539		0, 2032	▨						#149
☐	2011_MiKaWaCa_H2O	🟥	1	2	#150		1.83e+2, 184	1.83e+2, 38	1.85e-2, 540		▨							#150
☐	2013_LeMiMoKa_H2O	🟥	1	1	#156		1.81e-1, 207		1.20e-2, 292	▨								#156
☐	2014_ReOuMiWa_H2O	🟨	2	0	#159		3.44e-2, 304	3.55e-2, 692	▨									#159
☐	2015_CaMiLoKa_H2O	🟥	0	1	#160		1.20e+0, 735	▨										#160
☐	2016_MiKaMoGa_H2O				#162		▨											#162
☐	2016_Voronina_H2O_IUPAc				#163	▨												#163
	Data source		🟨	🟥	▨	#163	#162	#160	#159	#156	#150	#149	#146	#136	#132	#124	#123	▨

Figure 6. Fragment of the table summarizing the RMSDs for all pairs of data sources relating to the main isotopologue of the water molecule.

Transitions. Representation of binary relations between data sources

2014_ReOuMiWa_H2_18O Primary source

L. Regalia, C. Oudot, S. Mikhailenko, L. Wang, X. Thomas, A. Jenouvrier, P. Von der Heyden,
Water vapor line parameters from 6450 to 9400 cm^{-1},
Journal of Quantitative Spectroscopy and Radiative Transfer, 2014, Volume 136, Pages 119–136,
DOI: 10.1016/j.jqsrt.2013.11.019, http://dx.doi.org/10.1016/j.jqsrt.2013.11.019.
Annotation

2004_MaRoMiNa_H2_18O Primary source

Show Commentaries

Macko P., Romanini D., Mikhailenko S. N., Naumenko O.V., Kassi S., Jenouvrier A., Tyuterev Vl. G., Campargue A.,
High Sensitivity cw-Cavity Ring Down Spectroscopy of water in the region of the 1.5 µm atmospheric window,
Journal of Molecular Spectroscopy, 2004, Volume 227, Issue 1, Pages 90-108,
DOI: 10.1016/j.jms.2004.05.020.
Annotation

Comparison of the Data Sources

The Number of Transitions in Identical Ro-Vibrational Bands	$\Delta\omega_{max}$	Vacuum Wavenumbers RMSD	A_{00}	A_{01}	A_{10}
87	2.73e+0	2.92e-1	4	4	1

Comparison of the Data Sources on Separated Ro-Vibrational Bands

v_1^f	v_2^f	v_3^f	v_1^i	v_2^i	v_3^i	The Number of Transitions in Identical Ro-Vibrational Bands	$\Delta\omega_{max}$	Vacuum Wavenumbers RMSD	A_{01}	A_{10}
1	2	0	0	0	0	38	2.73e+0	4.43e-1	3	1
0	2	1	0	0	0	45	5.80e-3	1.70e-3	1	0
1	0	1	0	0	0	2	3.00e-3	2.20e-3	0	0
2	0	0	0	0	0	1	1.60e-3	1.60e-3	0	0
0	4	0	0	0	0	1	3.00e-4	3.00e-4	0	0

Figure 7. Tabular and graphical representations of the binary relations between data sources 2014_ReOuMiWa[23] and 2004_MaRoMiNa.[24]

7.1.4. Graphical representation of binary relations between all data sources

The tabular representation used for an analysis of the binary relations between data sources has one drawback. For a large number of data sources, for example, for the water molecule, the table in Fig. 6 is cumbersome, and it would take a lot of time to look it up. The help comes from a graphical representation of the results obtained from a quality analysis of the data. A plot can be constructed using an ontology of information resources to be employed in molecular water spectroscopy. We have used a dedicated software (see Refs.[25, 26]) for a graphical representation of the binary relations contained in the ontology code. Figure 8 is a plot whose edges represent values of the binary relations, while the vertices correspond to the primary measured and calculated data sources. The area of a circle representing a data source is proportional to the number of transitions it contains. To the large circles lying in the circumference of a larger diameter there correspond expert data sources. The data sources containing measurement results are in the internal circumference.

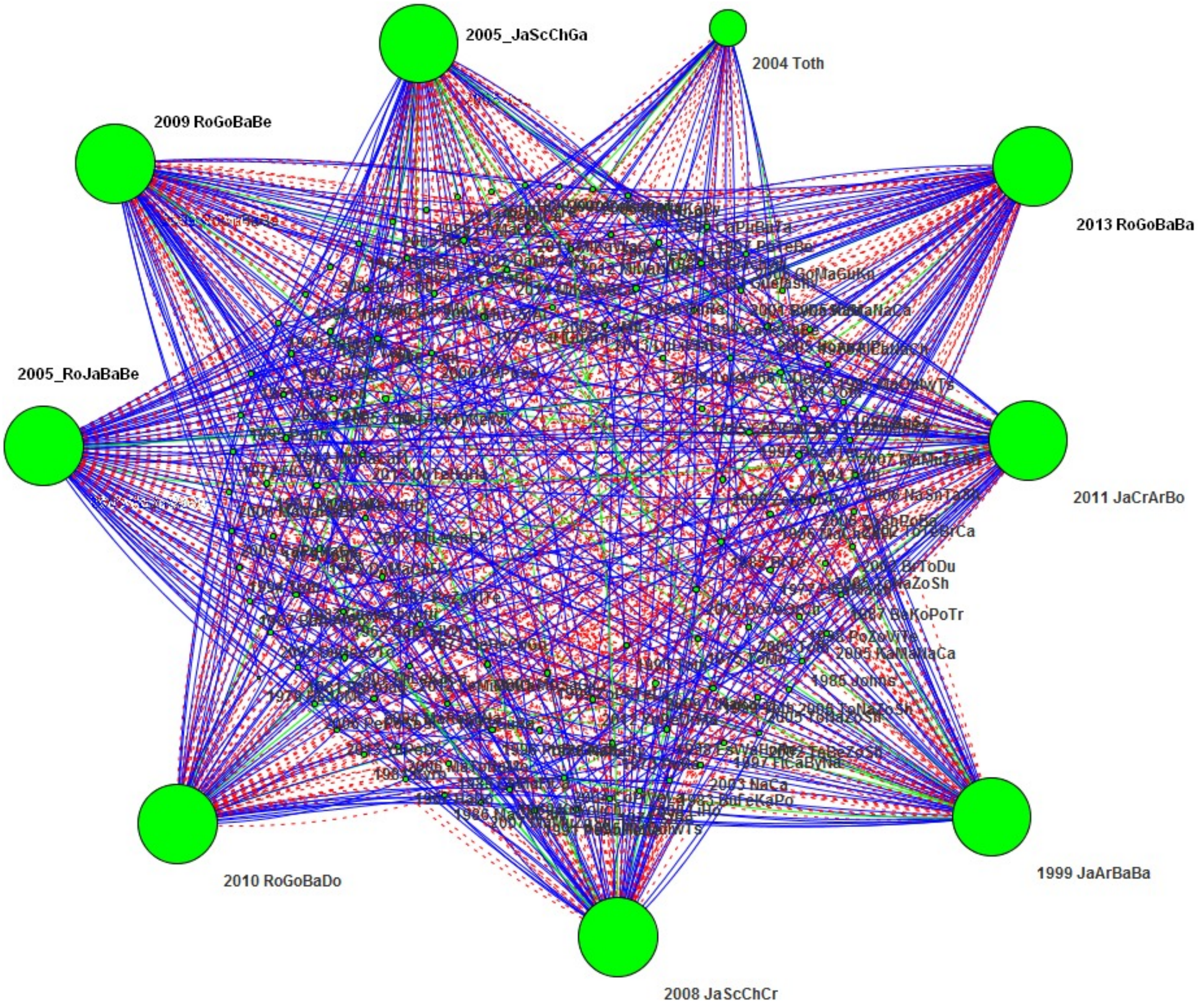


Figure 8. Plots describing the binary relations between pairs of primary and expert data sources for the water molecule.

The plot shows the maximum calculated difference between the wavenumbers of identical transitions (ΔF_i) and the RMSDs between the data sources containing identical transitions. The criteria for the consistency of data are of the form $\Delta F_i < 0.04$ cm^{-1} and the RMSDs are <0.1 cm^{-1}. The blue arcs are indicative of the fact that the data agree according to both of the criteria. The green arcs indicate the fulfillment of inequality $\Delta F_i > 0.04$ cm^{-1} and the RMSDs (A_1 and A_2)

are <0.1 cm^{-1}. The red arcs correspond to ΔF_i > 0.04 cm^{-1} and the RMSDs are >0.1 cm^{-1}. To the expert data sources shown in Fig. 8 there correspond nine green circles of larger diameter located closer to the edges of the figure.

8. SUMMARY

In the framework of an IUPAC-sponsored project,[4-8] a critical analysis of the wavenumbers of nine main water isotopologues was performed using a complete set of measured data about vacuum transitions. A great deal of work was devoted to the quality assurance of the energy levels and transitions. The resulting set of published data contained only consistent wavenumbers. In addition, a set of published transitions rejected by experts was presented, and the reasons why part of the data was declined by experts were given.

Subsequently, the W@DIS information system was built to provide data import and export, tabular and graphical representations, and quality analysis of spectral data sources for the water molecule. Over the years, W@DIS has become a public application for the systematization of published spectral data for a number of molecules, which made it possible to provide access for researchers (or programmable agents) dealing with spectroscopic data. The free access is not restricted to the data of interest but also involves data properties that allow for quality assessment of published data, states, and calculated and measured rovibrational transitions.

Development of the W@DIS information system was motivated by the need for storage and representation of primary data sources for the transitions involved in radiation transfer involving the water molecule and of the results obtained from a quality analysis of the transitions and reference energy levels computed by means of the MARVEL algorithm. The latter application relies on a consistent set of wavenumbers and has been applied to nine water isotopologues.

Published computed data for the isotopologues of the water molecule are also stored in W@DIS along with measured data. In addition, the W@DIS applications are employed for a critical evaluation of the spectral data for the hydrogen sulfide molecule, H_2S. The digital library used contains data sources (the data were acquired from the published literature on the subject) for more than fifty molecules. Matching the spectral data available in the published sources requires a great deal of joint efforts of different research groups.

ACKNOWLEDGEMENTS

The work was supported by Russian Foundation for Basic Research (Grant 13-07-00411) and by the COST Action CM1405, MOLIM: Molecules in Motion. AGC acknowledges the support of NKFIH under grant No. NK83583.

References

[1] Rothman, L.S., Gordon, I.E., Babikov, Y., Barbe, A., Benner, D. C., Bernath, P. F., Birk, M., Bizzocchi, L., Boudon, V., Brown, L. R., Campargue, A., Chance, K., Coudert, L., Devi, V.M., Drouin, B. J., Fayt, A., Flaud, J.-M., Gamache, R. R., Harrison, J., Hartmann, J.-M., Hill, C., Hodges, J. T., Jacquemart, D., Jolly, A., Lamouroux, J., LeRoy, R. J., Li, G., Longo, D., Mackie, C.J., Massie, S. T., Mikhailenko, S., Műller, H. S. P., Naumenko, O. V., Nikitin, A. V., Orphal, J., Perevalov, V., Perrin, A., Polovtseva, E. R., Richard, C., Smith, M. A. H., Starikova, E., Sung, K., Tashkun, S., Tennyson, J., Toon, G. C., Tyuterev, V. G., Auwera, J. V., Wagner, G. “The HITRAN 2012 Molecular Spectroscopic Database,” J. Quant. Spectrosc. Rad. Transfer. 130, 4–50 (2013).

[2] Jacquinet-Husson, N., Crepeau, L., Armante, R., Boutammine, C., Chédin, A., Scott, N.A., Crevoisier, C., Capelle, V., Boone, C., Poulet-Crovisier, N., Barbe, A., Campargue, A., Benner, C. D., Benilan, Y., Bézard, B., Boudon, V., Brown, L.R., Coudert, L.H., Coustenis, A., Dana, V., Devi, V. M., Fally, S., Fayt, A., Flaud, J.-M., Goldman, A., Herman, M., Harris, G. J., Jacquemart, D., Jolly, A., Kleiner, I., Kleinböhl, A., Kwabia-Tchana, F., Lavrentieva, N., Lacome, N., Xu L.-H., Lyulin, O. M., Mandin, J.-Y., Maki, A., Mikhailenko, S., Miller, C. E., Mishina, T., Moazzen-Ahmadi, N., Müller, H. S. P., Nikitin, A., Orphal, J., Perevalov, V., Perrin, A., Petkie, D. T., Predoi-Cross, A., Rinsland, C. P., Remedios, J.J., Rotger, M., Smith, M. A. H., Sung, K., Tennyson, J., Toth, R.A., Vandaele, A.-C., Auwera, J. V., “The 2009 edition of the GEISA spectroscopic database,” J. Quant. Spectrosc. Rad. Transfer. 112, 2395–2445 (2011).

[3] Bykov, A. D., Kozodoev, A. V., Privezentsev, A.I., Sinitsa, L.N., Tonkov, M. V., Filippov, N.N., Fazliev, A. Z., Tretyakov, M. Y., "Distributed information system on molecular spectroscopy," Proc. SPIE 6580, 65800W (2006).

[4] Tennyson, J., Bernath, P.F., Brown, L.R., Campargue, A., Carleer, M.R., Császár, A. G., Gamache, R. R., Hodges, J.T., Jenouvrier, A., Naumenko, O.V., Polyansky, O.L., Rothman, L.S., Toth, R. A., Vandaele, A. C., Zobov, N., Daumont, L., Fazliev, A.Z., Furtenbacher, T., Gordon, I. E., Mikhailenko, S. N., Shirin, S. V., "IUPAC Critical Evaluation of the Rotational-Vibrational Spectra of Water Vapor. Part I. Energy Levels and Transition Wavenumbers for $H_2{}^{17}O$ and $H_2{}^{18}O$," J. Quant. Spectrosc. Rad. Transfer. 110, 573–596 (2009).

[5] Tennyson, J., Bernath, P. F., Brown, L.R., Campargue, A., Császár, A. G., Daumont, L., Gamache, R. R., Hodges, J. T., Naumenko, O. V., Polyansky, O. L., Rothman, L. S., Toth, R. A., Vandaele, A. C., Zobov, N. F., Fally, S., Fazliev, A. Z., Furtenbacher, T., Gordon, I. E., Hu, S.-M., Mikhailenko, S. N., Voronin, B.A., "IUPAC critical evaluation of the rotational–vibrational spectra of water vapor. Part II: Energy levels and transition wavenumbers for $HD^{16}O$, $HD^{17}O$, and $HD^{18}O$," J. Quant. Spectrosc. Rad. Transfer. 111, 2160–2184 (2010).

[6] Tennyson, J., Bernath, P. F., Brown, L. R., Campargue, A., Császár, A. G., Daumont, L., Gamache, R. R., Hodges, J. T., Naumenko, O. V., Polyansky, O. L., Rothman, L.S., Vandaele, A. C., Zobov, N. F., Al Derzi, A. R., Fábri, C., Fazliev, A. Z., Furtenbacher, T., Gordon, I. E., Lodi, L., Mizus, I. I., "IUPAC critical evaluation of the rotational–vibrational spectra of water vapor, Part III: Energy levels and transition wavenumbers for $H_2{}^{16}O$," J. Quant. Spectrosc. Rad. Transfer. 117, 29–58 (2013).

[7] Tennyson, J., Bernath, P. F., Brown, L. R., Campargue, A., Császár, A. G., Daumont, L., Gamache, R. R., Hodges, J. T., Naumenko, O. V., Polyansky, O. L., Rothman, L. S., Vandaele, A. C., Zobov, N. F., Dénes, N., Fazliev, A. Z., Furtenbacher, T., Gordon, I. E., Hu, S.-M., Szidarovszky, T., Vasilenko, I. A., "IUPAC critical evaluation of the rotational–vibrational spectra of water vapor. Part IV. Energy levels and transition wavenumbers for $D_2{}^{16}O$, $D_2{}^{17}O$, and $D_2{}^{18}O$," J. Quant. Spectrosc. Rad. Transfer. 142, 93–108 (2014).

[8] Tennyson, J., Bernath, P. F., Brown, L. R., Campargue, A., Császár, A. G., Daumont, L., Gamache, R. R., Hodges, J. T., Naumenko, O. V., Polyansky, O. L., Rothman, L. S., Vandaele, A. C., Zobov, N. F., "A database of water transitions from experiment and theory," Pure Appl. Chem. 86, 71–83 (2014).

[9] Furtenbacher, T., Császár, A.G., "MARVEL: Measured active rotational–vibrational energy levels. II. Algorithmic improvements," J. Quant. Spectrosc. Rad. Transfer. 11, 929–935(2012).

[10] Furtenbacher, T., Csaszar, A. G., and Tennyson, J., "MARVEL: measured active rotational-vibrational energy levels," J. Mol. Spectrosc. 245, 115-125 (2007).

[11] Information system W@DIS, http://wadis.saga.iao.ru/

[12] Information system ReSpecTh, http://ReSpecTh.hu

[13] Dubernet, M. L., Boudon, V., Culhane, J. L., Dimitrijevic, M. S., Fazliev, A. Z., Joblin, C., Kupka, F., Leto, G., Le Sidaner, P., Loboda, P. A., Mason, H. E., Mason, N.J., Mendoza, C., Mulas, G., Millar, T. J., Nuñez, L.A., Perevalov, V. I., Piskunov, N., Ralchenko, Y., Rixon, G., Rothman, L.S., Roueff, E., Ryabchikova, T.A., Ryabtsev, A. S., Sahal-Brechot, Schmitt, B., Schlemmer, S., Tennyson, J., Tyuterev, V. G., Walton, N. A., Wakelam, V., Zeippen, C. J., "Virtual atomic and molecular data centre," J. Quant. Spectrosc. Rad. Transfer 111, 2151–2159 (2010).

[14] Dubernet, M. L., Antony, B. K., Ba, Y. A., Babikov, Yu. L., Bartschat, K., Boudon, V., Braams, B. J. Chung, H.-K., Daniel, F., Delahaye, F., Del Zanna, G., de Urquijo, J., Dimitrijević, M. S., Domaracka, A., Doronin, M., Drouin, B. J., Endres, C. P., Fazliev, A. Z., Gagarin, S. V., Gordon, I. E., Gratier, P., Heiter, U., Hill, C., Jevremović, D., Joblin, C., Kasprzak, A., Krishnakumar, E., Leto, G., Loboda, P. A., Louge, T., Maclot, S., Marinković, B. P., Markwick, A., Marquart, T., Mason, H. E., Mason, N. J., Mendoza, C., Mihajlov, A. A., Millar, T. J., Moreau, N., Mulas, G., Pakhomov, Yu., Palmeri, P., Pancheshnyi, S., Perevalov, V. I., Piskunov, N., Postler, J., Quinet, P., Quintas-Sánchez, E., Ralchenko, Yu., Rhee, Y.-J., Rixon, G., Rothman, L. S., Roueff, E., Ryabchikova, T., Sahal-Bréchot, S., Scheier, P., Schlemmer, S., Schmitt, B., Stempels, E., Tashkun, S., Tennyson, J., Tyuterev, V. L., Vujčić, V., Wakelam, V., Walton, N. A., Zatsarinny, O., Zeippen, C. J. and Zwölf, C. M., "The Virtual Atomic and Molecular Data Centre (VAMDC) Consortium for Astrophysics," J. Phys. B: At. Mol. Opt. Phys. 49, 074003 (2016)

[15] Akhlyostin, A. Yu., Kozodoev, A. V., Lavrentiev, N. A., Privezentsev, A., Fazliev, A. Z., "Computed knowledge base for description of information resources of molecular spectroscopy. 4. Software," Russ. Digit. Libr. J. 15 (3), (2012). http://elbib.ru/index.phtml?page=elbib/rus/journal/2012/part3/AKLPF (in Russian).

[16] Lavrentyev, N. A., Makogon, M. M., Fazliev, A. Z., "Comparison of the HITRAN and GEISA Spectral Databases Taking into Account the Restriction on Publication of Spectral Data," Atmos. Ocean. Optics 24, 436–451 (2011).

[17] Fazliev, A., Privezentsev, A., Tsarkov, D., Tennyson, J., "Ontology-Based Content Trust Support of Expert Information Resources in Quantitative Spectroscopy, In: Knowledge Engineering and the Semantic Web," Communications in Computer and Information Science, Springer Berlin Heidelberg, 394, 15–28 (2013).

[18] Császár, A. G., Fazliev, A. Z., Tennyson, J., "W@DIS – Prototype information system for systematization of spectral data of water," Abstracts the 20th Colloquium on High Resolution Molecular Spectroscopy, 270-271 (2007).

[19] Bykov, A., Naumenko, O., Sinitsa, L., Rodimova, O., Tonkov, M., Tvorogov, S., Fazliev, A., Filippov, N., "Information aspects of molecular spectroscopy," IAO Publshing House, Tomsk, 360 (2008).

[20] Tanaka, M., Naumenko, O., Brault, J., Tennyson, J. "Fourier transform absorption spectra of $H_2^{18}O$ and $H_2^{17}O$ in the 3ν+d and 4ν polyad region," J. Mol. Spectrosc. 34, 1–9 (2005).

[21] Voronina, S. S., Privezentsev, A. I., Tsarkov, D V., Fazliev, A. Z., "An Ontological Description of States and Transitions in Quantitative Spectroscopy," Proc. SPIE XX Int. Symp. Atmosph. Ocean. Optics: Atmosph. Phys. 9292, 92920C (2014).

[22] Akhlyostin, A. Yu., Lavrentiev, N. A., Privezentsev, A.I., Fazliev, A. Z., "Computed knowledge base for descrbing information resources in molecular spectroscopy. 5. Expert data quality," Russ. Digit. Libr. J. 16(4), (2013). http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2013/part4/AKLPF.

[23] Regalia, L., Oudot, C., Mikhailenko, S., Wang, L., Thomas, X., Jenouvrier, A., Von der Heyden, P., "Water vapor line parameters from 6450 to 9400 cm^{-1}," J. Quant. Spectrosc. Rad. Transfer. 136, 119–136 (2014).

[24] Macko, P., Romanini, D., Mikhailenko, S. N., Naumenko, O. V., Kassi, S., Jenouvrier, A., Tyuterev, Vl. G., Campargue, A., "High Sensitivity CW-Cavity Ring Down Spectroscopy of Water in the Region of the 1.5 μm atmospheric window," J. Mol. Spectr. 227, 90–108 (2004).

[25] Apanovich, Z. V., Vinokurov, P. S., Akhlyostin, A. Yu., Privezentsev, A. I., and Fazliev, A. Z., "Visualization of binary relations between data sources in quantitative spectroscopy," The 15th All Russ. Conf. Internet Modern Soc., 7–15 (2012).

[26] Apanovich, Z. V., Vinokurov, P. S., Kislitcina, T. A., "Visualization tools for large scientific portal content," J. Inform.Technol. Ser. Inform. Technol. 9, 5–14 (2011).