Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 171 (2017) 104–111

Contents lists available at ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

journal homepage: www.elsevier.com/locate/saa

Investigation of fluorescence methods for rapid detection of municipal wastewater impact on drinking water sources

Nicolas M. Peleato [a,*], Raymond L. Legge [b], Robert C. Andrews [a]

[a] Department of Civil Engineering, University of Toronto, 35 St. George St., Toronto, Ontario M5S 1A4, Canada
[b] Department of Chemical Engineering, University of Waterloo, 200 University Ave., Waterloo, Ontario N2L 3G1, Canada

ARTICLE INFO

Article history:
Received 28 March 2016
Received in revised form 13 July 2016
Accepted 16 July 2016
Available online 18 July 2016

Keywords:
Fluorescence spectroscopy
Organic matter characterization
Wastewater
Drinking water
Support vector machines

ABSTRACT

Fluorescence spectroscopy as a means to detect low levels of treated wastewater impact on two source waters was investigated using effluents from five wastewater facilities. To identify how best to interpret the fluorescence excitation-emission matrices (EEMs) for detecting the presence of wastewater, several feature selection and classification methods were compared. An expert supervised regional integration approach was used based on previously identified features which distinguish biologically processed organic matter including protein-like fluorescence and the ratio of protein to humic-like fluorescence. Use of nicotinamide adenine dinucleotide-like (NADH) fluorescence was found to result in higher linear correlations for low levels of wastewater presence. Parallel factors analysis (PARAFAC) was also applied to contrast an unsupervised multiway approach to identify underlying fluorescing components. A humic-like component attributed to reduced semiquinone-like structures was found to best correlate with wastewater presence. These fluorescent features were used to classify, by volume, low (0.1–0.5%), medium (1–2%), and high (5–15%) levels by applying support vector machines (SVMs) and logistic regression. The ability of SVMs to utilize high-dimensional input data without prior feature selection was demonstrated through their performance when considering full unprocessed EEMs (66.7% accuracy). The observed high classification accuracies are encouraging when considering implementation of fluorescence spectroscopy as a water quality monitoring tool. Furthermore, the use of SVMs for classification of fluorescence data presents itself as a promising novel approach by directly utilizing the high-dimensional EEMs.

1. Introduction

Source water protection and monitoring, especially when considering surface waters, are important components for reducing public health risk associated with drinking water. Unplanned indirect reuse of treated wastewater is commonplace and likely to intensify given trends in urbanization and increases in population density [7]. While microbial risk of adequately treated wastewater impacting treated drinking water quality is likely low, one or more treatment failures could result in intermittent contamination events [38]. Additionally, increased frequency and intensity of storm events due to climate change is anticipated to impact the occurrence of sewer overflow events into drinking water sources [31]. Furthermore, wastewater components refractory to treatment, such as pharmaceuticals, presents a relatively unknown level of risk [13]. Current approaches for the estimation of wastewater impacts on source waters rely on laborious measurement of anthropogenic micropollutants such as caffeine and artificial sweeteners [5,29], which are problematic for incorporation into control schemes. The expected transient nature of treated wastewater impacts due to events such as sewer overflows or changes in currents dictates that detection of wastewater impacts needs to occur rapidly within hours or less. Currently available online water quality measures, such as total organic carbon, conductivity, chlorine residuals, turbidity, and pH are only expected to detect extreme impacts (> 1% by volume) of recycled or treated wastewater [15]. As such, there is a need for systems which may provide more sensitive real time assessment of source water quality to better inform treatment and protection plans.

Fluorescence spectroscopy is becoming a popular water quality measure and natural organic matter characterization technique [16] with the potential for implementation as an online monitoring tool [35]. Fluorescence data is typically collected at iterated excitation and emission wavelengths, referred to as excitation-emission matrices (EEMs). Several fluorescence EEM analysis techniques are reported in the literature, the most simplistic of which tracks intensities at key wavelengths [25]. However, this approach neglects the majority of

---

Abbreviations: CV, cross-validation; EEM, excitation-emission matrix; FDR, false discovery rate; FN, false negatives; FNR, false negative rate; FP, false positives; FPR, false positive rate; MLR, multinomial logistic regression; NADH, nicotinamide adenine dinucleotide; PARAFAC, parallel factors analysis; RBF, radial basis function; ROC, receiver operating characteristics; RU, raman units; SVM, support vector machine; TN, true negatives; TOC, total organic carbon; TP, true positives; TPR, true positive rate; UV, ultraviolet; V, validation.

* Corresponding author.

E-mail address: nicolas.peleato@mail.utoronto.ca (N.M. Peleato).

http://dx.doi.org/10.1016/j.saa.2016.07.031
1386-1425/

information collected and since molecular structure is tied to peak position, which may have a degree of commonality among individual organic molecules, organic fluorescence peaks are highly convoluted [32]. To account for this overlapping nature of organic classes and the high-dimensional nature of EEMs, dimensionality reduction and deconvolution using multi-way PARAFAC analysis and similar approaches have been successful [25,37]. The application of PARAFAC elucidates underlying components which vary in the data set, however does not guarantee representation of specific organic compounds. For example, similar components between several studies have shown inconsistencies [18] and presence of all identified PARAFAC components in separated resin fractions found by [14] emphasizes the uniqueness of NOM characterization techniques. While each PARAFAC component possibly represents several organic compounds, it provides a functional grouping which is tied to variance between samples. As such, the process still provides useful information on how the fluorescence signals change while being tied to spectral regions which are related to distinct molecular structures [37].

Use of fluorescence spectroscopy for the detection of water quality changes associated with wastewater discharge events has been investigated by several authors [1,12,15,38]. In a review of potential monitoring tools for recycled water systems and identification of sewage contamination, Henderson et al. [15] described the use of protein-like to humic-like ratios determined by fluorescence as a possible measure for sewage impact detection. Other routine online monitoring parameters such as total organic carbon and conductivity were compared to the fluorescence approach. Stedmon et al. [38] reported a strong linear correlation for wastewater presence <2% by volume and fluorescence of a tryptophan-like component derived by PARAFAC. These results conform to observations that wastewaters would be rich in protein-like material due to intensive microbial activity while NOM in natural sources would be mainly derived from plant materials [17]. Previous studies have provided a preliminary assessment of possible features that can be used for wastewater impact detection, however further investigation into implementable methods and quantification of possible detection accuracies is needed.

Support vector machines (SVMs) are a popular classification technique which perform well using high-dimensional data sets [39]. They are described as boundary methods, where only edge cases between classes are used in modelling as opposed to all data, typical for most other chemometric methods [44]. The basis of classification using SVMs is to linearly separate classes in either the original space or a projected high dimensional feature space that can be efficiently non-linearly related to the original space using a kernel [2,24]. There have been several successful practical chemometric applications of SVMs [4,11,39]. With particular relevance, Thissen et al. [39] applied a support vector approach for the analysis of near-infrared (NIR) spectra and found superior performance to the commonly used partial least squares (PLS) algorithm. However, use of SVMs have since had limited attention for spectral analysis in chemometric applications or environmental sciences and to our knowledge have not been applied to fluorescence EEMs.

This study examined efforts to further establish the use of fluorescence spectroscopy for real time or rapid assessment of wastewater discharges to drinking water source waters. For this discussion we define rapid as feedback in 15 min or less. Using controlled mixing experiments, data is presented that shows treated wastewater impact detection under 2% by volume when considering 5 municipal wastewater effluents and 2 surface waters in Southern Ontario, Canada. Furthermore, the use of novel fluorescence measures for this application is identified including nicotinamide adenine dinucleotide (NADH)-like fluorescence. Use of NADH, a product of microbial metabolism, fluorescence has been investigated for monitoring wastewater treatment processes and is an indicator of microbial activity [9,42]. A comparison between expert supervised approaches (peak picking) and unsupervised dimensionality reduction using PARAFAC is made using multinomial logistic regression (MLR), a classification method comparable to linear regression. Furthermore, high-dimensional classification with SVMs using the full EEMs is presented as an alternative to requiring dimensionality reduction or feature selection prior to supervised classification.

2. Methods

2.1. Samples

Five wastewater effluents were sampled at the discharge to surface waters serving as drinking water source waters to best represent realistic conditions. Source waters were sampled from the intake of water treatment plants located close to points of wastewater discharge (approximately 3–15 km). Routine elevated influence of the treated wastewater was not expected for the drinking water intakes sampled, however the municipalities have expressed some degree of concern on possible conditions which may make wastewater impact possible. All water and wastewater operations conformed to applicable treatment and source water protection laws.

Wastewater treatments included conventional aerobic treatment, membrane bioreactors, and lagoon systems. Treatment plant 1 is a conventional system which relies on aerobic treatment and disinfected with chlorine prior to discharge. Treatment plant 2 also practices conventional aerobic treatment with phosphorous removal aided by ferrous chloride and disinfection prior to discharge. Treatment plant 3 is a sequential batch reactor plant with tertiary sand filtration followed by UV disinfection. Treatment plant 4 is an extended aeration plant with tertiary membrane filtration followed by UV disinfection. One lagoon system with alum treatment and seasonal discharge was also sampled. The two samples from the lagoon were taken on days when discharging to the receiving water body. Samples were collected approximately weekly for each treatment plant and source water between April and August 2015. Drinking water source waters were Lake Ontario (DOC: 2–2.5 mg/L) and Lake Simcoe (DOC: 4–5 mg/L). From historical monitoring data, wastewater effluent DOC is <10.0 mg/L indicating a 2% impact would be equivalent to <0.2 mg/L DOC change.

2.2. Mixing experiments

For each sampling day, controlled amounts of wastewater effluent were mixed in Lake Ontario or Lake Simcoe water, also collected on the same sampling day, at concentrations ranging from 0.10 to 15% by volume. Mixing was carried out at room temperature (21 ± 1 °C) in 20 mL amber glass vials. Fluorescence measurements were conducted immediately after the vials were vortexed for approximately 10 s. The focus of the mixing experiments was to identify low-level detection and therefore concentrations below 2% (0.10, 0.25, 0.50, 1.0, and 2.0) were principally analyzed. Select samples were analyzed at higher impact levels to understand the limitations of the relationships.

2.3. Fluorescence

Fluorescence spectra were collected primarily using an Agilent Cary Eclipse fluorescence spectrophotometer (Mississauga, Canada). Treatment plant 1 samples were analyzed using a Perkin Elmer LS50B fluorescence spectrophotometer (Woodbridge, Canada). Optimal instrument settings were determined based on previous studies and in-house testing [33]. Excitation and emission wavelength ranges were 250–380 nm (10 nm increments), and 250–600 nm (1 nm increments), respectively. With the stated instrument settings, data acquisition for one sample took approximately 13 min. Once predictive models have been established, the feedback of an online fluorescence system would be limited by sample acquisition time. In this case we have kept the instrument settings such that samples are acquired rapidly or under 15 min.