

А.Д. Быков, Б.А. Воронин, А.В. Козодоев, Н.А. Лаврентьев,
О.Б. Родимова, А.З. Фазлиев

Информационная система для решения задач молекулярной спектроскопии. 1. Структура информационных ресурсов

Институт оптики атмосферы СО РАН, г. Томск

Поступила в редакцию 15.09.2004 г.

Дано описание модели данных, использованной при создании информационной системы по молекулярной спектроскопии (<http://saga.atmos.iao.ru>). Выделены основные части модели данных, к числу которых относятся фундаментальные характеристики молекул, параметры спектральных линий и спектральные функции. Описана система ввода данных, применяемая пользователем для создания собственных наборов данных. На примере расчета коэффициента поглощения показаны связи между частями модели данных.

Введение

Молекулярная спектроскопия является одной из ключевых фундаментальных наук, используемых во многих прикладных исследованиях. Для атмосферных наук наиболее важными являются спектральные данные о десятках молекул. Стоит упомянуть, что только данные о молекуле воды содержат информацию о сотнях миллионов спектральных линий. Работа с такими объемами данных вне рамок их машинной обработки практически бессмысленна. Это обстоятельство делает необходимым использование автоматизированных информационных вычислительных систем (ИВС) в молекулярной спектроскопии.

Информационных систем в области молекулярной спектроскопии создано достаточно много, и большая часть из них использует в качестве данных параметры спектральных линий. В банках данных (БД) HITRAN [1] и GEISA [2] собрана информация о почти сорока молекулах. Они доступны в Интернете и распространяются на CD-дисках. Первая известная нам попытка организовать автоматизированную ИВС с прямым доступом в сети Интернет для просмотра параметров спектральных линий и расчетов спектральных функций была сделана в работах В. Головки и Ю. Бабилова [3–5]. Основой созданной ими ИВС SPECTRA (<http://spectra.iao.ru>) стали данные о параметрах спектральных линий из БД [1, 2] и оригинальные данные, полученные в ИОА СО РАН (<ftp://ftp.iao.ru/pub/CDS-296> и <ftp://ftp.iao.ru/pub/CDS-1000>).

В системах, доступных по сети Интернет при коллективной работе с информационными ресурсами, становится важным понимание модели данных, применяемой в ИВС. В нашей работе представлено описание модели данных, используемой при создании ИВС по молекулярной спектроскопии

(<http://saga.atmos.iao.ru>). Особенности данной ИВС являются расширение модели данных и предоставление пользователю возможности загрузки собственных данных в рамках представляемой модели данных.

По своему происхождению данные в спектроскопии делятся на два типа: экспериментальные и расчетные. Описания этих типов имеют принципиальные различия. При описании экспериментальных данных наряду со знанием условий эксперимента ключевым является знание параметров экспериментальной установки. В описании же расчетных данных ключевыми являются используемые методы расчета и употребляемые ими модели данных. В сложившейся практике создания ИВС для проведения расчетов с использованием параметров спектральных линий эти описания не используются. Обусловлено это, прежде всего, тем фактом, что машинная обработка таких описаний большинству исследователей в области спектроскопии пока не доступна: с одной стороны, из-за недостаточного понимания ими возможностей современных информационных технологий, а с другой — из-за увеличения затрачиваемых средств на проведение такого рода работ.

Далее в статье описываются структура информационных ресурсов, процедуры занесения и отображения данных в молекулярной спектроскопии. Стоит особо подчеркнуть, что используемый ниже термин «молекулярная спектроскопия» применяется не для описания спектроскопии как научной дисциплины во всем многообразии, а только для ее информационной модели. Информационная модель материализуется в форме информационных ресурсов [6], которые разделяются на данные и метаданные [7]. На примере вычисления коэффициента поглощения показывается уровень информационного сервиса, доступный пользователю ИВС.

Структура информационных ресурсов

Цикл работы состоит из нескольких стадий. Основными из них являются сбор, хранение, обработка и доставка данных пользователю. Сбор или ввод данных в ИВС осуществляются с помощью html-форм, доступных как администратору, так и пользователю. Материалы хранятся в базах данных. Наиболее трудоемкой при программной реализации этого цикла является обработка данных в силу того, что она требует детального знания их семантики. Доставка данных осуществляется по сети Интернет.

В использованной нами модели данных можно выделить три качественно разные структуры:

1. Фундаментальные характеристики молекулы.
2. Параметры спектральных линий.
3. Спектральные функции.

Порядок расположения структур в этой тройке важен при подготовке метаданных. Данные о строении молекул используются для расчета параметров спектральных линий, которые, в свою очередь, применяются для расчета спектральных функций. При расчете, например, коэффициента поглощения это означает, что при формировании метаданных строится иерархия, на нижнем уровне которой находятся метаданные для фундаментальных характеристик, на следующем уровне метаданные для параметров спектральных линий и, наконец, на последнем уровне метаданные для спектральных функций.

Фундаментальными молекулярными характеристиками обычно считаются параметры равновесной конфигурации, функция потенциальной энергии, функция дипольного момента или поляризуемости. В нашей модели данных под фундаментальными характеристиками молекул понимаются такие параметры молекул, которые позволяют рассчитать спектроскопические параметры линий – центры, интенсивности, полуширины, сдвиги давлением или параметры кросс-релаксации. Другими словами, – это характеристики, определяющие энергию молекулы. В зависимости от способа описания ими могут быть либо параметры полного молекулярного гамильтониана (потенциальная энергия, дипольный момент), либо параметры эффективного гамильтониана (вращательные, центробежные и резонансные постоянные, параметры эффективного дипольного момента). К ним, естественно, необходимо добавить квадрупольные, октупольные моменты молекул и другие параметры, характеризующие межмолекулярное взаимодействие в газах.

Использование параметров спектральных линий, перечень которых различен в разных БД, основано на предположении о применимости при рас-

четах гипотезы об изолированной линии. Эти данные ориентированы на применение в расчетах спектральных функций (коэффициентов поглощения, функции пропускания и т.д.).

Систематизация данных об экспериментальных значениях спектральных функций, в первую очередь, необходима для проведения сравнений их с рассчитанными значениями. Особое внимание при заполнении БД по спектральным функциям уделено экспериментам, выполненным при исследовании крыльев спектральных линий.

Как уже упоминалось выше, информационные ресурсы имеют две компоненты: данные и метаданные. Наиболее часто используется широкое определение метаданных: метаданные – это данные о данных. Оно не содержит ни критерия полноты метаданных для данных предметной области, ни указания на то, с позиции какой научной дисциплины рассматриваются метаданные. По этой причине метаданные, принятые нами для описания данных, определены ниже перечислением. К числу метаданных, имеющих, на наш взгляд, первостепенное значение в молекулярной спектроскопии, относятся данные об источнике информации для всех трех структур данных и ряд параметров, определяющих области применимости, величину ошибки, а также условия проведения экспериментов, методы расчетов и т.д. Их можно использовать для логического обоснования применимости получаемых результатов.

Данные различаются также по правам доступа к ним. В ИВС существуют два уровня доступа – административный и пользовательский. Интерфейс для ввода данных в ИВС для обоих уровней доступа одинаков. Различие только в том, что внесенные администратором данные становятся доступными всем пользователям, тогда как данные, внесенные пользователем, – только ему одному. На административном уровне осуществляются ввод данных и контроль их целостности, а на пользовательском – только ввод данных. В настоящее время можно вводить данные двух структур – параметры спектральных линий (минимальный набор – центры и интенсивности линий) и спектральные функции (коэффициент поглощения, функция пропускания и т.д.).

Для каждой структуры данных существуют собственные методы их обработки. Ниже описаны способы работы с параметрами спектральных линий и спектральными функциями.

Стадия обработки данных включает в себя технические действия (запросы к БД, графическое и табличное отображение) и предметную обработку. Техническую стадию мы опишем только для графического отображения. Предметная сторона обработки данных связана с вычислением коэффициентов поглощения смеси газов.

Отметим, что при работе в ИВС SPECTRA пользователи не могут вводить собственные данные о параметрах спектральных линий и проводить вычисления на основе этих данных. Единственная возможность для ввода данных, существующая в ИВС SPECTRA, ориентирована на ввод пользователем собственных данных в виде двумерного массива с целью сравнения на графике с результатами расчета.

Операции с информационными ресурсами

В этом разделе описана работа с данными, а именно процедура ввода данных в информационную систему и отображение данных для конечного пользователя. Рассмотрим две структуры: параметры спектральных линий и спектральные функции.

В соответствии с их физическим смыслом для каждой молекулы параметры спектральных линий разделены на несколько групп. К ним относятся:

- параметры изолированной спектральной линии (интенсивность, центр линии, энергия нижнего уровня, статистический вес верхнего и нижнего состояния, момент перехода и т.д.);
- параметры идентификации (колебательная и колебательно-вращательная идентификация);
- параметры, обусловленные столкновениями (полуширина, сдвиг давлением, температурная зависимость полуширины и т.д.).

Часть физических характеристик наделена атрибутами. В число атрибутов входят индикатор происхождения данных (эксперимент, расчет), тип шкалы (абсолютная, относительная) и единицы измерения физических величин.

Заполнение экспертами банков спектральных данных, как правило, осуществляется на основе множественности значений параметров линий. Частота обновления банков данных весьма изменчива от одного года до нескольких лет. Это обстоятельство создает определенные трудности в обновлении результатов расчетов спектральных функций. Решение этой проблемы находится на пути создания сервисов ИВС, позволяющих пользователю, с одной стороны, создавать в ней собственные информационные ресурсы, а с другой – включать в них ресурсы, имеющиеся в системе. В первом случае важна процедура ввода данных в ИВС, а во втором – процедура составления коллекций ресурсов.

Общим при вводе разных структур данных является создание источника данных. Описание источника включает в себя аббревиатуру, авторов, название статьи (монографии), название журнала

(включая том, номер, год и страницы), URL-адрес журнала и комментариев. Прежде чем занести данные, пользователь описывает источник данных или устанавливает связь между уже описанным в системе источником данных и загружаемыми данными.

Процедура ввода данных использует данные пользователя, подготовленные в виде файла, содержащего колонки данных. Порядок следования колонок произволен, но он синхронизируется с перечнем физических величин, составляемым пользователем при вводе данных. На первом этапе ввода пользователь создает источник данных. На втором этапе пользователь выбирает молекулу и интерактивно формирует перечень величин в соответствии с порядком колонок в файле. Обязательными являются колонки, содержащие центры и интенсивности спектральных линий. На заключительном этапе пользователь определяет значения атрибутов физических величин.

При вводе данных о коэффициенте поглощения большая часть параметров обязательна для ввода. К числу этих параметров относятся спектральные параметры (разрешение, длина пути), термодинамические параметры (температура и давление), взаимодействующие вещества (поглощающий и уширяющий газы), волновое число и коэффициент поглощения (единицы измерения) и способ получения коэффициента поглощения (эксперимент, расчет).

Просмотр диаграммы интенсивностей спектральных линий молекулы проводится в разделе «Молекула». Он организован в два этапа. На первом этапе выбирается молекула или изотопическая модификация молекулы из имеющегося перечня, задаются спектральный интервал, термодинамические условия и нижнее или верхнее значение интенсивности линий, по которому проводится выборка из базы данных. На втором этапе пользователь получает список источников данных, удовлетворяющих его запросу, и указывает необходимые ему источники из списка. Отметим, что список источников данных для каждого из них содержит число линий, попадающих в указанный пользователем спектральный интервал, и число спектральных линий, не попадающих в интервал, но относящихся к полосам, попавшим в интервал.

Визуализация данных осуществляется с помощью графического приложения в форме аплета. В рамках данного приложения возможны масштабирование изображения, выделение разных источников данных для их отображения и зеркальное попарное сравнение графиков. На рис. 1 показано сравнение двух наборов данных. Детальное описание возможностей по визуализации данных можно найти на сайте (<http://saga.atmos.iao.ru>).

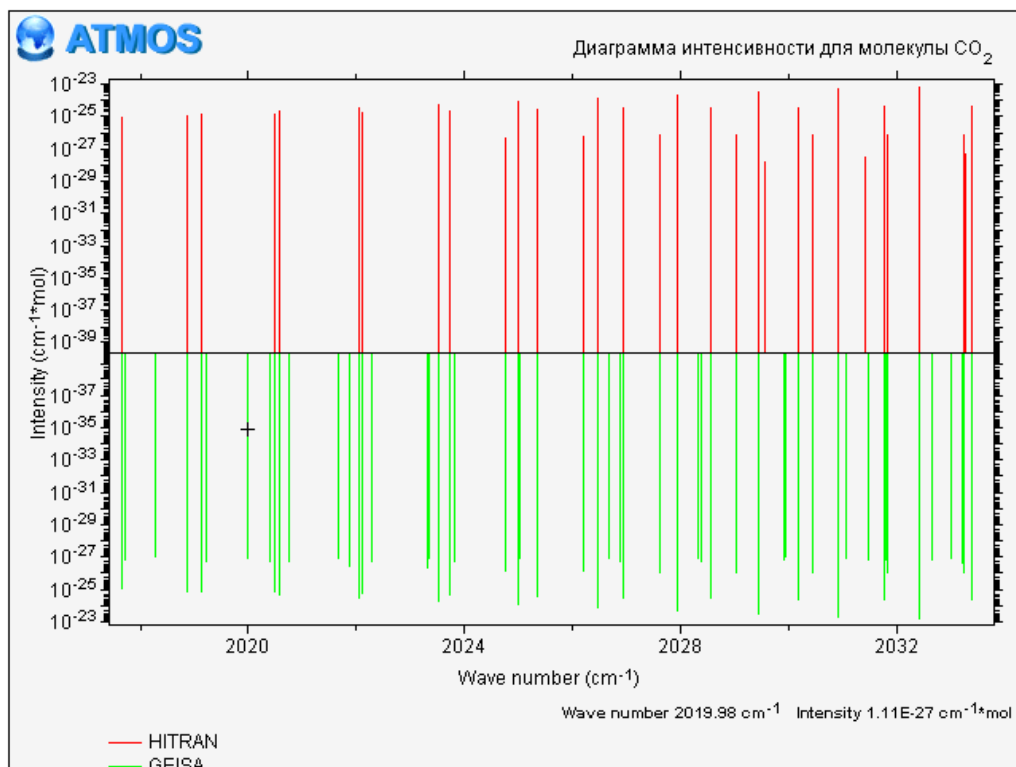


Рис. 1. Сравнение параметров спектральных линий для молекулы CO₂ в спектральном диапазоне 2018–2034 см⁻¹ по двум источникам информации (HITRAN (вверху) и GEISA). Крестом отмечено положение манипулятора на графике, и в нижней части графика отображены параметры отмеченной спектральной линии

Расчет коэффициента поглощения

При расчете спектральных функций используются параметры изолированной линии и параметры, обусловленные столкновениями. Для их расчета в ИВС выделены две физически разные ситуации: чистый газ и смесь газов. Ниже приведено описание данных, необходимых для вычисления коэффициента поглощения для чистого газа (однородного по составу). Обобщение на случай смеси газов в достаточной мере просто.

Как известно, коэффициент поглощения чистого газа s на частоте ω , обусловленный поглощением всеми линиями в интервале $[\omega - \Delta\omega, \omega + \Delta\omega]$, можно записать как

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

где суммирование проводится по всем линиям j в выбранном интервале $[\omega - \Delta\omega/2, \omega + \Delta\omega/2]$; ω – частота, на которой считается коэффициент поглощения, ω_j – частота центра j -й линии; $S_j^{(s)}$ – интенсивность j -й линии; $G^{(s)}(\omega, \omega_j, T, p^{(s)})$ – контур j -й линии при самоуширении, T – рассматриваемая температура, К, T_0 – температура, при которой заданы параметры линий в источнике информации о параметрах спектральных линий; $p^{(s)}$ – давление

рассматриваемого поглощающего газа, атм. Величина интервала $\Delta\omega$ задается из эвристических предположений и может зависеть от типа газа и термодинамических условий.

Интенсивность j -й линии при произвольной температуре T есть

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

где Q_V и Q_R – колебательная и вращательная статистические суммы соответственно; E_j – энергия нижнего уровня, с которого осуществляется переход, см⁻¹.

Таблицы величин

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

имеются, например, в HITRAN-92.

Первым этапом при подготовке расчета является выбор вещества, спектрального интервала, термодинамических условий и порогового значения интенсивности изолированной линии, ниже которого данные о линии в расчете не учитываются. На втором этапе пользователю по заданным им условиям предоставляется список источников информации о параметрах спектральных линий. После вы-

бора одного источника надо задать тип контура спектральной линии, интервал обрезания контура линии $\Delta\omega$ и точки, в которых проводится расчет. Задание точек можно осуществлять одним из трех способов: перечислением (в файле), указанием в html-форме для ввода непосредственно числа равномерных разбиений выбранного спектрального интервала (не меньше чем число линий) или фактора разбиения (число разбиений при этом равно произведению фактора разбиения на число спектральных линий, попавших в заданный спектральный интервал).

Для вычисления коэффициента поглощения используется следующий набор данных: центры и ин-

тенсивности спектральных линий, статистические суммы, коэффициент температурной зависимости полуширины, полуширина и сдвиг линии.

Результат расчетов по желанию пользователя может быть сохранен на сервере. После этого его можно будет сравнивать с имеющимися в БД результатами экспериментов или базовых расчетов. На рис. 2 приведено сравнение результатов расчета с экспериментом [8].

Система обеспечивает просмотр метаданных для хранящихся в базе данных результатов экспериментов (Winters_1964). Их можно просмотреть с той же страницы. На рис. 3 приведен пример представления метаданных.

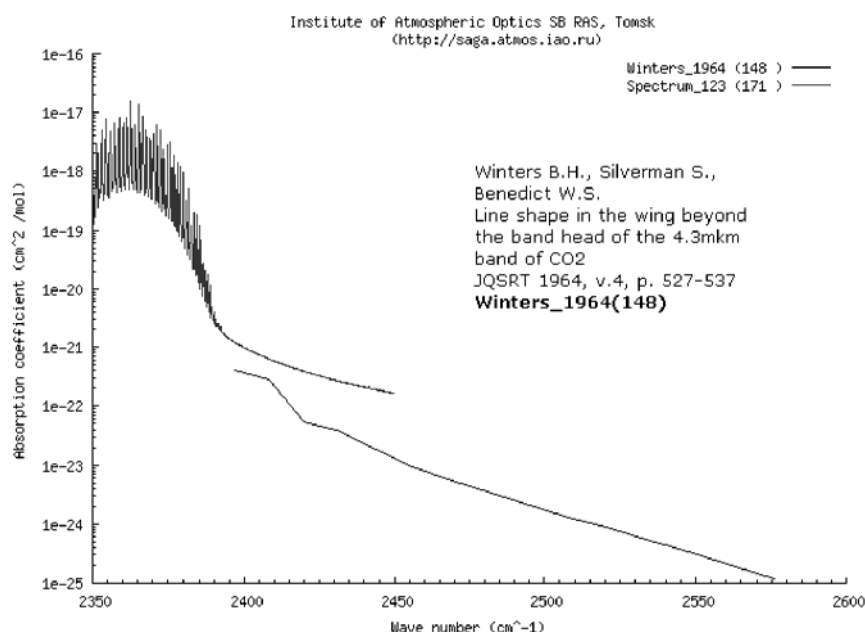


Рис. 2. Сравнение экспериментальных данных для CO₂ (Winters_1964, нижняя кривая) с результатом расчета (Spectrum_123)

Вещество		Термодинамические условия	
Поглощающий газ	CO2	Температура (°K)	296
Уширяющий газ	self	Давление (atm)	0.25
Массив данных		Давление уширяющего газа (atm)	0.25
Частоты перехода (число значений) (cm ⁻¹)	15	Спектральные параметры	
Коэффициент поглощения (число значений) (atm ⁻² cm ⁻¹ , exp)	15	Разрешение (cm ⁻¹)	0.6
Ошибки	Yes	Длина пути (m)	10
Литература		Интервал частот (cm ⁻¹)	2397-2576
Авторы	Winters B.H., Silverman S., Benedict W.S.		
Название статьи	Line shape in the wing beyond the band head of the 4.3microm band of CO2		
Журнал, год издания, том, страницы	JQSRT 1964, v.4, p. 527-537		
Комментарий			

Рис. 3. Метаданные по источнику информации Winters_148

Заключение

В статье описана структура информационных ресурсов, использованная при создании информа-

ционно-вычислительной системы «Атмосферная спектроскопия». Основное внимание уделено описанию структурированной модели данных. Стоит отметить, что большая часть данных молекулярной

спектроскопии в настоящее время может быть представлена в рамках слабоструктурированной или неструктурированной моделей данных и они содержатся в созданной ИВС. Описание операций с такими данными не входит в наши задачи и не представлено в данной статье.

Дальнейшее развитие модели данных, прежде всего фундаментальных характеристик молекул, будет выполнено при создании новой версии системы. Первые шаги в этом направлении уже сделаны [9].

Авторы благодарны чл.-кор. С.Д. Творогову (грант «Оптическая спектроскопия молекул и радиационные процессы в атмосфере», НИИ № 373.2003.5), а также РФФИ (грант № 02-07-90139) за поддержку данной работы.

1. <http://www.hitran.com>
2. <http://ara.ltd.polytechnique.fr>
3. Бабилов Ю.Л., Барб А., Головки В.Ф., Михайленко С.Н., Тютчев В.Г. Интернет-коллекции по молекулярной спектроскопии // Тр. RCDL. 2001. С. 183–187.

4. Babikov Yu.L., Fasliev A.S., Golovko V.F., Rodimova O.B. WEB information system: atmospheric spectroscopy // Proc. SPIE. 2000. V. 4341. P. 604–615.
5. Михайленко С.Н., Бабилов Ю.Л., Тютчев В.Г., Барб А. Банк данных по спектроскопии озона, доступный в Интернете (S&MPO) // Вычисл. технол. 2002. Т. 7. Спец. выпуск. С. 64–70.
6. Коголовский М.Р. Перспективные технологии информационных систем. М.: ДМК Пресс; Компания АйТи, 2003. 288 с.
7. Коголовский М.Р. Научные коллекции информационных ресурсов в электронных библиотеках // Тр. Первой Всерос. науч. конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции», 19–22 октября 1999. СПб.: Изд-во СПбГУ, 1999. С. 16–31.
8. Winters B.H., Silverman S., Benedict W.S. Line shape in the wing beyond the band head of the 4.3 μm band of CO_2 // J. Quant. Spectrosc. and Radiat. Transfer. 1964. V. 4. N 4. P. 527–537.
9. Козодоев А.В., Азизов Р.К., Быков А.Д., Пшеничников А.М., Фазлиев А.З. Распределенная информационная система по молекулярной спектроскопии // Сб. тезисов Междунар. симпоз. «Атмосферная радиация», 22–25 июня 2004 г. СПб.: Изд-во СПбГУ, 2004. С. 92–93.

A.D. Bykov, B.A. Voronin, A.V. Kozodoev, N.A. Lavrent'ev, O.B. Rodimova, A.Z. Fazliev. **Information system for molecular spectroscopy. I. Structure of information resources.**

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 separated to be the fundamental molecular characteristics, spectral line parameters, and spectral functions. The data input system, allowing the user to create custom data sets, is described. The relations between the parts of the data model are demonstrated with the calculation of the absorption coefficient as an example.