HYPERSPECTRAL DETECTION OF MARINE CLAY IN COASTAL WATERS USING THE SPECTRAL ANGLE METHOD

Jyrki Tuominen¹, Tarmo Lipping¹, Viljo Kuosmanen², and Sari Repka³

1. Tampere University of Technology, Signal Processing, Pori, Finland; {jyrki.tuominen / tarmo.lipping}@tut.fi
2. Geological Survey of Finland, Remote Sensing, Espoo, Finland; viljo.kuosmanen@gtk.fi
3. University of Turku, Marine Biology, Pori, Finland; sari.repka@utu.fi

ABSTRACT
Clay induced turbidity is the major cause of complex changes in the ecosystem of coastal water areas. The accuracy of the algorithms for the assessment of the clay concentration may suffer significantly from the presence of other substances like Algal chlorophyll or dissolved organic matter, highly influencing the optical properties of the water. The objective of this study was to develop an accurate and robust algorithm for the clay concentration retrieval using high spectral resolution hyperspectral data. In order to provide necessary information for the spectral analysis of the water samples, a controlled experiment was arranged. Carefully weighted clay portions were diluted in water and reflectance spectra were recorded using a field spectrometer. Traditionally simple algorithms are used to estimate the remotely sensed water quality variables. In this paper a novel algorithm based on the principle of spectral angle measure is presented. The accuracy of the proposed algorithm was just slightly better compared to the band-ratio algorithm, but it is more robust against the effects of other optically detectable substances and noise.

INTRODUCTION
Increasing water turbidity is one of the human-induced environmental changes in coastal areas. Human activity causes both organic (algal) and non-organic (clay) turbidity to rise in aquatic environments. Rivers and brooks transport clay particles to coastal waters from agricultural land; and other human activities such as landfill activities, dredging, gravel take, ships turbines, building etc. may cause non-organic turbidity to rise (1). The effects of increasing organic turbidity on the ecosystem and biochemical cycles are complex. Especially the underwater biological diversity may be harmed by increasing turbidity, e.g., Fucus vesiculosus is sensitive to the turbidity (2). It may also be possible that clay turbidity influences the phytoplankton community structure. The planktonic food web may change due to the increasing turbidity (3). Organic particles (alive or dead) enter the food web either by grazing or via decomposition by micro-organisms. Clay particles cannot provide energy to the grazers or bacteria. They may, however, contain nutrients and trace-elements that can be utilised by phytoplankton or bacteria (4). In lakes, the clay-turbidity has been shown to have serious effects on the food web and to prevent the improvement of the lake condition due to the restoration methods (5). These effects can be assumed to take place in coastal marine environments as well.

Information revealing the environmental effects of the non-organic turbidity in coastal areas is badly needed by the environmental administration for the Environmental Impact Analysis of projects that will increase the clay induced turbidity. Remote sensing has already demonstrated its ability to provide water quality relevant data. Excellent work has been performed to extract the essential water quality parameters such as chlorophyll content, turbidity and the total amount of suspended matter (6,7,8). Yet, it appears that there is no standard algorithm for the estimation of suspended sediments in coastal waters (9). From the remote sensing perspective, the water environment can be divided into two categories: Case 1 and Case 2. Case 1 refers to phytoplankton dominated waters (e.g. open oceans) whereas case 2 refers to coastal waters which contain re-suspended sediments, terrigenous particles and dissolved organic matter (yellow substances) (10). Remote sensing in the coastal zone has generally been far less successful than in other areas, such as open oceans or
terrestrial environment (11). This is mainly due to complex interactions of the three optically detectable substances: algal chlorophyll, suspended sediments and dissolved organic substances. These substances have significant impact on the water quality. Each substance has its own reflectance and absorption trends complicating the derivation of the clay concentration estimation algorithms based on the remotely sensed data.

Several studies have been conducted to address the impact of the suspended sediments on the spectral signature of the surface waters (12,13,14,15). Researchers have usually addressed the general trends in the spectral responses of water samples with varying concentrations of organic and inorganic suspension. Quibell (1992) showed that the addition of the particulates generally increases the reflectance at the wavelengths longer than 550nm (16). Gin et al. (2003) proposed a band ratio algorithm $R_{595}/R_{754}$ for the estimation of the marine clay concentration. The algorithm produced rather good results with the coefficient of determination $r^2$ varying from 0.53 to 0.86 according to the particle size of the marine clay.

Case 2 water quality retrieval algorithms traditionally use simple band-ratio or band-difference techniques. However, promising techniques utilising more information have been introduced in recent years. Gitelson et al. (2008) presented a three-band NIR-Red algorithm producing excellent results in chlorophyll-a concentration retrieval in turbid case 2 waters (17). Cococcioni et al. (2004) presented an approach to case 2 water quality retrieval based on the use of fuzzy logic (18). The proposed method produced promising results for the estimation of concentrations of optically active constituents of case 2 seawater. Doerffer & Schiller (2007) proposed the use of neural network in order to overcome the problems related to the complex nature of case 2 waters in water quality retrieval (19).

It is quite likely that the derivation of a general algorithm for the clay concentration estimation to be used globally is not feasible. This is mainly due to the different clay types and different particle sizes. In addition, the concentrations of the other optically detectable substances than clay vary too much in the coastal water areas around the world. Yet, an algorithm developed for a smaller geographic region such as the Baltic Sea can be used as a valuable tool in the studies of the clay’s re-suspension and biological effects. In this study, a controlled experiment was conducted to measure the reflectance spectra of seawater samples with varying clay concentration. The purpose of the spectral analysis was to develop a robust and accurate algorithm for the clay concentration estimation for the coastal water areas of the Baltic Sea using high resolution hyperspectral data.

**METHODS**

**Test data**

Reflectance spectra of seawater samples of varying clay concentrations were measured in a laboratory environment. The seawater was collected from board of a boat at an offshore location where both the organic and the non-organic suspension is very low. Total Suspended Matter (TSM) of the collected water was 5 mg/l and it can be assumed that the matter was mainly organic.

Marine clay was collected from the seaside of the town Pori in South-West Finland (61°36.842’N 21°26.765’E). The small portions of clay were carefully weighted using a precision laboratory scale. The spectral measurements were carried out in a glass container filled with 24 centimetres of seawater. The small portions of clay were ground into powder and diluted in the seawater in the container. The reflectance spectrum of each water sample was recorded using an ASD FS-2500 portable spectrometer. A total number of six water samples with added clay were measured. The clay concentration in the prepared water samples varied between 10.6 and 142 mg/l. The clay concentration values are typically in that range at the coast of the Baltic Sea. A black sheet painted with black Synthal paint was positioned under the container in order to ensure that the bottom reflectance does not influence the measured reflectance of the water sample. The Synthal paint has a very low reflectance throughout the measured wavelength range. The measured reflectance of the paint is under 0.03 in the range 350 to 2500 nm.

The container was illuminated with an Oriel research arc lamp source having a power consumption of 1000 W. The Oriel lamp is a laboratory equipment designed for spectral measurements providing
constant spectral characteristics. The Oriel lighting source was operated with a voltage stabiliser unit which ensured a constant illumination of the measured water sample. Black sheets were installed around the container in order to block out all stray light from other sources. It was necessary to illuminate the measured water sample through the glass because the straight mirror reflection from the water sample had to be avoided. In order to verify the absence of bottom effects the reflectance spectrum of the empty container bottom was measured. The measured reflectance was very low throughout the measured wavelength range. Reflectance peaks and absorption points of the bottom spectrum could not be observed in the water sample spectra. Hence it is safe to assume that there were no reflections from the bottom involved in water sample measurements. The arrangement of reflectance measurements is shown in Figure 1.

![Figure 1: The arrangement of reflectance measurements.](image)

Reflectance measurements of water samples in glass containers need a specific correction for the removal of the optical effect of the container. The correction coefficients were calculated using an empirical method where a white reference target was measured inside and outside the empty container. The number of wavelength channels is 2151 and is constant throughout the work. Therefore, indexing of the channels has been omitted in the notation.

The measured reflectances are as follows:
- \( T_p \) = Reflectance of the empty container
- \( P_p \) = Homogenous reference target in the same container
- \( P_i \) = Homogenous reference target outside the container

Denoting the apparent reflectance \( T_p \) as \( A_{T_p} \) and the corrected reflectance \( P_i \) as \( A_{P_i} \), it follows:

\[
A_{T_p} = c_1 \frac{P_p}{P_i} - c_2
\]

\[
A_{P_i} = \frac{c_1 P_p}{T_p + c_2}
\]

where \( c_1 \approx 4.233 \) and \( c_2 \approx 3.994 \).

The constants \( c_1 \) and \( c_2 \) were estimated by minimising the total sum of the absolute difference \( A_{T_p} - T_p \). The optical effect of the glass container can be removed from the reflectance \( R_w \) of a water sample measured in that container using the following relation:

\[
A_{R_w} = \frac{c_1 R_w}{T_p + c_2}
\]

where \( A_{R_w} \) denotes the corrected reflectance.
Algorithms

A linear regression is perhaps the most often used empirical method in the remote sensing of water quality. The use of channel ratios in the regression algorithm has been found to be suitable for the retrieval of many water quality parameters (20). Quite often the wavelength channels are called ocean colour bands in the literature addressing the remote sensing of water quality, so the ocean colour band ratio algorithms are as essential as the band ratio algorithms. The use of band-ratio algorithms for the estimation of the clay concentration has been proposed in the literature and therefore, they are used for a comparison in the evaluation of the spectral angle method proposed in this paper (8). The general form of a band ratio algorithm used in the assessment of clay concentration \( C \) is:

\[
C = a \frac{R_1}{R_2} + b
\]  

(4)

where \( R_1 \) and \( R_2 \) are the remotely sensed reflectances at predefined wavelengths and \( a \) and \( b \) are empirically determined regression parameters. High spectral resolution imaging spectrometers such as AISA dual can provide information on as many as 481 wavelength channels. When the simple band-ratio algorithms are calculated based on such data the majority of the information is not utilised. The use of more than two wavelength channels can lead to more robust and accurate algorithms.

Spectral Angle Mapper (SAM) is a spectral classification method that uses an \( n \)-dimensional angle to match remotely sensed pixels to a reference spectrum. The algorithm determines the similarity between two spectra by calculating the angle between the spectra, treating them as vectors in a space with a dimensionality equal to the number of bands (21). A small angle means close match to the reference spectrum. The spectral angle \( \alpha \) can be expressed as:

\[
\alpha = \cos^{-1} \left( \frac{\sum_{i=1}^{nb} t_i r_i}{\left( \sum_{i=1}^{nb} t_i^2 \right)^{\frac{1}{2}} \left( \sum_{i=1}^{nb} r_i^2 \right)^{\frac{1}{2}}} \right)
\]  

(5)

where \( nb \) is the number of bands, \( t \) is the target spectrum and \( r \) is the reference spectrum. The spectral angle between the target and the reference spectra in the case of a two-band image is shown in Figure 2.

The feasibility of the SAM method in clay concentration retrieval is not self-evident, the method has some shortcomings. When using the spectral angle method, a normalisation of spectra is carried out. Only the shape of these spectra, and not their brightness, is taken into account. Furthermore, SAM
does not distinguish between negative correlation and positive correlation as it uses absolute values.

In order to overcome the limitations of the SAM method, improved spectral methods such as Spectral Correlation Mapper (SCM) and Spectral Multiple Correlation Mapper (SMCM) have been proposed (22,23,24). The main difference between these methods concerns the criteria of similarity used for comparing the spectra. SCM and SMCM offer some benefits in certain situations but in this kind of application where the whole spectrum is not employed there is no clear advantage when compared to SAM. The only essential requirement for a spectrum-dependent variable in this application is a strong correlation between the variable and clay concentration.

In the proposed algorithm the spectrum of the water sample without added clay suspension was used as the reference. The spectral angle between the measured water sample spectrum and the reference spectrum representing water with a very low organic and non-organic suspension was calculated. The clay concentration $C$ was calculated using the following equation:

$$
C = a' \alpha + b'
$$

where $\alpha$ is the spectral angle between the reference and the measured spectra and $a'$ and $b'$ are empirically determined regression parameters. In this paper this equation is referred to as the spectral angle algorithm. It is quite usual that all the available wavelength bands are not used in the spectral angle based classification. It is very likely that some of the wavelength bands do not contain information relevant to the clay concentration of the water. There are several methods for dimensionality reduction of hyperspectral data presented in the literature such as feature selection, Principal Component Analysis (PCA) and Minimum Noise Fraction (MNF) (25). The PCA and MNF components are always image dependent and therefore cannot be used in general algorithms. The wavelength bands used in the algorithm proposed in this paper were chosen depending on the correlation analysis. The band combination that produces the strongest correlation between the spectral angle $\alpha$ and the clay concentration $C$ leads to the highest retrieval accuracy when using Eq. (6).

A very desirable property of a clay concentration estimation algorithm would be its robustness, i.e., insensitiveness to the presence of other substances than clay. An organic suspension is often present in case 2 waters. Therefore, the robustness of algorithms to withstand organic suspensions was tested. In our study the spectra of the combined clay and organic suspension were obtained by simulation. The simulated reflectance spectra of the combined suspension were calculated as a linear combination of the measured spectrum of a water sample with added clay and the spectrum of a water sample with strong organic suspension. Water samples with a strong organic suspension were manufactured by diluting green algae in water with very low suspension.

The simulated reflectance spectrum $R_s$ was calculated using the following equation:

$$
R_s = R_c \frac{100 - P}{100} + R_o \frac{P}{100}
$$

where $R_c$ is the spectrum of the water sample with added clay, $R_o$ is the spectrum of the water sample with the organic suspension and $P$ is the percentage of the water with organic suspension in the mix. $R_s$ is the simulated spectrum of the water sample containing both suspensions, i.e., clay and organic. An example of $R_s$ is shown in Figure 3. $R_c$ is the measured spectrum of the water sample with a clay concentration of 67.4 mg/l and $R_o$ the measured spectrum of the water sample with organic suspension. The percentage of the water with organic suspension in the simulated spectrum is 20%. Hence, 20% of the simulated reflectance spectrum $R_s$ is determined by $R_o$ and 80 % is determined by $R_c$. The reflectance peak around 715 nm and absorption around 680 nm due to chlorophyll can clearly be seen in the reflectance spectrum of the water sample with organic suspension. This indicates a significant chlorophyll concentration although the exact value could not be measured. The Total Suspended Matter (TSM) of seawater used as a reference (with very low suspension) was 5 mg/l before the addition of clay, so there is a small amount of organic suspension; however, this amount is very small compared to the amount of added clay or the amount of organic suspension in the water sample represented by the green line.
In many cases hyperspectral data have noisy channels due to the poor alignment of the hyperspectral sensor or other factors related to signal quality (26). In the worst cases channels containing excessive noise have to be excluded from the dataset before processing the hyperspectral data. Generally, the signal-to-noise ratio of satellite-borne hyperspectral sensors is rather poor (27). It is important that an algorithm used to retrieve clay concentration is not sensitive to noise. The sensitivity of algorithms to noise was evaluated by generating simulated reflectance spectra where random noise was added to the measured reflectance of water samples with the known clay concentration. The simulated reflectance spectra were calculated using the following equation:

$$R_s = R_c + \frac{N}{SNR}$$

where $R_c$ denotes the spectrum of a water sample with added clay, $N$ is a random noise signal which has the same mean value as $R_c$ and $SNR$ is the signal-to-noise ratio of $R_s$.

**RESULTS**

The measured reflectance spectra of the water samples containing varying amounts of the added clay are shown in Figure 3. Ten repeated measurements were used to calculate the spectra for each water sample. Original measurements covered the wavelength range from 350 to 2500 nm, but the range above 1150 nm was excluded in this study because the reflectance of the water at that range is close to zero. All the measured reflectance spectra have two local maxima and one local minimum. The reflectance maxima are centred at the wavelengths 590 and 815 nm. The local reflectance minimum is centred at the wavelength 760 nm. The increase of the reflectance around 815 nm seems to have a quite linear relationship with the clay concentration of the water, but the increase of reflectance around 590 nm saturates when the clay concentration becomes higher. The decrease of the reflectance at 760 nm also saturates at higher clay concentrations.

In order to determine the best wavelength channels to be used in a band-ratio based algorithm, the correlation between clay concentration and hyperspectral band ratio was calculated for all possible wavelength combinations. The band ratio $R_{709}/R_{585}$ produced the best result with the coefficient of determination $r^2 = 0.941$. The band ratio $R_{995}/R_{754}$ proposed in the literature had a weaker correlation giving the coefficient of determination $r^2 = 0.519$. The correlation between $R_{754}$ and the clay concentration is not strong. This is most likely because the type or particle size of the added clay was differ-
ent in the measurements described in this paper compared to the referenced paper. The regression parameters \( a \) and \( b \) in Eq. (4) were determined and the clay concentration \( C \) in milligram per litre was calculated with the following equation:

\[
C = 8702 \frac{R_{708}}{R_{585}} + (-6498)
\]  

Figure 4: Reflectance spectra of water samples containing varying amounts of added clay.

Not all wavelength bands contain information relevant for clay concentration. Therefore, only a small number of bands was selected to be used in the spectral angle calculation. Band selection was done using a correlation analysis. The correlation between the spectral angle and the clay concentration was calculated for all possible band combinations. The correlation was calculated using different numbers of wavelength bands in order to determine how many bands should be used in the spectral angle calculation. The correlation between the spectral angle and the clay concentration is not an adequate criterion for the band selection without restrictions. There are band combinations that provide a good correlation between the spectral angle and the clay concentration, but the change in the spectral angle is very small with respect to the change in the clay concentration. This makes the algorithms based on such band combinations very sensitive to noise and other possible interferences. Therefore, additional restrictions were imposed on a band selection: the spectral angle corresponding to the lowest clay concentration 10.6 mg/l cannot be higher than 10% of the spectral angle corresponding to the highest clay concentration 142 mg/l. The best obtained coefficients of determination \( r^2 \) using a different number of wavelength bands are shown in Table 1.

Table 1: Results of correlation analysis using different number of wavelength bands.

<table>
<thead>
<tr>
<th>Number of bands</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r^2 )</td>
<td>0.932</td>
<td>0.967</td>
<td>0.962</td>
<td>0.954</td>
<td>0.944</td>
<td>0.933</td>
<td>0.922</td>
</tr>
</tbody>
</table>

The strongest correlation was obtained using five wavelength bands. Only a certain number of bands correlate strongly with the clay concentration. When more bands are used, bands with a weaker correlation decrease the correlation between spectral angle and the clay concentration. The correlation weakens gradually when more than 5 bands are employed, but the weakening is not dramatic. The wavelengths that produced the strongest correlation were: 437, 637, 685, 749 and 952 nm. The regression parameters \( a' \) and \( b' \) in Eq. (6) were determined and the clay concentration \( C \) in mg/l was calculated using the following equation:
\[ C = 643.7 \cdot \cos^{-1} \left( \frac{\sum t_i r_j}{\left( \sum t_i^2 \right)^{1/2} \left( \sum r_j^2 \right)^{1/2}} \right) + 73.9 \]  

(10)

where \( t \) denotes the remotely sensed target spectrum, \( r \) denotes the reference spectrum of the water sample without clay suspension, and \( i = \{437, 637, 685, 749, 952\} \). The root mean square error (RMSE) of the clay concentration estimate was calculated according to:

\[ \text{RMSE} = \sqrt{\frac{1}{N-2} \sum_{i=1}^{N} (C_{IV,i} - C_{RS,i})^2} \]  

(11)

where \( N \) is the number of measurements at different clay concentrations, \( C_{IV,i} \) is the measured \textit{in vivo} clay concentration and \( C_{RS,i} \) is the estimated clay concentration. RMSE values calculated for the band-ratio and spectral angle algorithms measurements at different concentrations are shown in Table 2.

\textit{Table 2: The percentage of RMSE error of the mean clay concentration for band-ratio and spectral angle algorithm.}

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Band-ratio</th>
<th>Spectral angle</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>14.5%</td>
<td>14.1%</td>
</tr>
</tbody>
</table>

The RMSE was slightly lower for the spectral angle algorithm as was expected on the basis of the correlation analysis.

The sensitivity of the algorithm to the presence of organic suspension was evaluated using simulated reflectance spectra representing water samples containing both the clay and the organic suspension. RMSE values for both presented algorithms corresponding to the varying percentage of organic suspension are shown in Table 3.

\textit{Table 3: RMSE values for the band ratio and spectral angle algorithms calculated using varying percentage of organic suspension.}

<table>
<thead>
<tr>
<th>Organic suspension</th>
<th>0%</th>
<th>10%</th>
<th>20%</th>
<th>30%</th>
<th>40%</th>
<th>50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE (Band-ratio)</td>
<td>14.5%</td>
<td>17.1%</td>
<td>20.4%</td>
<td>24.6%</td>
<td>30.1%</td>
<td>38.7%</td>
</tr>
<tr>
<td>RMSE(Spectral angle)</td>
<td>14.1%</td>
<td>14.8%</td>
<td>17.4%</td>
<td>20.8%</td>
<td>25.3%</td>
<td>32.2%</td>
</tr>
</tbody>
</table>

As can be seen from Table 3, RMSE values are significantly lower (about 15%) for the spectral angle algorithm. When the simulated water sample contained equal volumes of water with different suspension, i.e., clay and organic, the RMSE for the spectral angle algorithm was 32.2%. The result indicates that the spectral angle algorithm can provide reasonable results even when other optically detectable substances than clay are present in the water area under study.

In order to evaluate the sensitivity of the algorithms to noise, random noise was added to the reflectance spectra and the RMSE caused purely by the noise was measured. In order to get statistically representative RMSE values this process was repeated 10,000 times and the average RMSE for all measurements was calculated. The relationship between the RMSE of the algorithms and the SNR of the hyperspectral data is shown in Table 4.

The average RMSE for the spectral angle algorithm was about 10% smaller compared to the band-ratio algorithm. This indicates the robustness of the spectral angle algorithm although the difference between the algorithms was quite small.
### Table 4: The relationship between the RMSE of algorithms and SNR of hyperspectral data.

<table>
<thead>
<tr>
<th>SNR</th>
<th>100</th>
<th>30</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE(Band-Ratio)</td>
<td>1.61%</td>
<td>4.89%</td>
<td>16.25%</td>
</tr>
<tr>
<td>RMSE(Spectral angle)</td>
<td>1.45%</td>
<td>4.41%</td>
<td>14.6%</td>
</tr>
</tbody>
</table>

### CONCLUSIONS

In this paper a spectral angle based algorithm for clay concentration retrieval is presented. The proposed algorithm was tested using reflectance spectra of water samples with added clay suspension. A controlled experiment was arranged in the laboratory environment in order to record the necessary reflectance spectra. The band-ratio algorithm proposed earlier in the literature was compared against the band-ratio algorithm using optimal wavelength bands. The algorithm using optimal wavelength bands produced significantly better results, which demonstrates the need for a local calibration when developing an algorithm. The optimal number of wavelength bands used in the SAM algorithm was quite small, indicating that only a few bands have a very strong correlation to the clay concentration. The correlation decreases quite slowly when the number of bands increases from the optimal value.

The difference in the retrieval accuracy between the SAM and band-ratio algorithms was small, although in favour of the SAM algorithm. The results clearly demonstrate the robustness of the proposed SAM algorithm. The retrieval error caused by the presence of organic suspension in the water sample is smaller using the SAM algorithm when compared to the band-ratio algorithm. The retrieval error caused by the presence of noise in the measured spectrum was also smaller.

The results of this study clearly show the potential of the spectral angle algorithm, but more research should be done in order to understand the optical properties of optically detectable substances in case 2 seawater areas. Future work will contain controlled experiments where water samples with combined suspension are prepared and the reflectance spectra are recorded. A comprehensive spectral library would provide necessary information on the mixed suspension water spectra, which would make it possible to derive more accurate algorithms for the estimation of clay concentration.

### ACKNOWLEDGEMENTS

The authors would like to thank the University of Turku for the use of their facilities at the Reposaari Environmental Research Centre and for their assistance in the collection of water samples.

### REFERENCES


