

Информационная система GrafOnto. Статистика, поиск и интеграция научных графиков, контроль качества цитирования

Н.А. Лаврентьев, О.Б. Родимова, А.З. Фазлиев

*Институт оптики атмосферы СО РАН, пл. академика В.Е.Зуева,
Томск, Россия*

Аннотация. Создана ИС GrafOnto, содержащая опубликованные графики и рисунки, отобранные экспертами в предметных областях «Континуальное поглощение» и «Слабосвязанные молекулярные комплексы». Система содержит графическое представление свойств 19 молекул, 44 комплексов и 75 их смесей. Свойства характеризуются 110 физическими функциями и 32 аргументами. Ключевым понятием при формировании коллекции является примитивный график. В настоящее время в коллекции содержится 6294 примитивных графиков, из которых сформированы ранее опубликованные 1367 составных графиков и 237 составных рисунков.

В ИС GrafOnto решены информационные задачи расчета статистики графиков и функций, атрибутивного, контекстного и семантического поиска примитивных графиков, количественной оценки качества цитирования, интеграции графиков в форме составных графиков и автоматического пересчета функций и аргументов в совместимые единицы измерений.

Ключевые слова: Континуальное поглощение, слабосвязанные молекулярные комплексы, информационная система GrafOnto, качество цитирования графиков

GrafOnto Information System. Statistics, search and integration of scientific plots, citation quality control

N.A. Lavrentiev, O.B. Rodimova, A.Z. Fazliev

*Institute of Atmospheric Optics SB RAS, Academician V.E. Zuev Square, Tomsk,
Russia*

Abstract. The GrafOnto IS has been created, containing published graphs and figures selected by experts in the subject areas of "Continuous absorption" and "Weakly bound molecular complexes". The system contains a graphical

representation of the properties of 19 molecules, 44 complexes and 75 of their mixtures. The properties are characterized by 110 physical functions and 32 arguments. The key concept in forming the collection is a primitive graph. Currently, the collection contains 6294 primitive graphs, from which 1367 previously published composite graphs and 237 composite figures have been formed.

The GrafOnto IS solves information problems of calculating graph and function statistics, attribute, contextual and semantic search for primitive graphs, quantitative assessment of citation quality, integrating graphs in the form of composite graphs and automatically recalculating functions and arguments into compatible units of measurement.

Keywords: Continuous absorption, weakly bound molecular complexes, GrafOnto information system, plot citation quality

1. Введение

Поверхность Солнца имеет температуру порядка 6000 К, максимум приходящего излучения на поверхность Земли приходится на УФ диапазон. Земля, в то же время, является источником излучения. Так как температура ее поверхности сравнительно невелика, максимум уходящего излучения приходится на ИК диапазон. Препятствием и для приходящего и для уходящего излучения является атмосфера Земли, так как молекулы, содержащиеся в ней, могут поглощать излучение. Основным поглотителем УФ излучения является молекула озона, и озоновый слой, входящий в состав атмосферы, защищает поверхность Земли от этого вредного излучения. Основная часть уходящего ИК излучения покидает Землю в диапазоне 8-20 мкм благодаря тому, что в этом участке спектра практически нет сильных полос поглощения имеющих в атмосфере молекул. Этот участок спектра называют «окном прозрачности» атмосферы. Понятно, что точные расчеты уходящего излучения очень важны в климатических задачах, так как знание баланса между приходящим и уходящим излучением непосредственно влияет как на восстановление долговременной климатической истории Земли, так и на современный прогноз погоды.

Основные составляющие атмосферу газы – кислород и азот – не поглощают в ИК диапазоне. Здесь вступают в игру так называемые малые газовые составляющие атмосферы, среди которых особую роль играет водяной пар. Диапазон частот, отвечающий окну прозрачности атмосферы и находящийся между двумя сильными полосами водяного пара, является предметом пристального исследования физиков более ста лет. И более ста лет назад было обнаружено, что, несмотря на отсутствие сильных линий воды, поглощение в этом окне присутствует, хотя слабое и не имеющее спектральной структуры, и обусловлено оно именно водяным паром. Это поглощение называли континуальным. Как оказалось, оно существует не только в этом, но и в других промежутках между полосами поглощения, а также в пределах самих полос. Дискуссии об его происхождении,

продолжающиеся более ста лет, лишь в последнее время стали склоняться к некоторой приемлемой общей точке зрения.

Первой попыткой объяснения природы континуального поглощения было предположение, что оно обусловлено поглощением далекими крыльями сильных линий [1]. В то время мнение о том, что контур линии является лоренцевским, было общепринятым и незыблемым. Однако объяснить с его помощью все закономерности континуального поглощения оказалось невозможным. В другом варианте объяснения континуальное поглощение приписывалось димерам водяного пара [2], присутствующим в атмосфере. У этой точки зрения были свои слабости, тем более, что спектры димеров в условиях атмосферы обнаружены не были. Тем временем в экспериментах было обнаружено, что контур линий в далеких крыльях радикально отличается от лоренцевского. Первые теоретические исследования в этом направлении показали, что континуальное поглощение может быть объяснено как обусловленное суммой линий, имеющих специфическое поведение в крыльях [3]. Эта позиция, аргументированная теоретически, выступает сейчас как асимптотическая теория крыльев линии (АТКЛ) и активно применяется для описания континуального поглощения. До некоторой степени альтернативой АТКЛ является так называемая статическая теория крыльев линий, имеющая распространение на западе [5].

Развитию димерной гипотезы способствовали экспериментальные исследования, обнаружившие спектр димеров и подтвердившие их наличие в атмосфере [6,

7]. Прогресс в многоэлектронных расчетах позволил получить спектр димеров воды в широком диапазоне частот [8]. Кроме того, классические траекторные расчеты детализировали картину димерного поглощения, выделив помимо стабильных квазистабильные димеры и т.н. «свободные пары» [9]. В итоге в ряде случаев все континуальное поглощение приписывалось различным типам димеров [10]. Такой позиции придерживается и появившиеся недавно исследования, использующие константы скоростей реакций образования димеров как средство описания поглощения димерами разных типов [11].

Следует отметить, что АТКЛ по своему построению принимает во внимание все парные взаимодействия, за исключением тех, что приводят к образованию стабильных димеров. Поэтому она позволяет оценивать поглощение стабильными димерами [12]. Сложившаяся ситуация заставляет предположить, что описание континуума должно представлять некую комбинацию конкурирующих подходов. Как нам кажется, в пользу такого положения свидетельствует и факт обнаружения в континуальном поглощении в пределах полос некоего «неопознанного» поглощения, которое остается после вычитания поглощения, обусловленного димерами разных типов [13].

Научные графики используются в статьях тех предметных областей, в которых табличное представление данных не эффективно при представлении результатов исследований. В частности, это касается массивов, представляющих кривые с максимумами и минимумами. При исследовании континуального поглощения и молекулярных комплексов значительная часть информации представлена в графиках. В информационной системе (ИС) GrafOnto содержатся и поддерживаются графики, содержащие базовую информацию об этих предметных областях. Пользователь может выбрать эти данные и сравнивать их между собой.

Работа по созданию ИС GrafOnto была начата в 2015 году, в 2020 году была анонсирована Виртуальным центром атомных и молекулярных данных в статье [14] и заканчивается в 2025 году.

2. Особенности графических ресурсов по спектроскопии

Графики и рисунки, опубликованные в статьях по континуальному поглощению молекул, смесей молекул и комплексов, а также молекулярным комплексам, содержат от одной до сотни точек для результатов экспериментов и непрерывные кривые для расчетов. В статьях, как правило, большая часть рисунков содержит несколько результатов исследований в одной системе координат. Иногда на одном рисунке совмещается пара систем координат. Такие составные графики составляют значительную часть ресурсов научных графиков.

1.1. Типы графиков в ИС GrafOnto

1.1.1. Примитивный график (оригинальный, цитирующий, экспертный)

Оригинальный примитивный график состоит из двух компонент: графического представления одного числового массива и метаданных этого графика. Метаданные представляют свойства числового массива, их значения, название графика и библиографическую ссылку на публикацию. Цитирующий примитивный график состоит из трех компонент: двух вышеописанных компонент оригинального примитивного графика и цитируемой библиографической ссылки [15].

Одни и те же физические величины в публикациях имеют разные названия и единицы измерения. Это приводит к сложностям при сравнении графиков. По этой причине в ИС GrafOnto использованы названия функций, наиболее часто используемые в статьях.

При сравнении физических величин перевод от одной размерности физических величины к другой осуществляется автоматически.

1.1.2. Составной график (оригинальный, цитирующий, многостатейный)

Составной график содержит одну систему координат, в которой находится более одного примитивного графика. Оригинальный составной график не содержит цитирующих примитивных графиков. Метаданными оригинального составного графика являются все метаданные примитивных графиков, входящих в него, и подпись к рисунку, извлеченная из публикации.

Цитирующий составной график содержит хотя бы один цитирующий примитивный график. Метаданные такого графика характеризуют одну систему координат и содержат все библиографические ссылки на цитируемые графики.

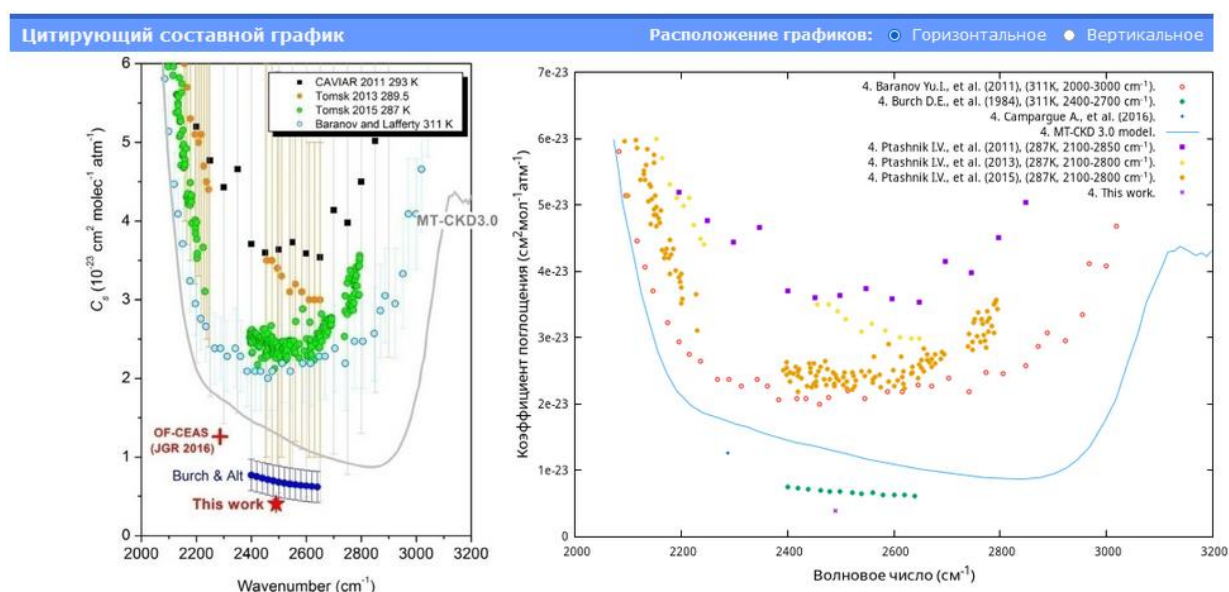


Рисунок 1. Пример цитируемого составного графика.

На рисунке 1 справа график, взятый из публикации, а справа его аналог, построенный в ИС GrafOnto.

В многостатейном составном графике цитирующий примитивный график представлен в паре с оригинальным (цитируемым) графиком из коллекции GrafOnto. Такая пара описывается тремя компонентами: парный график, таблица и свойства парного графика.

2.1.3. Составной рисунок

Составной рисунок содержит набор примитивных графиков и составных графиков, их метаданные и подпись к рисунку.

3. Информационная система GrafOnto

В ИС GrafOnto в графической форме представлены связанные между собой значения физических величин, характеризующих вещество (молекулы, слабосвязанные молекулярные комплексы и смеси молекул и комплексов). Онтологическое представление информации и знаний связано с набором свойств, описывающих термодинамические условия, библиографическую ссылку на публикацию, в которой были размещены рисунки или таблицы, тип измерений или расчетов, методы измерения или расчетов и т.д.

ИС освобождает пользователей от рутинной работы поиска статей (например, при поиске необходимого для цитирования графика) и классических графиков, используемых при сравнении, а также дает количественную оценку расхождения сравниваемых графиков.

3.2. Информационные задачи

Основные информационные задачи, решаемые в системе GrafOnto: атрибутивный, контекстный и семантический поиск графических ресурсов, расчет статистических характеристик ИС, интеграция примитивных графиков в составной график и рисунок, предоставление количественной оценки качества цитирующих графиков и формирование онтологического описания графических ресурсов коллекций графиков и рисунков.

3.2.1. Расчёт статистики графических ресурсов (Физические величины, вещества, функции и аргументы, число графиков разного типа)

Таблица 1. Фрагмент статистики коэффициентов поглощения ($\text{см}^2\text{мол}^{-1}\text{атм}^{-1}$) для разных веществ входящих в примитивные графики и составные графики и составные рисунки

Смеси	ПГ	СГ	СР	Молекулы	ПГ	СГ	СР
Ar + CO ₂	4	1		H ₂ O	1153	211	19
H ₂ O + CO ₂	1	1					
H ₂ O + H ₂ O-H ₂ O	10	6	1	Комплексы	ПГ	СГ	СР
H ₂ O + N ₂	178	51	3	H ₂ O-H ₂ O	74	32	5

В Таблице 1 для коэффициентов поглощения молекул и молекулярных комплексов, а также смесей молекул и комплексов показаны количества трех типов графиков.

В Таблице 2 представлены спектральные функции для смеси $\text{CO}_2 + \text{He}$. Во второй колонке указано число примитивных графиков для каждой из функций, в третьей колонке число составных графиков, содержащих функцию из строки и в четвертой колонке число встречаемости функции в строке в составном рисунке.

Таблица 2. Фрагмент расчёта статистики графических ресурсов для смеси $\text{CO}_2 + \text{He}$.

$\text{CO}_2 + \text{He}$. Примитивные графики (ПГ): 208. Составные графики (СГ): 71. Составные рисунки (СР): 19.	ПГ	СГ	СР
Заданная функция	28	10	8
Коэффициент поглощения (Км^{-1})	6	2	
Коэффициент поглощения (см^{-1})	83	27	11
Коэффициент поглощения ($\text{см}^{-1}\text{Амага}^{-1}$)	18	6	
Коэффициент поглощения ($\text{см}^{-1}\text{Амага}^{-2}$)	37	13	3

Рост числа графиков в коллекции GrafOnto графиков ИС по годам показан в Таблице 3.

Таблица 3. Статистика графических ресурсов (число графиков разного типа)

Тип графика	Временной интервал формирования коллекции						
	2018	2019	2020	2021	2022	2023	2024
Примитивные графики	2358	4471	4683	5787	6288	6785	6866
Оригинальные графики	-	3012	3163	4049	4481	4915	4979
Цитирующие графики	558	1199	1160	1417	1481	1277	1296
Экспертные графики	-	299	321	321	330	333	332
Составные графики	526	961	1000	1250	1376	1517	1526
Оригинальные составные графики	-	531	552	708	812	-	-
Цитирующие составные графики	107	430	269	362	237	-	-
Многостатейные графики	-	148	79	180	327	-	-
Составные рисунки	96	163	168	205	242	279	281

3.2.2. Атрибутивный поиск примитивных графиков и интеграция примитивных графиков из коллекций GrafOnto

Атрибуты разных типов графиков описаны в работах [16]. Выбор наборов графиков определяется ограничениями на атрибуты примитивных

графиков, содержащихся в коллекции. В таблицах 2-4 показаны интерфейсы для поиска графиков и результат интеграции примитивных графиков в составной.

На рисунке 2 показаны атрибуты, значениями которых можно ограничивать поиск требуемых графических ресурсов. При выборе обязательными атрибутами являются физические величины по осям абсцисс и ординат, вещество (молекула, смесь или молекулярный комплекс), тип источника, тип данных и происхождение данных (оцифрованные или оригинальные).

Формирование критериев запроса и поиск запрошенных примитивных графиков				
Графические характеристики				
☒	Ось абсцисс (X) (Размерность)*	Температура (K) (513) ▾	☒	Ось ординат (Y) (Размерность)*
Вещество. Физические характеристики вещества и среды				
☒	Вещество*	H ₂ O (2222) ▾	☐	γ*
☐	Температура	270 - 296 K ☒ Температура не задана	☐	Метод измерения (решения), Модель, Аппроксимация*
☐	Давление (полное)	0 - 10 атм ☒ Полное давление не задано	☒	Тип данных*
☒	Тип источника*	Первичный (4980) ▾	☒	Происхождение данных*
Библиографическая ссылка				
☐	Поиск библиографической ссылки	Введите ключевое слово или имя автора	☐	Временной интервал публикации (годы)
☐	Номер рисунка в публикации		☐	Подпись к графику
Найти запрашиваемые графики * В выкидных меню в скобках (N) указано число загруженных графиков, относящихся к выбранной величине				

Рис. 2. Задание значений атрибутов для поиска примитивных графиков

На рисунке 2 показана часть метаданных примитивного графика, по которой пользователь ищет требуемые ему графики. В левой части строки можно зафиксировать галочкой необходимые графики.

☒		Рисунок 10. This work. Spectral dependence of the self-continuum cross-section, C_s , at room temperature derived in this work.	H ₂ O T = 296 K P = 0.0131579 атм Точек: 19	γ. Коэффициент поглощения (см ² мол ⁻¹ атм ⁻¹) 3.18082e-25 - 2.73242e-24 X. Волновое число (см ⁻¹) 5876.95 - 6524.55	D. Mondelain, A. Aradj, S. Kassi, A. Campargue, The water vapour self-continuum by CRDS at room temperature in the 1.6 μm transparency window, Journal of Quantitative Spectroscopy and Radiative Transfer, 2013, Volume 130, Pages 381-391, DOI: 10.1016/j.jqsrt.2013.07.006. Annotation
-------------------------------------	---	---	---	---	--

Рис. 3. Типовая строка для просмотра и выбора примитивных графиков

На рисунке 4 показаны все три выбранных графика по континуальному поглощению в разных спектральных интервалах для паров молекулы воды при температуре 296K.

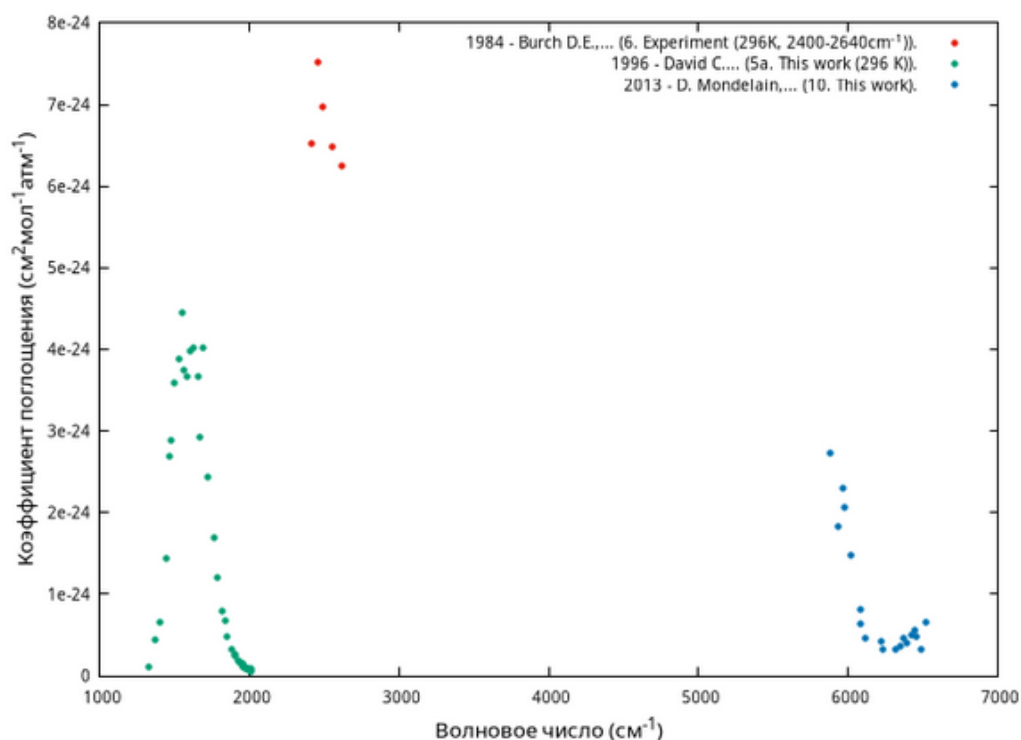


Рис. 4. Интеграция примитивных графиков в составной график

3.2.3. Качество цитирования графиков

Основная проблема нахождения цитируемого графика для цитирующего графика состоит в том, что библиографическая ссылка не всегда достаточна для его точного определения. На практике подавляющее большинство авторов при ссылке на график не указывают номер рисунка. Естественно возникают трудности в поиске однозначно не определенного цитируемого графика. В статье [17] была предложена методика оценки близости графиков, позволившая количественно определять величину близости.

На рисунке 5 приведен пример оценки близости цитируемого и цитирующего графиков. На рисунке цитирование частичное, при этом в совпадающем интервале расхождение графиков составляет 1.167%.

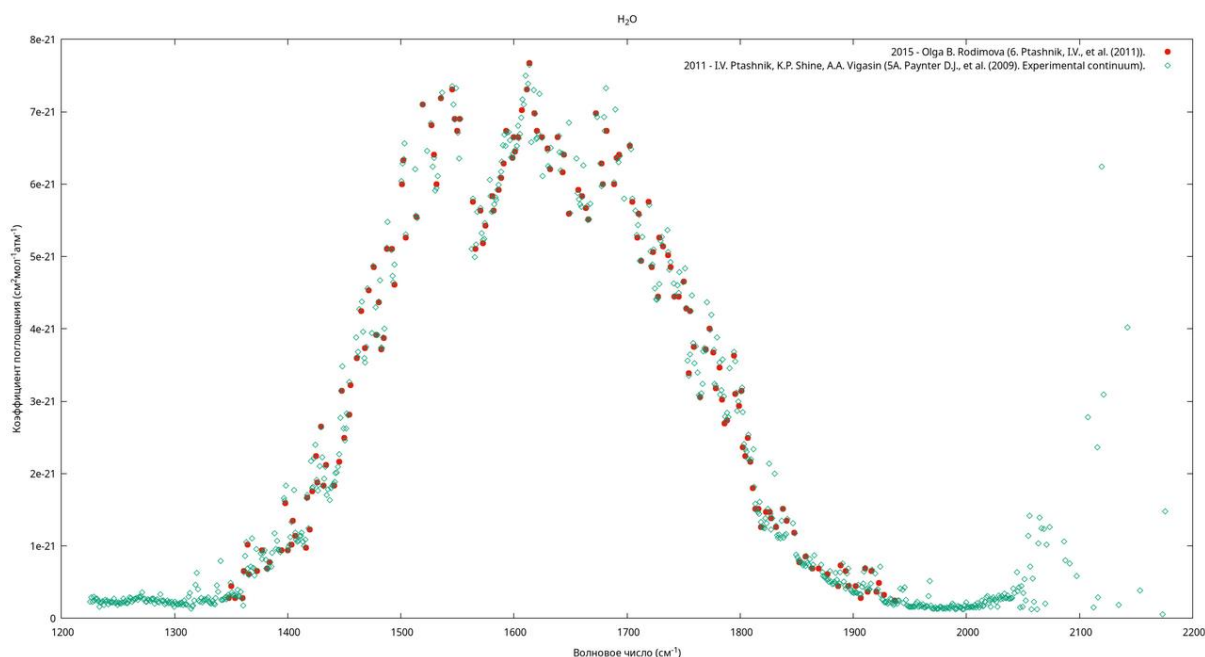


Рис.5. Выбор цитируемого графика для цитирующего графика

Работа выполнена в рамках государственного задания ИОА СО РАН.

Литература

1. Elsasser W.M. Far infrared absorption of atmospheric water vapor // *Astrophys. J.* 1038. V.87. P. 497-507.
2. Penner S.S., Varanasi P. Spectral absorption coefficients in the rotation spectrum of water vapor // *JQSRT* 1967. V.7. No.4. P.687-690.
3. Творогов С.Д., Несмелова Л.И. Радиационные процессы в крыльях полос атмосферных газов // *Изв. АН СССР, сер. ФАО* 1976. Т. 12. №6. С. 627 – 633.
4. Несмелова Л.И., Родимова О.Б., Творогов С.Д. Контур спектральной линии и межмолекулярное взаимодействие. Новосибирск: Наука. 1986. 216 с.
5. Ma Q., Tipping R.H., Leforestier C. Temperature dependences of mechanisms responsible for the water-vapor continuum absorption: 1. Far wings of allowed lines // *J. Chem. Phys.* 2008. 2008V.128. P.124313
6. Ptashnik I.V., Smith K.M., Shine K.P., Newnham D.A. Laboratory measurements of water vapour continuum absorption in spectral region 5000–5600 cm⁻¹: Evidence for water dimers // *Quarterly Journal of the Royal Meteorological Society* 2004. V.130. Iss.602. Part A. P.2391–2408.
7. Serov E.A., Koshelev M.A., Odintsova T.A., Parshin V.V., Tretyakov M.Yu. Rotationally resolved water dimer spectra in atmospheric air and pure water vapour in the 188–258 GHz range // *Phys. Chem. Chem. Phys.* 2014. V. 16, N 47. P. 26221-33-1–26221-33-15.

8. Kjaergaard H., Garden A., Chaban G., Gerber R., Matthews D., Stanton J. Calculation of vibrational transition frequencies and intensities in water dimer: Comparison of different vibrational approaches // *J. Phys. Chem. A*. 2008. V. 112. P. 4324–4335.
9. Epifanov S.Y., Vigasin A.A. Subdivision of phase space for anisotropically interacting water molecules // *Mol. Phys.* 1997. V. 90. P. 101–106.
10. Ptashnik I.V., Shine K.P., Vigasin A.A. Water vapour self-continuum and water dimers. 1. Review and analysis of recent work // *JQSRT*. 2011. V. 112. P. 1286–1303.
11. Simonova A.A., Ptashnik I.V., Shine K.P., Semi-empirical water dimer model of the water vapour self-continuum within the IR absorption bands // *JQSRT* 2024. V.329. P.109198.
12. Богданова Ю.В., Родимова О.Б. Соотношение между поглощением мономерами и димерами водяного пара в пределах вращательной полосы H_2O // *Оптика атмосферы и океана*. 2018. Т.31. №5. С.341-348.
13. Simonova A.A., Ptashnik I.V., Elsey J., McPheat R.A., Shine K.P., Smith K.M. Water vapour self-continuum in near-visible IR absorption bands: Measurements and semiempirical model of water dimer absorption // *JQSRT* 2022. V.277. P.107957.
14. Damien Albert, Bobby K. Antony, Yaye Awa Ba. et al.. A Decade with VAMDC: Results and Ambitions // *Atoms* 2020. V. 8(4). P.76. <https://doi.org/10.3390/atoms8040076>.
15. Ахлѣстин А.Ю., Лаврентьев Н.А., Фазлиев А.З. Систематизация научных графических ресурсов по молекулярной спектроскопии // *Научный сервис в сети Интернет: труды XIX Всероссийской научной конференции (18-23 сентября 2017г., г. Новороссийск)*. - М.: ИПМ им. М.В.Келдыша, 2017. С. 34-42. <http://keldysh.ru/abrau/2017/39.pdf> doi:10.20948/abrau-2017-39.
16. Lavrentiev N.A., Privezentsev A.I., Fazliev. A.Z. Tabular and Graphic Resources in Quantitative Spectroscopy // In: Manolopoulos Y., Stupnikov S. (eds) *Data Analytics and Management in Data Intensive Domains. DAMDID/RCDL* 2018. Communications in Computer and Information Science. Springer, Cham. V.1003. P. 55-69. (2019). DOI https://doi.org/10.1007/978-3-030-23584-0_4.
17. Лаврентьев Н. А., Родимова О. Б., Фазлиев А. З. Систематизация опубликованной научной графики, представляющей характеристики континуального поглощения водяного пара. II. Публикации 1981–2000 гг. // *Оптика атмосферы и океана*. 2022. Т. 35. № 03. С. 217–231. DOI: 10.15372/AOO20220307.