WATER IN PLANETARY AND COMETARY ATMOSPHERES:

H_2O / HDO TRANSMITTANCE AND FLUORESCENCE MODELS

G. L. Villanueva[1,2]

M. J. Mumma[1]

B. P. Bonev[1,2]

R. E. Novak[3]

R. J. Barber[4]

M. A. DiSanti[1]

[1] Solar System Exploration Division, Mailstop 690.3, NASA Goddard Space Flight Center, Greenbelt, MD, 20771, USA. Geronimo.Villanueva@nasa.gov
[2] Department of Physics, Catholic University of America, Washington, DC 20064.
[3] Department of Physics, Iona College, New Rochelle, NY 10801.
[4] Department of Physics and Astronomy, University College London, UK.

Abstract

We developed a modern methodology to retrieve water (H_2O) and deuterated water (HDO) in planetary and cometary atmospheres, and constructed an accurate spectral database that combines theoretical and empirical results. Based on a greatly expanded set of spectroscopic parameters, we built a full non-resonance cascade fluorescence model and computed fluorescence efficiencies for H_2O (500 million lines) and HDO (700 million lines). The new line list was also integrated into an advanced terrestrial radiative transfer code (LBLRTM) and adapted to the CO_2 rich atmosphere of Mars, for which we adopted the complex Robert–Bonamy formalism for line shapes. We then retrieved water and D/H in the atmospheres of Mars, comet C/2007 W1 (Boattini), and Earth by applying the new formalism to spectra obtained with the high-resolution spectrograph NIRSPEC/Keck II atop Mauna Kea (Hawaii). The new model accurately describes the complex morphology of the water bands and greatly increases the accuracy of the retrieved abundances (and the D/H ratio in water) with respect to previously available models. The new model provides improved agreement of predicted and measured intensities for many H_2O lines already identified in comets, and it identifies several unassigned cometary emission lines as new emission lines of H_2O. The improved spectral accuracy permits retrieval of more accurate rotational temperatures and production rates for cometary water.

1. Introduction

Water is among the most searched molecules in the Universe, owing to its role as a hydrogen repository and its strong connection with life on Earth. Even though it comprises just 0.02% of Earth's mass, all known forms of life depend on water through its role in metabolism. Our definition of "habitability" is thus strongly linked to water abundance, ultimately driving its search across the Universe.

Water has been sought and found in extremely diverse astronomical environments, from the cold interiors of comets [1] to the hot atmospheres of exoplanets [2] and the even hotter atmospheres of stars [3]. The recent discovery of water in proto-planetary disks [4], and further isotopic measurements in comets [5], are used to probe the beginnings of our Solar System. As a strongly polar molecule, water has a relatively high sublimation temperature when compared to the common apolar gases (H_2, CH_4, NH_3, H_2S, CO_2, O_2, N_2, etc.). This property molded our Planetary System, separating the formative zone of rocky terrestrial planets (Mercury, Venus, Earth, Mars) from that of the gas-rich Jovian planets by the frost line (at ~2.7 AU, [6]) where water ice first becomes stable.

Our capabilities to measure water remotely have expanded greatly with the recent advent of powerful high-resolution spectrometers atop high altitude mountains and in space. In concert with this evolution, the detection of high temperature water in sunspots has driven exponential growth of spectral databases that characterize the

radiative properties of water, now reaching more than 1,200 million lines for its isotopologues (see Tables 1 and 2).

Remote sensing of water has not been restricted to abundance measurements, but includes the characterization of its spin-isomeric (ortho, para) and isotopic forms (H_2O, HDO) that trace the environment in which the water molecules formed. The nuclear spin temperature (derived from the ortho/para abundance ratio of water, hereafter OPR) may be sensitive to formation temperatures lower than 50K; above this value the spin-isomer populations are described by statistical equilibrium (OPR = 3, see Fig. 1 of [7]). Moreover, models of nebular gas-phase chemistry (e.g., [8]) predict important enrichments of deuterated water at low temperatures ($T < 80K$). Since both ratios (nuclear spin species, and isotopologues) are very sensitive to the temperature at which the molecules initially formed, they are now being used to better understand the primordial conditions of our Planetary System.

Water is (by far) the most abundant primary volatile in cometary nuclei, and the abundances of other species are expressed relative to it (C_2H_6/H_2O, CH_3OH/H_2O, HCN/H_2O, etc.). These "mixing ratios" with respect to water are the principal metrics for the currently developed taxonomic classification of comets based on primary volatile composition (i.e., H_2O is the "baseline" for the taxonomy, and therefore its accurate modeling is critical).

1.1. Detection of cometary water by fluorescent emissions.

Historically, the high opacities at the core of telluric water lines prevented measurements of water and its isotopologues from ground-based observatories. However, atmospheric transmittance improves rapidly with increasing altitude owing to steadily decreasing water abundance. This fact drove the development of high altitude observatories and ultimately airborne platforms (e.g., the Kuiper Airborne Observatory, KAO), that allowed the exploration of new spectral regions and enabled many unique discoveries. A prime example is the detection and characterization of water vapor in Halley's comet [1,9,10] using the high-resolution Fourier Transform Spectrometer ($\lambda/\Delta\lambda \sim 10^5$) onboard the KAO. This pioneering work not only led to the first measurement of water vapor and its nuclear spin species in a comet, but it also revealed the complexities in the excitation of cometary water.

Detailed models suggested that fluorescence driven by direct solar pumping is the main mechanism responsible for emission by cometary water at infrared wavelengths [11,12]. This emission process enables the search for water at infrared wavelengths using ground-based observatories. As vibrationally excited water molecules cascade downward, they emit photons in fundamental bands and also in "hot-bands" at frequencies where telluric opacities are reduced. Although photons in the emitted fundamental bands are absorbed by atmospheric water in its ground vibrational state, the emitted hot-band lines terminate on higher vibrational levels

with much smaller populations at atmospheric temperatures, and thus they permit measurements of water in remote objects from ground-based observatories.

Spectral lines of water detected in 1P/Halley with KAO included many predicted lines of the targeted ν_3 fundamental band near 2.7 μm, but three lines of a hot-band (011-010, or $\nu_2 + \nu_3 - \nu_2$) were detected unexpectedly [9]. Solar pumping of other hot-bands (100–010; 001-010) was later modeled by [13]. The decision to de-commission the KAO in 1995 created an urgent need for alternative means of water detection from the ground, stimulating the development of a general hot-band method for detecting cometary water. Its first successful application was to the $\nu_2 + \nu_3$ hot-band near 2μm ($\nu_1 + \nu_2 + \nu_3 - \nu_1$, or 111-100) in comets C/1991 T2 (Shoemaker-Levy) and 6P/d'Arrest [14], and then in C/1996 B2 Hyakutake [15]. Three additional hot-bands (100-010, 001-010, and 021-010) were identified later in C/1995 O1 Hale-Bopp [16] and in C/1996 B2 Hyakutake [17], and six more hot-bands were detected in C/1996 H1 (Lee) [18,19].

Today, we measure many such lines simultaneously in a network of ten (or more) vibrational hot-bands, and we perform an intensity analysis that characterizes the rotational temperature for the emitting water population. These developments, in concert with advances in infrared detectors, have allowed measurements of water beyond Earth's atmosphere with unprecedented sensitivity and accuracy.

1.2. Recent model developments and applications to water in comets and on Mars.

Although simple in concept, the retrieval of water abundance from these measurements is far from trivial when considering non-resonant (i.e., "hot-band") fluorescent emission. Computation of line-by-line fluorescence efficiencies (g-factors) entails construction of a full quantum mechanical model for the molecule. This requires a complete characterization of the rotational structure (energy levels) for all vibrational levels involved (both high-energy levels pumped by sunlight, and lower levels involved in the subsequent cascade), along with statistical weights, selection rules, perturbations (e.g., Coriolis effects, splittings, tunneling) and band emission rates. Not only is this task extremely complex, but also information for most hot-bands is not available in community spectral databases (e.g., HITRAN, GEISA). The main driver for these community databases is the precise characterization of our own atmosphere (200-300K), so they often omit information about hot-bands, which are only populated significantly at high temperatures (> 1000K). Consequently, models of hot-band fluorescence (e.g., in the 011-010, or 100-010 bands) relied on the harmonic oscillator approximation to estimate the radiative properties of these bands (e.g., [13]). Until now, this assumption limited the ultimate accuracy of retrieved rotational temperatures, column densities, and production rates for water in comets.

Recent developments in molecular variational techniques for solving the nuclear motion problem have revolutionized the field of molecular spectroscopy. These methods are capable of generating accurate spectral parameters for millions of lines

when considering a precise potential energy surface (PES). Some successful examples include: H_2O [20], HDO [21], NH_3 [22,23], HCN and HNC [24]. Recently, Rothman et al. [25] combined contemporary spectroscopic data with *ab initio* theoretical information for certain molecules, leading to a very complete spectral database for H_2O, CO_2, CO, NO and OH (HITEMP2010, more details for the H_2O component are provided in Section 2.1).

Here, we present a full non-resonant fluorescent cascade model (with realistic solar pumping) for H_2O (considering 500 million lines) and HDO (700 million lines), along with line-by-line computed fluorescence efficiencies for both isotopologues. The new model utilizes an updated set of optimized spectral parameters that draws upon newly obtained experimental and theoretical data (we also corrected several inconsistencies found in existing databases). We integrated the new line list into an advanced radiative transfer model (LBLRTM, used for computing atmosphere transmittances and radiances). We then synthesized spectra for Earth and Mars for comparison with spectra recorded with the high-resolution spectrograph at Keck II (NIRSPEC) atop Mauna Kea, Hawaii. We show selected examples that demonstrate the excellent agreement attained between measurement and model.

As a final validation, we compare measured and modeled fluorescent intensities for water in comets. The new model reproduces the measured intensities of emission lines already identified as H_2O in the cometary spectrum, and it also correctly identifies several additional lines that were previously unassigned. The confidence

limits for retrieved rotational temperatures, column densities, and production rates for water are greatly improved by use of the new model.

2. Spectral databases for water (H_2O and HDO)

2.1 Available compilations

In the early sixties, Gates et al. [26] published a compilation of line parameters and computed spectra for water vapor bands at 2.7 μm. This initiated an era of shared spectroscopic databases that now contain more than 50 molecules and over a hundred isotopologues. In 1973, Garing & McClatchey reported the first compilation of multiple molecules (~100,000 lines, [27]) and in 1983 Rothman et al. [28] reported a molecular database containing ~181,000 lines, which set the foundations of the now widely used HITRAN database (~3 million lines, [29]). A complementary effort was started in 1976 at Laboratoire de Météorologie Dynamique (LMD) in France [30], resulting in the GEISA database, now containing spectral parameters for 50 molecules (111 isotopologues) with almost 4 million lines [31]. Fed by results from numerous laboratory spectroscopists, these compilations have allowed a continuing revolution in remote sensing of planetary atmospheres.

Extrapolation of these line parameters to other environments has not always been straightforward, in particular because the reported broadening coefficients are given for a nitrogen-rich atmosphere and the line completeness is normally restricted to bands that are strong enough at telluric temperatures (< 400K). New

laboratory experiments are now exploring other collision partners such as CO_2 (e.g., [32]), the main atmospheric constituent of Mars and Venus.

Enhancement of these databases to make them applicable at higher temperatures (>1000K) is extremely complex. At higher temperatures many more energy levels are populated, exponentially increasing the complexity of the measured laboratory spectra and ultimately constraining the ability of spectroscopists to extract line-by-line parameters. The solution to this problem emerged from theoretical studies, where the complete molecular motions are modeled and spectral line parameters are synthesized. The first steps for these studies were challenging, leading to spectral parameters of moderate accuracy and limited temperature and dynamic range (e.g., [33,34]). With the advent of powerful computers and better characterized potential-energy surface (PES) and dipole moment surface (DMS) descriptions, these theoretical studies can now synthesize extremely accurate spectral parameters for millions of spectral lines (e.g., [20]). The recent high-energy compilation of water in HITEMP [25], provides a hybrid approach in which the precision of the energy description of the BT2 database is further refined by integrating semi-empirical information present in HITRAN and in other high-temperature linelists (e.g., [35,36]).

As presented in Figure 1, terrestrial databases (HITRAN and GEISA) generally provide a complete description of the main spectral lines for H_2O at 296K, even though they contain only a small subset of the lines contained in BT2 (~eighty per million). We refer to HITRAN and GEISA as terrestrial databases, since they are

mainly intended to synthesize terrestrial spectra (temperatures near 296K and for a N_2 rich atmosphere). For HDO, both terrestrial databases contain practically the same spectral information, but they are relatively incomplete when compared to the *ab initio* VTT database [21]. Recently, a reliable list of energies for HDO became available [37] which combines information from 76 sources, and was done as a collaborative effort to provide reliable spectroscopic parameters for different isotopologues of water by the International Union of Pure and Applied Chemistry (IUPAC). We use this comprehensive survey (HDO-IUPAC) to correct the *ab initio* parameters (see Section 2.2).

The great value of the *ab initio* databases is nicely revealed in Figs. 1 and 2, with the lower panel of Fig. 1 showing densities of up to ~1,000,000 lines per 10 cm^{-1} for *ab initio* databases (in comparison to ~100 per 10 cm^{-1} of terrestrial atlases) and Fig. 2 showing that the *ab initio* databases have almost a complete characterization of the high energies (up to 25,000-30,000 cm^{-1} for both isotopologues). Not surprisingly, the biggest limitations of terrestrial databases appear at higher temperatures where high energy levels become populated. Terrestrial databases are severely limited for lines associated with energies higher than 4000 cm^{-1} (Figure 2). This limitation is particularly problematic when synthesizing spectra for environments with temperatures higher than 1000K, or when computing radiation fields involving non-LTE excitation of high-lying levels and subsequent cascade (e.g., cometary fluorescence). Consequently, synthesis of spectra in these regimes requires a new compilation and validation of information contained in terrestrial and high-energy databases.

2.2 An improved compilation of Einstein coefficients and ro-vibrational energies

Most of the spectral parameters contained in the terrestrial databases have been obtained using extremely high spectral resolutions, and consequently the reported energies and frequencies in these lists are normally of high precision. In contrast, such precision with *ab initio* methods is particularly difficult, owing to micro perturbations in the molecular motion not entirely described by the available potential-energy surfaces (PES). The big advantage of theoretical lists arises from the fact that all lines are computed from a single self-consistent solution, while terrestrial databases are in general a collection of values obtained from diverse experiments employing different calibration techniques which consequently give rise to inconsistencies in these databases. In general, terrestrial databases suffer mainly from inconsistencies in the reported intensities (see section 2.1 (H_2O) in [29]), while *ab initio* databases are impacted by imprecise frequencies and energies.

We also encountered certain discrepancies in the reported symmetries and line identifications for both types of databases (see Tables 1 and 2). We tested the accuracy of reported symmetries (ortho and para, for water) by comparing the reported statistical weights to values computed from the reported ro-vibrational quanta. Ortho (para) levels correspond to an even (odd) sum of the quantum numbers K_a, K_c, and n_3 (the number of ν_3 quanta). The ro-vibrational identifications (ID) were verified by comparing the tabulated total energy of the levels ($E=E_{rot}+E_{vib}+E_{elec}$) with the corresponding vibrational energies (E_{vib}; the vibrational energies were identified by searching the levels with J=0). Lines accessing levels

with energies lower than the corresponding vibrational energy (i.e., negative rotational energies) were flagged as having incorrect quanta, while lines having incomplete upper (or lower) level IDs were flagged as having unknown ID. Most inconsistencies were found in the theoretical databases (see Tables 1 & 2).

The labeling of levels resulting from the *ab initio* solutions is particularly challenging since the theoretical models provide data only for J and the symmetry block of each energy level [20]. Modelers employ complex identification schemes, and more levels can be identified (depending on how specifically these algorithms are defined) but with a higher probability for error. For instance the BT2 line list (with a conservative labeling scheme) contains only modest errors but leaves more than 97% of the lines unidentified, while the more aggressive algorithm considered in the VTT database leaves only 20% of lines unidentified but with an 11% rate of error.

Line IDs are particularly important for non-LTE investigations (e.g., cometary fluorescence) when accounting for rotational/vibrational contributions in cascade, and when comparing and correcting synthetic line parameters with spectral information obtained in the laboratory. Errors in the symmetry labeling are particularly important for our measurements of water spin temperatures in comets, since the symmetry information is necessary for retrieving the ortho-to-para ratio (OPR). Fortunately, we found relatively few inconsistencies in the BT2 database for this parameter (see Table 1).

Combining information from multiple databases is not trivial. A clear set of rules is needed to determine which lines and spectral parameters from each source should be included in the new compilation (e.g., see the discussion of HITEMP, [25]). Our compilation and correction approach emphasized the self-consistency of the spectral parameters. In particular, fluorescence branching ratios must be computed correctly for use in non-LTE fluorescence models, in turn requiring that line intensities and Einstein 'A' coefficients be consistent among the different bands. Consequently, we adopted the Einstein coefficients of the self-consistent BT2 database for our calculations, but corrected the BT2 energies/frequencies with information contained in the terrestrial databases (HITRAN, GEISA) and in the Semi-Empirical Line Positions (SELP, [35,36]) atlas at the core the H_2O-HITEMP atlas. The BT2 database is composed of two components: 1) the energy table containing the energies of all ro-vibrational levels, and 2) a spectral table containing the Einstein-coefficients for transitions between the levels established in the energy table. On the other hand, the GEISA, HITRAN and HITEMP databases report line frequencies but lack information on the global ro-vibrational structure, and therefore different lines accessing the same level may indicate different energies for that level.

We tested the consistency of the energy information contained in the terrestrial databases by computing statistics (standard deviation and weighted mean) on the reported energies of lower (E_{low}) and upper ($E_{up} = \nu + E_{low}$) quantum states for all lines. For ro-vibrational states that appear in multiple terrestrial databases, the listed level energies of most (94-98%) are consistent within 0.01 cm^{-1} or better.

However, the rate decreases to 86% when we require a consistency limit better than 0.001 cm^{-1} (see Tables 1 and 2). By selecting the levels (from the terrestrial databases) having a consistency better than 0.3 cm^{-1}, we corrected the energy tables of BT2 and VTT for all matching ro-vibrational levels.

Because of the problems in the labeling of levels in BT2 and VTT, we define a "match" between a terrestrial and a theoretical level only when both IDs have the same 6 quantum numbers (n_1, n_2, n_3, J, K_a, K_c) and the difference between the empirical and theoretical energies is less than 0.3 cm^{-1} (to avoid correcting mislabeled levels). The quantum numbers n_x indicate the number of quanta of the ν_x vibrational mode (e.g., $3\nu_3$ corresponds to n_3=3). We have found differences in the energies reported in HITRAN, GEISA and in SELP. The energies in SELP and HITRAN are very consistent, but we have found more significant differences between HITRAN/SELP and GEISA. Considering that SELP is more extensive (by almost a factor of two with respect to the terrestrial databases), we consider mainly the SELP energy information for our compilation. Based on the SELP/HITRAN compilation of energies, we corrected the majority of the low energy levels (<10,000 cm^{-1}) in the BT2 database, and obtained excellent agreement between the frequencies of our fluorescence model and cometary data (see discussion in Section 5.1 and Fig. 6). For HDO, we correct the VTT energies using the comprehensive IUPAC survey [37]. Histograms of the corrections (theoretical – empirical) are presented in Figure 2.

Unlike HITEMP, which replaces *ab initio* frequencies by empirically-derived ones in those cases where both upper and lower states are known empirically, we have

additionally substituted experimentally-derived energy levels for the *ab initio* ones in cases where the energy of only one state (almost always the lower) is known empirically. Our approach is based on two premises: empirically-derived energy levels are in general more accurate than computed ones, and systematics are not a significant source of error in the BT2/VTT line lists. The first of these assumptions would appear to be reasonable, having regard to the relative accuracies that are claimed for the experimental and theoretical data. The second premise is supported by Fig. 2, in which the differences between the *ab initio* and empirical energies appear to be quasi-random, indicating that systematics are unlikely to be a major source of error in the computed values.

It is easily demonstrated that the practice of substituting for only the lower (and hence the more accurate of the two levels) will on average remove exactly as much error as it introduces, reducing it in cases where A and B are in error in different directions, and increasing it in those cases where the errors on A and B are in the same direction. Overall improvements in accuracy are achieved by substituting empirical data for the less accurate (normally the upper state) *ab initio* value, but in practice this seldom happens. The main advantage that results when only the lower level of a pair of *ab initio* levels is replaced by an empirically-derived energy is not that it leads to a reduction in the sum of the absolute value of all the errors, but rather that it narrows the spread of the errors, that is to say it reduces the root mean square of the errors of all the transitions in the set and leads to a more self-consistent database, which is important when matching to experiment. The

approach has enabled us to achieve excellent agreement between observation and our fluorescence model in the frequency ranges that we have examined.

3. Modeling of water in cometary atmospheres (non-LTE case)

Water is the most abundant volatile in cometary atmospheres. It radiates at infrared wavelengths via non-resonance fluorescence excited by solar radiation, a full non-LTE process. Collision partners in cometary atmospheres usually lack sufficient energy to excite vibrational transitions and the rate of quenching collisions is much smaller than the radiative decay rates for (infrared active) excited states. Thus, the vibrational manifold is not populated in LTE (local thermodynamic equilibrium). Instead, solar radiation pumps the molecules into an excited vibrational state, which then de-excites by rapid radiative decay (all three vibrational modes of water are infrared active). Infrared photons are emitted through decay to the ground vibrational state, either directly (resonant fluorescence) or through branching into intermediate vibrational levels (non-resonant fluorescence).

For the computation of non-resonance fluorescence, the ro-vibrational structure of the molecule must be very well characterized up to very high energies. With an effective blackbody temperature of 5778K the Sun emits radiation over a wide range of frequencies (see Figure 1), pumping the water molecule into highly excited states. In the case of a cometary atmosphere at 100K, 99% of the Solar pumping occurs via transitions with energies (E_{up}) lower than 7500 cm^{-1}, and thus the spectral databases used for computing fluorescence rates should be fully complete up to this energy limit. However, this requirement is not satisfied for terrestrial databases.

In the hypothetical case of a 296K pumping source, this requirement is greatly reduced to $E_{up} < 1100$ cm^{-1}; levels and transitions for this regime are very well described in HITRAN and GEISA. Before the advent of BT2/VTT and other high-energy spectral databases, fluorescence modelers had assumed that different modes of vibration are not coupled and they ultimately relied on the harmonic oscillator approximation to estimate the strength of the hot-bands from cold-bands (e.g., [13]). This approximation leads to imprecise branching ratios, fluorescence rates and line positions. Nonetheless, it allowed investigators to overcome the limitations of existing spectral databases for two decades, obtaining reasonable results for this complex problem (e.g., [1,13], and references therein).

In the detailed work by Dello Russo et al. [38], the modeling of H_2O fluorescence was advanced by adding rotational branching ratios for four bands (200-100, 200-001, 101-100, 101-001) obtained from the BT2 database, while all the vibrational branching ratios (with the exception of 200 level) were computed considering the Born-Oppenheimer approximation and the latest available spectroscopic information at the time. Because of the limitations of the considered spectral databases and in the assumed solar field (see next paragraph), this previous model did not fully describe the complexity of the water spectrum (Fig. 5), although it allowed retrieval of rotational temperatures and ortho-para-ratios (OPR) from high-resolution spectra with practical accuracy.

An important recent development has been the very accurate measurement of the solar radiation field. Most current cometary fluorescence models often assume the

source for solar pumping is simply a blackbody continuum with the effective temperature of the Sun. This approximation is to some extent correct for the continuum flux at certain wavelengths (2900 – 3300 cm^{-1}), but it introduces important inaccuracies beyond this spectral range (see Figure 5). In addition, the existence of Fraunhofer lines in the impinging solar radiation leads to changes in the fluorescence pumps as a function of the heliocentric velocity of the comet (the Swings effect). Omitting this effect introduces not simply a relative error, but can lead to biased retrievals of rotational and spin temperatures, since these are derived from line-by-line intensity ratios. For the pumping radiation field, we developed a synthesis of the solar spectrum using a combination of empirical parameters from the solar spectrum [39,40] calibrated with a stellar continuum flux model [41]; see Appendix B of [42].

Fluorescence emission rates were computed following a four-step process: 1) we updated the energy tables of BT2 and VTT using values retrieved from the SELP/IUPAC databases (see Tables 1 and 2); 2) we synthesized a high-spectral resolution Solar spectrum at the defined cometary heliocentric velocity; 3) we calculated Solar fluorescence pumps for 500 million lines for H_2O (BT2) and 700 million lines for HDO (VTT); 4) fluorescence emission rates (g-factors) were computed for 1200 million lines (BT2+VTT) considering the appropriate branching ratios for each ro-vibrational level (see details of our General Fluorescence Model (GFM) in [42]). Vibrational pumps and cascades are presented in Figures 3 and 4, while the principal factors involved in computing line-by-line and level-by-level g-factors are presented in Figure 5.

These calculations are extremely intense, owing to the complexity of the computation and the enormous amount of data involved. To increase the computational speed, we parallelized our fluorescence model to run on multi-processor computers, ultimately leading to substantial increases in computational efficiencies. We also explored a reduction in the spectral databases considered. As presented in the cascade figures, pumps and cascades in comets with upper state energies higher than 12,000 cm^{-1} are negligible, since these have band rates lower than 10^{-7} s^{-1}. Introducing this cutoff in energy greatly reduces the size of the database for our calculations, by a factor of ~380 for H_2O (1.3 million lines) and a factor of ~90 for HDO (~8 million lines). We validated this approach by comparing results based on the complete and reduced databases and observed no significant differences in predicted g-factors. These optimizations allow us to compute a complete set of cometary g-factors in a matter of hours, instead of days/weeks. As shown in Figure 5, the new model agrees very well with high-resolution data of comet Boattini obtained with NIRSPEC at Keck II (see detailed discussion in Section 5.1).

4. Modeling of water in planetary atmospheres (LTE case)

When synthesizing spectra of water (H_2O and HDO) on Earth and Mars (at 200-300 K), HITRAN and GEISA generally contain most of the strongest lines at infrared wavelengths. As shown in Figures 3 and 4, HITRAN and GEISA are mostly complete at 296 K for H_2O, although some spectral regions contain no empirical information for HDO. Unlike non-LTE problems, when synthesizing LTE spectra we are mainly

concerned about the "local" consistency and precision of the parameters within a selected spectral range. For instance, missing or inconsistent intensities at a frequency of 5,000 cm^{-1} (i.e., at 2 μm in wavelength) do not affect our LTE modeling at 3000 cm^{-1} (wavelength, 3.33 μm), in contrast to cometary fluorescence emissions originating from cascades of transitions associated with diverse energies and frequencies. Consequently, we use the "globally" consistent BT2 intensities (see previous section) for our computations of fluorescence efficiencies, while for LTE modeling we prefer the highly precise "local" spectral intensities and frequencies contained in HITRAN and GEISA.

For high temperature LTE problems (e.g., exoplanets), HITEMP is the suitable choice since it combines the local precision of HITRAN and other semi-empirical databases with the comprehensiveness of BT2 for high temperature computations. However, HITEMP does not yet contain the VTT information for HDO (see Figure 1). Rothman et al. [25] have also assigned line broadening coefficients to the BT2 lines in HITEMP by extrapolating the procedure in [43] to high rotation-vibration levels. We tested both types of compilations (HITRAN and HITEMP) in our radiative transfer modeling of H_2O on Earth and Mars, and found no significant differences between them. This is expected, as HITEMP adds high-energy spectral lines of H_2O that become significant only at temperatures much higher than those present in the atmospheres of Earth and Mars. Consequently, for the LTE radiative transfer modeling of these atmospheres we will restrict our discussion to the HITRAN spectral database.

4.1 Telluric transmittances and radiances

An accurate spectral database can serve numerous LTE purposes, such as the synthesis of transmittances and radiances of planetary atmospheres by integrating these lists into advanced radiative transfer models. For instance, we compute terrestrial transmittances to assist in the analysis of spectroscopic data collected using ground-based observations (e.g., taken using NASA-IRTF, Keck II, VLT), and in particular, we spectrally calibrate our infrared data by matching the synthetic spectra to the observed telluric spectral radiance.

To that end, we tested the performance of three radiative transfer models (LBLRTM, GENLN2 and SSP). The Spectrum Synthesis Program (SSP, [44]) is a robust model, although with limitations when synthesizing spectra at extremely high spectral resolutions. The latest version (v4) of GENLN2 [45] provides highly realistic and Doppler-limited spectral synthesis of the terrestrial atmosphere, but we encountered problems in the calculation of spectral line shapes since the model incorrectly accounts for pressure spectral-shifts [46]. In addition, GENLN2 is no longer supported by NCAR (National Center for Atmospheric Research). For these reasons we adopted the new efficient Line-By-Line Radiative Transfer Model (LBLRTM, [47]) maintained by AER (Atmospheric and Environmental Research).

We tested the LBLRTM model extensively, and obtained excellent results (e.g., [42]). For the terrestrial layering scheme, we consider the standard tropical profile, modified to describe the local observatory conditions through two temperature parameters (T1 and T2), a pressure-scaling factor (PF) and abundance factors (AF).

The given pressure profile is scaled following P'(z) = P(z) x PF, while the temperature profile is divided into tropospheric (affected by T1) and stratospheric (affected by T2). The abundance profiles are scaled by a molecular multiplier relative to the tropical value [48]. These parameters are fitted for each dataset (typically of 10 minutes) to accommodate for variations in airmass and in properties of the atmosphere (see details in [42]). We have verified that these parameters are consistent (within uncertainties) with local meteorological data (pressure, temperature).

One limitation of LBLRTM, and other radiative transfer models, is that the vertical profiles of isotopologues cannot be parameterized. This is particularly problematic for the heavier isotopes of H_2O (in particular for HDO) since these have lower vapor pressures than standard $H_2{}^{16}O$. This difference in the phase curve of the isotopes introduces important divergences in fractionation, causing the heavier water isotopes to preferentially condense and precipitate ultimately leading to their relative depletion with altitude.

The deuterium abundance is normally quantified relative to the reference value (D/H = 1.5576 x 10^{-4}, VSMOW Vienna Standard Mean Ocean Water) in parts per mil (per thousand, ‰), and consequently (for example) a parcel with 40% of its HDO removed would be described with δD = -400 per mil. The fractionation in the troposphere is also strongly dependent on atmospheric dynamics (e.g., see convective/subsiding results in [49], and formation of clouds and atmospheric

microphysics in [50]), resulting in highly variable deuterium enrichments with respect to altitude, time and position on the planet (e.g., [51]).

In the troposphere, the ratio of HDO to H_2O is controlled mainly by Rayleigh fractionation, which explains the rapid decrease of D/H with increasing altitude in the first 15 km above the surface (Fig. 7). Water is transported from the troposphere to the stratosphere through direct convective injection and clouds [52,53], while additional water is created in the stratosphere by oxidation of CH_4 and H_2 by OH, Cl and $O(^1D)$ [54]. In a similar fashion, HDO is formed in the stratosphere from deuterated methane (CH_3D). Because CH_3D has shorter lifetime than CH_4, the D/H in water increases with altitude above the tropopause (see Figure 7). Information on deuterium fractionation at higher altitudes was lacking until recently, when Sandor & Clancy [55] measured mesospheric HDO and $H_2{}^{18}O$ at radio wavelengths (200-226 GHz), using the 12-m telescope at Kitt Peak (32°N, 112°W). Their measurements revealed extremely high variability in deuteration (δD ranged from -580 to -3 ‰) in the lower mesosphere (50-70 km), which they associated with differential photolysis rates of HDO and H_2O combined with atmospheric transport.

These measurements demonstrate the importance of using realistic vertical profiles for H_2O and HDO when synthesizing terrestrial spectra, especially in spectral regions where both isotopologues are radiatively active (e.g., in the NIR: $2\nu_2$ of H_2O and ν_1 of HDO sampled simultaneously with NIRSPEC). Accordingly, we compiled δD data from several sources [52,55-60], and created a "standard" water fractionation profile (Figure 7). This vertical profile represents the "median" value measured at

each altitude. Considering the previously mentioned high variability of δD, we expect this profile to change significantly with time and space. We integrated this profile into our layer-by-layer radiative transfer model. This is more realistic than simply adopting a constant value (VSMOW) at all altitudes above the tropopause, although, caution is still needed owing to the highly variable ratio of D/H in stratospheric and mesospheric water. On the other hand, the contribution to the total water column from the stratosphere and mesosphere is minimum. For instance if the water column above Mauna Kea (4,200 m) is 5 mm PWV, the contribution from above 14,000 m is only 4 μm PWV (less than 0.1%). Nevertheless, the parameterization of water above the tropopause is important when modeling certain low-energy lines that are stronger at lower temperatures, and when modeling the water spectra as measured from airborne/balloon observatories (e.g. SOFIA).

Because LBLRTM does not permit a separation of the different isotopologues with altitude (layer-by-layer), we integrated the HDO vertical profile into the model by treating HDO as a distinct molecular species in the input spectral database and inside the LBLRTM routines. Apart from a realistic atmospheric description, a complete and accurate spectral database is essential for achieving correct results. The latest HITRAN version contains almost 3 million lines from 42 molecules, yet the database is still insufficient in some spectral regions. Ethane (C_2H_6) provides a good example. Even though C_2H_6 is present in the terrestrial atmosphere at only trace amounts (0.1–2 ppbv), the lines of its strongest band (ν_7) are prominent in high-resolution atmospheric spectra. We thus extended the HITRAN database by

adding 5610 spectral lines of C_2H_6 and 1780 lines of 4 bands of isotopic CO_2 [42,46,61].

We also explored performing fine corrections to the spectral parameters of the strongest lines of water present in HITRAN. For this purpose, we used high signal-to-noise ratio data of Mars collected on 26 April 2010 using NIRSPEC at Keck II. As shown in Figure 8, after removing telluric, solar and Martian lines from our spectrum, the residuals (trace 'd') still show plenty of structure. The majority of the residual structure present in trace 'd' appears to be associated with incorrect modeling of telluric lines (see trace 'a').

In this spectral region, HITRAN2008 contains 213 lines of water, although only 24 lines are needed to describe the main spectral morphology. These 24 lines have reported uncertainties of 10% for their tabulated intensities and of 0.01 cm^{-1} for their center frequencies (Rothman et al. 2009). Typical signal-to-noise ratios exceed 500 for our Mars spectra (and sometimes approach 2000), so our spectral analyses require much higher precisions for intensities (of order 30 ppm) and line frequencies (0.0004 cm^{-1}). Of course these precisions are "relative" within the spectral range, since we do not have an absolute spectral calibrator (aside from the solar lines). By performing a Levenberg-Marquardt optimization study on the line parameters, we retrieved "fine" corrections to the spectral parameters tabulated in HITRAN, all within the reported uncertainties (see Table 3 and final residuals in trace 'f' of Figure 8). This method is particularly advantageous for the analysis of our ground-based observations, since these corrections can be applied to all our

datasets, including those obtained with other instruments (e.g., CSHELL at NASA-IRTF, CRIRES at VLT).

4.2 Planetary transmittances and radiances: the Mars case

Extrapolating the modeling presented in the previous section (section 4.1) to other planetary atmospheres is far from trivial. The layering, geometry, radiative transfer model, and the spectral lists need to be tailored to the completely different conditions present on other planets. In particular, the line broadening parameters contained in HITRAN have been retrieved for typical pressures and temperatures present on Earth and its N_2 rich atmosphere. As demonstrated by laboratory experiments (e.g., [32,62]) these parameters are significantly different for a carbon dioxide atmosphere (e.g., Mars and Venus).

Planetary scientists have used different methods to take this effect into account, some by scaling the N_2-broadening parameters by a constant factor (***1.3***: [63], ***1.6***: [64], ***1.3***: [65], ***1.5***: [66], ***1.7***: [67]), while others (e.g., [68,69]) use the CO_2-H_2O line list presented by [70] for lines in the 200-900 cm^{-1} spectral range. The formalism in [70] considered only the real part of the Robert–Bonamy formulation, but it was recently found that the imaginary terms could change the value of the half-widths by as much as 25% and that the full complex calculations gave much better agreement with laboratory measurements [32,62]. We therefore calculated half-widths and line-shifts for a CO_2 atmosphere (suitable for Mars and Venus) by applying the methodology presented in [62], with the experimental coefficients presented in

[32], assuming that the half-widths are not dependent on vibration state. This approximation should be reasonable for the half-width and its temperature dependence.

We integrated this "adapted" spectral database into our Martian radiative transfer model (MRTM) based on the terrestrial LBLRTM code (see above). The transformation of LBLRTM to synthesize Martian spectra was achieved by disabling the telluric layering scheme contained in the package and by setting up a "user-defined" scheme that considers the correct geometric observing conditions and the necessary corrections that take into account the sphericity of the planet.

We consider 30 layers and a maximum altitude of 100 km for the layering scheme, with layer thickness increasing from 10 meters near the surface to 10 km for the upper atmosphere. Our approach follows a similar scheme used in [71] and that used for water retrievals from MGS/TES data [72] and MEX/PFS data [73]. The optical path of each layer was corrected to compensate for the curved atmosphere relative to the planetary radius (3396 km for Mars) following the methodology described in [74]. Local atmospheric conditions and temperature profiles at each (surface) footprint were estimated using the Mars Climate Database (v4.3, [75]) for the appropriate solar longitude and Martian local time. Effects of dust and ice aerosols were not included in the radiative transfer.

We explored the possibility of integrating the MGS-TES scattering model [66] into our atmospheric model, but the high degree of uncertainty in the input parameters (dust vertical profile, dust properties, dust abundances, ice clouds parameterization,

etc.) prevented us from achieving reliable results. Fortunately, the broad spectral grasp of CRIRES (VLT) and NIRSPEC (Keck II) allowed us to measure H_2O simultaneously with a newly discovered band of CO_2 [61], permitting us to retrieve accurate mixing ratios for H_2O and other trace species at every footprint on Mars (with topographic corrections automatically included). Results using this new model are presented in Figures 8 and 9 (see also Section 5.2).

5. Discussion

5.1 Modeling of water fluorescence (non-LTE): comparison to cometary data

We validated our water fluorescence model by comparing it to infrared data of comet C/2007 W1 (Boattini) acquired in July 2008 using the Near-InfraRed Echelle SPECtrograph (NIRSPEC, [76]) at the Keck II Telescope in Hawaii. In addition to H_2O and HDO, we sampled 11 other gases using three instrument settings (KL1, KL2, MWA). All settings sample water lines, and therefore mixing ratios of trace species (relative to water) are retrieved with high confidence. On July 9, we sampled HDO, CH_3OH, C_2H_6 and H_2O with the KL1 setting, and HCN, C_2H_2, NH_3, CH_4, C_2H_6, H_2CO and H_2O with the KL2 setting. On July 10, we repeated the KL1 setting and then sampled the fundamental band of CO (5 lines, v = 1 – 0) along with H_2O (3 lines) in the MWA setting near 4.7 μm. The cross-dispersed frames were processed and calibrated following our standard methodology for analyzing two-dimensional (spectral and spatial) data [5].

From each spectral order, we extracted nucleus-centered spectra after binning 9 spatial rows along the slit, together spanning ± 0.9 arc-sec or ± ~ 230 km from the nucleus. Rotational temperatures (T_{rot}) were then obtained for water by correlation and excitation analyses [77-79]. This process requires flux measurements for multiple lines spanning a broad range of excitation energies, along with line-by-line fluorescence models tailored for a range of temperatures. For the rotational analysis, we included lines measured in both L-band settings. KL1 samples lines of fairly low rotational excitation that usefully constrain the lower bounds for T_{rot}, while KL2 samples a number of lines of high rotational excitation thereby bounding the high temperature end of acceptable T_{rot} values. From the excitation analysis we also extract the abundance of each spin species, and the ortho-para ratio (see details in [7]).

For each dataset (July 9th and 10th 2008), we computed fluorescence models using the cometary ephemerides at the time of the observations: R_h = 0.893 - 0.899 AU (heliocentric distance), v_h = 9.77 - 10.36 km s^{-1} (heliocentric velocity of the comet), Δ = 0.348 - 0.356 AU (geocentric distance), v_D = 12.91 - 12.98 km s^{-1} (geocentric velocity of the comet). The fluorescence models were then multiplied by telluric transmittances synthesized with the methodology presented in section 4.1. From 26 water lines (10 para and 16 ortho) measured on 9 July, we retrieved a nucleus-centered rotational temperature of 80 ± 2 K and an OPR of 2.87 ± 0.15 (T_{spin} > 34K). Using data taken on both nights (9 and 10 July UT, with a total of 40 minutes of integration), we obtained an upper limit of 8.3 VSMOW (3-sigma) for D/H in water in Comet Boattini (see results for other comets in [5] and refs. therein).

The modeled and measured intensities for individual lines agree remarkably well, and the model successfully describes the general morphology of the water spectrum in both (L-band and M-band) settings (Figure 6). The frequency corrections applied to the BT2 energies (section 2.2) are of critical importance, as shown in the 'b' and 'c' panels. With a high-resolution spectral sampling of 0.04 cm^{-1}/pixel, corrections up to 0.3 cm^{-1} (see Fig. 2) would represent significant shifts up to 7 pixels. These energy corrections, in addition to the use of a self-consistent spectral database, and the introduction of a realistic solar fluorescence pumping, have led to considerable improvements in the agreement between synthesized and measured line-by-line intensities.

When compared to the retrievals with the previous water fluorescence model, the new model increases the precision of the retrieved production rate by a factor or ~2, and leads to an overall correction of ~15% to the previous values. For instance, with the current model we obtain $Q(H_2O) = 116.25 \pm 9.47$, while the previous model resulted in $Q(H_2O) = 133.45 \pm 20.15$ (both in units of 10^{26} molecules s^{-1}). Note the difference in both the values of the retrieved production rates and in the ratio of production rate and error (the confidence limits are improved by about a factor of two by the new model, in this example). The main reason for the improvement in the latter is the improved agreement between predicted and observed line fluxes, with correspondingly smaller confidence limits in the retrieved production rate. For instance, consider the two lines originating from the $200–2_{21}$ level ($200–2_{21}$ to $100–3_{30}$ at 3394.1 cm^{-1}; and $200–2_{21}$ to $001–3_{22}$ at 3372.8 cm^{-1}); the notation is ($n_1\, n_2\, n_3$ – $J\, K_a\, K_c$). Previously these lines were systematically excluded from quantitative

analyses because older models predicted their intensities to be ~1.5 – 2 times weaker than observed (see Figure 2). A more accurate modeling of both the solar pump rate, spectral completeness and ro-vibrational branching ratios resulted in very good agreement with observations, and permitted their inclusion in our retrievals of rotational temperatures, ortho-para ratios, and production rates.

5.2 Modeling of water in planetary atmospheres (LTE): application to Mars data

We validated our model for planetary atmospheres by analyzing NIRSPEC spectra of Mars. Measurements of water on Mars (with a temperature similar to Earth's) are only possible when Mars is at a high relative Doppler-shift (> 10 km s^{-1}), and thus we scheduled our water observations in April 2010 when the relative velocity between Mars and Earth reached its maximum during this apparition. On 26 April 2010, we oriented the NIRSPEC slit (0.144" x 24") East-West over the center of the planetary disk, and sampled mid-latitudes north of Valles Marineris.

The cross-dispersion capability of NIRSPEC allowed us to sample a wide range of frequencies across six spectral orders in the L-band. Order #21 of our KL1 setting sampled multiple lines of HDO (and CO_2) near 3.7 μm, while order #25 sampled multiple lines of H_2O (and CO_2) near 3.1 μm (see Figs. 8 and 9). The simultaneous measurement of CO_2 with the isotopologues of water allowed us to retrieve high-confidence absolute mixing ratios for H_2O and HDO relative to CO_2, since all gases sampled the same mean topography (footprint) on Mars. The combination provided highly accurate D/H ratios for the sampled region on Mars. The fine

corrections applied to the spectral parameters allowed us to obtain high quality residuals for H_2O and HDO, and ultimately for D/H in water (compare traces 'e' and 'f' of Figure 8, and traces 'c' and 'd' of Figure 9). In the presence of significant aerosol opacities (e.g., global dust storm), the retrieved D/H will be only of the column above the aerosol cloud. We can infer the existence of aerosols, by comparing our observed atmospheric columns (derived from the CO_2 measurements) with those predicted by the GCM. In general, we obtain very good agreement with the predicted pressures estimated by GCM models (e.g., [61,80]), implying a general small impact to the derivation of D/H by aerosols.

Considering a spectral extract of data taken on 26 April from the center of the planetary disk (19°N and 55°W, 9:30AM local time on Mars, L_s 82° of Mars Year 30), we retrieved a vertical column density (correcting for observing geometry) of 4.636 ± 0.235 pr-μm for H_2O, and of 10.127 ± 0.367 pr-nm for HDO, leading to a ratio of 7.0 ± 0.4 VSMOW for D/H in water at this position on the planet. The retrievals were based on the residuals presented in Figs. 8 and 9, using spectral parameters of H_2O and HDO adapted to Mars (CO_2 atmosphere) (see Section 4.2). Our localized deuterium enrichment factor (D/H = 7.0 ± 0.4 VSMOW) is larger than the full-disk results of [81] (5.0 ± 0.2 VSMOW) and [82] (5.8 ± 2.6 VSMOW). This difference is actually expected, since the full disk measurements also sample regions of low deuterium as observed by [83] and modeled by [84]. Three-dimensional modeling of the D/H cycle on Mars predicts large depletions of deuterium at high-latitudes, mainly due to the formation of water ice clouds in which HDO is strongly enhanced [84].

This important deuterium enrichment of water on Mars is probably indicative of a significant loss of water, because of the preferential escape of the lighter form of water over geological times. How much water was lost and when this loss mainly occurred are (once again) topics of intense debate. The two main elements used to infer the loss of water over time (D/H ratios in present atmospheric water, and in ancient water from Martian meteorites) are based on highly disputed results. Previous studies of Martian meteorites have shown highly variable D/H and lower values (~2 VSMOW) than current atmospheric values (e.g., [85]). However, Greenwood et al. [86] reported a D/H of 4 VSMOW for the ancient ALH84001 meteorite (4.5—3.9 Ga) and 5.6 VSMOW for the young shergottites (0.17 Ga), and suggested that the earlier measurements may have been biased by significant terrestrial contamination. Greenwood et al. ultimately concluded that Mars lost the majority of its water by 3.9 Ga. However, little is known about the current reservoirs of water on Mars and their D/H content. The fact that we observe strong geographical and seasonal variability of D/H on Mars [83,87] may indicate multiple water reservoirs of varying sizes, that gain and lose water to the atmosphere as functions of time [88].

6. Conclusions

We developed a comprehensive methodology to measure water (H_2O) and deuterated water (HDO) in diverse environments and under a wide range of excitation conditions. The models make use of the latest spectroscopic databases (of both empirical and theoretical origins) for H_2O and HDO, and they incorporate corrections retrieved from high-resolution spectra obtained using NIRSPEC at Keck

II. Because most observations of water in the Universe are performed using ground-based telescopes, precise modeling of telluric transmittances is of crucial importance. The new model incorporates the strong vertical fractionation of deuterated water found in our own atmosphere, which was not considered in previous terrestrial radiative transfer models.

We addressed this important issue by integrating our spectral compilation of water, CO_2 and C_2H_6 lines into an advanced radiative transfer model of the terrestrial atmosphere (LBLRTM), which features a rigorous line-by-line, layer-by-layer radiative transfer analysis and includes realistic atmospheric conditions, abundance and isotopic profiles, and geometrical conditions. The terrestrial model was also extended for use on Mars by modifying the layering scheme and by calculating half-widths and line-shifts for a CO_2 atmosphere using the complex Robert-Bonamy formalism.

Modeling of water in non-LTE environments was achieved by developing a full non-resonance fluorescence model with cascade (500 million H_2O lines and 700 million HDO lines). This model utilizes a novel approach to synthesize the solar pump and marries the comprehensiveness of the BT2 database with the spectral precision of semi-empirical databases (SELP/IUPAC/HITRAN/GEISA). Energy tables and other lists are available at http://astrobiology.gsfc.nasa.gov/Villanueva/spec.html.

We successfully validated the methodology by retrieving water and D/H on Mars, in C/2007 W1 (Boattini) and on Earth, from data obtained using the high-resolution spectrograph NIRSPEC/Keck II atop Mauna Kea (Hawaii). The new model accurately

describes the complex morphology of the water bands and greatly increases the accuracy of the retrieved abundances (and the D/H ratio in water) with respect to previously available models. We expect that this newly developed methodology for retrieving H_2O and HDO in planetary atmospheres will assist in unraveling the true history of water in our Solar System and beyond.

Acknowledgements

GLV acknowledges support from NASA's Planetary Astronomy Program (08-PAST08-0034) and NASA's Planetary Atmospheres Program (08-PATM08-0031). NASA's Planetary Astronomy Program (RTOP 344-32-07) and NASA's Astrobiology Program (RTOP 344-53-51) supported MJM. BPB acknowledges support from the NSF Astronomy and Astrophysics Research Grants Program (AST-0807939). NSF-RUI supported REN through Grant (AST-0805540). We thank Dr. Alan Tokunaga and Alain Khayat for assisting with the acquisition of Mars data in 2010, and Dr. Jonathan Tennyson and Dr. Iouli Gordon for providing critical assistance in the interpretation of the HITEMP and SELP databases. We greatly acknowledge the valuable insights of the reviewers that led us to an improved manuscript. We thank the staff of the W. M. Keck Observatory (operated as a scientific partnership among CalTech, UCLA, and NASA) for their exceptional support throughout our long Mars and cometary observing Programs. The authors wish to recognize and acknowledge the very significant cultural role and reverence that the summit of Mauna Kea has always had within the indigenous Hawaiian

community. We are most fortunate to have the opportunity to conduct observations from this mountain.

1

Table 1. Summary of spectral lines and energy level information retrieved from the four principal repositories of spectroscopic parameters for water (H_2O).*

	GEISA	%	HITRAN	%	BT2	%	HITEMP	%
Number of H_2O lines	**41,447**	100.0	**37,432**	100.0	**505,806,255**	100.0	**114,209,395**	100.0
Flagged symmetry (a)	N/A	N/A	0	0.0	84,699	0.0	13,507	0.0
Flagged line ID (b)	69	0.2	69	0.2	5,423	0.0	677	0.0
Unknown line ID (c)	2,422	5.9	882	2.4	491,191,064	97.1	109,568,331	95.9
Number of levels	**8,072**	100.0	**8,762**	100.0	**221,097**	100.0	**28,598(f)**	100.0
Energy consistency (d)								
better than 0.3 cm^{-1}	7,841	97.1	8,674	99.0				
better than 0.1 cm^{-1}	7,792	96.5	8,670	99.0				
better than 0.01 cm^{-1}	7,686	95.2	8,598	98.1				
better than 0.001 cm^{-1}	6,931	85.9	7,581	86.5				
BT2 energy levels (e)								
corrected using GEISA					**7,044**	3.2		
corrected using HITRAN					**7,486**	3.4		
corrected using SELP(f)					**11,989**	5.4		

*The four repositories are: GEISA [31], HITRAN [29], BT2 [20], HITEMP [25].

(a) Incorrect symmetries were determined by comparing the reported symmetry (statistical weight in HITRAN and HITEMP) and the corresponding symmetry derived from the reported level quanta (K_a, K_c and n_3).

(b) Lines accessing levels with energies lower than the corresponding vibrational energy (i.e. negative rotational energies) were flagged as having incorrect quanta.

(c) Lines where the upper or the lower level was not fully known were flagged as having unknown ID.

(d) GEISA and HITRAN are repositories of line positions (not of energy levels), and thus different lines accessing the same line report slightly different energies. The consistency of the energy information contained in these terrestrial databases was tested by computing statistics (standard deviation and weighted mean) on the reported lower and upper energies of all lines.

(e) The theoretically calculated energies of BT2 were corrected using semi-empirical information when the difference between the empirical and theoretical values was lower than 0.3 cm^{-1}.

(f) The majority (55% = 15,606/28,598) of the H_2O energy information of HITEMP is based on the semi-empirical line position (SELP) atlas that contains 15,606 energy levels. SELP is based on the database of experimental energy levels reported by Tennyson et al. [35], and includes recent updates obtained using high-temperature experiments (e.g., [36]).

2

1

Table 2. Summary of spectral lines and energy level information retrieved from the four principal repositories of spectroscopic parameters for deuterated water (HDO).*

	GEISA	%	HITRAN HITEMP	%	VTT	%
Number of HDO lines	**11,980**	100.0	**13,238**	100.0	**697,450,825**	100.0
Flagged symmetry (a)	N/A	N/A	0	0.0	N/A	N/A
Flagged line ID (b)	1	0.0	1	0.0	79,042,333	11.3
Unknown line ID (c)	147	1.2	2	0.0	145,936,238	20.9
Number of levels	**2,050**	100.0	**3,350**	100.0	**163,491**	100.0
Energy consistency (d)						
better than 0.3 cm^{-1}	2,043	99.7	3,350	100.0		
better than 0.1 cm^{-1}	2,037	99.4	3,349	100.0		
better than 0.01 cm^{-1}	1,959	95.6	3,283	98.0		
better than 0.001 cm^{-1}	1,820	88.8	2,871	85.7		
VTT energy levels (e)						
corrected using GEISA					**1,534**	0.9
corrected using HITRAN					**2,539**	1.6
corrected using IUPAC					**5,287**	3.2

* The three main repositories for HDO are: GEISA, HITRAN, VTT [21] and the IUPAC HDO survey [37].

Labels (a), (b), (c), (d) and (e) are as described in Table 1.

2

1

Table 3. Corrections applied to the HITRAN database, based on telluric transmittances observed with NIRSPEC at Keck II (Mauna Kea, Hawaii).*

Upper v_1 v_2 v_3	Upper J Ka Kc	Lower J Ka Kc	Line Strength S_{OLD}	Line Strength S_{NEW}	Diff (%)	σ (%)	ν_{OLD} (cm^{-1})	ν_{NEW} (cm^{-1})	Diff (cm^{-1})	σ (cm^{-1})
0 2 0	3 2 1	3 1 2	124.70	123.60	-0.874	0.002	3219.3835	3219.3828	-0.0008	0.0002
0 2 0	4 2 2	4 1 3	35.90	36.10	0.569	0.003	3220.4421	3220.4430	0.0009	0.0002
0 2 0	2 2 0	2 1 1	31.14	31.88	2.365	0.002	3222.0347	3222.0371	0.0024	0.0002
0 2 0	7 2 5	8 1 8	0.54	0.48	-9.850	0.019	3223.3258	3223.3260	0.0003	0.0018
1 0 0	7 1 6	8 4 5	4.00	4.07	1.709	0.005	3225.7061	3225.7105	0.0044	0.0004
1 0 0	7 0 7	8 3 6	4.00	4.01	0.280	0.005	3226.0686	3226.0719	0.0033	0.0004
0 2 0	5 2 3	5 1 4	67.37	61.96	-8.026	0.002	3227.4646	3227.4678	0.0032	0.0002
0 2 0	3 1 3	2 0 2	61.72	62.85	1.831	0.002	3229.9003	3229.9021	0.0018	0.0002
0 0 1	8 4 4	9 6 3	3.28	3.32	1.086	0.009	3230.4202	3230.4210	0.0009	0.0007
1 0 0	7 7 0	8 8 1	3.53	3.63	2.950	0.010	3230.9833	3230.9849	0.0016	0.0005
0 2 0	4 1 3	3 2 2	10.65	10.92	2.510	0.003	3232.2735	3232.2744	0.0009	0.0002
0 2 0	4 0 4	3 1 3	52.85	53.02	0.314	0.002	3233.0194	3233.0174	-0.0020	0.0002
0 2 0	2 2 1	2 1 2	71.41	72.88	2.065	0.002	3236.6489	3236.6480	-0.0009	0.0002
0 2 0	5 1 4	5 0 5	45.59	42.92	-5.856	0.002	3240.1067	3240.1054	-0.0013	0.0002
0 2 0	6 2 4	6 1 5	10.89	10.79	-0.950	0.003	3241.7733	3241.7739	0.0006	0.0002
1 0 0	8 2 7	9 3 6	4.36	4.31	-1.181	0.005	3243.0450	3243.0464	0.0014	0.0005
0 2 0	5 2 3	4 3 2	7.14	7.11	-0.431	0.005	3244.4053	3244.4054	0.0001	0.0003
0 2 0	4 1 4	3 0 3	181.50	184.30	1.524	0.003	3244.9426	3244.9428	0.0002	0.0005
0 2 0	3 2 2	3 1 3	26.79	27.60	3.014	0.007	3245.4021	3245.4037	0.0016	0.0004
0 0 1	7 4 4	8 6 3	5.33	5.31	-0.352	0.005	3247.3633	3247.3638	0.0005	0.0004
0 2 0	5 0 5	4 1 4	145.60	146.10	0.375	0.002	3254.1481	3254.1465	-0.0016	0.0002
1 0 0	7 6 1	8 7 2	7.58	7.29	-3.914	0.006	3256.0857	3256.0886	0.0029	0.0003
0 2 0	4 2 3	4 1 4	63.34	62.45	-1.401	0.002	3257.2261	3257.2267	0.0006	0.0002
1 0 0	8 5 4	9 6 3	3.90	3.87	-0.797	0.006	3258.0741	3258.0752	0.0011	0.0005

*All lines originate from the ground-vibrational state, and have reported uncertainties of 10% (error code 5, see Rothman et al. 2005) for the intensities, and 0.01 cm^{-1} for the frequencies (error code 3). The line intensities (S_{OLD}, S_{NEW}) are given in units of 10^{-23} cm^{-1}/(molecule cm^{-2}) at 296K. All required corrections are within 1σ of the HITRAN reported uncertainties. See retrieval results in Figure 8.

2

1

Table 4. Water on Earth and Mars, and in comet C/2007 W1 (Boattini).

Molecule	Column abundance	D Enrichment Factor, w.r.t. VSMOW [a]
Mars – Water column at 19°N and 55°W (9:30AM) at L_s 82° (MY 30), 26 April 2010		
H_2O (water)	4.636 ± 0.235 pr-μm	7.01 ± 0.44
HDO (deuterated water)	10.127 ± 0.367 pr-nm	
Earth – Water column above Mauna Kea (Hawaii) at 4,145 m, 26 April 2010 [b]		
H_2O (water)	1.304 ± 0.070 pr-mm	0.52 ± 0.03
HDO (deuterated water)	0.212 ± 0.017 pr-μm	

Isomer / isotopologue	Production rate (molecules s^{-1})	Production rate (molecules s^{-1}) and OPR	D Enrichment Factor, w.r.t. VSMOW [a]
Comet C/2007 W1 (Boattini), 9 July 2008, NIRSPEC at Keck II [c]			
H_2O – Ortho	(0.89 ± 0.02) x 10^{28}	**(1.20 ± 0.02) x 10^{28} s^{-1}**	< 8.3 (3σ)
H_2O – Para	(0.31 ± 0.01) x 10^{28}	OPR = 2.87 ± 0.15	
HDO	< 2.73 x 10^{25} (3σ)		

(a) The Vienna Standard Mean Ocean Water (VSMOW) is the standard defining the water isotopic composition of Earth's oceans - $(D/H)_{VSMOW}$ = 155.76 ppm.

(b) The reported uncertainties for the columns and isotopic ratios reflect the level of variability of the water columns during the observations on 26 April 2010 (05:30 - 08:42 UT) using NIRSPEC at Keck II atop Mauna Kea. Systematic uncertainties originating from the assumed Martian atmospheric conditions (temperature/pressure estimated by the MCD model) are smaller (<3%) than the intrinsic precision of the measurement.

(c) Details on the chemical composition and retrievals for comet Boattini are presented in [79].

2

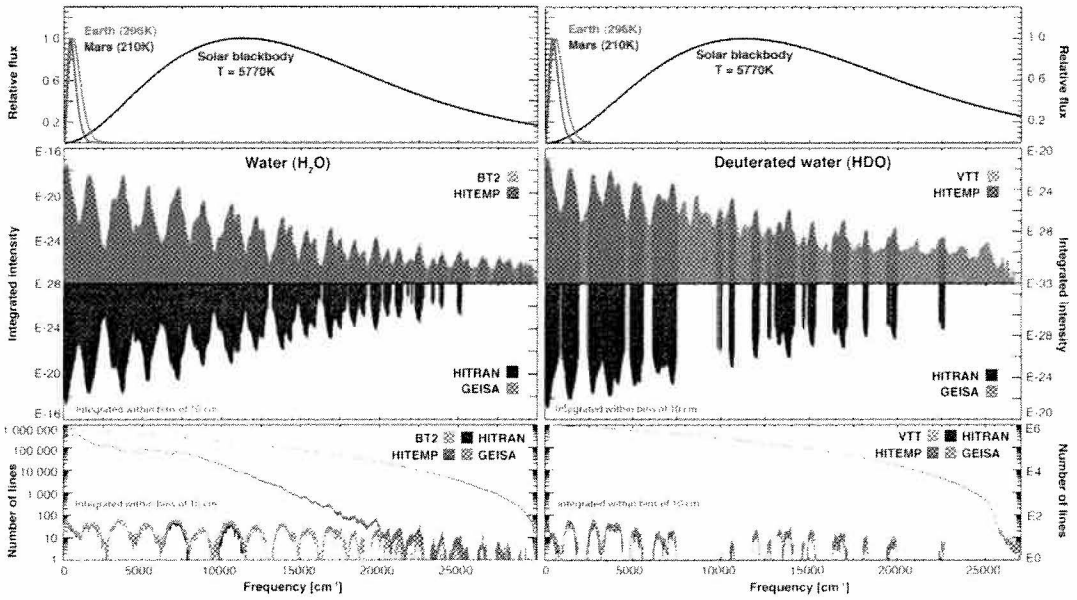

Figure 1. Comparison of the five principal spectroscopic databases of water (H_2O; left) and of deuterated water (HDO; right). The upper panel shows normalized blackbody radiation curves for characteristic temperatures of Mars, Earth and the Sun. The middle panel shows integrated line intensities (cm^{-1}/(molecule cm^{-2})) for the four databases within 10 cm^{-1} spectral bins at a temperature of 296K. The lowest panel shows number of lines within 10 cm^{-1} spectral bins.

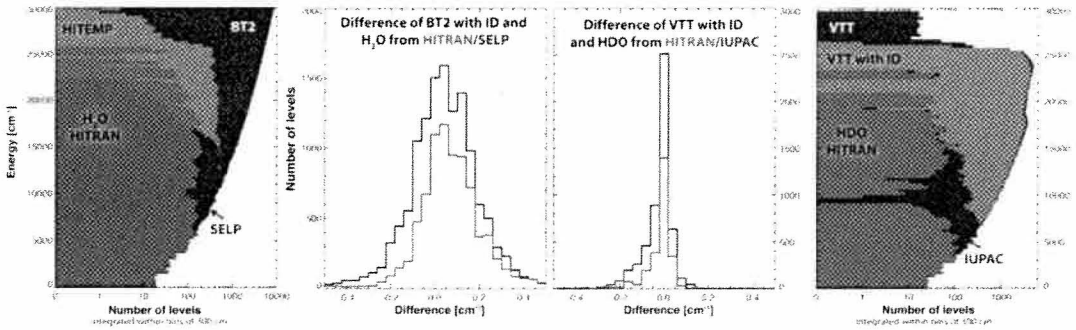

Figure 2. Comparison of the energy-level characterization of HITRAN, SELP, HITEMP, BT2, VTT and HDO-IUPAC databases. The leftmost and rightmost panels show the level density from 0 to 30,000 cm^{-1} considering 300 cm^{-1} bins for H_2O and HDO respectively. The middle panels show histograms of the difference in energy between BT2/VTT and semi-empirical values where the databases have matching ro-vibrational identifications (see text for details). The theoretically calculated energies of BT2 and VTT were corrected used information from the semi-empirical databases when the difference was lower than 0.3 cm^{-1} (dotted lines) – see Tables 1 and 2.

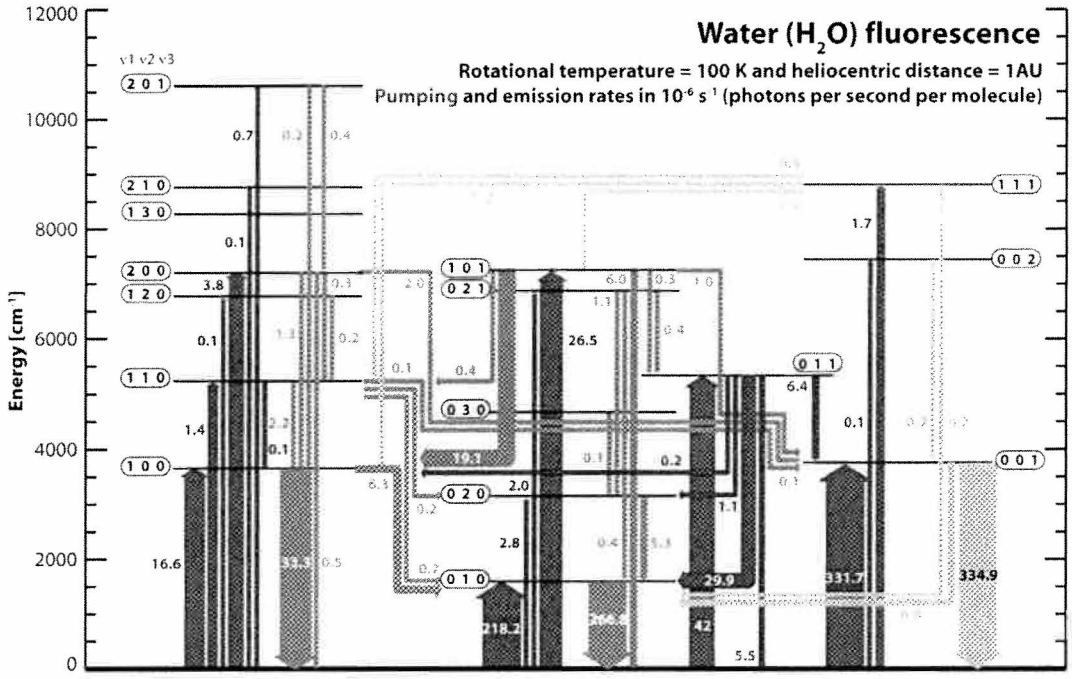

Figure 3. Diagram showing full non-resonance fluorescence for H_2O in a comet at 1 Astronomical Unit (AU) and with a rotational temperature of 100 K. The pumping rates (shown in blue) were calculated considering a realistic Solar model, and the emission rates (shown in red/green/purple/yellow colors) were calculated by subsequent cascade down to the ground-vibrational level and considering line-by-line and level-by-level branching ratios which take into account all 500 million transitions. Only pumps/emissions with vibrational rates higher than 10^{-7} s^{-1} are shown.

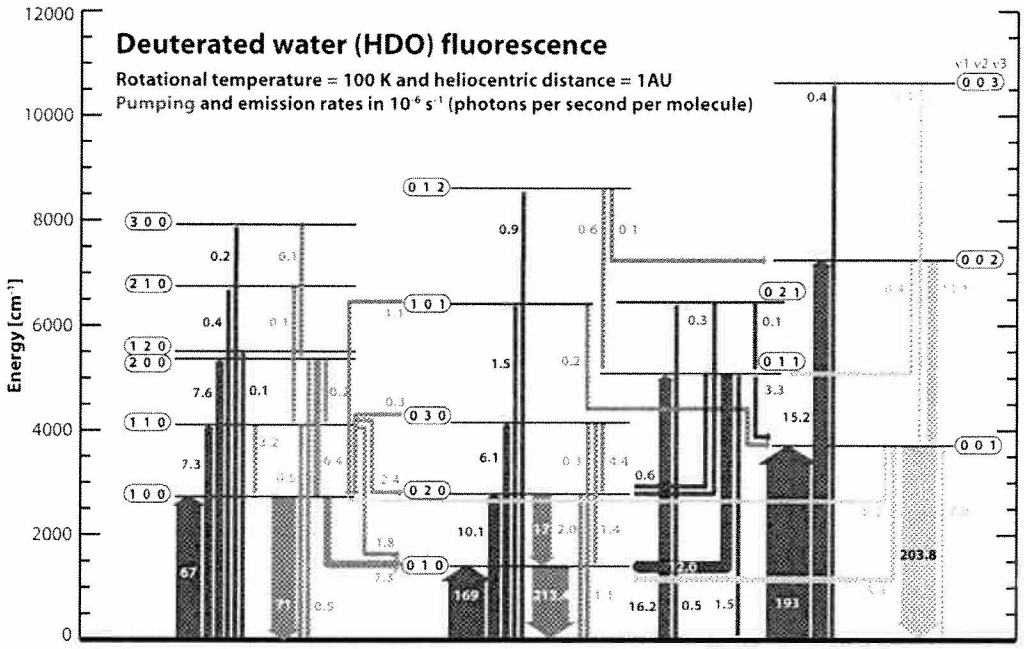

Figure 4. Diagram showing full non-resonance fluorescence for deuterated water (HDO) in a comet at 1 Astronomical Unit (AU) and with a rotational temperature of 100 K. Only pumps/emissions with rates higher than 10^{-7} s^{-1} are shown, and the rates are in photons per second per HDO molecule.

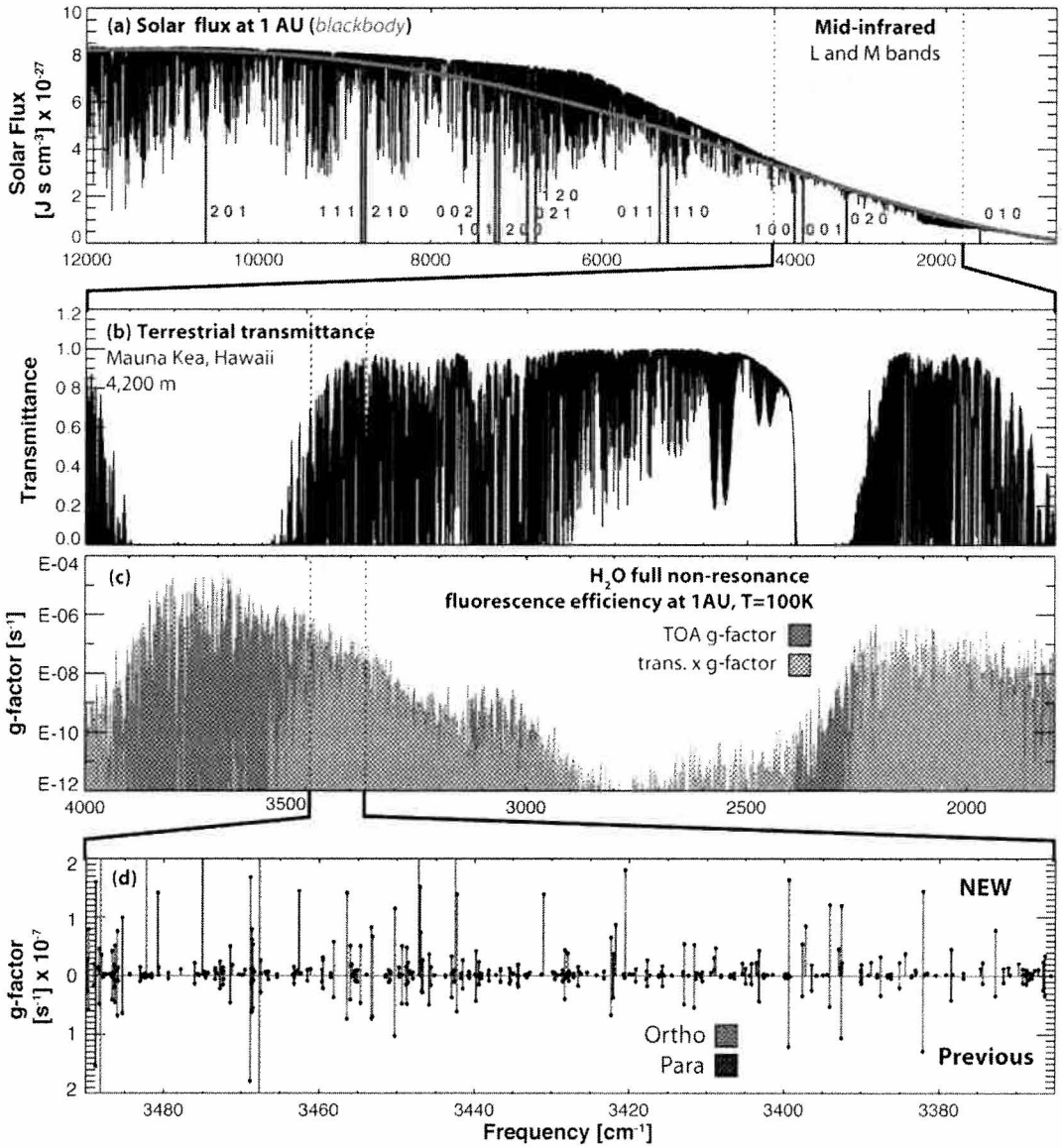

Figure 5. Diagram showing the main elements involved in the modeling of the water fluorescence emission. (a) Solar spectrum, which is a combination of a theoretical continuum model and a highly precise solar line list. The red trace shows the historically considered fluorescence pump for comets (a 5778 K blackbody), and the blue vertical traces the position of the main H_2O pumps presented in Figure 3. (b) Atmospheric transmittance atop Mauna Kea, synthesized using LBLRTM [42,47], showing strong H_2O and CO_2 absorptions at 3700 cm^{-1} and 2350 cm^{-1}. (c) Stick spectrum of the newly developed water fluorescence spectrum, where the 'red' trace indicates fluorescence efficiencies observable only at the Top Of the Atmosphere (TOA), and 'green' trace indicates the observable flux with a telescope atop Mauna Kea (note the logarithmic ordinate scale). (d) A comparison of the new model and the previous model [38] for a rotational temperature of 100K.