Determination of water supply sources in the Barcelona distribution system by UV spectrophotometry and PLS
S. Platikanov, V. Garcia, E. Landeros, R. Devesa, L. Matía and R. Tauler

ABSTRACT
A new method for the water source apportionment of the Barcelona (Spain) water distribution system is proposed. The method is based on the combined use of UV spectrophotometric measurements in the wavelength from 190–230 nm, and multivariate data analysis using the Partial Least Squares (PLS) chemometric method. From the differences in the organic matter content of the different water sources and of their corresponding UV spectral features, PLS was able to determine the relative amounts of the two main river water sources in samples of tap water from the different locations of the Barcelona city water distribution system. The extension of the method to determine the relative amounts in water blends, prepared from samples from five water treatment plant sources of the same city’s distribution system, required the combined use of some other parameters. In particular, the distinction and apportion of the water coming from a desalinisation plant could be successfully achieved once concentrations of Boron were included in the analysis.

Key words | PLS, source apportionment, UV, water distribution system

INTRODUCTION
Quality control of drinking water is a major concern everywhere. Natural organic matter (NOM) is a characteristic feature of every type of water source. The presence of NOM in drinking water is of particular importance since it affects aesthetic water qualities such as taste, colour and odour (Spellman 2007). Every different type of NOM has particular features associated with its geographical origin, vegetation, soil, etc. (Schäfer 2000). For example, aquatic algae deliver dissolved organic compounds with high amounts of nitrogen and low contents of aromatic carbon and phenolic groups. On the other hand, terrestrial derived dissolved organic matter (DOC) has relatively low nitrogen content but a large amount of aromatic and phenolic groups. The contribution of each organic matter source is also seasonally dependent (Sharp et al. 2006), and the hydrological and biogeochemical processes involved can alter the chemical composition and physical structures of the NOM. However, NOM is a complex mixture of heterogeneous chemical compounds and it is not possible at present to fully describe its chemical structure in the natural environment (Croue 2004). NOM profiles can be used as a diagnostic marker or as a typical fingerprint based on its structure and features. The more frequently used analytical technique for the direct investigation of NOM in water systems is UV/VIS spectroscopy (Thomas & Burgess 2007).

The drinking water in the distribution network system (WDS) of Barcelona is distributed by the AGBAR (Aigües de Barcelona) group of companies. The water in the WDS originates from more than one water source and is processed in different treatment plants. The two main water sources are, at present, Ter and Llobregat river systems (see Figure 1), together with local groundwater sources. Additionally, a large water desalinisation plant has recently been built (Gueguen et al. 2008) to provide drinking water from the
Mediterranean Sea. These different raw water sources show great physicochemical differences in the water quality before and after treatment processes. Another possible cause of variation in the water quality is due to the different treatment processes implemented in the five treatment plants present in the area (see Figure 1). The plant disinfection processes comprise conventional disinfection procedures, electrodialysis reversal and seawater desalinisation procedures. Hereafter the plants will be encoded as:

WTP1: Cardedeu water treatment plant using conventional disinfection procedures to treat water from Ter River;
WTP2: Sant Joan Despí water treatment plant using conventional disinfection procedures to treat water from Llobregat River;
WTP3: Abrera water treatment plant using conventional disinfection procedures to treat water from Llobregat River;
WTP4-EDR: Abrera water treatment plant using electrodialysis reversal treatment of water from Llobregat River;
SWRO: El Prat reverse osmosis seawater plant treats water from the Mediterranean Sea, close to the Barcelona area.

See Figure 1 for a map of the different locations of these water plants. The finished water obtained at the exit of each treatment plant can vary significantly due to its NOM content (also because of its natural origin, Llobregat, Ter or Mediterranean Sea) and to its mineral/elemental content (also because of its different natural origin and/or because of the plant treatment procedures implemented).

The identification of the source of drinking water in any supply location of the water distribution system (WDS) is a challenge for the operational maintenance, repair and management of the system. Knowledge about water source apportionments (% of each water source in mixtures), at a specific location of the WDS, is of interest for a better understanding and detection of possible sources in leak accidents and in establishing legal property rights. Also, it is very important to know the sources of the waters and percentages of the blends in order to understand the origin of the problem and how to act when a quality problem appears in the distribution system.

A proper inspection has a significant impact on the operation and maintenance cost and on the effectiveness of the systems. A method for water source determination has been proposed based on the solid-phase microextraction GC/MS spectroscopic assay of the chlorination by-products (Dufresne et al., http://www.thermo.com/eThermo/CMA/PDFs/Articles/articlesFile_12035.pdf.). It was associated with a usage of specific software performing pattern recognition on the MS profiles of different water sources. However, there is a need to develop inexpensive reliable methods for the in situ identification and determination of water sources and of their apportionment at each point of the WDS.

UV spectrophotometry is one of the most appealing approaches for broader analytical analyses. This spectrophotometric method is frequently used for the investigation of NOM content in drinking water. Usually organic matter content is measured at UV254 nm (USEPA Method 415.3), but also UV/VIS measurements can be expanded to the entire spectral range from 190–1100 nm. UV spectra of complex water blends will result in extensive spectral overlap due to the high number of absorbing components in this wavelength range in natural water samples. UV spectrophotometric measures can be coupled with advanced data analysis chemometric techniques to improve their resolution power. Chemometrics is at present a well established field in...
chemical data analysis (Brown 2000) and has also recently been recognized in the analysis of water science data (Nollet 2007). There is a large number of chemometric techniques suitable for the study of complex multivariate water data sets, like principal component analysis (PCA) (Simeonova & Simeonov 2006) or PCA based regression methods (Wentzell & Lohnes 1999; Narasimhan & Shah 2008) which were shown to achieve very good predictive abilities in spectral regression problems, and partial least squares regression (PLS) (Platikanov et al. 2007) and several other methods not well known outside chemometrics that may play an important role. In summary, the main goal of this study was to establish PLS models for water source apportionment in blends on the basis of their NOM spectral profiles and physicochemical parameters.

METHODS

Methodology

Multivariate regression techniques reveal the relationship between two different data blocks (matrices) of chemical data. The main goal of this study is to build multivariate linear models able to describe, explain and predict the apportionment of the different water sources inside the Barcelona water distribution system and in house tap water blends (y block of variables) obtained from them as a function of their UV spectra together with other possible physicochemical parameters like the concentrations of the chemical elements present in the water samples (X block of variables). This involves finding an adequate mathematical model or function that relates the variables in these two data blocks, i.e. $y = f(X)$, where the y block gives the apportionment of each source in each water blend and the X block gives the UV spectra and some additional physicochemical parameter, like the elemental concentrations of the blends. This modeling follows a two-step approach. The first step is the model calibration. The calibration step is followed by a second prediction step in which this model is used to estimate unknown apportions of blend samples from their UV spectra. Once experimental data were properly arranged in data matrices, they were mean centered (considering only UV spectra as predictors) or autoscaled (column mean centering and scaling of UV spectra and one additional physicochemical parameter at a time). Group scaling could have been considered as an alternative pre-processing technique. This would be appropriate when the full wavelength spectra and only one or two element concentrations were used as predictors. In this study however, only wavelengths where the absorption was high (between 190–250 nm) were selected for data analysis and noisy wavelengths were excluded. This assured reducing of the possible noise amplification in the scaling operation. This preliminary data treatment eliminated offsets, changes in measurement units and focused the analysis on proper modeling of observed variances in measured variables. This data pretreatment is frequently used in multivariate data analysis (Massart et al. 1998).

In particular, the Partial Least Squares (Geladi & Kowalski 1986; Massart et al. 1998) multivariate linear regression (PLS) method has been used and evaluated for the modeling of the water sources apportionment in their blends. PLS attempts to maximize the covariance between $X$ and $y$ data blocks. PLS searches for a common factor subspace most congruent to both data blocks, and its predictions are usually better than using other multilinear regression methods such as the ordinary multilinear regression method (MLR), especially if the X variables are highly correlated, like in the spectroscopic measurements used in this work. PLS transforms the high number of original variables (spectra) into a smaller number of orthogonal variables called “components”, “factors” or “latent variables”, which are linear combinations of the original variables. The first latent variables contain useful information about the major covariance sources between the two blocks of variables, whereas the last ones basically represent the uncorrelated variance and noise, which has to be discarded and is not considered in the modeling. A new matrix of weights (reflecting the covariance structure between the X predictors and y response variables) is obtained from PLS analysis, which provides rich information about the nature of the different covariance sources. The selection of the optimal number of components (the number of latent variables) in PLS is performed using internal cross validation (leaving out one sample at a time) and external validation for optimal prediction of y values in new samples not used in the calibration step. In this work, for method/model validation, an external set of samples (water
blends with known source apportionments), not participating in the model calibration, was used.

For brevity, in this study the relative errors of water source apportionment in percentage are reported for both calibration and prediction steps.

Relative errors of concentrations in percentage, for both calibration and prediction steps, are calculated as follows:

\[
\text{Rel. error in } \% = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}}{\frac{1}{n} \sum_{i=1}^{n} (y_i)^2} \times 100
\]

**Preparation of water samples, instrumentation, chemical analysis and software**

This work consisted of two studies. The first one was intended to identify the sources/apportionment of tap water collected from four of Barcelona’s districts. UV spectra were used as a data set of water blends. Three of these city districts, Fondo, Gracia and Horta (D1–D3), were supplied with drinking water treated in WTP1 (see Figure 1), representing the original Ter River water source. The fourth district, Les Corts (D4), is supplied with water, treated in WTP2 and WTP3, originating from Llobregat River water. WTP4-EDR and SWRO plants were not involved in this first study.

In a second study, water samples were provided directly from the 5 water treatment plants. The feasibility of the apportionment of the five water sources was studied from blends of these water samples, also using their UV spectra, elemental analysis and chemometrics.

In both cases, water blends were prepared according to a three-level experimental design based on a Box-Behnken design. In the first experiment, 24 samples were prepared according to a Box-Behnken statistical experimental design (Box & Behnken 1960) and after their random sampling two new data subsets were selected, one for calibration and the other for validation. The calibration data set comprised 18 samples, and the remaining 6 samples were selected for external validation. In the second experiment, new water blends were prepared using a similar experimental design to the one just described but prepared using 22 water blends with 16 of them used for the calibration step and the other 6 used for the external model validation.

Apportioning of water sources in the water blends were done in the range of 0–100% with predominance in the middle range values such as 11, 20 and 50% of each water source in the blend. In this study apportionments were never predicted below 0 or over 100%. If this was the case, an alternative regression method like PLS logistic regression (Bastien et al. 2005) could have been used.

UV spectra were recorded in the range of 190–1100 nm in a 1 cm quartz cell, using an Agilent HP8453 diode array spectrophotometer. Wavelengths range used was between 190–230 nm.

In the first experiment two replications were performed during a month. Both times, water samples were collected during the morning and after allowing the water to run out from the tap for some minutes.

In the second experiment water samples were prepared externally by the technician staff working in the Barcelona WDS. They were delivered to the laboratory the same day of production and they were appropriately mixed. A preliminary study of water samples’ evolution over time did not show any significant changes in its spectrophotometric properties within a week. This was expected, since water remains in the distribution systems for one week or less and, therefore, the experimental conditions used in this work can be considered appropriate for emulating what occurs in the Barcelona distribution system.

Aluminium, Barium, Boron, Calcium, Copper, Chromium, Strontium, Iron, Magnesium, Manganese, Nickel, Potassium, Silicon, Sodium and Zinc concentrations were determined by Inductively Coupled Plasma Optical Emission Spectrometry, ICP-OES, (Perkin Elmer Optima 4300 DV). Bicarbonate was analyzed by a robotic titrosampler with conductivity module 855 and 856. Chloride, Nitrate and Sulfate concentrations were estimated by Ionic Chromatography (Dionex ICS-2000). All these analytical determinations (except bicarbonates, which are not included in EC Drinking Water Directive 98/83/EC), were ISO17025 accredited.

The variables in $X$ and $y_i$ data sets were initially arranged using EXCEL (Microsoft, Redmont, WA, USA) and subsequently transferred to the MATLAB computer workspace environment (MATLAB version 6.5, The Mathworks, Natick
Chemometrics modeling was performed using PLS Toolbox 4.2 (Eigenvector Research, Manson, WA, USA). Statistical experimental design was carried out by using the software Unscrambler 9.8 (CAMO PROCESS AS, Oslo, Norway).

RESULTS AND DISCUSSION

UV spectrophotometric analysis of tap water samples

Before multivariate modeling of water blend spectra, the analysis of the spectra of different Barcelona district water samples led to the preliminary conclusion that D1–D3 district waters had practically the same UV spectra profile and that they had the same original source of water (from Ter River) as a result of having a similar organic matter content. Figure 2 shows the spectral difference between the two main sources of water for Barcelona, i.e. Ter River (D1 blue lines) and Llobregat River (D4 red lines). The sample spectra were obtained from two district water locations, each one replicated after 12 hours (solid and dashed lines). The replicates in the plot cannot be clearly distinguished, because they were highly overlapped due to the lack of variation in their chemical content for this period. The UV spectra of the tap water from the D4 district were clearly different from the UV spectra of the other tap water samples, leading to the conclusion that the D4 water sample was the only one coming from Llobregat River. This fact was confirmed by the technical staff of the water distribution company.

This preliminary study led to the conclusion that the two main water sources (Ter and Llobregat water river sources) could be simply distinguished spectrophotometrically by means of their UV spectra. Consequently, the apportionments of the different district water sources from D1 to D3 were summed up and recalculated as a single Ter River source against the amount of D4 (Llobregat River sample). PLS modeling was then performed using a calibration subset of 18 blends. In the case of the modeling of the Llobregat River water apportionment, two latent variables were selected (after cross-validation) to explain almost 99% of the variance/information in the data and leading to very low prediction errors (6% prediction error in the calibration). This calibration model was then used to predict the Llobregat River source content of the external subset of the 6 blends selected for validation, resulting in only 7% prediction errors. Similarly, low prediction errors were also achieved for Ter water source apportionment in the blends by PLS modeling. These low errors obtained in the validation were achieved because of the reproducible intrinsic patterns in the organic matter of the different sampling locations in the distribution system. The model that was based on the 2 latent variables was reliable enough and offered good prediction properties.

To summarize, the separation between the two main river sources was possible in this first study due to their UV absorption spectra and because of their distinct organic matter patterns, which differ considerably between the two river water sources. The next, more difficult, step was to try to distinguish among the five water treatment plant sources (see Figure 1) and the relative apportionment of their blends.

UV spectrophotometric analysis plus elemental and mineral content analysis of AGBAR water blend samples

Like in the first experiment, UV spectra of the five different water source samples were first registered. Figure 3 shows the normalized spectra of the five water sources provided by the company. Very large similarities among some of the spectra...
can be noticed, especially between the spectra of WTP2 and WTP3 water sources.

The highest pairwise correlation coefficient was found to be of 0.999 between WTP2 and WTP3 water sources. These two water sources come from Llobregat River and the two treatment plants are employing similar conventional water treatment procedures. Therefore, samples from these two water sources are expected to have very similar organic matter content and give analogous UV absorption spectra. On the contrary, the sample from WTP4-EDR water source (a plant which implements reverse electrodialysis filtering as water treatment) gave an absorption spectrum rather different to the previous two absorption spectra mentioned, despite their common origin from the Llobregat River. This means that the reverse electrodialysis process had changed significantly the organic matter content of the Llobregat water, leading to water with a different UV spectrum. Moreover, the WTP4-EDR water spectrum showed a similar spectrum with a high pairwise correlation coefficient 0.98 to the WTP1 river source spectrum. On the other hand, SWRO water (from the desalinization plant) spectrum did not show any characteristic spectral feature in the UV region. It only absorbed UV light in a very narrow range from 190–205 nm, confirming that this type of water had very low organic matter content and that modeling this source on the sole basis of UV absorption will probably not be possible. Although the pairwise correlation coefficients of its spectrum with the spectra of the other water sources was very low, ranging from 0.6 to 0.88 (because it had nearly no absorption, whereas the others did), it only had a dilution effect; moreover, it was very difficult to distinguish it from the others and the prediction of its apportionment on the sole basis of its UV absorption was not possible.

Therefore, other physicochemical properties, such as conductivity and mineral content, were considered to describe the five water sources and their relative content. The results obtained from this mineral content analysis, together with the spectrophotometric data, were arranged in the same table data matrix (independent variables X) to be subjected to PLS modeling. Elemental and mineral analysis comprised 21 parameters routinely monitored by the water company quality control laboratory. Table 2 gives the results of these analytical determinations and their basic statistics.

Some of the parameters analyzed did not show any significant feature for the characterization of the water blend samples analyzed and were therefore not useful for the purpose of this study. The parameters showing significant changes in their values were incorporated one by one into the X data table, together with the spectral measurements, and were then submitted to new PLS modeling. It was obvious that WTP2 water showed many of the maximum parameter changes, which was in contrast to the SWRO water (from the desalinization plant), which showed minimum changes in most of its parameter values. Finally, only the boron concentration was found to be at its maximum concentration in SWRO water samples. This can be explained by its original sea water source and the low effect the water treatment had on this parameter.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Pairwise correlation coefficients among the 5 AGBAR water treatment plant spectra</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WTP1</td>
</tr>
<tr>
<td>WTP1</td>
<td>1</td>
</tr>
<tr>
<td>WTP2</td>
<td>1</td>
</tr>
<tr>
<td>WTP3</td>
<td>1</td>
</tr>
<tr>
<td>WTP4-EDR</td>
<td>1</td>
</tr>
<tr>
<td>SWRO</td>
<td></td>
</tr>
</tbody>
</table>
Figure 4 depicts the prediction errors finally obtained for calibration and validation when spectral data and one of the additional physicochemical parameters at a time were considered. PLS analysis of UV spectra, together with some of the physicochemical parameters considered, gave satisfactory predictions. In particular, the best parameters proved to be Barium (Ba) and Boron (B). Prediction errors for the external validation water samples were from 10–12%. Results obtained in the modeling of the WTP4-EDR samples showed that Boron was very important to correctly predicting the apportionment of this source, giving just a 5% error in its prediction for the validation samples. Very good results were also found in the modeling and prediction of the desalination plant (SWRO) water source, showing again that the concentration of Boron was the most important parameter to consider (prediction errors below 10%). However, less satisfactory results were achieved for the differentiation of WTP2 and WTP3 sources, both in the calibration and in the validation step.

Figure 5 shows the results obtained for the prediction of WTP2 and WTP3 water sources when they were considered individually, as separate sources, and jointly, as coming from the same source, e.g. Llobregat River. When they were considered separately, the predictions of their relative contents were wrong; the results yielded high prediction errors both in calibration and in validation, regardless of the physicochemical parameter added to the spectral data. In contrast, when WTP2 and WTP3 were considered jointly, i.e., when they were considered from the same source, low prediction errors were then achieved. Prediction errors for the external validation samples were usually below 10%. Again, Boron was the most important parameter included in the model.

Table 3 summarizes one example of the results obtained in the complete modeling of the whole experimental data set. Boron was confirmed as the most important parameter in the data analysis, together with UV spectral data, for water source apportionment. Also, a possible good combination of parameters to be considered together with spectral data was Boron and conductivity. The addition of conductivity increased the reliability and robustness of the PLS models and made them more versatile for the different conditions in the Barcelona water distribution system.

Table 2

<table>
<thead>
<tr>
<th></th>
<th>Conductivity 20 °C</th>
<th>Bicarbonate</th>
<th>Chloride</th>
<th>Sulfate</th>
<th>Nitrate</th>
<th>Sodium</th>
<th>Potassium</th>
<th>Calcium</th>
<th>Magnesium</th>
</tr>
</thead>
<tbody>
<tr>
<td>WTP1</td>
<td>382</td>
<td>145</td>
<td>38</td>
<td>50.6</td>
<td>4.15</td>
<td>15.4</td>
<td>5</td>
<td>49.5</td>
<td>9.1</td>
</tr>
<tr>
<td>WTP2</td>
<td>1194</td>
<td>233</td>
<td>214</td>
<td>137</td>
<td>11.5</td>
<td>111</td>
<td>19.8</td>
<td>102.3</td>
<td>25.3</td>
</tr>
<tr>
<td>WTP3</td>
<td>1100</td>
<td>217</td>
<td>189</td>
<td>132</td>
<td>10.7</td>
<td>96.9</td>
<td>18.7</td>
<td>99.4</td>
<td>23.1</td>
</tr>
<tr>
<td>WTP4-EDR</td>
<td>365</td>
<td>101</td>
<td>55</td>
<td>24.4</td>
<td>2.71</td>
<td>49.2</td>
<td>6.5</td>
<td>17.3</td>
<td>4.8</td>
</tr>
<tr>
<td>SWRO</td>
<td>311</td>
<td>8.1</td>
<td>97</td>
<td>&lt;5</td>
<td>&lt;0.5</td>
<td>54.5</td>
<td>&lt;5</td>
<td>&lt;5</td>
<td>1.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Boron</th>
<th>Strontium</th>
<th>Iron</th>
<th>Manganese</th>
<th>Nickel</th>
<th>Barium</th>
<th>Silicon</th>
<th>Aluminium</th>
</tr>
</thead>
<tbody>
<tr>
<td>WTP1</td>
<td>27</td>
<td>0.45</td>
<td>111</td>
<td>6</td>
<td>11</td>
<td>31</td>
<td>1.7</td>
<td>34</td>
</tr>
<tr>
<td>WTP2</td>
<td>82</td>
<td>1.48</td>
<td>5</td>
<td>&lt;1</td>
<td>6</td>
<td>55</td>
<td>2.19</td>
<td>&lt;24</td>
</tr>
<tr>
<td>WTP3</td>
<td>45</td>
<td>1.37</td>
<td>7</td>
<td>&lt;1</td>
<td>&lt;5</td>
<td>53</td>
<td>2.15</td>
<td>&lt;24</td>
</tr>
<tr>
<td>WTP4-EDR</td>
<td>34</td>
<td>0.24</td>
<td>&lt;5</td>
<td>&lt;1</td>
<td>6</td>
<td>10</td>
<td>1.92</td>
<td>&lt;24</td>
</tr>
<tr>
<td>SWRO</td>
<td>742</td>
<td>0.05</td>
<td>&lt;5</td>
<td>&lt;1</td>
<td>&lt;5</td>
<td>9</td>
<td>0.5</td>
<td>&lt;24</td>
</tr>
<tr>
<td>Min</td>
<td>27</td>
<td>0.05</td>
<td>&lt;5</td>
<td>&lt;1</td>
<td>&lt;5</td>
<td>9</td>
<td>0.5</td>
<td>&lt;24</td>
</tr>
<tr>
<td>Max</td>
<td>742</td>
<td>1.48</td>
<td>111</td>
<td>6</td>
<td>11</td>
<td>55</td>
<td>2.19</td>
<td>34</td>
</tr>
<tr>
<td>Average</td>
<td>186</td>
<td>0.718</td>
<td>26.6</td>
<td>2</td>
<td>6.6</td>
<td>31.6</td>
<td>1.692</td>
<td>26</td>
</tr>
</tbody>
</table>

*a* values recorded in mg/L.

* values recorded in µg/L.

*c* values below the limit of quantitation

---

Figure 4 depicts the prediction errors finally obtained for calibration and validation when spectral data and one of the additional physicochemical parameters at a time were considered. PLS analysis of UV spectra, together with some of the physicochemical parameters considered, gave satisfactory predictions. In particular, the best parameters proved to be Barium (Ba) and Boron (B). Prediction errors for the external validation water samples were from 10–12%. Results obtained in the modeling of the WTP4-EDR samples showed that Boron was very important to correctly predicting the apportionment of this source, giving just a 5% error in its prediction for the validation samples. Very good results were also found in the modeling and prediction of the desalination plant (SWRO) water source, showing again that the concentration of Boron was the most important parameter to consider (prediction errors below 10%). However, less satisfactory results were achieved for the differentiation of WTP2 and WTP3 sources, both in the calibration and in the validation step.

Figure 5 shows the results obtained for the prediction of WTP2 and WTP3 water sources when they were considered individually, as separate sources, and jointly, as coming from the same source, e.g. Llobregat River. When they were considered separately, the predictions of their relative contents were wrong; the results yielded high prediction errors both in calibration and in validation, regardless of the physicochemical parameter added to the spectral data. In contrast, when WTP2 and WTP3 were considered jointly, i.e., when they were considered from the same source, low prediction errors were then achieved. Prediction errors for the external validation samples were usually below 10%. Again, Boron was the most important parameter included in the model.

Table 3 summarizes one example of the results obtained in the complete modeling of the whole experimental data set. Boron was confirmed as the most important parameter in the data analysis, together with UV spectral data, for water source apportionment. Also, a possible good combination of parameters to be considered together with spectral data was Boron and conductivity. The addition of conductivity increased the reliability and robustness of the PLS models and made them more versatile for the different conditions in the Barcelona water distribution system.
Special attention has been paid to the prediction of SWRO apportionments. This was shown not to be possible using only UV spectra. Including boron and conductivity greatly improved the prediction of SWRO water content for the water blends investigated. Indeed, boron was found at its highest concentrations in SWRO water source samples. An example of SWRO modeling when UV spectra, Boron and conductivity were simultaneously considered as parameters is given in detail in Figure 6. Five latent variables were found to explain more than 99% of the $y$ block variance, assuring as low prediction errors as 5.3% for the calibration samples and 10.9% for the external validation samples. Graphically, these predictions were compared to the actual concentrations in Figure 6.

Future research should be focused on the development of an experimental system for on-line monitoring of the water distribution system based on the combined use of UV, some rapid elemental analysis for boron, and chemometrics.

**CONCLUSIONS**

The following were the main conclusions derived from the present work:

- UV spectrophotometric analysis coupled with chemometrics has been shown to be a powerful tool for the differentiation of different raw water sources having
different natural organic matter contents. In the case of tap water from the Barcelona distribution system, it was possible to distinguish between water sources from two different rivers (Ter and Llobregat) in their blends and to estimate their relative apportionments with prediction errors of around 7%.

- The compositions of water blends from 5 different Barcelona water treatment plants were not predicted well on the basis of UV analysis only, due to overlapping spectra and a lack of natural organic matter content in some of these water sources. Especially difficult when predicting the composition of water blends was the estimation of the individual water apportionments from electrodialysis and desalinization plants. In contrast, the clear differences among the nature and amounts of organic matter in Llobregat and Ter rivers allowed successful prediction of their apportionments.
- The addition of information from elemental/mineral analysis of the different water sources, apart from their UV spectra, allowed successful prediction of all the water sources in the distribution system.
- Boron concentration was the most important parameter, apart from UV spectra, for the correct prediction of water sources from the desalinization plant. The prediction errors were below 10% for the external validation samples. Therefore, Boron can be considered as a very important variable in cases where a desalination plant is part of the distribution system.
- Two conventional water treatment plants, situated in the same river system and located near each other, were impossible to differentiate either on the basis of UV spectra profile analysis or their mineral content. However, their joint contribution could be estimated as a sum of apportionments, i.e. as coming from the same source.

**ACKNOWLEDGEMENTS**

The methodology described in this article has been developed within the research project QUALITAX, leaded by CETaqua Water Technologic Centre with the financial support of
AGBAR. The project has been carried out by Aigües de Barcelona (AGBAR Group) and the Department of Environmental Chemistry, IDAEA-CSIC.

REFERENCES


Box, G. & Behnken, D. 1960 Some new three level designs for the study of quantitative variables. Technometrics 2, 455–475.


USEPA Method 415.3 2005 Determination of Total Organic Carbon and Specific UV Absorbance at 254nm in Source Water and Drinking Water.