Appendix A
Prepared by Emmett J. Ientilucci and John R. Schott

**Evaluation of the utility of the new blue band for water quality assessment**

**Introduction**
This study was aimed at a quick evaluation of the potential of the new blue band (443 nm) to improve water quality assessment using Landsat. As such, we took an ideal approach assuming perfect atmospheric correction (i.e., the assessment used surface reflectance values) and no noise. The assessment used data generated by the Hydrolight in-water radiative transfer model (c.f. Mobley and Fairbanks, et. al.). The model is widely used by the ocean science community to predict surface reflectance and surface leaving radiance as a function of the optical properties of water and contaminating constituents. In our case we included the three most common coloring agents. Chlorophyll-a (CHL) was included based on its concentration in mg/m$^3$. Total suspended solids (TSS) were included as concentration in g/m$^3$. Colored dissolved organic material (CDOM), also known as Gelbstoffe, was included as a scalar based on the absorption of filtered water at 350 nm. The Hydrolight model was run to generate spectral data using a factorial design scheme described below. For comparison purposes a hyper-spectral analysis was run along with the three and four band Landsat cases. The spectral data was convolved with the nominal Landsat bands to generate effective in-band reflectance values. Several regression analyses were then conducted to determine how well the water quality parameters of interest could be independently predicted by the observations.

**Approach**
Before we start our analysis, we first explain the nature and format of the data sets under test. The “Y” variable to be estimated is made up of a specific water constituent and can be of dimension 1x640. The constituents can be one of 3 parameters. Namely, CDOM (colored dissolved organic matter), TSS (total suspended solids), and CHL (chlorophyll-a), as can be seen in Figure 1.

The “X” matrix data set is made up of observed reflectance values. This matrix is of dimension 45x640 (assuming column major notation). The 45 columns originate from the fact that there are 45 spectral points that make a reflectance curve (355-795nm, delta=10nm). Each row in the matrix is the result of a specific run of Hydrolight. To generate the various reflectance curves from Hydrolight, the input water constituents were varied in a factorial manor. The total number of observations was 640. This is because there were 8 variations of CDOM, 8 variations of TSS, and 10 variations of CHL (i.e., 8x8x10=640), as can be seen in Table 1. A sample of these combinations can also be seen in Figure 1a.
Figure 1  a) Column vectors showing the combinations of CDOM, TSS, and CHL, respectively, used to generate reflectance curves from Hydrolight. This particular matrix is of dimension 3x640. b) Reflectance curves as a result of the constituent combinations. The spectral values are along the columns. This matrix is of dimension 45x640.

Table 1 Scalars used for various constituents in generating reflectance curves from HYDROLIGHT.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDOM</td>
<td>0.015</td>
<td>0.03</td>
<td>0.145</td>
<td>0.55</td>
<td>2.0</td>
<td>8.0</td>
</tr>
<tr>
<td>TSS</td>
<td>0.015</td>
<td>0.03</td>
<td>0.145</td>
<td>0.55</td>
<td>2.0</td>
<td>8.0</td>
</tr>
<tr>
<td>CHL</td>
<td>0.03</td>
<td>0.3</td>
<td>0.525</td>
<td>0.75</td>
<td>2.0</td>
<td>5.0</td>
</tr>
</tbody>
</table>

Spectral Analysis

The presented study used a classical multivariate linear regression model (LRM) for estimation of various water constituents. The model is of the form,

\[ Y = X \beta + \varepsilon \]

Where \( Y \) is the response, \( X \) is made up of predictors, \( \beta \) is made up of regression coefficients, and \( \varepsilon \) is the associated error term. This is the approach of Ordinary Least Squares (OLS) regression. We need to determine \( \beta \), however. This is done in a least-squares sense where the estimate of \( \beta \) is given by

\[ \hat{\beta} = (X'X)^{-1}X'Y \]
where $\beta_{\text{hat}}$ is a matrix representing a projection of $X$ onto an explanatory space where the information in $X$ about $Y$ is maximally exploited. The LRM is used to estimate CDOM, TSS, and CHL simultaneously. This can be done by simply including all three water constituent parameters in the $Y$ variable/matrix, as was seen in Figure 1a.

**Estimation of CDOM (Using 45 Bands)**

When estimating the water constituent CDOM, we first mean center the CDOM values. This takes care of the “intercept” value. We then generate an estimate of $Y$ while adding back in the mean. That is,

$$Y_{\text{est}} = Xc \cdot \beta + \text{mean}(Y)$$

Finally, we compare the CDOM estimates to the original CDOM values, as can be seen in Figure 2.

![Figure 2 a) Plot comparing the first 150 observations of CDOM. Notice how CDOM cycles every eight observations, as can be seen with observation number 7 labeled (assuming the first observation is number zero). b) Plot comparing all 640 observations.](image)

We should also look at the residuals for all the observations. This is simply calculated using the relation,

$$\text{error\_vector} := Y - Y_{\text{est}}$$

A plot and histogram of the CDOM residuals for all the observations can be seen in Figure 3.
In analyzing Figures 2 and 3 we should recall that the structure of the data is due to the factorial nature of the experiment. When moving from lowest to highest observations, we cycle first through incrementing CDOM, then TSS, and finally CHL, thus the structure in Figure 2. The structure indicates that we have large errors in CDOM predictions in the unlikely case of high concentrations of CDOM concurrent with very low concentrations of TSS and CHL.
Estimation of TSS (Using 45 Bands)

Figure 4 a) Plot comparing the first 150 observations of TSS. Notice how TSS repeats its cycle every 64 observations as can be seen by the label in the figure (assuming the first observation is zero). b) Plot comparing all 640 observations.

Figure 5 a) Plot of the residuals for TSS vs. observation numbers. b) Histogram of the TSS residuals.

An assessment of the TSS shows low residual errors with a pattern of larger errors with very high CDOM values (which occurs every 64 observations).
**Estimation of CHL (Using 45 Bands)**

Figure 6  
(a) Plot comparing the first 150 observations of CHL.  
(b) Plot comparing all 640 observations.

Figure 7  
(a) Plot of the residuals for CHL vs. observation number.  
(b) Histogram of the CHL residuals.

An assessment of the CHL errors also shows very low errors with a trend toward higher errors with very large values of TSS (again note TSS peaks every 64 observations).

**Landsat Convolved Reflectances**

The above experiments were all repeated using simulated Landsat bands. The 45 spectral reflectance bands were convolved down to 4 bands. The band centers are 443 (new blue), 482, 562, and 655nm. The experiments were then repeated assuming we had only the 3 historic red, green, and blue bands followed by studies including the proposed new blue band. The results are reviewed below.
Estimation of CDOM (Using 3 Bands)

Figure 8  a) Plot comparing the first 150 observations of CDOM.  b) Plot comparing all 640 observations.

Figure 9  a) Plot of the residuals for CDOM as a function of observation number.  b) Histogram of the CDOM residuals.
Estimation of CDOM (Using 4 Bands)

Figure 10 a) Plot comparing the first 150 observations of CDOM. b) Plot comparing all 640 observations.

Figure 11 a) Plot of the residuals for CDOM as a function of observation number. b) Histogram of the CDOM residuals.

An assessment of Figures 8-11 shows that the overall errors have, as expected, increased considerably with the smaller numbers of bands. However, this is an encouraging trend most obviously seen when comparing Figure 8a with Figure 10a that shows slightly better tracking of the CDOM variability with the 4 band data (this is confirmed numerically in the results section).
Estimation of TSS (Using 3 Bands)

Figure 12 a) Plot comparing the first 150 observations of TSS. b) Plot comparing all 640 observations.

Figure 13 a) Plot of the residuals for TSS vs. observation numbers. b) Histogram of the TSS residuals.
Estimation of TSS (Using 4 Bands)

Figure 14 a) Plot comparing the first 150 observations of TSS, b) Plot comparing all 640 observations.

Figure 15 a) Plot of the residuals for TSS vs. observation numbers. b) Histogram of the TSS residuals.

An assessment of the TSS errors again shows the expected increase in errors in going from a spectroscopic to a band pass system, however, once again there is an increase in performance with four bands that is most obvious in the details in Figures 12b and 14b. Moreover, the overall errors in TSS are encouragingly low.
Estimation of CHL (Using 3 Bands)

Figure 16 a) Plot comparing the first 150 observations of CHL  b) Plot comparing all 640 observations.

Figure 17 a) Plot of the residuals for CHL vs. observation number.  b) Histogram of the CHL residuals.
Estimation of CHL (Using 4 Bands)

a) Figure 18 a) Plot comparing the first 150 observations of CHL. b) Plot comparing all 640 observations.

b) Figure 19 Plot of the residuals for CHL vs. observation number. b) Histogram of the CHL residuals.

An assessment of the error trends in the CHL data shows the same overall trend as with CDOM and TSS with 4 bands out performing 3 and more importantly the errors appear to be significantly reduced in the lower (more common) concentrations of TSS.
Results

This section summarizes the results from all the tests in the form of root mean squares (RMS) of the residual errors between the input value and the prediction. We should point out that this simple study only used simple linear predictive estimation and only dealt with one variable at a time treating each constituent as an independent variable. More sophisticated approaches involving sequential estimation may yield better overall results. However, given the time and budget constraints we did not investigate those approaches.

The equations used to compute the single and total RMS errors can be seen below.

\[
\text{rms} := \frac{1}{N} \sum_{i=0}^{N-1} (Y_i - \text{Yest}_i)^2
\]

\[
\text{rms}_{\text{total}} := \frac{1}{N} \sum_{i=0}^{N-1} \left( (\text{YCDOM}_i - \text{YestCDOM}_i)^2 + (\text{YTSS}_i - \text{YestTSS}_i)^2 + (\text{YCHL}_i - \text{YestCHL}_i)^2 \right)
\]

Note that the total RMS equation is combining terms with different units and should only be taken as the crudest measure of aggregate performance.

The RMS errors of Table 2 illustrate that the fourth band shows an improvement in all constituents with a significant improvement in the CHL estimates. The overall errors are quite low for TSS given the range of TSS expected in coastal and fresh waters. The error values for CDOM and CHL are still quite large and would limit the data’s utility.

Table 2. RMS errors from linear predictions.

<table>
<thead>
<tr>
<th></th>
<th>CDOM</th>
<th>TSS</th>
<th>CHL</th>
<th>Total Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 bands</td>
<td>7.8</td>
<td>2.5</td>
<td>18.3</td>
<td>20.0</td>
</tr>
<tr>
<td>4 bands</td>
<td>7.4</td>
<td>2.4</td>
<td>15.7</td>
<td>17.5</td>
</tr>
<tr>
<td>45 bands</td>
<td>3.5</td>
<td>0.01</td>
<td>0.02</td>
<td>3.5</td>
</tr>
</tbody>
</table>

It is important to recognize that this study used a modeling process to estimate reflectance as a function of constituent concentration and the results are limited by the integrity of the model. Furthermore, we used a very wide range of concentrations in the study and did all predictions on the entire data set. It is clear that in most cases predictive equations would be used that span a narrower range of concentrations appropriate to a particular waterway. This would significantly reduce the magnitude of the expected errors. This is illustrated in Table 3 where the model was used on a reduced data set of 441 samples with the highest concentrations for each constituent removed from the factorial design. The results show a significant overall improvement in the expected errors. It can be seen that the TSS and CDOM errors are quite low and the CHL errors are also reduced to where they could be useful for some coastal water studies. Finally, we see that the new blue band has a positive impact reducing the error in all terms.
Table 3. RMS errors from linear predictions, using sub-set data.

<table>
<thead>
<tr>
<th></th>
<th>CDOM</th>
<th>TSS</th>
<th>CHL</th>
<th>Total Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 bands</td>
<td>2.1</td>
<td>0.9</td>
<td>6.7</td>
<td>7.3</td>
</tr>
<tr>
<td>4 bands</td>
<td>1.9</td>
<td>0.7</td>
<td>5.6</td>
<td>5.9</td>
</tr>
<tr>
<td>45 bands</td>
<td>0.7</td>
<td>0.01</td>
<td>0.03</td>
<td>0.7</td>
</tr>
</tbody>
</table>

**Conclusion**

The new blue band appears to significantly improve the ability to predict CHL. It also has a positive impact on TSS but limited impact on CDOM (see caveats on this study as discussed in the results section). In particular this study suggests that the use of a fourth band may permit meaningful quantitative assessment of CHL concentrations. We should point out that this study did not assess the characteristics of the new blue band (i.e., band location, width, etc.) only the potential for a band centered at 442 nm to improve the performance of the Landsat Data Continuity Mission (LDCM) for water quality assessment. In closing it is important to recognize that in coastal and fresh waters the ranges of constituents are significantly larger than in ocean waters and errors such as those indicated here, which would be unacceptable in open ocean studies, can still provide a great deal of useful information in tracking sediment transport and algal blooms. It is finally noted that for coastal and fresh water studies, the relative low TSS errors are particularly encouraging.

**References**

