A global, high accuracy *ab initio* dipole moment surface for the electronic ground state of the water molecule

Lorenzo Lodi,[1] Jonathan Tennyson,[1,a)] and Oleg L. Polyansky[2]
[1]*Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, United Kingdom*
[2]*Institute of Applied Physics, Russian Academy of Sciences, Uljanov Street 46, Nizhniy Novgorod, Russia 603950*

(Received 12 May 2011; accepted 8 June 2011; published online 20 July 2011)

A highly accurate, global dipole moment surface (DMS) is calculated for the water molecule using *ab initio* quantum chemistry methods. The new surface is named LTP2011 and is based on all-electron, internally contracted multireference configuration interaction, including size-extensivity corrections in the aug-cc-pCV6Z basis set. Dipoles are computed as energy derivatives and small corrections due to relativistic effects included. The LTP2011 DMS uses an appropriate functional form that guarantees qualitatively correct behaviour even for most high energies configuration (up to about 60 000 cm^{-1}), including, in particular, along the water dissociation channel. Comparisons with high precision experimental data show agreement within 1% for medium-strength lines. The new DMS and all the *ab initio* data are made available in the supplementary material. © *2011 American Institute of Physics*. [doi:10.1063/1.3604934]

I. INTRODUCTION

The foremost importance of the rotational-vibrational spectrum of the water molecule in many areas of science, including atmospheric physics and astrophysics has long been established.[1,2] Accurate knowledge of water spectra is not only required for models of global radiative transport and the earth's energy budget, but are also essential for more detailed studies. For example, retrievals of column densities and profiles of other species by remote sensing satellites, such as SCISAT-1 and ENVISAT, are strongly dependent on the ability to make reliable models of water spectra in the windows studied. The demands of modern remote sensing satellites require water line intensities to 1% or better. This demand is not primarily for monitoring water columns, but because water absorbtion is so ubiquitous that its lines interfere with many other retrievals that are being attempted or planned. Failure to model water absorption accurately thus introduces a major source of error into these observations, severely degrading their usefulness. This level of accuracy, which is our aim here, is hard to achieve experimentally, except for specially designed, high-precision experiments which can only cover a limited spectral range.

Another source of motivation for the construction of a new water global dipole moment surface (DMS) is given by recent studies probing the spectrum of water up to dissociation[3–5] and even beyond it.[6] Theoretical models of such studies[7,8] require a global dipole moment surface.

Line intensities can be calculated within the Born-Oppenheimer approximation following a well-established two-step procedure.[9] In the first step the molecular potential energy surface (PES) is used to compute nuclear-motion rotational-vibrational wavefunctions. Then, in the second step, line intensities are obtained by computing the matrix element of the molecular DMS between two nuclear-motion wavefunctions. While the representation of the PES is often improved using spectroscopic data, tests suggest that the most reliable DMSs are produced *ab initio*.[11]

In the case of water very accurate *ab initio*[5,7,12] and semi-empirical[13–15] PESs are available and the error in computed line intensities is determined to a large extent by the molecular DMS, which therefore plays the key role in the whole procedure. Several water DMSs have been published, the most recent and accurate being those of Refs. 16–21. We refer to Refs. 10, 16, and 18 for more background on the theoretical calculation of the water rotational-vibrational spectrum.

In this work we present a new, global, *ab initio* DMS for water which we test against both high precision measurements[24] and the latest version of the HITRAN database.[25] In contrast to the earlier DMSs referred to above, this DMS is based upon an *ab initio* approach explicitly directed to produce a high accuracy DMS rather than a PES.

II. CALCULATION OF THE *AB INITIO* DATA

A. Electronic structure method

The electronic structure method employed is internally contracted multi-reference configuration interaction (IC-MRCI) as implemented in the program MOLPRO.[26] All ten electrons were correlated. The acronyms IC-MRCI+Q and IC-MRCI+P will be used to indicate, respectively, the (renormalised) Davidson-corrected and Pople-corrected energies (relaxed-reference values were used). For an overview of the electronic structure methods used in this work and references to the original papers, see Ref. 10.

An 8-electron, 10-orbital complete active space was used, comprising the first lowest-energy 8 a' and 2 a'' orbitals

a)Electronic mail: j.tennyson@ucl.ac.uk.

(C_s symmetry labels; for C_{2v} symmetries the specification is 5 a_1, 3 b_1, and 2 b_2 orbitals). This reference space will be referred to as [8, 10]. The lowest-energy a' orbital, identifiable as the oxygen 1*s* atomic orbital, was correlated at the IC-MRCI stage but was not included in the active space. Although this IC-MRCI treatment was explicitly developed to compute the DMS presented here, the corresponding water PES, which is a byproduct of the calculation, has already been used with excellent results.[5,7] For example, the IC-MRCI+Q dissociation energy, 41 108 cm^{-1}, was estimated[7] to be only about 38 cm^{-1} lower than the experimental value of 41 145.94 $\pm$ 0.15 cm^{-1}.[3]

Other tests performed during the course of this work also point to the high accuracy of our chosen IC-MRCI treatment. Such tests include comparison with full configuration interaction and high-order (up to fifth) coupled cluster energies for key geometries, calculation of the water equilibrium dipole moment (see Tables 3 and 4 of Ref. 10), and comparison of the equilibrium geometry and equilibrium force constants (see Sec. 3.1 of Ref. 7).

B. Details on the computation of the dipoles

Dipole moments may be computed in two ways. In the most straightforward approach molecular dipole moments are computed as the expectation value of the dipole-moment operator; alternatively, they may be computed as the derivative of the energy, considered as a function of an external, weak, uniform electric field, for zero field strength. If the exact wave function and energy were known, the two methods would yield exactly the same value. In the actual case of approximate wave functions and energies, dipoles computed as expectation values (XP) or as energy derivatives (ED) usually differ, and there is a good reason to believe that ED dipoles are more accurate than the corresponding XP ones, see Sec. 3.12 of Ref. 10 and references therein. It was therefore decided to employ ED dipoles as a basis for our new DMS.

The necessary energy derivatives were computed using the two-point central-difference formula

$$\mu_\alpha = \frac{E(\lambda_\alpha) - E(-\lambda_\alpha)}{2\lambda_\alpha} + \mathcal{O}(\lambda_\alpha^2), \qquad (1)$$

where the external field strength λ_α ($\alpha = x, y$; see Sec. III for a precise specification of the orientation of the axes) was fixed to $\lambda_\alpha = 3 \times 10^{-4}$ a.u. This value was determined as optimal by comparing complete active space self-consistent field (CASSCF) XP and ED dipoles, as for this particular method XP and ED dipoles coincide exactly and discrepancies are due to numerical error in the ED value. The chosen value for λ_α should guarantee a numerical derivative error of less than 10^{-5} a.u. for all geometries.

A significant advantage arising from the use of ED dipoles is that it is in this case possible to use size-consistency-corrected energies, i.e., IC-MRCI+Q or IC-MRCI+P. The drawback of the ED approach is that four independent runs are necessary to determine both components of the dipole.

Relativistic corrections to the dipoles were obtained in a similar manner by computing the derivatives with respect to the external electric field strength of the mass-velocity, one-electron Darwin (MVD1) relativistic corrections to the IC-MRCI energies. More specifically, relativistic corrections are computed by MOLPRO as expectation values of the MVD1 Hamiltonian.

The grid of geometries for which dipoles were computed was devised by starting from an older one used in previous studies[12,16] and supplementing it with further points in a quasi-random fashion. Candidate new (r_1, r_2, θ) values were generated at random and then accepted into the extended grid if, first, they lay in the desired energetic region and, second, if they lay further than a given threshold from points already in the grid.

For each geometry a total of 7 independent IC-MRCI runs were performed: 4 runs in the presence of electric field in the $\pm x$ and $\pm y$ directions in the aug-cc-pCV6Z basis set to obtain the finite-difference dipoles and 3 runs with no external field in the aug-cc-pCV*n*Z (n = Q, 5, 6) basis sets. Each aug-cc-pCV6Z run took about 90 GB of disk and 12 h real time on a workstation with two dual-core 3.00 GHz Intel Xeon 5160 central processing units and 4 GB of RAM.

By using the wealth of data available for each geometry it was possible to extensively cross-check the calculated energies against possible problems. For example, plotting the difference between expectation-value and finite-difference IC-MRCI dipoles as a function of the energy identified problems for about 200 geometries, which were dropped from further analysis. Finally, a data set corresponding to 2628 geometries, which passed all the tests, was obtained and used to build the DMS. All points were computed using C_s symmetry.

C. Surface intersections

The ground $\widetilde{X}\ ^1A_1$ electronic state of water goes though a so-called conical intersection for linear geometries of the H–O–H type when one of the OH bonds is stretched more than about 3 a_0. More information including correlation diagrams of low-lying electronic states and plots of the surface intersection can be found in Refs. 27–29.

When the intersection is crossed the lowest eigenvalue of the Hamiltonian, which is the quantity we compute, experiences a derivative discontinuity and the corresponding dipoles are discontinuous. This surface intersection occurs at energies higher than about 42 000 cm^{-1} and its effects are visible in some of our computed dipoles. For our present purposes we are not interested in such high-energy configurations and, as detailed in Sec. III, the corresponding dipoles were not included in the fit. It was therefore possible at this time to pragmatically ignore the issues related to this surface intersection.

We will mention for completeness that the water ground electronic state experiences two surface intersections for very high energies: when dissociating to H_2 + O and for linear geometries of the HHO kind. Our *ab initio* dataset does not include any of such configurations.

D. Estimation of the uncertainty of the computed dipoles

It is interesting to analyze the errors in the computed *ab initio* dipoles due to the various approximations introduced.

To put things into perspective, let us first mention some features of the computed water dipoles. The ED, aug-cc-pCV6Z dipoles evaluated at the equilibrium geometry[22] $r_e = 1.81002$ a_0, $\theta_e = 104.485^\circ$ are 0.72980 a.u. for IC-MRCI+Q [8,10] and 0.72992 a.u. for IC-MRCI+P [8,10] (see p 147 of Ref. 23). The magnitude of the modulus of dipoles varies from zero (for linear symmetric geometries of the H–O–H type) up to a maximum of about 1.0 a.u. for small-angle geometries with $\theta \approx 40^\circ$. The modulus of the relativistic correction is of 1.7×10^{-3} a.u. at equilibrium and varies between zero and about 3×10^{-3} a.u. The Born-Oppenheimer diagonal correction (also known as adiabatic correction) at equilibrium was computed by Hobson *et al.*[30] and amounts to 8×10^{-4} a.u. Because of its small magnitude, it was not included in the present study.

We now proceed to analyze the error due to the various approximations, in increasing order of accuracy. The modulus of the differences of dipoles computed in two specified approximations A and B, $|\vec{\mu}_A - \vec{\mu}_B|$, is reported in Fig. 1. Acronyms of quantum chemistry methods will be appended with ∼XP or ∼ED to indicate that dipoles were computed as expectation values or as energy derivatives. Unless otherwise specified, dipoles were computed in the aug-cc-pCV6Z basis set and using the [8, 10] complete reference space.

1. Dynamical electron correlation (no plot)

For energies up to about 10 000 cm^{-1} CASSCF dipoles differ from IC-MRCI ∼XP ones by 6–8 $\times 10^{-3}$ a.u. For higher energies the effect of dynamical correlation rises quickly and differences often reach up to 2×10^{-2} a.u. for energies up to 20 000 cm^{-1} and can be as high as 6×10^{-2} a.u. for energies up to 40 000 cm^{-1}.

2. Method of computation of the dipoles – plot (a)

IC-MRCI dipoles computed by XP and ED differ by about 3×10^{-3} a.u. near equilibrium and typically up to about 6×10^{-3} a.u. for higher energies.

3. Effect of size-extensivity corrections – plot (b)

Dipoles based on Davidson-corrected energies differ with respect to uncorrected IC-MRCI∼ED dipoles by 2×10^{-3} a.u. and up to about 4×10^{-3} a.u. for higher energies.

4. Basis-set incompleteness error – plot (c)

IC-MRCI∼XP dipoles computed in the aug-cc-pCV5Z and aug-cc-pCV6Z basis sets show differences of 5 $\times 10^{-4}$ a.u. near equilibrium and up to about 1.5×10^{-3} a.u.

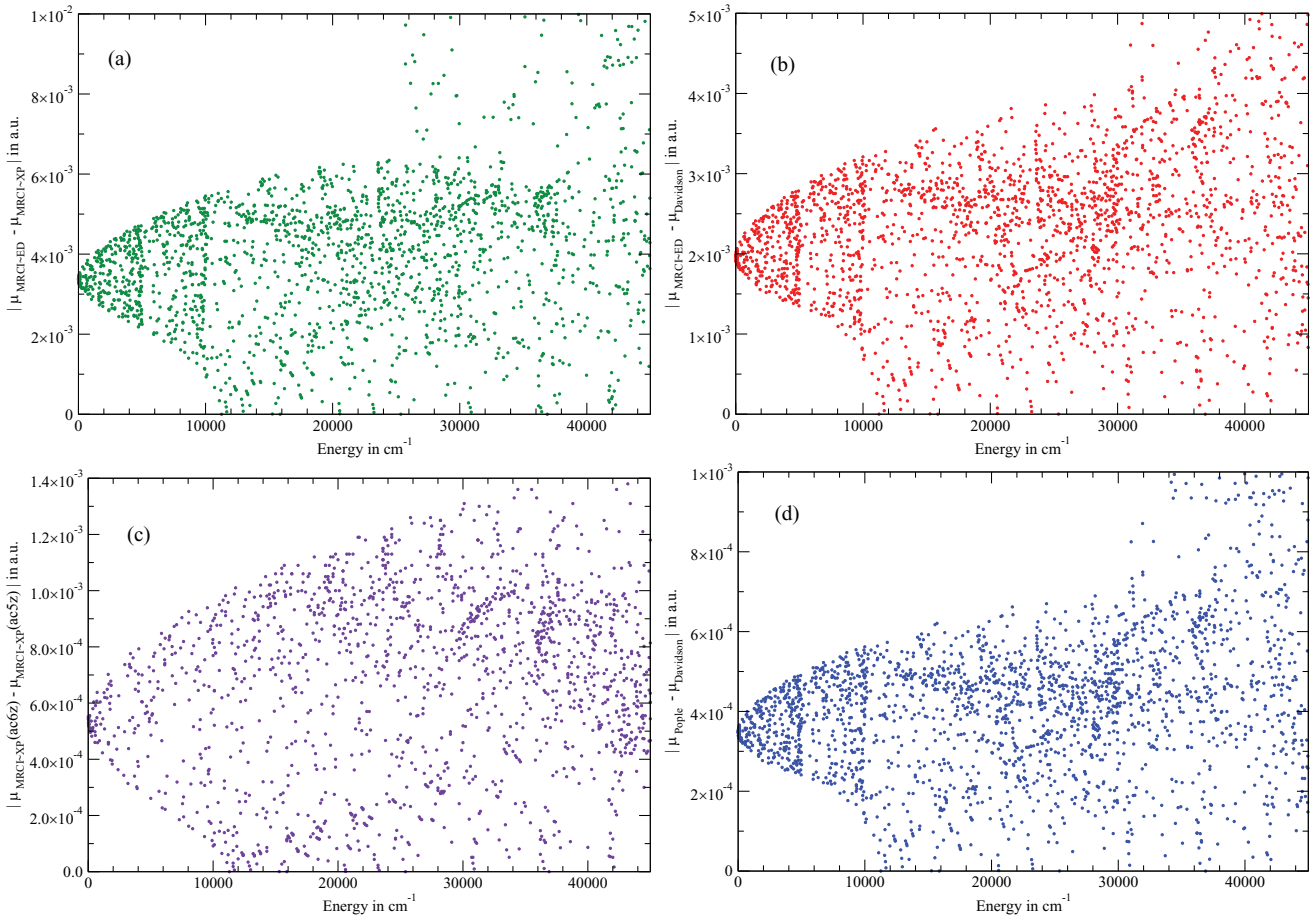

FIG. 1. Plots of the modulus of the difference between dipoles computed using different approximations (see the label of each plot's vertical axis). Unless otherwise specified the basis set is aug-cc-pCV6Z.

for higher energies. The error due to basis set incompleteness therefore appears to be considerably smaller than the one due to points (*a*) and (*b*) above. Furthermore, ED dipoles are expected to converge faster with respect to the basis set size than XP ones. We therefore estimate that the typical basis set incompleteness error of aug-cc-pCV6Z∼ED dipoles is of the order of 3×10^{-4} a.u. near equilibrium.

5. Effect of further correlation effects – plot (d)

The difference between Davidson-corrected and Pople-based dipoles may be taken as a lower-bound estimate of the correlation-treatment error. Near-equilibrium Davidson and Pople dipoles differ by about 3×10^{-4} a.u., and the difference grows up to about 7×10^{-4} a.u. for higher energies. These values are only slightly smaller than the basis-set error (*c*) above.

These considerations led to an estimate for the accuracy of the *ab initio* dipoles of about 5×10^{-4} a.u. close to equilibrium, of the order of 10^{-3} a.u. for energies up to about 25 000 cm^{-1} and indicatively of about 2×10^{-3} a.u. for higher energies. This analysis was useful to set a target accuracy for the fitted surface.

III. FITTING PROCEDURE

The LTP2011 dipole moment surface is a function of the two OH bond lengths r_1 and r_2 and of the $\widehat{HOH}$ bond angle θ. The molecule is taken to lie on the xy plane, with the oxygen atom at the origin, the x axis bisecting the bond angle θ, and hydrogen atoms 1 and 2, respectively, in the 1st and 4th quadrant. It follows that μ_x is the component parallel to the bond-angle bisector and μ_y is the one perpendicular to it.

Schwenke and Partridge (SP2000) (Ref. 17) suggested fitting both dipole components at the same time by the use of an auxiliary scalar function q by writing

$$\begin{aligned} \mu_x &= \cos\left(\frac{\theta}{2}\right)(r_1\, q(r_1, r_2, \theta) + r_2\, q(r_2, r_1, \theta)), \\ \mu_y &= \sin\left(\frac{\theta}{2}\right)(r_1\, q(r_1, r_2, \theta) - r_2\, q(r_2, r_1, \theta)). \end{aligned} \tag{2}$$

The function q has the physical dimension of a charge and can be interpreted as representing fictitious point-charges on the H_1, H_2, and O atoms of magnitude, respectively, $q(r_1, r_2, \theta)$, $q(r_2, r_1, \theta)$, and $[-q(r_1, r_2, \theta) - q(r_2, r_1, \theta)]$. SP2000 proceeded by fitting q with a linear least-square strategy, and the same strategy was also used in the core-valence-relativistic (CVR) dipole surface.[16]

In practice, this approach requires very high-order polynomials in the $(\cos\theta - \cos\theta_e)$ (up to 18th order in Ref. 17) to obtain a satisfactory fit. Such high-order polynomials cause spurious oscillations, which, in turn, can have a catastrophic effect on computed line intensities. An earlier fit to same data as SP2000 used a different functional form and suffered even more from these problems.[18]

High-overtone bending modes[16,17] are, particularly, sensitive to these small oscillations. Past studies using formula (2) (Refs. 16 and 17) supplemented the *ab initio* data by values obtained by the one-dimensional spline interpolation, with a view to damping down the unwanted oscillations.

It was found more fruitful for the LTP2011 surface to give up formula (2) and to fit separately the μ_x and μ_y components in a very straightforward way. Having introduced the fitting variables $x_1 = r_1 + r_2$, $x_2 = r_2 - r_1$, and $x_3 = \pi - \theta$, each component μ_α, $\alpha = x, y$ was fitted to the simple polynomial expansion:

$$\mu_\alpha = \sum_{i,j,k} c_{ijk}^{(\alpha)}\, x_1^i x_2^j x_3^k. \tag{3}$$

For both components we imposed $0 \le i + j \le n_r$, $0 \le k \le n_\theta$, where n_r and n_θ are chosen values. The indices ijk included in the sum were further restricted in such a way as to reproduce the symmetry properties of the μ_x or μ_y components. μ_x has the symmetry properties $\mu_x(r_1, r_2, \theta) = \mu_x(r_2, r_1, \theta)$ and $\mu_x(r_1, r_2, \pi) = 0$; the indices were therefore selected by $j = 0, 2, 4, \ldots, n_r$; $k = 1, 2, 3, \ldots, n_\theta$. For μ_y, the symmetry property is $\mu_y(r_1, r_2, \theta) = -\mu_y(r_2, r_1, \theta)$ and the indices were restricted to $j = 1, 3, 5, \ldots, n_r$; $k = 0, 1, 2, \ldots, n_\theta$. The linear fitting coefficients c_{ijk} were then obtained by a weighted least-square linear-fit using the freely available LSQ FORTRAN routine.[31] Other fitting variables were also tried, including $\sqrt{r_2 \pm r_1}$ and $\log(1 + \alpha(r_2 \pm r_1))$; these alternative choices sometimes gave smaller residuals for small fits (n_r and n_θ up to about 5), but there seems to be no advantage with respect to the simpler form given by Eq. (3) for larger fits.

A flexible weighing function, $\lambda(E)$, was employed, where weights depending on the energy were used. $\lambda(E)$ is parametrized by a cut-off energy E_c and a cut-off width E_w and is defined by

$$\lambda(E) = \frac{1 - \tanh(t)}{2} \quad \text{with} \quad t = \frac{E^2 - E_c^2}{2 E E_w}. \tag{4}$$

$\lambda(E)$ goes exponentially to 1 when $E < E_c - E_w$, is exactly 1/2 for $E = E_c$, and goes exponentially to zero for $E > E_c + E_w$. Various values for E_c were tried and finally the value of 25 000 cm^{-1} was chosen. The fits were not found to be very dependent on the cut-off width, and a small value $E_w = 1000$ cm^{-1} was used. In our final fit it was also decided to include a minimum weight of 10^{-4}, so in practice $\max(\lambda(E), 10^{-4})$ was used as the weighing function. The minimum weight was included to improve slightly the surface for energies greater than E_c.

A. Asymptotic behaviour

A damping function $\omega(r)$ such that $\lim_{r \to +\infty} \omega(r) = 0$ was introduced in order to force the dipole moment function to a reasonable behaviour in regions having energies higher than E_c and therefore not covered by the *ab initio* points included in the fit. In particular, some care was spent to check that the dipole function had a reasonable behaviour along the lowest-energy water dissociation channel, corresponding to the breaking of one O–H bond.

When one hydrogen is pulled away the dipole moment should reduce to the dipole of the OH diatomic, appropriately projected according to our xy axes convention. This led to the

functional form

$$\mu_x = \left(\sum_{ijk} c_{ijk}^{(x)} x_1^i x_2^j x_3^k\right) \omega(r_1)\omega(r_2) + \cos(\theta/2)\{\mu_{\rm OH}(r_1)[1-\omega(r_2)]+\mu_{\rm OH}(r_2)[1-\omega(r_1)]\},$$

$$\mu_y = \left(\sum_{ijk} c_{ijk}^{(y)} x_1^i x_2^j x_3^k\right) \omega(r_1)\omega(r_2) + \sin(\theta/2)\{\mu_{\rm OH}(r_1)[1-\omega(r_2)]-\mu_{\rm OH}(r_2)[1-\omega(r_1)]\}. \quad (5)$$

Functions such as e^{-r} or e^{-r^2} are often employed as damping functions to force certain terms to zero at large values of r.[16,17] However, in the set-up of the LTP2011 surface we found it useful to devise a different form, such that it should provide an infinitely strong damping after a certain $r > r_{\rm max}$ while being exactly equal to one for $r < r_{\rm min}$. We therefore settled for the infinitely smooth function $\omega(r)$ defined for $r_{\rm min} < r < r_{\rm max}$ by

$$\omega(r) = [1+\exp(4\tan(t))]^{-1}, \quad (6)$$

where

$$t = \pi\left(\frac{r-r_{\rm min}}{r_{\rm max}-r_{\rm min}} - \frac{1}{2}\right). \quad (7)$$

For $r \le r_{\rm min}$ we set $\omega(r) = 1$ and for $r \ge r_{\rm max}$ we set $\omega(r) = 0$.

A very accurate O–H dipole curve $\mu_{\rm OH}(r)$ is not necessary for our purposes. The one employed in the LTP2011 surface, relative to the OH ground electronic state $X\ ^2\Pi$, is based on the internally contracted, averaged quadratic coupled cluster (IC-AQCC) method[32] in a full-valence, (7-electron, 5-orbital) complete active space comprising 3 orbitals of a_1 symmetry, one orbital of b_1 symmetry, and one orbital of b_2 symmetry (C_{2v} symmetry labels). The oxygen 1*s* orbital was not correlated. The aug-cc-pV5Z basis set was employed, dipoles were computed by expectation value, and the program MOLPRO (Ref. 26) was used. Fifty-nine points for $r = 1.15$ to $r = 4.05$ a_0 with spacing 0.05 a_0 were computed. The IC-AQCC method was chosen over IC-MRCI or IC-ACPF because tests at the near-equilibrium bond length $r = 1.90$ a_0 in the aug-cc-pVTZ and aug-cc-pCVTZ basis sets showed that IC-AQCC [7,5] frozen-core expectation-value dipoles agree best with more accurate dipoles computed as derivative of the energy, including core and relativistic corrections and using a larger 7-electron, 7-orbital active space. Computed dipoles are estimated to be accurate to about 2×10^{-3} a.u.

Inspired by the suggestions by Bytautas *et al.*,[33] the OH dipole curve was fitted to a sum of nine even-tempered gaussians, multiplied by r^2:

$$\mu_{\rm OH}(r) = r^2 \sum_{i=0}^{8} a_i\, e^{-\alpha\beta^i r^2}. \quad (8)$$

The r^2 factor was found to improve the fit somewhat and also correctly forces $\mu_{\rm OH}$ to zero for $r \to 0$. The final fit has a root-mean-square fitting error (rms) of only 4×10^{-4} a.u. A FORTRAN subroutine computing $\mu_{\rm OH}$ is included in the LTP2011 surface.

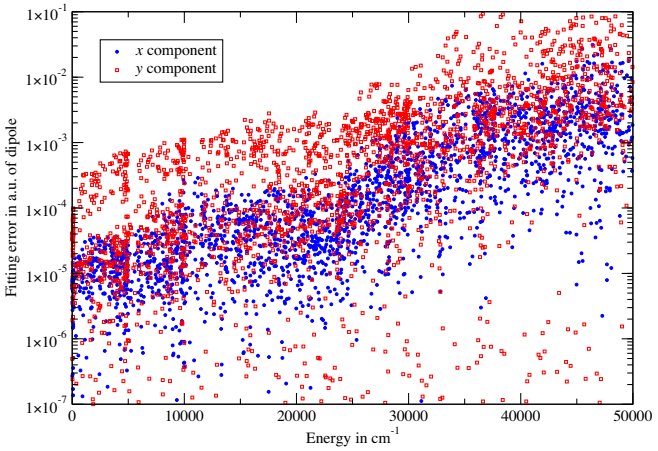

FIG. 2. Fitting residuals for the LTP2011 dipole moment surface. The fit used $n_r = n_\theta = 8$ for the μ_x component and $n_r = 9$, $n_\theta = 8$ for the μ_y component.

B. Final surfaces

The LTP2011 surface uses $n_r = n_\theta = 8$ for the μ_x component and $n_r = 9$, $n_\theta = 8$ for the μ_y one, corresponding, respectively, to 200 and 225 linear fitting parameters. It is based on the IC-MRCI+Q dipoles supplemented by the MVD1 relativistic corrections, computed as described in Sec. II B. The residuals of the fit are reported in Fig. 2. It is apparent from this figure that the *y* component is much more difficult to fit accurately.

The weighted rms, defined by

$$\text{rms}^2 = \frac{1}{N}\sum_{k=1}^{N} W_k\left(\mu_\alpha\left(r_1^{(k)}, r_2^{(k)}, \theta^{(k)}\right) - \mu_\alpha^{(k)}\right)^2 \quad (9)$$

is of 3×10^{-5} a.u. for the *x* component and 4×10^{-4} a.u. for the *y* component. As discussed in Sec. II D the accuracy of the *ab initio* dipoles can be estimated to 5×10^{-4} a.u. close to equilibrium and is probably a factor 2–5 larger for higher energies; keeping these values into account, the fitting accuracy achieved can be considered satisfactory.

Besides LTP2011 three more surfaces were produced and are made available, mainly for testing purposes:

1. LTP2011P: As LTP2011 but based on Pople-corrected dipoles ("P" for Pople).
2. LTP2011NR: As LTP2011 but does not include the relativistic corrections ("NR" for "non relativistic").
3. LTP2011S: As LTP2011 but includes fewer fitting parameters: $n_r = 6$, $n_\theta = 7$ for the μ_x component and $n_r = n_\theta = 7$ for the μ_y component ("S" for "small"). The rms errors for the *x* and *y* components are 5×10^{-5} a.u. and 5×10^{-4} a.u.

All surfaces are expected to give very similar results. For example, in many of our internal tests line intensities computed using LTP2011P were on average 0.1% stronger than intensities computed with LTP2011, with a scatter around the

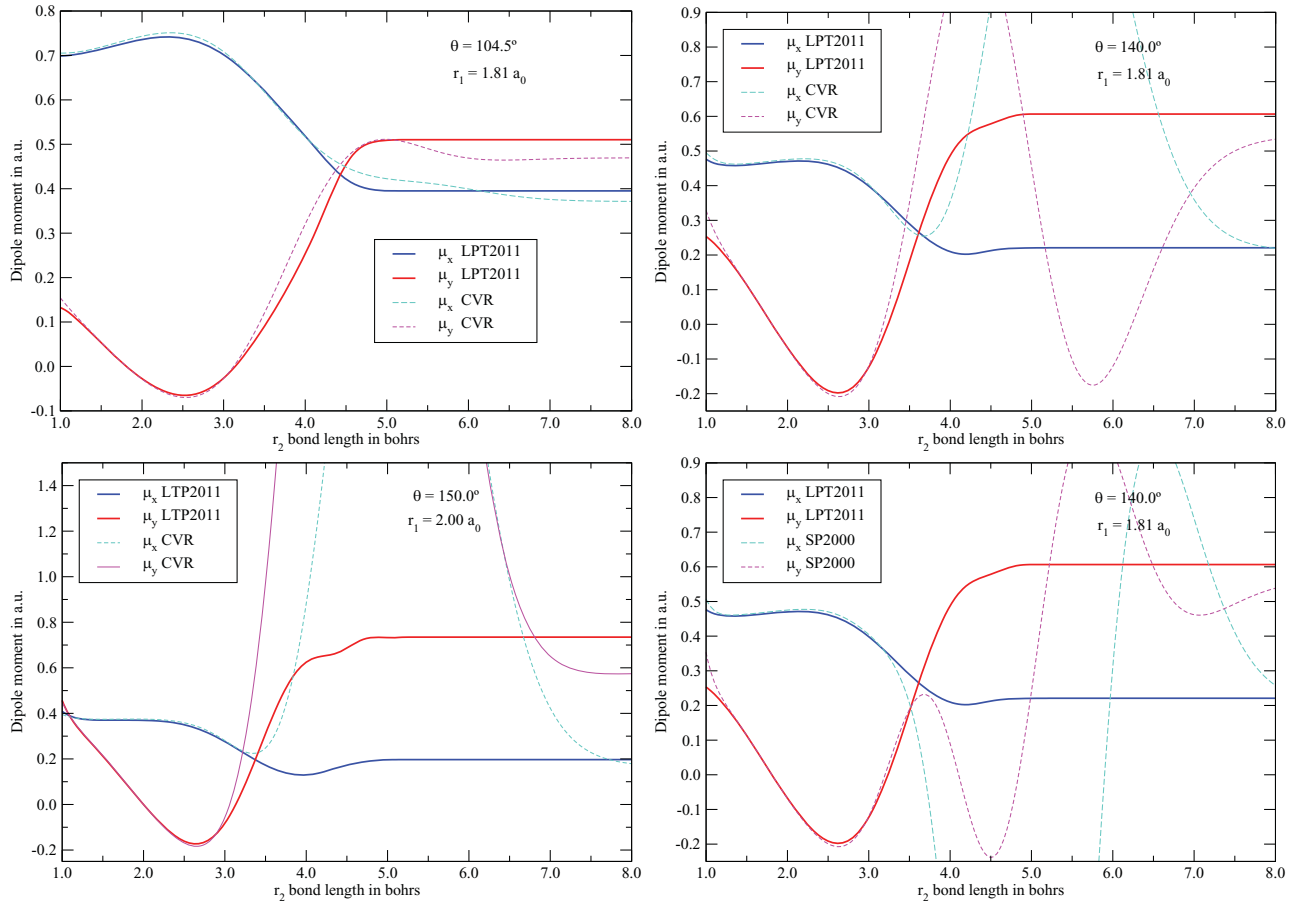

FIG. 3. One-dimensional cuts of the new LTP2011 DMS and of the older CVR (see Ref. 16) and SP2000 (see Ref. 17) DMSs. The older surfaces show large, spurious oscillations at long bond lengths.

average of ±0.2%. Only for a few lines-differences went up to ±0.5%. Line intensities for strong lines computed with LTP2011NR are expected to be about 0.5% stronger than LTP2011-based ones. As discussed below, the LTP2011S surface is particularly appropriate for transitions involving a very high degree of excitation.

As discussed in Sec. IV A some line intensities show an extreme sensitivity to the DMS used and/or to the potential energy surface used to compute the nuclear-motion wavefunctions. For such sensitive lines an accurate theoretical value cannot be computed at present. Computing line intensities using different, high-quality DMSs, such as those provided above is a good way to test for such sensitivity.

The new damping function given by Eq. (6) helped in significantly reducing undesirable, unphysical features at higher energies. Regions of the DMS for energies higher than about 25 000 cm^{-1} are generally not important for the calculations of room-temperature water line intensities, as transitions involving such highly excited states would be too weak to be measured. However, novel experimental techniques capable of probing very high energy transitions[3,5,8] show that good behaviour at high energies is highly desirable. The LPT2011 surface should be considered an improvement over previous surfaces, such as SP2000 (Ref. 17) and CVR (Ref. 16) in regions with energies higher than about 25 000 cm^{-1}, see Fig. 3 for some indicative plots.

The large oscillations visible in Fig. 3 will result in spurious transition dipoles between high-lying states. Conversely, fits which contain even very small oscillations in the equilibrium region will result in significantly overestimated line intensities for very high overtone transitions. To show this effect we computed line intensities for transitions from the ground rotational-vibrational state to $J = 1$ levels of increasing energy. Results are reported in Fig. 4 for a series of DMS.

The expected approximate behaviour of line intensities displayed in Fig. 4 is to decay exponentially with respect to the energy of the upper energy level. It is clear from the plot that intensities calculated with the CVR surface start deviating from this trend when the energy of the upper state is greater than about 20 000 cm^{-1}; after this value computed line intensities vary only weakly and are therefore likely to be too strong. On the other hand, intensities computed with the older DMS due to Gabriel *et al.*,[20] which is fitted with a simple, low-order polynomial (up to 4th order), show the expected qualitative behavior. Comparisons with our surfaces show that, while the high accuracy LTP2011 DMS extends the range of reliability over CVR, it still probably overestimates transition intensities at the higher energies. Conversely, the less aggressively fit LTP2011S surface joins that of Gabriel *et al.* in showing the expected exponential behavior over the entire energy range considered. We therefore recommend the

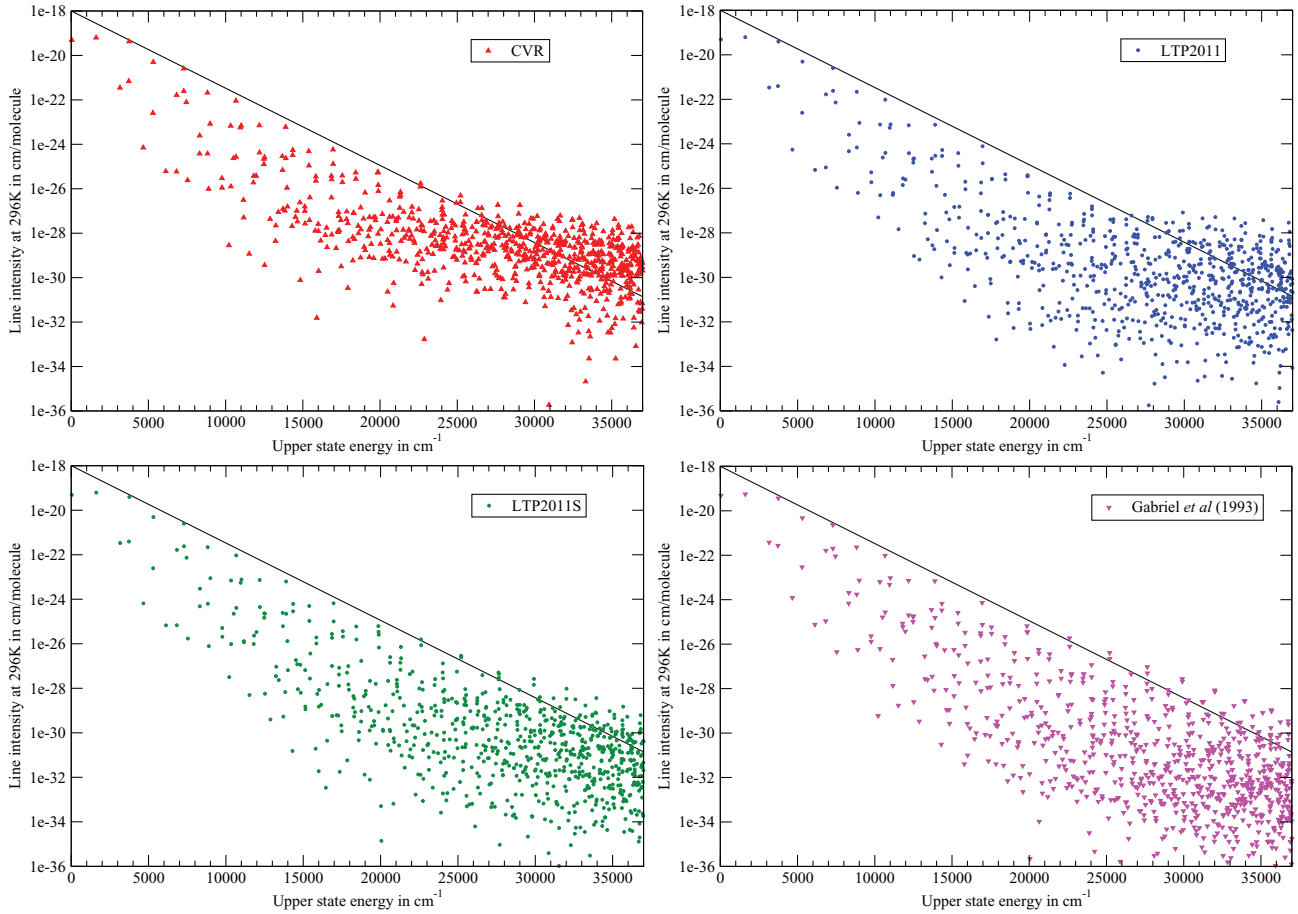

FIG. 4. Line intensities from the ground rotational-vibrational state to $J = 1$ excited states computed with the CVR, (see Ref. 16) the LTP2011, the LTP2011S, and the Gabriel *et al*. (see Ref. 20) dipole surfaces. The straight line links the strongest transitions in the HITRAN database, (see Ref. 25) which extend to 23 000 cm^{-1}; they provide a guide for the eye illustrating the expected exponential trend.

use of the LTP2011S fit for studies involving high overtone transitions.

Finally, we present as Fig. 5 a low-resolution simulation of the rotation-vibration spectrum of water in the near ultraviolet. This simulation was made using the LTP2011S DMS and wavefunctions generated from the recent PES of Bubukina *et al*.,[15] since this potential was fitted to high-lying levels (up to 26 000 cm^{-1}). The figure does not give a full simulation; instead it is based on vibrational band intensities and a notional Gaussian profile with $\sigma = 100$ cm^{-1} for each band. This is sufficient to show that the total absorption is weak and therefore unlikely to be of significance for the earth's atmosphere.[34] However, it could be important in environments with a high column density of water or a particularly strong flux in the near ultraviolet.

IV. COMPARISON WITH ACCURATE EXPERIMENTAL DATA

In this section our dipole moment surfaces will be compared with accurate experimental data. Following common practice, we will call integrated effective cross-sections "line intensities,"[9] measured in cm^{-1}/(molecule/cm^2) (or, for short, cm/molecule). This unit is also sometimes called the "HITRAN unit" as it is used in the HITRAN spectroscopic database.[25,35] This choice of units leads to very small numerical values, as the cross section of a single molecule is, at most, of the order of 100 pm^2. The DVR3D suite of programs[36] was used to compute nuclear-motion wavefunctions and transition intensities. Unless otherwise stated all nuclear-motion wave functions were produced using the accurate, semi-empirical water potential energy surface by Shirin *et al*.[13] (Shirin 2003). Another set of wave functions was calculated using the

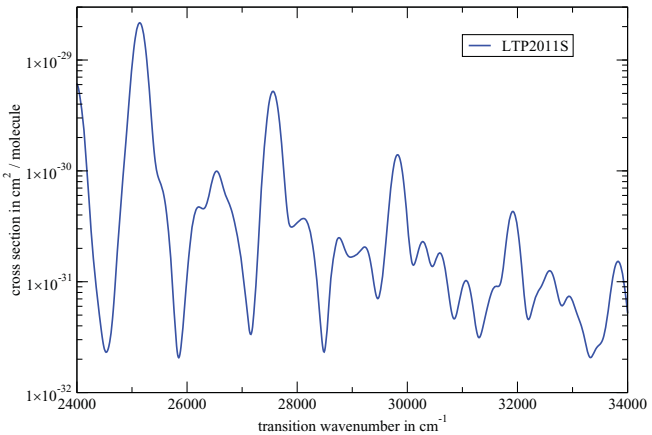

FIG. 5. Simulated low-resolution spectrum of water in the near ultraviolet.

ab initio CVRQD PES of Barletta *et al.*[12] and was used for sensitivity tests. Wave functions with rotational angular momentum J up to 13 were used.

A. Line sensitivity

Water line intensities I of interest span an enormous range: the strongest lines have intensities $I \approx 10^{-18}$ cm/molecule, while the weakest lines considered have $I \approx 10^{-28}$ cm/molecule or even less. Lines with intensities greater than about 10^{-23} cm/molecule can generally be computed reliably. On the other hand, weaker lines are sometimes extremely sensitive to details of the PES used to compute the wavefunction, sometimes to such an extent that, because of the intrinsic errors of practical computations, their intensities cannot be calculated reliably. In part, this situation is not surprising. Several so-called resonance effects (Coriolis, Fermi, Darling-Dennison) are present in water[37] and it is to be expected that weak line intensities involving such resonant levels will not be stable upon small changes in the PES used to compute the rotational-vibrational wavefunctions. Weak lines may also show high sensitivity to the DMS, particularly, if there are spurious effects introduced in the fitting. Because of this situation it is inevitable that a subset of computed *ab initio* weak lines will always remain relatively inaccurate, even when using very high-quality PESs and DMSs.

The approach taken in this study was to ascertain the sensitivity of computed line intensities with respect to changes both of the nuclear-motion wave functions and of the DMS. This was done by using different combinations of high-quality PESs and DMSs. This strategy allowed us to easily identify lines which showed high sensitivity (see Sec. IV C for details).

B. Comparison with accurate line intensities

Lisak, Harvey, and Hodges[24] recently reported very accurate measurements of a small set of 15 water line intensities. Of these, a lone outlier at 7 179.75201 cm^{-1} is consistently stronger than theoretical calculations by about 22% and was excluded from the analysis. This experimental line probably corresponds to the blending of two resonant transitions. To appropriately compare with theoretical intensities, the experimental intensities were divided by the value used by Lisak *et al.* for the natural abundance of the major water isotopologue, 0.997317.

A plot of the ratios is reported as Fig. 6. The error bars in the figure correspond to the stated experimental uncertainties. See also Fig. 8 of Ref. 24 for a similar plot including other high-quality dipole surfaces. In particular, the best theoretical lines reported by Lisak *et al.*[24] were those calculated with the CVR dipole surface. Intensities computed using the CVR surface are on average 1.3% too strong and have a scatter of ±0.7% (Lisak *et al.* did not correct for isotopologue abundance and therefore quoted the slightly larger values of 1.4% and ±0.8%). Lines computed with the LTP2011 dipole surface are slightly too strong by 0.4% and have a reduced scatter of ±0.6%.

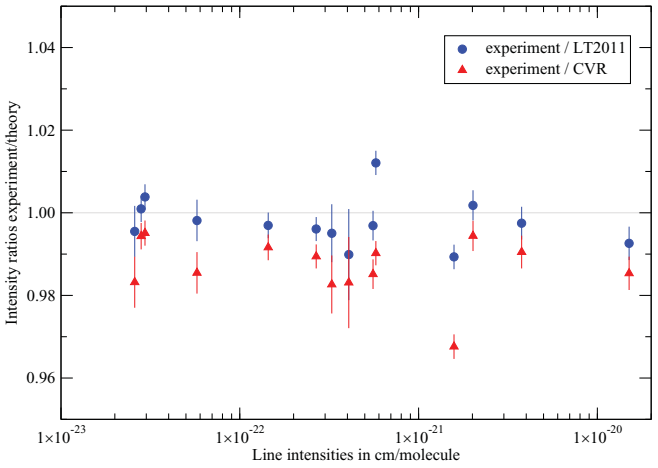

FIG. 6. Comparison of the ratios of line intensity calculated with the LTP2011 dipole surface and the CVR surface with those reported by Lisak *et al.* (See Ref. 24.)

Line intensities were also calculated using nuclear-motion wave functions produced using the CVRQD energy surface by Barletta *et al.*[12] and the LTP2011 dipole surface. The results are very similar to the ones discussed above, with LTP2011 lines still too strong by 0.4% and a slightly larger scatter of 1.0%.

These should be regarded as exceptionally good results for *ab initio* line intensities and are the best ever produced by this kind of computations.

C. Comparison with HITRAN2008

The 2008 edition of the HITRAN spectroscopic database[25] contains 36 550 fully assigned transition lines for the $H_2{}^{16}O$ water isotopologue. Of these, 34 569 transitions involve levels with $J \leq 13$ and were selected for comparison with theory. As already done in previous studies,[38] the natural logarithm of the ratios of line intensities, ϱ, and the standard deviation of the ratios, σ, were found useful quantities in comparing data sets. In the following, the term "average of the ratios" will be used to indicate the geometrical mean of the ratios and the term "scatter" with a percentage sign will indicate the quantity 100σ.

Line intensities were computed for the following combinations of potential/dipole surfaces: Shirin2003/LTP2011, Shirin2003/LTP2011S, CVRQD/LTP2011. The scatter between the various computed line intensities permitted us to identify lines which showed great sensitivity. More specifically, all 3 possible ratios between the 3 datasets above were considered and a line was marked as "sensitive" if the standard deviation of the natural logarithm of ratios was more than 0.20 (corresponding to a scatter of about ±20%). Following this criterion 24 676 lines were labelled as "stable" and 2 281 lines were labelled as sensitive. Results of the intensity ratios LTP2011/HITRAN2008 are graphically displayed in Fig. 7.

A total of 26 957 HITRAN lines with $J < 13$ were matched using quantum number assignments to the various theoretical datasets. The agreement with the strongest lines ($I \geq 5 \times 10^{-19}$ cm/molecule) is exceptionally good, with the

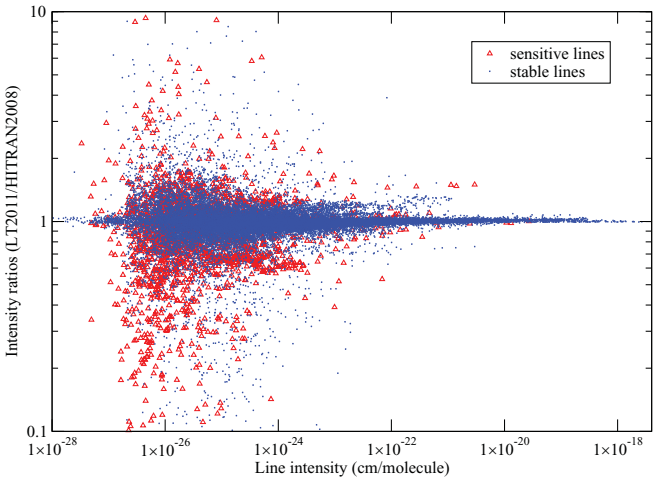

FIG. 7. Ratios of line intensity calculated with the LTP2011 dipole surface and those reported in the 2008 version of the HITRAN database.

average of the intensity ratios equal to 1.001 and a scatter of only ±0.5%. This is an improvement over the CVR dipole moment surface,[16] which was known to overestimate strong lines by about 2%. The various improvements adopted in the electronic structure treatment at the basis of the LTP2011 surface appear to have eliminated this small systematic bias.

For medium-to-strong stable lines with $I \geq 10^{-22}$ cm/molecule, the overall agreement is still very good, with an average ratio of 1.01 and a scatter of ±4.5%.

About 65% of the remaining 21 877 weaker stable lines agree with the theoretical lines within ±10%, 82% of them agree within ±20% and 92% agree within ±50%. Stable lines which differ by more than ±50% are typically very weak lines with intensities $I < 10^{-25}$ cm/molecule where accurate measurements are difficult.

About 90% of the 2238 lines denominated sensitive are weak lines with $I < 10^{-25}$ cm/molecule. Fifty percent of these sensitive lines agree with theory within ±25%, while 84% agree within a factor of 2.

Overall, the agreement with HITRAN data is very good, especially keeping in mind that HITRAN is mostly based on measurements for which uncertainties in the intensities are generally in the 5–10 % range for good measurements, and in some cases are much worse than this. Furthermore, HITRAN data can be affected by systematic problems. For example, it is evident that there is some structure in Fig. 7 (horizontal swarms of points), very probably indicating systematic errors in some of the HITRAN line intensities. This problem with HITRAN has been noted before[38,39] and is by no means the only one.[40]

V. SUMMARY

A new, high accuracy DMS for the water molecule is presented. The surface is based on all-electron IC-MRCI calculations in a large 8-electron, 10-orbital reference space in the aug-cc-pCV6Z basis set. *Ab initio* dipole moments were calculated as derivatives of the Davidson-corrected energies, thus approximately incorporating the effect of quadruple excitations. Relativistic corrections are based on the MVD1 Hamiltonian.

The LTP2011 dipole moment surface is the final step in a series of high-quality water DMSs and solves some of the defects which afflicted its predecessors. For most medium and strong lines with $I > 10^{-22}$ cm/molecule, LTP2011 is expected to produce line intensities with errors of 1% or less. About 10% of weaker lines show high sensitivity to the DMS and the PES used for their computation and cannot be calculated very reliably. Running multiple calculations with different combinations of high-quality DMSs and PESs is of great use in identifying such sensitive lines. A database of room-temperature water line intensities computed using the LTP2011 is in preparation.

The LTP2011 DMS, and, in particular, the LTP2011S fit, also significantly improves the behaviour for high overtone transitions in the ultraviolet at high energies ($E > 25\,000$ cm^{-1}) and is expected to be appropriate for computing line intensities for transitions between very highly excited states. All the DMSs and the *ab initio* data on which they are based are made available in the supplementary material.[41]

ACKNOWLEDGMENTS

We thank UCL research computing for help in performing the electronic structure calculations reported here and Attila Császár for helpful discussions. This work was supported by the UK Natural Environment Research Council (NERC) and by the Royal Society.

[1] J. Tennyson, Phys. Scr. **76**, C53 (2006).
[2] P. F. Bernath, Phys. Chem. Chem. Phys. **4**, 1501 (2002).
[3] P. Maksyutenko, T. R. Rizzo, and O. V. Boyarkin, J. Chem. Phys. **125**, 181101 (2006).
[4] P. Maksyutenko, J. S. Muenter, N. F. Zobov, S. V. Shirin, O. L. Polyansky, T. R. Rizzo, and O. V. Boyarkin, J. Chem. Phys. **126**, 241101 (2007).
[5] M. Grechko, O. V. Boyarkin, T. R. Rizzo, P. Maksyutenko, N. F. Zobov, S. V. Shirin, L. Lodi, J. Tennyson, A. G. Császár, and O. L. Polyansky, J. Chem. Phys. **131**, 221105 (2009).
[6] M. Grechko, P. Maksyutenko, T. R. Rizzo, and O. V. Boyarkin, J. Chem. Phys. **133**, 081103 (2010).
[7] A. G. Császár, E. Mátys, T. Szidarovszky, L. Lodi, N. F. Zobov, S. V. Shirin, O. L. Polyansky, and J. Tennyson, J. Quant. Spectrosc. Radiat. Transf. **111**, 1043 (2010).
[8] N. F. Zobov, S. V. Shirin, L. Lodi, B. C. Silva, J. Tennyson, A. G. Császár, and O. L. Polyansky, Chem. Phys. Lett. **507**, 48 (2011).
[9] P. F. Bernath, *Spectra of Atoms and Molecules*, 2nd ed. (Oxford University Press, New York, 2005).
[10] L. Lodi and J. Tennyson, J. Phys. B **43**, 133001 (2010).
[11] A. E. Lynas-Gray, S. Miller, and J. Tennyson, J. Mol. Spectrosc. **169**, 458 (1995).
[12] P. Barletta, S. V. Shirin, N. F. Zobov, O. L. Polyansky, J. Tennyson, E. F. Valeev, and A. G. Császár, J. Chem. Phys. **125**, 204307 (2006).
[13] S. V. Shirin, O. L. Polyansky, N. F. Zobov, P. Barletta, and J. Tennyson, J. Chem. Phys. **118**, 2124 (2003).
[14] S. V. Shirin, N. F. Zobov, R. I. Ovsyannikov, O. L. Polyansky, and J. Tennyson, J. Chem. Phys. **128**, 224306 (2008).
[15] I. I. Bukukina, N. F. Zobov, O. L. Polyansky, S. V. Shirin, and S. N. Yurchenko, Opt. Spectrosc. **110**, 160 (2011).
[16] L. Lodi, R. N. Tolchenov, J. Tennyson, A. E. Lynas-Gray, S. V. Shirin, N. F. Zobov, O. L. Polyansky, A. G. Császár, J. N. P. van Stralen, and L. Visscher, J. Chem. Phys. **128**, 044304 (2008).
[17] D. W. Schwenke and H. Partridge, J. Chem. Phys. **113**, 6592 (2000).
[18] H. Partridge and D. W. Schwenke, J. Chem. Phys. **106**, 4618 (1997).
[19] G. S. Kedziora and I. Shavitt, J. Chem. Phys. **106**, 8733 (1997).

[20]W. Gabriel, E. A. Reinsch, P. Rosmus, S. Carter, and N. C. Handy, J. Chem. Phys. **99**, 897 (1993).
[21]U. G. Jørgensen and P. Jensen, J. Mol. Spectrosc. **161**, 219 (1993).
[22]A. G. Csãszãr, G. Czakã, T. Furtenbacher, J. Tennyson, V. Szalay, S. V. Shirin, N. F. Zobov, and O. L. Polyansky, J. Chem. Phys. **122**, 214305 (2005).
[23]L. Lodi, Ph.D. dissertation, University College London, 2008, also see http://www.tampa.phys.ucl.ac.uk/ftp/eThesis/.
[24]D. Lisak, D. K. Harvey, and J. T. Hodges, Phys. Rev. A **79**, 052507 (2009).
[25]L. S. Rothman, I. E. Gordon, A. Barbe, D. Chris Benner, P. F. Bernath, M. Birk, V. Boudon, L. R. Brown, A. Campargue, J.-P. Champion, K. Chance, L. H. Coudert, V. Danaj, V. M. Devi, S. Fally, J.-M. Flaud, R. R. Gamache, A. Goldmanm, D. Jacquemart, I. Kleiner, N. Lacome, W. J. Lafferty, J.-Y. Mandin, S. T. Massie, S. N. Mikhailenko, C. E. Miller, N. Moazzen-Ahmadi, O. V. Naumenko, A. V. Nikitin, J. Orphal, V. I. Perevalov, A. Perrin, A. Predoi-Cross, C. P. Rinsland, M. Rotger, M. Åimeäkovãi, M. A. H. Smith, K. Sung, S. A. Tashkun, J. Tennyson, R. A. Toth, A. C. Vandaele, J. Vander Auwera, J. Quant. Spectrosc. Radiat. Transf. **110**, 533 (2009).
[26]MOLPRO, a package of *ab initio* programs designed by H. J. Werner and P. J. Knowles, version 2009.1, R. Lindh, F. R. Manby, M. Schütz, *et al*..
[27]R. A. Gangi and R. F. W. Bader, J. Chem. Phys. **55**, 5369 (1971).
[28]G. Durand and X. Chapuisat, Chem. Phys. **96**, 381 (1985).
[29]R. Polak, I. Paidarova, and P. J. Kuntz, J. Chem. Phys. **87**, 2863 (1987).
[30]S. L. Hobson, E. F. Valeev, A. G. Császár, and J. F. Stanton, Mol. Phys. **107**, 1153 (2009).
[31]A. J. Miller, J. Royal Stat. Soc. C **41**, 458 (1992); LSQ, A FORTRAN 95 module for weighted unconstrained linear least-squares calculations, 2002. Also see http://jblevins.org/mirror/amiller/.
[32]P. G. Szalay and R. J. Bartlett, J. Chem. Phys. **103**, 3600 (1995).
[33]L. Bytautas, N. Matsunaga, T. Nagata, M. S. Gordon, and K. Ruedenberg, J. Chem. Phys. **127**, 204313 (2007).
[34]J. D. Haigh, private communication (2011).
[35]L. S. Rothman, J. Quant. Spectrosc. Radiat. Transf. **111**, 1565 (2010).
[36]J. Tennyson, M. A. Kostin, P. Barletta, G. J. Harris, O. L. Polyansky, J. Ramanlal, and N. F. Zobov, Comput. Phys. Commun. **163**, 85 (2004).
[37]A. Bykov, O. Naumenko, L. Sinitsa, B. Voronin, J.-M. Flaud, C. Camy-Peyret, and R. Lanquetin, J. Mol. Spectrosc. **205**, 1 (2001).
[38]L. Lodi and J. Tennyson, J. Quant. Spectrosc. Radiat. Transf. **109**, 1219 (2008).
[39]L. H. Coudert, G. Wagner, M. Birk, Y. I. Baranov, W. J. Lafferty, and J. M. Flaud, J. Mol. Spectrosc. **251**, 339 (2008).
[40]L. Tallis, M. Coleman, T. Gardener, I. V. Ptashnik, and K. P. Shine, "Assessment of the consistency of H2O line intensities over the near-infrared using sunpointing ground-based Fourier transform spectroscopy," J. Quant. Spectrosc. Radiat. Transf. (in press).
[41]See supplementary material at http://dx.doi.org/10.1063/1.3604934, for a FORTRAN 95 routine containing the LTP2011, LTP2011S, LTP2011NR, and LTP2011P dipole surfaces and for the *ab initio* data on which they are based.