Anal. Chem. **2001,** *73,* 3153−3161

Studies on the Structure of Water Using Two-Dimensional Near-Infrared Correlation Spectroscopy and Principal Component Analysis

V. H. Segtnan,[†,‡] Š. Šašić,[†,§] T. Isaksson,[‡] and Y. Ozaki*[,†]

Department of Chemistry, School of Science, Kwansei-Gakuin University, Uegahara, Nishinomiya 662-8501, Japan, and Department of Food Science, Agricultural University of Norway, N-1432 Ås, Norway

The structure of water molecules in the pure liquid state has been subjected to extensive research for several decades. Questions still remain unanswered, however, and no single model has been found capable of explaining all the anomalies of water. In the present study, near-infrared spectra of water in the temperature region 6−80 °C have been analyzed by use of principal component analysis and two-dimensional correlation spectroscopy in order to study the dynamic behavior of a band centered around 1450 nm at room temperature, which is due to the combination of symmetric and antisymmetric O−H stretching modes (first overtone) of water. It has been found that the wavelengths 1412 and 1491 nm account for more than 99% of the spectral variation, representing two major water species with weaker and stronger hydrogen bonds, respectively. A third species located at 1438 nm, whose concentration was relatively constant as a function of temperature, is also indicated. A somewhat distorted two-state structural model for water is suggested.

Due to its great importance and various anomalies, water is unquestionably one of the most studied compounds in physics and chemistry. The structure of water has been subjected to a large number of studies, resulting in proposals of many different structural models.[1,2] However, no single model has yet been widely accepted as the one true model explaining all the properties of water.[3] The two perhaps most important classes of structural models are mixture models, describing an equilibrium mixture of discrete species,[1] and continuum models, which assume that water molecules are involved in almost complete hydrogen bonding (H-bonding).[2] The former embodies the concept of intermolecular H-bonds being momentarily concentrated in bulky clusters of water molecules that exist in dynamic equilibrium with other more dense species—momentarily meaning 10^{-11} s. The latter suggests that H-bonds are distributed uniformly throughout the sample and that many of the bonds existing in ice simply become distorted rather than broken when ice is melted. This permits the existence of a dynamic continuous network of water molecules.[3] A water molecule has four sites where other water molecules can connect through H-bonding, i.e., one at each of the H-atoms and two at the oxygen atom. The cooperativity of H-bonds is a fundamental property of liquid water, where the H-bonds are ∼250% stronger than in a water dimer.[4] The number, nature, and strength of such H-bonds make the basis for all structural studies concerning water.

A number of spectroscopic studies using mid-infrared (IR), Raman, and near-infrared (NIR) spectroscopies have been performed in attempts of unraveling the structural properties of water. The number of IR studies on water has increased significantly since the attenuated total reflection (ATR) technique was introduced, reducing the problem of the unfavorably strong water absorptions in this region. Libnau et al.[5] used partial least-squares (PLS) regression and evolutionary curve resolution to investigate temperature-dependent IR spectra of water. Their results support a two-state model structure suggested by Benson and Siebert,[6] where clusters consisting of polycyclic octamers or decamers are transformed into clusters of monocyclic tetramers or pentamers at increasing temperatures. Devlin et al.[7] recently investigated the IR spectra of aerosol ice samples at 100 K in order to determine the structure of large water clusters. Raman studies on water structure are numerous, most reports discussing the presence of isosbestic points to discriminate between continuum models and mixtures of two or more species.[8−14]

Relatively few spectroscopic studies on water structure have been performed using the NIR region. However, the informational value of NIR spectra is potentially high. One of the major

* To whom correspondence should be addressed. E-mail: ozaki@kwansei.ac.jp. Fax: +81-798-51-0914.

† Kwansei-Gakuin University.

‡ Agricultural University of Norway.

§ Present address: Spectroscopy Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139-4307.

(1) Luck, W. A. P. In *Water Activity: Influences on Food Quality*; Rockland, L. B., Stewart, G. W., Eds.; Academic Press: New York, 1981; pp 407−434.
(2) Némethy, G. In *Structure of Water and Aqueous Solutions*; Luck, W P., Ed.; Verlag Chemie: Weinheim/Bergstrasse, Germany, 1974; p 7.
(3) Chaplin, M. F. *Biophys. Chem*. **1999**, *83*, 211−221.
(4) Luck, W. A. P. *J. Mol. Struct*. **1998**, *448*, 131−142.
(5) Libnau, F. O.; Toft, J.; Christy, A. A.; Kvalheim, O. M. *J. Am. Chem. Soc*. **1994**, *116*, 8311−8316.
(6) Benson, S. W.; Siebert, E. D. *J. Am. Chem. Soc*. **1992**, *114*, 4269−4276.
(7) Devlin, J. P.; Joyce, C.; Buch, V. *J. Phys. Chem. A* **2000**, *104*, 1974−1977.
(8) Ratcliffe, C. I.; Irish, D. I. *J. Phys. Chem*. **1982**, *86*, 4897−4905.
(9) Giguére, P. A. *J. Raman Spectrosc*. **1984**, *15*, 354.
(10) Walrafen, G. E.; Hokmabadi, M. S.; Yang, W.-H. *J. Chem. Phys*. **1986**, *85*, 6964−6969.
(11) Walrafen, G. E.; Hokmabadi, M. S.; Yang, W.-H *J. Phys. Chem*. **1988**, *92*, 2433−2438.
(12) Hare, D. E.; Sorensen, C. M. *J. Chem. Phys*. **1990**, *93*, 25−33.
(13) Hare, D. E.; Sorensen, C. M. *J. Chem. Phys*. **1990**, *93*, 6954−6961.
(14) Robinson, G. W.; Cho, C. H.; Urquidi, J. *J. Chem. Phys*. **1999**, *111*, 698−702.

10.1021/ac010102n CCC: $20.00
Published on Web 05/24/2001

advantages of the NIR region over the IR region for studies of water and aqueous solutions is that the latter demands the use of an ATR crystal or a very thin transmittance cell to reduce the absorbance level, which introduces the problem of adsorption of molecules to the ATR or cell surface. Another important advantage is the greater separation of the overtone and combination bands compared with the fundamental bands. Lin and Brown[15] made calibrations of various chemical and physical properties of water against temperature-dependent NIR spectra, indicating that the spectra are very rich in information also about structural properties. Angell and Rodgers[16] measured NIR spectra of water and H_2O-D_2O solutions at various temperatures and suggested a disrupted network model for normal and supercooled water, described by exchange of energy between weak and strong H-bonds. Iwamoto et al.[17] and later Maeda et al.[18] proposed structural assumptions based mainly on the second derivatives of NIR spectra of water at different temperatures. It is worth noticing that NIR spectroscopy has also been used in numerous calibration studies and investigations of aqueous solutions. In most of these studies, very few comments on the water structure are given. The water structure should, however, deserve more attention since it can help in better understanding of various phenomena resulting from water influence.

Though all the above-mentioned NIR, Raman, and IR studies concern highly overlapped spectra that are composed of an unknown number of component bands, they do not employ chemometric techniques, which have proven to be very powerful in the analysis of similar data structures.[19] The only exception is the study by Libnau et al.,[5] where PLS regression and curve resolution were utilized. Lin and Brown[15] employed principal component regression (PCR) and multilinear regression (MLR) in their water studies, but only for the purpose of calibrating physical and chemical properties of water against temperature-dependent NIR spectra of water. The lack of use of robust and objective techniques in the analysis of water spectra is even more prominent if it is taken into account that subtle spectral changes are investigated mainly just to prove or falsify the existence of isosbestic points, which appears to be a crucial factor for verifying or rejecting either of the global models proposed. The need for simple, reliable, and objective tools for analyzing water spectra is apparent. In the present paper, we will apply purely statistical techniques in order to analyze a set of temperature-dependent NIR water spectra, in an attempt to explore the structure of water.

Two modern and popular techniques for analysis of spectral data are used: two-dimensional (2D) correlation spectroscopy[20,21] and principal component analysis (PCA).[22] The former has proven to be powerful in unraveling complicated vibrational spectra of biological, polymeric, or industrial materials. The latter technique is well established in statistics and chemometrics and has in the recent years received much attention for investigating a variety of chemical problems. In the present study, we will investigate the abilities of 2D spectroscopy, which is a strong visualization tool and directly related to the spectral features and concentrations, for characterizing the dynamics of temperature-perturbed water spectra. PCA, which gives a precise mathematical estimation of the changes along the sample and variable vectors, is utilized in order to verify and complement the results of the 2D correlation analysis. Our primary aim is to report an objective analytical study of temperature-dependent NIR spectra of water. We also aspire at investigating the possibilities of spectral changes being assigned to specific temperature regions, an approach that to our knowledge has not been reported for spectroscopic studies of water structure. We do not attempt to propose a specific structural model based on our results, because we believe that vibrational spectroscopy alone has too limited a capability to derive the exact structure of water. However, as for all the papers cited here, the results can be compared with theoretical models and offer crucial experimental support for the theoretical proposals. Hence, the last part of the present study discusses overall agreement of our results with structural models currently in dispute.

Analytical Techniques. PCA[22] is a data compression technique that is well established in many fields of spectroscopy, and any further description of the technique will not be given here. 2D correlation spectroscopy is a more recently developed method, and we will therefore give a brief description of this technique. The technique was first introduced by Noda[20] in 1986 and later developed into the more applicable generalized 2D correlation spectroscopy by the same author.[21] Enhanced spectral resolution can be achieved by spreading the data over the second dimension, and detailed inter- and intramolecular interactions can be studied by 2D correlation spectroscopy. A data matrix **X** composed of mean centered spectra can be described by eq 1, where **X**

$$\mathbf{X} = \mathbf{CP} \tag{1}$$

represents the spectra in rows, **C** is a matrix with species concentration profiles in columns, and **P** consists of pure component spectra ordered in rows. Synchronous correlation spectra can be calculated by eq 2,[23] where s and v are the numbers of

$$\mathrm{VV}_{\mathrm{synchronous}} = \frac{1}{s-1}\mathbf{X}^{\mathrm{T}}\mathbf{X} \tag{2A}$$

$$\mathrm{SS}_{\mathrm{synchronous}} = \frac{1}{v-1}\mathbf{X}\mathbf{X}^{\mathrm{T}} \tag{2B}$$

samples and variables, respectively. Equation 2A or B is used depending on whether we want to investigate in-phase correlations between the spectral intensity changes (VV = variable-variable), which is the traditional approach presented by Noda,[21] or between the samples (SS = sample−sample), i.e., spectra themselves, an approach recently suggested by Šašić et al.[24] Asynchronous correlation spectra can be calculated by eq 3, where **H** is the

(15) Lin, J.; Brown, C. W. *Appl. Spectrosc.* **1993**, *47*, 1720−1727.
(16) Angell, C. A.; Rodgers, V. *J. Chem. Phys.* **1984**, *80*, 6245−6252.
(17) Iwamoto, M.; Uozumi, J.; Nishinari, K. In *Proceedings of the International Near Infrared Spectroscopy/Near Infrared Technology Conference*, Hello, J., Kaffka, K. J., Gonczy, J. L., Eds.; Akademiai Kiado: Budapest, Hungary, 1986; pp 3−12.
(18) Maeda, H.; Ozaki, Y.; Tanaka, M.; Hayashi, N.; Kojima, T. *J. Near Infrared Spectrosc.* **1995**, *3*, 191−201.
(19) Adams, M. J. *Chemometrics in Analytical Spectroscopy:* The Royal Society of Chemistry: Cambridge, U.K.,1995.
(20) Noda, I. *Bull. Am. Phys. Soc.* **1986**, *31*, 520.
(21) Noda, I. *Appl. Spectrosc.* **1993**, *47*, 1329−1336.
(22) Vandeginste, B. G. M.; Massart, D. L.; Buydens, L. M. C.; De Jong, S.; Lewi, P. J.; Smeyers-Verbeke, J. *Handbook of Chemometrics and Qualimetrics: Part B;* Elsevier Science B. V.: Amsterdam, The Netherlands, 1998; pp 88−104.
(23) Šašić, S.; Ozaki, Y. *Appl. Spectrosc.*, in press.
(24) Šašić, S.; Muszynski, A.; Ozaki, Y. *J. Phys. Chem. A* **2000**, *104*, 6380−6387.

$$VV_{synchronous} = \frac{1}{s-1}\mathbf{X}^{T}\mathbf{H}\mathbf{X} \quad (3A)$$

$$SS_{synchronous} = \frac{1}{\nu-1}\mathbf{X}\mathbf{H}\mathbf{X}^{T} \quad (3B)$$

Hilbert transform matrix.[25] The Hilbert transform matrix orthogonalizes the matrix of experimental data, and eqs 3A and B represent the scalar products between original data and their orthogonalized forms. Equations 2 and 3 represent the main features of generalized 2D correlation spectroscopy. The sample−sample approach offers the possibilities of analyzing the concentration changes of the species as a function of the external perturbation and is particularly powerful in combination with the variable−variable approach.[26]

EXPERIMENTAL SECTION

The sample was prepared by passing tap water through activated charcoal and reverse osmosis filters. The water was then distilled and purified by an Ultrapure Water System (model GSR-200, Advantec). The resistance of the finally prepared water was greater than 18.1 MΩ cm^{-2}. NIR spectra of water were measured from 11 000 (910 nm) to 4000 cm^{-1} (2500 nm) at an 8-cm^{-1} resolution by a Nicolet FTIR/NIR spectrometer (Magna Series 560, Nicolet Instruments Corp., Madison, WI), using a 1-mm quartz cell. The cell holder was connected to a water bath, and the sample temperature was adjusted by circulating temperature-controlled water. The sample temperature was calibrated by use of a digital thermometer (TNA-20, Tasco Japan Co. Ltd., Osaka, Japan), by measuring the sample temperature before and after each spectral measurement. This system guaranteed a temperature accuracy of ±0.1 °C at 25 °C and ±0.2 °C at 6 and 80 °C. A total of 512 scans were coadded for each temperature spectrum. The sample was heated from 6 to 80 °C at 2 °C steps, giving 38 sample spectra.

Temperature-dependent spectral measurements were also performed with an empty cell in order to remove possible temperature effects of the cell on the spectra. However, the results were virtually identical before and after the correction of the spectra, the only difference being a slightly lower signal-to-noise ratio after the correction. For these reasons, we chose to use the spectra as they were originally measured, with air as reference.

All 2D correlation calculations and plots were made in MATLAB (version 6, The MathWorks, Inc., Nattick, MA). Second-derivative spectra were calculated using the Savitzky−Golay derivation[27] with a gap size of 15 points, i.e., here 15 nm. Calculation of the derivative spectra and principal component analyses were performed in The Unscrambler v.7.5 (CAMO ASA, Trondheim, Norway).

RESULTS AND DISCUSSION

Panels A and B of Figure 1 show NIR spectra at every 10 °C in the region 1300−1600 nm of water before and after mean normalization, respectively. The dominating band centered at 1440 nm is due to the combination of symmetric and antisymmetric

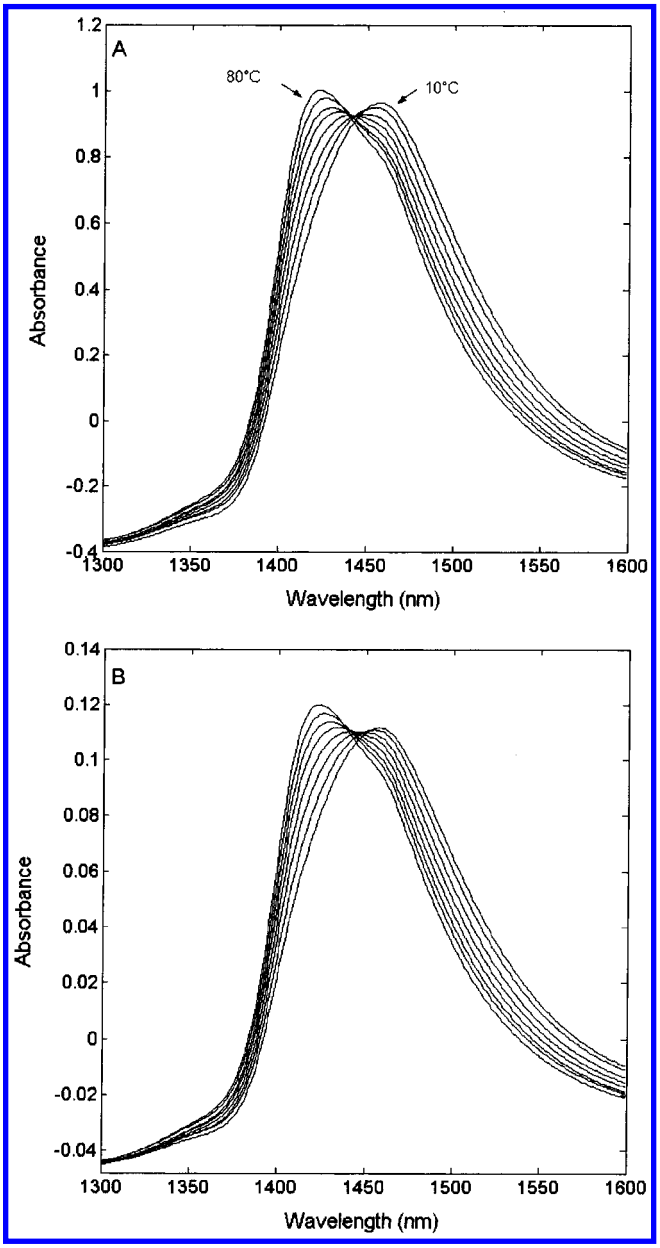

Figure 1. (A) Thirty-eight untreated NIR spectra of water measured over a temperature range of 6−80 °C at 2 °C increments. (B) The same spectra after mean normalization.

O−H stretching modes (first overtone). In Figure 1A, the band apparently shifts from 1460 to 1424 nm as the temperature is increased from 6 to 80 °C. Note that the spectra give an isosbestic point at ∼1446 nm. The apparent peak shift is attributed to the fact that the intermolecular H-bonds are generally weakened with temperature, which strengthens the covalent O−H bonds and consequently causes them to vibrate at higher frequencies. The total peak area decreases by more than 6% upon going from the lowest (6 °C) to the highest (80 °C) temperature, and the spectra were therefore mean normalized (to ensure fair integration, the region ranging from 1260 to 1760 nm, which are the wavelengths where the curves meet, was used). The decrease in overall area is in line with results previously reported.[5,8,10,11,16] In the process of mean normalization, all spectra are divided by their mean values in order to equalize the influence of all the variables by setting

(25) Noda, I. *Appl. Spectrosc.* **2000**, *54*, 994−999.
(26) Šašić, S.; Muszynski, A.; Ozaki, Y. *J. Phys. Chem. A* **2000**, *104*, 6388−6394.
(27) Savitzky, A.; Golay, M. J. E. *Anal. Chem.* **1964**, *36*, 1627−1638.

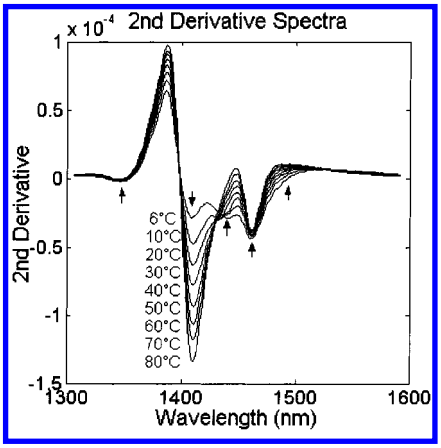

Figure 2. Second derivatives of selected spectra in Figure 1B.

the absorption coefficients of the species in the system to be the same.

Figure 2 shows the second derivatives of selected spectra at temperatures ranging from 6 to 80 °C, indicating peaks at 1346, 1411, 1441, and 1462 nm. It also shows a separation of the spectra around 1490 nm, and a broad negative feature can be found around 1650 nm (not shown here). Iwamoto et al.[17] and Maeda et al.[18] tentatively assigned these features to a combination involving a rotational mode of water (1346 nm)) and water molecules with different numbers of H-bonds (1411, 1441, 1462, 1490, and 1650 nm), respectively. Upward pointing arrows in Figure 2 indicate decreasing peak intensity with temperature, while the downward pointing arrow indicates increasing intensity. It should be noted from this figure that the temperature-dependent intensity changes are greatest at 1411 nm, followed by the feature around 1490 nm and the peak at 1462 nm. Changes can also be seen at the 1441-nm peak, which develops into a shoulder and eventually disappears with increasing temperature. The feature at 1346 nm changes little with temperature, and the broad feature centered at 1650 nm shows no changes whatsoever during temperature elevation. Derivation gives an indication of where original spectral features can be found that comprise the spectra of Figure 1. However, the derivative spectra are interpretable only in a visual fashion, as was also the case for the original spectra. To perform an objective analysis of the spectral dynamics, 2D correlation spectroscopy is utilized. The spectral region above 1600 nm was found not to be contributing to any results and was therefore omitted from further calculations.

2D Correlation Spectroscopy. Panels A and B of Figure 3 show the synchronous and asynchronous maps, respectively, of the 2D variable−variable correlation analysis of the spectral region 1300−1600 nm (derived by eqs 2A and 3A). The synchronous map shows autopeaks at 1412 and 1491 nm, indicating that the spectral features at these positions vary in phase with each other. The appearance of negative cross-peaks between the autopeaks and their signs reveals that the changes are correlated and occur in opposite directions. The asynchronous map shows two features with maximum and minimum values at 1412 and 1438 nm, indicating that out-of-phase spectral changes occur at these wavelengths and that the change at 1412 nm occurs before or at a higher rate than that at 1438 nm. The asynchronous map in Figure 3B is not very convenient for extracting information due to high overlapping of the slice spectra of which it is composed.

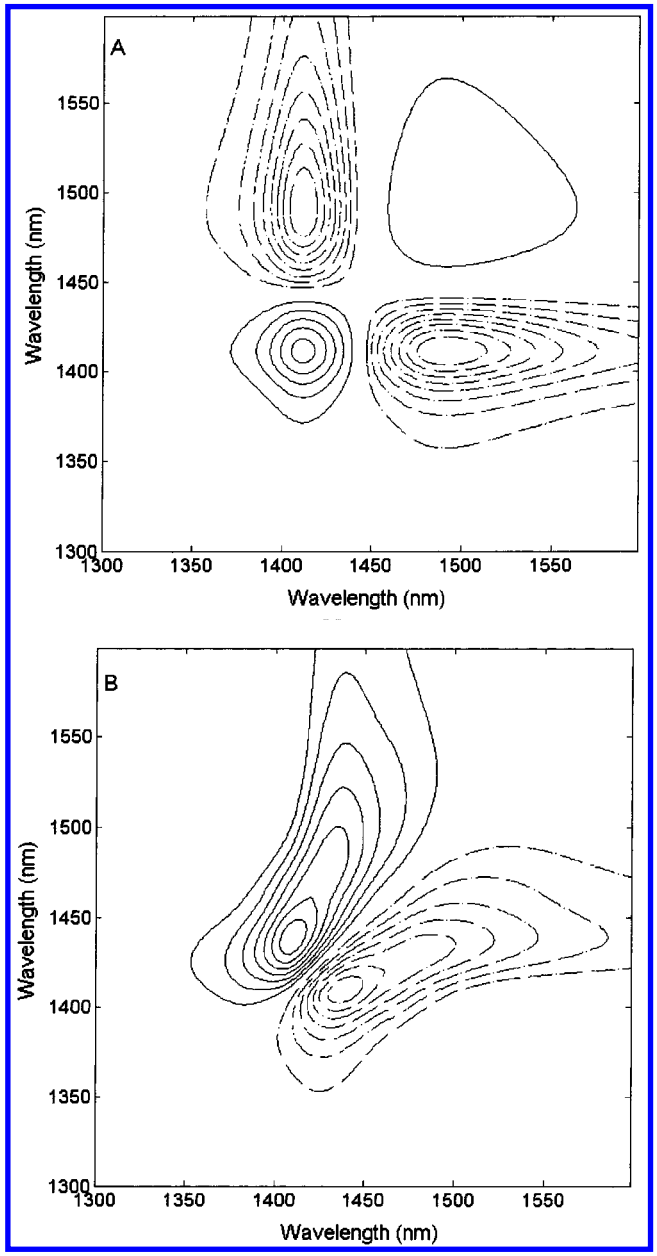

Figure 3. Synchronous (A) and asynchronous (B) variable-variable two-dimensional correlation spectra, generated from the spectra shown in Figure 1B.

To improve the clarity of the asynchronous correlations, we have plotted the most intensive slice spectra. Figure 4A shows slice spectra at 1412 (solid line) and 1438 nm (dashed line) derived from the asynchronous map in Figure 3B. The 1412-nm slice shows a negative peak at 1438 nm, flanked by smaller positive peaks at 1396 and 1524 nm. The 1438-nm slice spectrum consists of a positive peak at 1411 nm and a negative feature at 1483 nm. For comparison, the asynchronous slices at the two most prominent peaks at 1412 and 1491 nm in the synchronous map are illustrated in Figure 4B. The slices at 1412 (solid line) and 1491 nm (dashed line) show peaks at the exact same positions, which indicates that the same changes are correlated to the changes at both these wavelengths. What might be the most interesting finding in this illustration is that the correlation intensities at 1412 and 1491 nm are approximately equal to zero. This means that the changes at these two wavelengths are

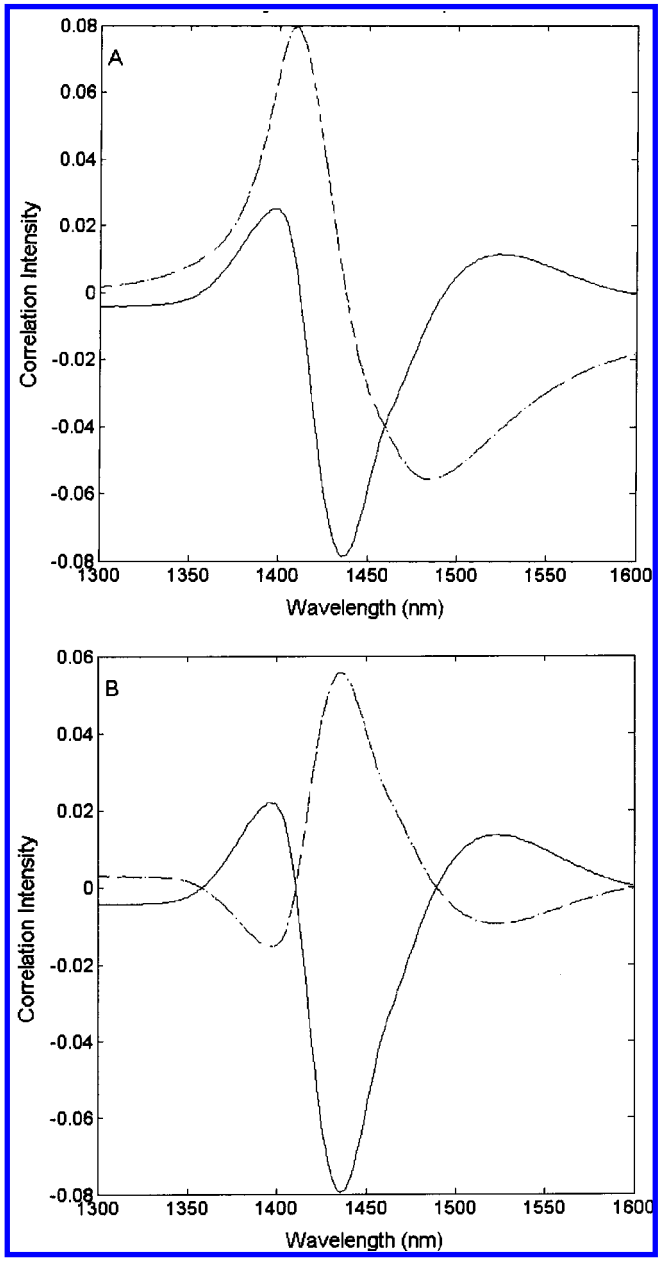

Figure 4. (A) Slice spectra from the asynchronous map shown in Figure 3B. Solid line represents the slice at 1412 nm, and dotted line represents the slice at 1438 nm. (B) The slices at 1412 (solid line) and 1491 nm (dashed line).

perfectly correlated, since peaks appear at these positions in the synchronous spectrum and not in the asynchronous spectrum.

What are the implications of the VV 2D correlation analysis? The synchronous spectra simply cover the spectral variances that dominate the temperature-dependent spectral system. If the system contained only two components, it would have only one degree of freedom after mean normalization and centering (the mean centering is always the first step of the calculation of 2D maps[21]), and the whole system would be sufficiently represented by the variation of only one vector, which is rendered as a synchronous spectrum. Moreover, it is clear from eq 3A that for a simple two-component system an asynchronous spectrum should not develop any peak, because the asynchronous spectrum is the scalar product of the dynamics of only one vector and the orthogonalized form of its dynamics. It is, however, beyond doubt that the asynchronous spectrum in Figure 3B (and consequently also Figure 4) is not noiselike. It shows that, in addition to the bands at 1412 and 1491 nm, at least one more band exists, with the one at 1438 nm being most prominent. In other words, the intensity changes at 1412 and 1491 nm are in phase with each other and different from those at 1438, 1396, and 1524 nm. The asynchronous spectra lead us to conclude that the NIR spectra of water are mainly determined by the bands at 1412 and 1491 nm that represent two dominating water species. However, three additional bands exist, none of which changes in phase with those at 1412 and 1491 nm. Unambiguous appearance of the band at 1438 nm shows the asset of the 2D correlation analyses compared with the second-derivative spectra, where this band is just slightly discernible.

Panels A and B of Figure 5 show the synchronous and asynchronous slice spectra from the sample−sample 2D correlation analysis (SS 2D, derived from eqs 2B and 3B), respectively. Since slice spectra normally are more illustrative than 2D maps for SS 2D spectroscopy, only the slices will be illustrated. The synchronous spectra show linear changes in the correlation intensities for the whole temperature region, which is typical for a two-component system where one component increases in concentration while the other decreases at a constant rate.[24] Figure 5A reveals that the spectra at 6 and 80 °C are the most negatively correlated and that the correlations increase as the two temperatures approach each other. The figure reflects that the spectrum at the mean temperature in the experiment (40 °C) roughly corresponds to the mean of all the spectra examined. The asynchronous slice spectra in Figure 5B show some distinct features, indicating that small out-of-phase changes take place. The highest correlation intensities are found between the sample pairs (6, 38), (6, 80), and (52, 80) °C. Despite the pattern formed by the asynchronous spectra, there unfortunately is no clear turning point or separations that can relate certain spectral species to specific temperatures. The most interesting observation in this plot is that a pattern does exist, indicating that temperature-related changes take place that cannot be explained by the continuous two-component change seen in Figure 5A. More work is needed to explain these phenomena.

2D spectroscopy has up to this point given us substantial information on the dynamic behavior of the water spectra. It cannot, however, quantify the significance of the features elucidated by the synchronous and asynchronous spectra. To solve this problem, and also to verify the results of the 2D analysis, we perform a PCA on our data.

Principal Component Analysis. Panels A−D of Figure 6 show the first four loadings and scores, respectively, from the PCA of the spectral region 1300−1600 nm. The eigenvalues and the variance explanation of the first six components are shown in Table 1. The first principal component accounts for 99.426% of the total spectral variance, and its loading is dominated by a positive peak at 1412 nm and a broader negative feature centered at 1492 nm. The score vector of the first component is simply a straight line, indicating that the changes in the spectra that are extracted by this component occur in a continuous fashion throughout the whole temperature region. With a variance explanation of 99.426%, it can be considered that the first principal component explains all spectral changes of importance. However,

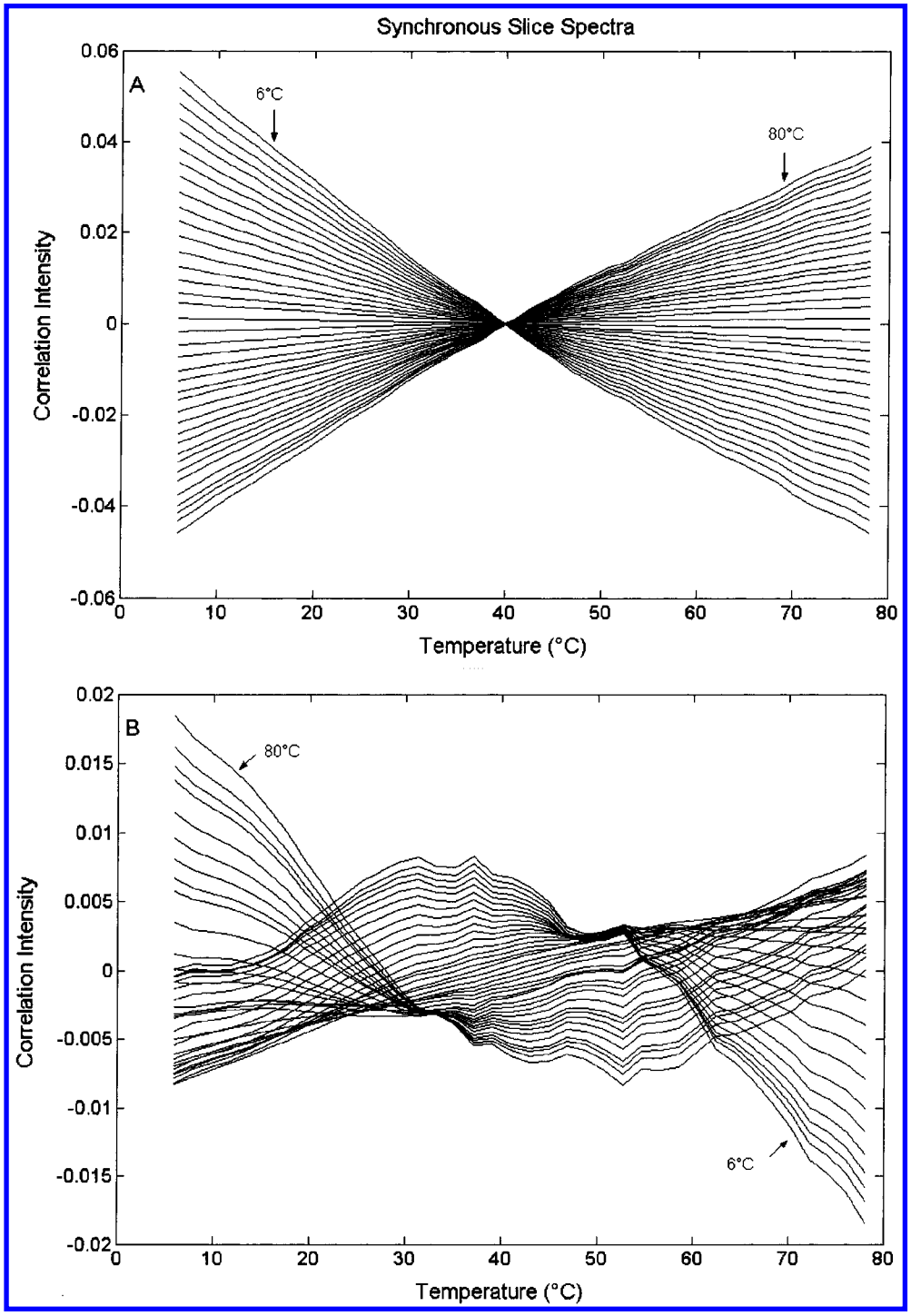

Figure 5. All synchronous (A) and asynchronous (B) slice spectra of the sample−sample two-dimensional correlation analysis.

in this study we are also interested in investigating whether any changes occur beyond the dominating phenomena. The second principal component accounts for 0.434% of the total variance, and its loading shows a dominating positive peak at 1438 nm and a negative feature at 1387 nm. Its score vector has a parabolic shape, with 38 °C at the top position, which indicates that the spectral changes captured by the second loading have a turning point around this temperature. As can be seen from the results, the VV 2D synchronous spectra offer basically the same information as the loading of the first principal component, while the W 2D asynchronous spectra roughly correspond to the loading of the second principal component. These relations have recently been discussed by Šašic et al.[28] The third component accounts for 0.138% of the total variance, and the loading gives a line with all its points above 0, indicating little more than a baseline shift. It does, however, display a small negative feature at 1420 and a more prominent one at 1469 nm. The third score vector in Figure 6C seems to have some structure up to about 26−30 °C, but above this temperature region, the vector loses it structure and gets noisy. It should be noted that the second-derivative spectra in

(28) Šašić, S.; Muszynski, A.; Ozaki, Y. *Appl. Spectrosc.*, in press.

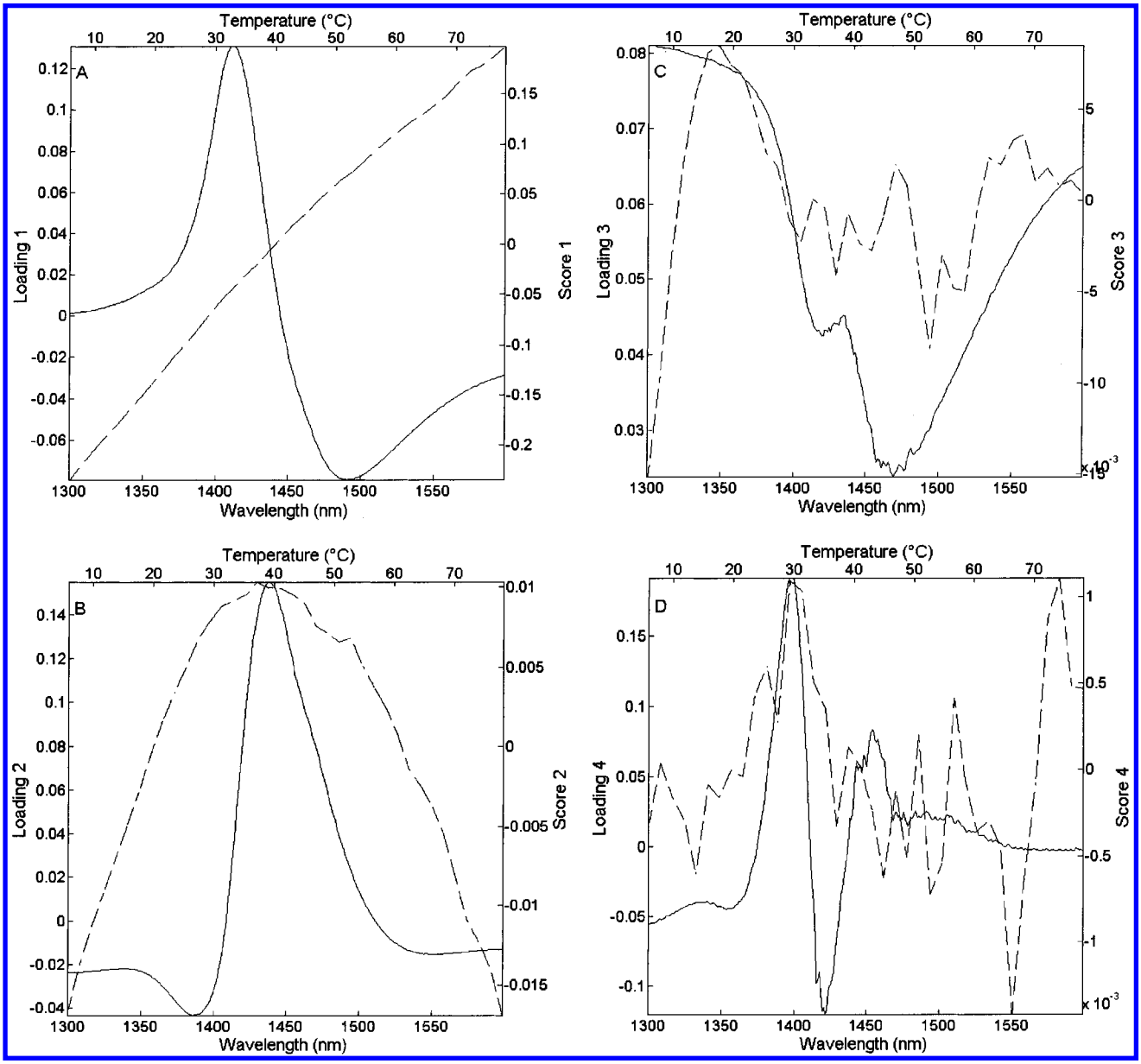

Figure 6. Loadings and scores of the first four principal components: (A) first, (B) second, (C) third, and (D) fourth. Bottom and left axes represent loadings, plotted in solid lines. Top and right axes represent scores, plotted in dashed lines.

Table 1. Eigenvalues (EV) and Percentage of Variance Captured (VC) for the First Six Principal Components (PC), for Original, Normalized, and Autoscaled Spectra of Water

		PC1	PC2	PC3	PC4	PC5	PC6
original spectra	EV	1.53×10^{-1}	3.03×10^{-4}	2.30×10^{-4}	1.80×10^{-6}	1.21×10^{-6}	1.88×10^{-7}
	VC	99.650	0.197	0.151	0.001	0.001	0.000
normalized spectra	EV	1.70×10^{-2}	7.42×10^{-5}	2.36×10^{-5}	3.00×10^{-7}	5.64×10^{-8}	1.55×10^{-8}
	VC	99.426	0.434	0.138	0.001	0.001	0.000
autoscaled spectra	EV	3.59×10^{1}	8.14×10^{-1}	4.34×10^{-1}	6.15×10^{-4}	2.46×10^{-4}	1.30×10^{-4}
	VC	96.637	2.192	1.168	0.001	0.001	0.000

Figure 2 show a distinct peak at 1462 nm, which is reasonably close to the feature at 1469 nm in the third loading. This indicates that we have a spectral feature whose concentration changes up to around 26−30 °C and subsequently decreases below the limit of detection or stabilizes. In the case of Figure 6D, both vectors seem to contain a lot of noise, which could be expected from the component's low eigenvalue and variance explanatory value.

We believe that the results presented in Figure 6 and Table 1 definitely put toward the presence of three principal components. This is perhaps easiest seen from Table 1, where the eigenvalues unambiguously corroborate such a conclusion by the gap between the variances of the fourth and lower factors compared with the third one. This becomes particularly noteworthy after autoscaling, which is a purely statistical pretreatment, is carried out. We can

reach the same conclusion by investigating the loadings and scores, differentiating information and noise according to their shapes. The presence of the three components allows us to quantify the evidence obtained by considerably simpler analytical techniques (2D spectroscopy and second derivation). The contribution to the total spectral variances of all the bands, except for those at 1412 and 1492 nm, is lower than 0.5%, and the same percentage is valid for all the concentration contributions that disarrange the conversion of the species giving the band at 1492 nm into the species showing the band at 1412 nm. Thus, though the NIR spectra of water apparently represent a multicomponent system, their dynamic behavior can be approximated to that of a two-component system without a severe loss of information.

To investigate the temperature dependence of the NIR spectra of water in more detail, we divided the spectra into three different temperature regions and performed PCA and 2D correlation analyses for each region separately. The 2D correlation maps showed no distinct differences between the three regions. The PCA results for the three regions were also similar to the results obtained for the whole temperature region, the only difference being that the third loading vector (Figure 6C) changed places with the second loading (Figure 6B) for the lowest temperature region. This vector also explained a larger amount of the total variance for the lowest temperatures, a tendency that was already indicated by the score vector in Figure 6C. These results suggest that there is no hidden structure in our data that cannot be found by analysis of the whole temperature region, and the results from this division are therefore not illustrated.

Water Structure. Our results indicate that there are two major spectral changes occurring as a function of temperature, taking place at about 1412 and 1491 nm. These bands are clearly observed also in the difference spectra of water.[18] The intensity changes at these two wavelengths occur in phase with each other in opposite directions. The fact that both PCA and 2D correlation results give distinct bands at these positions, which are found clearly outside the apparent peaks in the original spectra in Figure 1 (at 1424 and 1460 nm), suggests that we are dealing with two spectroscopically different components, where one converts into the other as a function of temperature. In other words, our results roughly suggest a two-state mixture model, where the two states represent weaker and stronger H-bonding at 1412 and 1491 nm, respectively.

Another finding that is in disfavor of a simple continuum model is the presence of an approximate isosbestic point in the original spectra (Figure 1A). A perfectly continuous system would change by shifting the peak position gradually while preserving the peak shape, leaving a broad area of overlapping spectra.[16] This is clearly not the case here. However, our results also show that distinct features are present in the spectral data, which cannot be explained by a simple two-component model. The most prominent deviation is found at 1438 nm both in the second principal component's loading (Figure 6B) and in the VV asynchronous map (Figure 3B). What we can say about this feature at this point is that its dynamic changes are out of phase with the changes at 1412 nm and that it has only minor impact on the total changes of the system. The band at 1438 nm could be assigned as representative of a third state of water, whose concentration changes only little with temperature. However, we have no clear proof that such an intermediate state exists and we can simply state that this wavelength represents small changes that cause deviations from a two-state model. As seen in Figures 4A, 6B, and 6C, features are also indicated just below 1400 nm, at 1469 nm, and just above 1500 nm. These broad features are much less intense than the one around 1438 nm, and their positions are also less precise. These features complicate the structural picture further, but their impact is too diffuse to allow interpretation in favor of specific species or structures. Some similarities can, however, be seen between the feature at 1469 nm, whose score vector in Figure 6C points to specific changes below 26−30 °C, and a Raman study by D'Arigo et al.[29] that concluded that a two-state approximation was gradually violated as the temperature decreased below 20 °C.

Hare and Sorensen[13] carried out a Raman study of HOD in liquid H_2O in the temperature range −31.5 to +160 °C and suggested a two-state mixture model for water, consisting of species at a higher and a lower energy state, designated NHB and HB, respectively. They emphasized that the model was only two-state in a broad sense, indicating that NHB and HB were clearly two distinct species but that the changes within the HB species could be described by a continuum model. They found that a rigid two-state model could not be supported by their data. However, after introducing a weak relaxation of the restrictions, they found that a two-state model was justifiable. It should be noted that expressions such as NHB and HB, and also 3HB and 4HB that are found in the literature, do not necessarily imply that H-bonds are broken. The bonds are better described as distorted, i.e., weakened through bending or stretching. A very similar conclusion was reached by Angell and Rodgers,[16] who suggested a disrupted network model described as a "quasi-lattice with broken bonds". Although they stated that the ground state of water model is a "random network structure that is certainly continuum with respect to H-bond strength", they later pointed out that the structure lay somewhere between those described by a continuum and a two-state mixture model. The most important piece of evidence on which the previous conclusions were based was the appearance of a weak band near the isosbestic point. This band was estimated to remain unchanged during heating. In the present study, the existence of this band is supported, and its exact position (1438 nm) and contribution to the overall spectral variations are estimated.

A better insight into the rather wide conception of NHB and HB was given by Walrafen et al.[11] They reported a Raman study of water over a temperature range of 3−95 °C, which suggested a water structure composed of fully H-bonded (4-HB) and partially H-bonded (3-HB) molecules. They reported results to be in accordance with a bifurcated H-bond model presented by Giguére[9] (The term bifurcated implies that both hydrogen atoms on one water molecule are H-bonded to the same oxygen atom on another molecule). This model involves two different kinds of H-bonds: the linear bond found in ice and the weaker bifurcated bond between three adjacent molecules. These three reports suggested structure models that are slightly different from each other, but they all suggested structures that can be considered somewhat distorted two-state models. In an attempt to precisely identify the

(29) D'Arigo, G.; Maisano, G.; Mallamace, F.; Migliardo, P.; Wanderlingh, F. *J. Chem. Phys.* **1981**, *75*, 4264.

sources of Raman scattering, Walrafen et al.[11] defined the positions and widths of the scattering species in water by means of curve fitting.

Finally, the present study just partially concurs with the findings reported by Libnau et al.[5] We present unambiguous evidence that the NIR spectra of water comprise more than two components. This statement is in a slight disagreement with the water structure suggested by Benson and Siebert,[6] which was supported by Libnau et al.[5] We do not deny, however, the Benson and Siebert proposal. Considering solely NIR spectra (as well as other types of vibrational spectra) without support from thorough theoretical studies cannot give an adequate answer to the question on how water molecules are arranged. The spectroscopic difference between (1) gradually weakened H-bonds during heating and (2) conversions between 3H- and 4H-bonded water molecules seems to be so wage that a single suggestion is very hard to propose. The main asset of this study is that it offers an objective analysis of a rather complicated, though visually fairly simple, set of spectra. Combined with several physicochemical studies, the aftermaths of this paper are in the same line as the quasi-two-state model. No temperature region has been found that can be explicitly related to any specific changes, except for the monotonic conversion of strongly H-bonded molecules into weakly H-bonded molecules with increasing temperature. Further work is needed to establish the nature of the species found in our study.

CONCLUSIONS

The statistical analyses of the NIR spectra of water in the 1300−1600-nm region at temperatures between 6 and 80 °C show that the spectral variation can be described mainly by the two bands at 1412 and 1491 nm, which arise from weakly and strongly H-bonded water molecules, respectively. This conclusion is reached by two analytical techniques: 2D correlation spectroscopy and PCA. The PCA results quantify the significance of the variances and reveal that the bands at 1421 and 1491 nm account for more than 99% of the total data variance. These two bands clearly represent different species that are perfectly correlated; i.e., the concentration of one species increases on the expense of the concentration of the other when the temperature is changed, and in this sense our results are in accordance with a two-state mixture model. However, the features located around 1438 nm, and to a lesser degree below 1400 nm, around 1469 nm, and above 1500 nm, show that the simple two-state model is not adequate. The presence of a third species that has very small impact on the overall spectral dynamics, whose concentration changes very little during temperature elevation, is established. This is, to our knowledge, the first time that clear evidence for the existence of more than two NIR-absorbing species of water is reached. This result significantly differs from many others by the fact that it was reached using solely objective statistical techniques and direct sampling. Referring to the previously proposed models of water leads us to conclude that, according to its temperature-dependent NIR spectra, water can be portrayed as a quasi-two-component mixture. Sample−sample correlation analyses and PCA scores illustrate that there is no indication of the existence of characteristic temperature ranges. Further work is needed to establish the nature of the species found in our study.

ACKNOWLEDGMENT

Dr. R. J. Berry of Kwansei-Gakuin University is acknowledged for composing the program for the 2D correlation spectroscopy in MATLAB.

Received for review January 23, 2001. Accepted April 9, 2001.

AC010102N