On the feasibility of near infrared spectroscopy to detect contaminants in water using single salt solutions as model systems

CrossMark

A.A. Gowen [a,b,*], F. Marini [c], Y. Tsuchisaka [b], S. De Luca [c], M. Bevilacqua [c], C. O'Donnell [a], G. Downey [a,d], R. Tsenkova [b]

[a] School of Biosystems Engineering, University College Dublin, Dublin, Ireland
[b] Biomeasurement Laboratory, Kobe University, Kobe, Japan
[c] Department of Chemistry, University of Rome "La Sapienza", Rome, Italy
[d] Teagasc Food Research Centre, Ashtown, Dublin, Ireland

ARTICLE INFO

ABSTRACT

This research work evaluates the feasibility of NIRS to detect contaminants in water using single salt solutions as model systems. Previous research has indicated the potential of near infrared spectroscopy (NIRS) for detecting solutes in water; however, a comprehensive investigation of the limit of detection of this technique has not been carried out. Near infrared transmittance spectra of aqueous salt solutions in the concentration range 0.002–0.1 mol L^{-1} (equivalent to 117–13,334 ppm or 0.0001–0.01% mass/mass) were investigated. The first overtone region of the near infrared spectrum (1300–1600 nm) was found to be the most effective wavelength range for prediction of salt concentration in aqueous solutions. Calibration models built using this wavelength range and employing the extended multiplicative scatter spectral pre-treatment resulted in root mean squared error of prediction values ranging from 0.004 to 0.01 mol L^{-1}. The limit of detection (LOD) was estimated to be of the order of 0.1% (mass/mass) or 1000 ppm. Within the framework of Aquaphotomics, it was possible to examine the effect of different salts on the NIR spectra of water in the first overtone range. Our results were confirmed through test experiments at various geographical locations employing dispersive and Fourier transform type NIRS instruments.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Water quality monitoring is a complex multi-spatial and multi-temporal problem, extending from safety monitoring of drinking water to operational monitoring of waste waters. Standard methods for water quality analysis involve intensive sampling regimes and multi-step sample preparation, requiring manual inputs which prohibit their integration into continuous monitoring systems. Consequently, there is currently a need for the development of low-cost, robust, reliable monitoring techniques that can be easily integrated into water flow systems [1]. Vibrational spectroscopic techniques, including near infrared (NIR) and mid infrared (MIR) spectroscopy, have been proposed as potential tools to improve the speed and reduce the cost of water quality testing [2]. Such techniques have become standard quality analysis tools in the food and pharmaceutical industries, in which they are commonly implemented as process analytical technologies for continuous quality monitoring.

Fundamental frequencies for molecular vibrations (e.g. stretching, bending) occur in the MIR region while their overtones and combinations appear in the NIR spectral range. Absorption bands tend to overlap in the NIR, resulting in a broad spectral response; consequently, the NIR range contains less accessible molecule-specific information than does the MIR. Analysis of aqueous samples in the MIR is challenging, due to the strong absorbance of MIR light by water, leading to a low signal to noise ratio. To overcome this problem, innovative sample presentation techniques, such as attenuated total reflectance (ATR), have been developed; however volume of sample that can be measured at any instant is inherently limited by this approach. The weaker absorption of NIR light by water facilitates the direct measurement of greater volumes of aqueous samples. For this reason, NIRS is an appealing technique.

The so-called water absorbance peaks of the NIR wavelength range are broad peaks located around 1300–1600 and 1900–2200 nm. These represent overlapped overtone and combination vibrations of fundamental stretching, bending and libration vibrations of intramolecular

OH bonds, and are also sensitive to changes in the intermolecular hydrogen bonds. They represent the most prominent features in the NIR spectrum of aqueous samples and they can dominate other, more subtle, absorbance features arising from other molecules in the solution. Although solutes such as metals and salts do not absorb NIR light, their presence can be detected via perturbation of water's hydrogen-bonded network. When solutes are added to water, intermolecular hydrogen bonds rearrange their orientation, resulting in changes in the shape and peak position of the water peaks in the NIR spectrum. The extent of these changes depends upon the properties of the solute, such as size and ionic strength [3]. It has previously been demonstrated that this effect could be used for the prediction of solute concentration in water [4,5], However, the limit of detection of the technique was not investigated in those studies. More recently, Sakudo and co-workers [6] investigated the NIR spectra of metals (Copper, Magnesium, Sinc and Iron) in 0.1 M Nitric Acid solutions, using a range of concentrations from 0 to 10 $mg\,L^{-1}$ and estimated the limit of detection of their technique to be 1–4 $mg\,L^{-1}$.

Aquaphotomics has recently emerged as framework to examine the behavior of water in various systems (e.g. biological systems such as plants and animals) through analysis of their NIR spectra [7]. It has been proposed that Aquaphotomics has potential for contaminant detection in aqueous systems. However, no studies have yet investigated this comprehensively [8]. In this study, we investigate the potential of NIRS for detection of low concentrations of contaminants in water, using salts as model contaminants. Salts were used for the following reasons: they represent relatively simple systems, exhibit no NIR absorption spectrum (therefore the resulting observed effects are solely due to changes in the water absorbance peaks), have relevance to salinity measurements and are comparable to previous studies. Although previous studies have investigated the effects of salts on water using NIRS, this is the first work employing low concentrations, which resemble those expected for major contaminants (0.002–0.1 $mol\,L^{-1}$, which corresponds to 0.01–1.3% w/w, or 117–13,334 ppm).

The main objectives of the work were to:

1. Evaluate the performance of NIR spectroscopy combined with chemometric modelling for prediction of salt concentration in water.
2. Evaluate the robustness of the results obtained using multiple geographical locations for testing.
3. Investigate the effects of low concentrations of salts on water structure using NIR spectroscopy.

2. Materials and methods

NIR transmittance spectra were measured over the wavelength region of 1000–2500 nm using quartz cuvette cells with 1 mm optical path length. Each experiment was repeated on three different days using independently prepared samples on each day (where "experiment" refers to the acquisition of NIR spectra for one salt solution at all concentrations). Aqueous solutions of 4 salts (NaCl, KCl, $MgCl_2$, $AlCl_3$) were prepared using Millipore water (Millipore, Molsheim, France, resistance ~20 m Ω). These salts were selected to create a set of solutions with different ionic strengths while keeping the anion constant. Ten concentrations (0.002, 0.004, 0.006, 0.008, 0.01, 0.02, 0.04, 0.06, 0.08, 0.1 $mol\,L^{-1}$) were prepared for each salt. NIR transmittance spectra of these samples were collected in an order randomised with respect to concentration. In order to monitor interfering signals (e.g. shifts in water absorbance bands temperature variations), control measurements of Millipore water were taken at the beginning, middle and end of each experiment. Ambient temperature over the three locations ranged from 21 to 26 °C.

In order to investigate the robustness of the results obtained, the experiment described below was carried out by three different researchers (1st, 2nd and 3rd authors) using different spectrometers, in three different geographical locations (Dublin (Ireland), Kobe (Japan) and Rome (Italy) respectively). NIR spectra in locations Ireland and Japan were obtained using NIR System 6500 spectrophotometers (Foss NIR-System, Laurel, USA), while those in Rome were obtained using a Nicolet 6700 FT-NIR instrument (Thermo Scientific Inc., Madison, WI).

2.1. Data analysis

Calibration models were built to predict salt concentration using PLS regression (PLSR), employing various spectral pre-treatments (standard normal variate (SNV), SNV+Savitsky Golay 1st and 2nd derivative smoothing (using 15 point smoothing window and 3rd order polynomial) and SNV+extended multiplicative scatter correction (EMSC, using 1st principal component loading of control (Millipore water) spectra as an interferent spectrum).

PLSR models were developed on various wavelength regions of the NIR spectrum (1000–2500 nm, 1000–1300 nm, 1300–1600 nm, 1600–1900 nm, 1900–2200 nm, 2200–2500 nm). In addition, two variable selection methods were applied to the data to evaluate their usefulness in terms of predictive model performance, as follows:

2.1.1. Genetic algorithm—Backward interval PLS (GA-biPLS)
Genetic algorithms are well known as powerful variable selection techniques. However, one of the main problems when dealing with GA is the risk of overfitting, i.e. to select, as meaningful, variables which are only by chance correlated with the responses. As the probability of chance correlation increases with the number of variables, one way to overcome this problem is to use a limited number of variables as an input to the GA selection. For this purpose, we chose a modified version of backward interval PLS (biPLS) [9]. The general concept of biPLS is to identify the spectral regions less meaningful for prediction and iteratively remove them from the data set in a backwards fashion. In its "normal" implementation, the number of intervals is defined only once (e.g. 50 intervals) and the procedure is stopped when a minimum in cross-validation error is reached. When used for pre-selection of spectral intervals prior to GA, backward elimination is performed up to the point when a 'maximum number of variables retained' criterion is met. In our case, a maximum of 300 variables was selected. GA was carried out using the PLS-Genetic Algorithm toolbox for Matlab [10] and biPLS was applied using functions in the iToolbox for Matlab [11].

2.1.2. Ensemble Monte Carlo variable selection (EMCVS)
This procedure selects the most informative variables in a dataset based on PLS regression coefficients. Compared to other state of the art variable selection methods, EMCVS increases consistency of variable selection and reduces processing time. This method is described in more detail elsewhere [12].

All models were constructed by sequentially leaving one experimental day out as a test set and using the remaining two experimental days as a calibration set. The number of latent variables for inclusion in each PLSR model was estimated by performing 2-fold cross validation on the calibration set, where the cancelation groups coincided with the different measurement days. Accordingly, to build the model, every second spectrum was left out of the calibration set, and then the model was applied to the remaining spectra. The rank corresponding to the minimum of the root mean squared error of cross validation was then selected as the optimal number of latent variables for each model. This method was selected as a fair means of comparison between competing models.