**kspectrum: an open-source code for high-resolution molecular absorption spectra production**

V. Eymet, C. Coustet, B. Piaud  
Mésostar SAS, 8 rue des pêchers, 31410 Longages, France  
E-mail: vincent.eymet@meso-star.com

**Abstract.** We present the *kspectrum* scientific code that produces high-resolution synthetic absorption spectra from public molecular transition parameters databases. This code was originally required by the atmospheric and astrophysics communities, and its evolution is now driven by new scientific projects among the user community. Since it was designed without any optimization that would be specific to any particular application field, its use could also be extended to other domains. *kspectrum* produces spectral data that can subsequently be used either for high-resolution radiative transfer simulations, or for producing statistic spectral model parameters using additional tools. This is a open project that aims at providing an up-to-date tool that takes advantage of modern computational hardware and recent parallelization libraries. It is currently provided by Mésostar (http://www.meso-star.com) under the CeCILL license, and benefits from regular updates and improvements.

1. **Introduction**

The production of synthetic high-resolution molecular spectra is a central concern in many scientific domains involving radiative transfer: in atmospheric sciences, the inversion of high-resolution spectral signals measured from space is used in order to retrieve composition and temperature profiles. In astrophysics, the measurements of emitted radiation can provide insightful information about the composition of nebulae and stellar atmospheres. Recently, astrobiologists have been interested in retrieving atmospheric composition and ground temperatures on distant exoplanets from radiation that is both emitted and reflected by the planet.

In study domains that involve high temperatures, radiation solvers both rely on high-resolution spectra [1] and statistical models [2] such as the Malkmus transmittivity model [3] that has historically been heavily used. Data sets that can be used in order to obtain Malkmus’ model parameters \( \bar{k} \) and \( \phi \) (respectively, the average absorption coefficient and the overlap parameters within a given narrowband) for a wide range of thermodynamic conditions (pressure, temperature and concentrations) for the main radiatively active species in flames (\( \text{H}_2\text{O}, \text{CO}_2 \) and CO) have been derived from measurements [1, 5]. This work is more than ever relevant [6] in the context of new theoretical developments on average transmittivity models that can also take into account the spatial heterogeneities of the participating medium [7]. The validation of radiative transfer results obtained with such spectral models must be performed using reference results, obtained with high-resolution spectra. In more recent approaches, it has been shown that both the representation of heterogeneities and the spectral integration could be achieved using null-collision algorithms [8, 9] that require using a modified absorption coefficient field.
that should be as close as possible from an upper-bound of the real absorption coefficient field within the medium; this modified absorption coefficient field has been estimated using high-resolution spectra until now, but this work can certainly also take advantage of statistical spectral models approaches.

Building synthetic spectra is a computationally heavy task, since transition databases keep growing, both in the number of molecular species that are taken into account and in the number of transitions that are identified for each species: transition databases originally gathered information mainly about H\textsubscript{2}O and CO\textsubscript{2}. Within the HITRAN database, the number of molecular species grew from 39 in the 2004 version [10] to 42 in the 2008 version [11], and the 2012 version [12] counts 47 species. Ambitious projects such as ExoMol [13] are specialized in gathering transition parameters for many complex molecules that are found in planetary and stellar atmospheres.

The \textit{kspectrum} project was initiated by the community of climate study, during a modelization effort on the atmosphere of Venus [14]: the spectral data used for the gas mixture, while the best available at that time, had hardly been referenced and it was not possible to produce it for different temperature or composition profiles. Since then, the evolution of the code was driven by the needs of its users [15, 16] and the various scientific projects it has been involved in. The code has been used in climate modeling of both solar-system planets [17, 18] and exoplanets [19, 20, 21]. Recently, it was also used in order to better define the boundaries of the habitability zone around main-sequence stars [22]. And a new study of Venus atmosphere that should be published in the next weeks [23] also uses high-resolution spectra produced by \textit{kspectrum}.

\textit{kspectrum} is an open-source project, aiming at giving a complete control and traceability about spectral data production, and that is progressively being enriched by the experience acquired while interacting with the users. The code is distributed under the CeCILL license and maintained by Méso-Star (http://www.meso-star.com/en_Products.html). The main purpose of \textit{kspectrum} is to provide a framework that can be easily modified and adapted to specific needs. The code has been designed in order to be flexible in terms of architecture and reliable, and can easily be adapted to various computational hardware and new parallelization libraries.

When computing atmospheric absorption spectra, an adapted line profile truncation is commonly used. Then, an ad-hoc continuum is added to the resulting spectra. This continuum accounts for both the underestimation of the absorption coefficient because of the truncation, and for specific additional opacity (namely, collision-induced absorption, CIA) that is specific to the study case. This results in very well validated codes. In \textit{kspectrum}, we privileged versatility: even though the code can perform some CIA computation, it is preferred not to use any additional continuum in molecular absorption spectra computations. Then, the user has the possibility to add its own continuum in order to account for specific opacity that has not been taken into account (and possibly to correct truncation effects). Since nothing has been hard-coded for specific applications, the code can be used in a wide variety of cases and scientific fields.

\section{Description of the code}

The main purpose of \textit{kspectrum} is to compute high-resolution molecular absorption spectra for a wide range of thermodynamic conditions, using the 47 molecular species referenced in the 2012 version of the HITRAN database [12]. However, the code is not restricted to using the HITRAN database, and the user can control whether or not high-temperature data is used when it is available for these molecular species.

\subsection{Physics}

The \textit{kspectrum} code has been designed in order to address a wide variety of applications. It can handle modern public transition databases: HITRAN [10, 11, 12], HITEMP [24] and CDSD
high-temperature versions (HITEMP-2010 for H$_2$O and CDSD-4000 for CO$_2$) are supported as well. In terms of line profiles, the user can choose between using a plain Lorentz profile and the Voigt line profile, and several sub-lorentzian corrections have been included for the CO$_2$ molecule in the 1.18 $\mu$m and 2.7 $\mu$m bands [27], in the 2.3 $\mu$m band [28] and the 4.3 $\mu$m band [29] (see figure ??). Some collision-induced absorption continuum parameterization has also been included [30, 31], but CIA continuum will preferably be computed separately from molecular absorption. Finally, the code can handle a separate abundance for each isotopologue of each molecular species.

2.2. Computational aspects

kspectrum was originally designed from the project needs with two goals in mind:

(i) to ensure a given uncertainty over each value of $k(\nu)$ that is computed

(ii) to ensure a given uncertainty when using a linear interpolation between two consecutive values of $k(\nu)$

The first point (i) imposes to be able to take into account the contribution of each known transition to the total absorption coefficient at a given wavenumber $\nu$, even lines whose center is defined at a (possibly) infinite distance of the considered value of $\nu$. The fact that the line profile is, in practice, poorly known after a few tens of inverse centimeters from the line center, should not be a constraint at this point: being able to compute synthetic spectra with and without truncation at least provides some insights about the boundaries where the “real” (physical) spectrum lies; this point is further developed in section 3. The second point (ii) imposes that the spectral step (between two consecutive values of $\nu$) has to be computed in order to fulfill this condition. The method used is conceptually very similar to the idea of multiple spectral grids developed in [32].

In terms of computation time, synthetic spectra computed with such a full control of the numerical uncertainty are possible, but difficult to achieve. The code was subsequently modified in order to enable faster computations. One major source of computation time gain is the spectral discretization: in addition to the reference spectral discretization, other algorithms have been implemented, with decreasing computational requirements:

- using a given number of discrete wavenumbers for representing line centers and inter-line regions
- using only line centers as the spectral grid
- constant spectral step
- pre-computed spectral grid

Additionally, the user has the possibility to remove least intense lines for the spectral discretization step. This feature is particularly useful when using high-temperature transition databases that include several hundred millions of lines. Of course, the spectral grid obtained with one of these faster spectral discretization algorithms or when using a limited number of transitions does no longer satisfies condition (ii).

Further computational gains have been found by the implementation of:

- Line profile truncation, either at a fixed distance from the line center, or at a fixed multiple of the line width that differs from one line to the other. As illustrated in figure 2, the effects of truncating line profiles can result in a under-estimation of the absorption coefficient, and possibly the presence of unrealistic steps in the spectrum.

- Computing each value of $k(\nu)$ using a convergent approach: the contribution of most intense lines is computed first. The expected convergence of the sum of line contributions as line intensities decrease is periodically checked, and the algorithm stops computing line contributions when a convergence criterion is satisfied.
Figure 1. Synthetic spectra of CO$_2$ at a temperature of 500K and a pressure of 10atm, computed with and without sub-Loerentzian line profile correction in the (a) 4.3 µm, (b) 2.7 µm and (c) 1.18 µm windows.
The consequence of using either of these features is that the accuracy over each value of $k(\nu)$ is no longer guaranteed (see condition (i)).

When possible, the various stages involved in spectra computation have been parallelized in \textit{ksppectrum}. These stages include:

- tabulation of the Voigt function, required by the reference spectral discretization algorithm that must satisfy condition (ii).
- the identification of transition parameters location in each LBL file that should be used, for a number of pre-defined spectral intervals. This step is required for a faster subsequent access to this data.
- the spectral discretization of any spectral interval.
- finally, the computation of $k(\nu)$ values.

The computational load of each process is balanced using a very basic but efficient algorithm: the total computational load (for instance, computing a set of $N_k$ values of $k(\nu)$ once the spectral discretization has been completed) is divided into $N_{task}$ tasks (with $N_{task}$ much greater than the number of processes). Each process will then be allocated one task at a time. In this example, the elementary task involves the computation of $\frac{N_k}{N_{task}}$ values of $k(\nu)$. Once the task is over, the process sends back its results to the master process, and is allocated a new task, until all tasks have been performed. This rudimentary heuristic ensures that each process is near-to optimally used: the number of computing processes is generally equal to the number of cores available. If, for any reason, one or several cores are required by system processes or other codes, \textit{ksppectrum} processes that are running on these cores will naturally manage fewer tasks than processes that run on fully available cores. This results in globally faster computations than if the number of tasks is equal to the number of computing processes.

Figure 2. Synthetic spectrum of $\text{H}_2\text{O}$ at a temperature of 500K and a pressure of 1 atm, computed using the full Voigt profile at an infinite distance from line centers, compared with the same spectrum obtained with a truncation of the line profile at 25 cm$^{-1}$ from the line centers.
Handling the latest transition databases has also been a challenge: the HITEMP-2010 database includes 114 millions of transitions for H$_2$O, for a total of 18 Gb. Even if all parameters are not required for spectra computations, it is not yet possible to load all the required information in RAM at once on a mainstream computer. This is even worse for CSDS-4000, that holds 73 Gb of data over 2768 files for 626 millions of CO$_2$ transitions. When not using the convergence algorithm (computing first the contribution of most intense transitions), the contribution of every line has to be added for every value of k(ν). Since loading all the required data at once is not possible, kспектrum uses a different strategy: it loads only the parameters for a limited number of transitions, and computes their contributions to the required value of k(ν). This procedure is repeated until the parameters of all transitions have been read and all contributions have been computed. This multi-step approach results in a slower computation, but one major advantage in terms of flexibility is that the code can run with very limited RAM capacity. The lower is the RAM capacity, the greater will be the computation time for a given spectrum because of a greater number of successive steps. Figure 3 illustrates the impact of using a high-temperature transition database on a H$_2$O spectrum.

Finally, kспектrum performs frequent backups of the current state of the computation (in both spectral discretization and actual spectrum computation stages), so that the computation can be resumed from the latest backup in the event of a power outage or a maintenance reboot. This feature may seem anecdotic, but has always been very appreciated by users.

3. Academic test case
As far as we know, using a Lorentz or a Voigt profile at an infinite distance from line centers is not correct, because in practice line profiles are not well known after a few tens or hundreds of inverse centimeters. So taking into account the contribution of every transition, even at an infinite distance from the considered wavenumber, results in an overestimation of the spectrum. Similarly, truncating every line profile at a fixed distance of 25 cm$^{-1}$ results in an underestimation
Figure 4. Specific intensity emitted by a one-meter homogeneous gas column containing $\text{H}_2\text{O}$, $\text{CO}_2$ and CO (all three species at a 0.1 molar fraction), at 1000K and 1atm. High-resolution results have been averaged over 25 cm$^{-1}$ spectral intervals. The black line has been obtained using a full line profile and taking into account the contribution of every known transition, and the red line shows the same result, obtained using a 25 cm$^{-1}$ line profile truncation and rejecting least intense lines.

of the spectrum. Without speaking of additional opacity sources such as collision-induced absorption, the real absorption spectrum lies somewhere between these two limits.

Figure 4 illustrates the kind of tests that can be performed with ksppectrum: two synthetic spectra have been computed for a 1000K, 1atm participating media, containing $\text{H}_2\text{O}$, $\text{CO}_2$ and CO with equal molar fractions of 0.1. Both spectra have been obtained from the 2012 version of the HITRAN database. The first spectrum (black line) was computed using a full line profile, and taking into account every known transition in the database for these species. The second spectrum (red line) was obtained on the basis of a 25 cm$^{-1}$ truncation of the line profile, and using the convergence algorithm for neglecting least intense transitions. Then the specific intensity emitted by a one-meter homogeneous gas column was computed over the high-resolution spectral grid, and has been averaged over 25 cm$^{-1}$ spectral intervals in order to obtain the results displayed in figure 4. Off course, the emphasis has been put on a spectral region where the difference between the two results is significant: in most parts of the infrared region, the two spectra match perfectly. The overall effect can be quantified by a comparison of the total emitted specific intensity, integrated in the $[150;9325]$ cm$^{-1}$ longwave region: it is 4564.47 W/m$^2$/sr using a full line profile and 4366.91 W/m$^2$/sr when a truncation is applied; the effect of using a truncation in this example is at most 5\% over the integrated specific intensity, even if it can locally be responsible of much more important differences in specific spectral regions.

4. Conclusion

ksppectrum has been progressively improved over the last few years. It has already been used for a number of scientific projects in the atmosphere and astrophysics communities: climate
modeling of Saturn’s atmosphere \cite{18} and of exoplanets \cite{21,15}, confirming hypothesis about exoplanets observed by space telescopes \cite{20}, simulate the climatic evolution of primitive Earth-like atmospheres \cite{19} and better defining the boundaries of habitable zones \cite{22}. The code has been thought to be robust and reliable, and can run as well on cheap desktop computers as on dedicated servers. Users have been interested in both the possibility to compute absorption spectra without compromises, and the fact that nothing specific to a given application has been hard-coded, making \textit{kspectrum} a very versatile tool.

In terms of perspective, there is already a clearly identified need to make \textit{kspectrum} compatible with transitional data available in the ExoMol database (\url{http://www.exomol.com/}). This would be very helpful for the astrophysics and astrobiology communities. This would imply some development time, mainly because of the large number of molecular species involved and the numerous formats used to provide various transition parameters within this database.

\textit{kspectrum} results can be used directly as input for radiative transfer codes, or may be used by the following companion codes:

- \textit{kdistribution}: does exactly what its name says: to produce k-distribution data sets from \textit{kspectrum} high-resolution spectra, for later use in radiative transfer solvers that are using the correlated-k spectral model \cite{33}. Using directly high-resolution spectra within radiative transfer solvers is also very useful; however, since the default behaviour of \textit{kspectrum} is to produce unstructured spectral grids, the radiative transfer solver has to be able to interpolate absorption coefficients from its input spectra over the spectral grid used for radiative transfer computations.

- \textit{gas\_mix}: this is a code that produces high-resolution spectrum for any gas mixture, in any thermodynamic conditions (just like what \textit{kspectrum} does, but much faster) using a database of pre-computed high-resolution spectra over a given (P/T/x) grid, for a number of molecular species. This code is able to take into account new spectra, as they are added to the database: the (P/T/x) grid over which spectra are computed does not need to be structured, and can be progressively improved. This idea has been initiated some years ago within the Starwest group and led to the development of the “High Resolution Spectra” database (\url{http://www.starwest.ups-tlse.fr/vincent-eymet/high-resolution-spectra/}). This first database was computed for a limited number of pressure levels and molecular species, which restricted its use mainly to combustion applications, and is now outdated in the sense that it was computed using the HITRAN-2008 database. But the main idea behind this project is still valid, and could eventually result in a online tool for providing fast access to spectral data for radiative transfer applications, similarly to what can be obtained using \url{http://spectra.ioa.ru/}

References

[1] R. Tao and F.M. Modest. A hybrid wavenumber selection scheme for line-by-line photon monte carlo simulations in high-temperature gases. \textit{Journal of Heat Transfer}, 135(8), 2013.

[2] Gopalendu P. and Modest M.F. A narrow band-based multiscale multigroup full-spectrum k-distribution method for radiative transfer in nonhomogeneous gas-soot mixtures. \textit{Journal of Heat Transfer}, 132(2), 2009.

[3] W. Malkmus. Random lorentz band model with exponential-tailed $S^{-1}$ line-intensity distribution function. \textit{Journal of Optical Society of America}, 57:323–329, 1967.

[4] J. Taine and A. Soufiani. Gas ir radiative properties: from spectroscopic data to approximate models. \textit{Advances in Heat Transfer}, 33:295–414, 1999.

[5] A. Soufiani and J. Taine. High temperature gas radiative property parameters of statistical narrow band model for H$_2$O, CO$_2$ and CO and correlated-k model for H$_2$O and CO$_2$. \textit{Int. J. Heat and Mass Transfer}, 40(4):987–991, 1997.

[6] Ph. Rivière and A. Soufiani. Updated band model parameters for H$_2$O, CO$_2$, CH$_4$ and CO radiation at high temperature. \textit{International Journal of Heat and Mass Transfer}, 55:3349–3358, 2012.
Tashkun, J. Tennyson. HITEMP, the high-temperature molecular spectroscopic database. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 111:2139–2150, 2010.

[25] S.A. Tashkun, V.I. Perevalov, J.-L. Teffo, A.D. Byko, and N.N. Lavrentieva. CDSD-1000, the high-temperature carbon dioxide spectroscopic databank. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 82:165–196, 2003.

[26] S.A. Tashkun and V.I. Perevalov. CDSD-4000, High-resolution, high-temperature carbon dioxide spectroscopic databank. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 112, issue9:1403–1410, 2011.

[27] J.B. Pollack, J.B. Dalton, D.H. Grinspoon, R.B. Wattson, R. Freedman, D. Crisp, D.A. Allen, B. Bezard, C. DeBergh, L.P. Giver, Q. Ma, and R. Tipping. Near-infrared light from Venus’ nightside: A spectroscopic analysis. *Icarus*, 103:1–42, 1993.

[28] M.V. Tonkov, N.N. Filipov, V.V. Bertsev, J.P. Bouanich, Nguyen Van-Thanh, C. Brodbeck, J.M. Hartmann, C. Boulet, F. Thibault, and R. LeDoucen. Measurements and empirical modeling of pure CO2 absorption in the 2.3-µm region at room temperature: far wings, allowed and collision-induced bands. *Applied Optics*, 35 No. 24:4863–4870, 1996.

[29] M.Y. Perrin and J.M. Hartmann. Temperature-dependant measurements and modeling of absorption by CO2 – N2 mixtures in the far line-wings of the 4.3µm CO2 band. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 42:311–317, 1989.

[30] M. Gruszka and A. Borysow. Computer simulation of the far infrared collision induced absorption spectra of gaseous CO2. *Molecular physics*, 93 No. 6:1007–1016, 1998.

[31] Y. Baranov, W.J. Lafferty, and G.T. Fraser. Infrared spectrum of the continuum and dimer absorption in the vicinity of the O2 vibrational fundamental in O2/CO2 mixtures. *Journal of Molecular Spectroscopy*, 228:432–440, 2004.

[32] C. Caliot. Modélisation et simulation de l’émission énergétique et spectrale d’un jet réactif composé de gaz et de particules à haute température issu de la combustion d’un objet pyrotechnique. PhD thesis, Thèse de doctorat, Institut National Polytechnique de Toulouse, 2006.

[33] A. A. Lacis and V. Oinas. Description of the correlated-k distribution method for modeling nongray gaseous absorption, thermal emission, and multiple scattering in vertically inhomogeneous atmospheres. *Journal of Geophysical Research*, 96:9027–9063, 1991.