Chapter 8

Over Constrained Linear Matrix Inversion with Statistical Selection

Emmanuel Boss and Collin Roesler

8.1 General Description

Semi-analytic inversions of remotely-sensed reflectance have been available since 1995 (Roesler and Perry, 1995). However, a procedure that provides an uncertainty of the inverted parameter for each individual spectrum based on uncertainties in the remote-sensing data and the model has only recently been devised (Wang *et al.*, 2005).

We use the same model philosophy as in Wang *et al.* (2005) with a slight modification (we use a single phytoplankton absorption spectrum). We assume a known relationship between r_{rs} and the absorption and backscattering coefficients (Gordon *et al.*, 1988):

$$r_{rs}(\lambda) = \frac{L_u(\lambda, 0^-)}{E_d(\lambda, 0^-)} = 0.0949\frac{b_b(\lambda)}{a(\lambda) + b_b(\lambda)} + 0.0794\left(\frac{b_b(\lambda)}{a(\lambda) + b_b(\lambda)}\right)^2. \quad (8.1)$$

The quadratic form is important for high $r_{rs}(\lambda)$ values (Garver and Siegel, 1997). Gordon *et al.* (1988) estimated that the model errors in Equation 8.1 are less than 10%.

The total absorption coefficient is partitioned as follows:

$$a(\lambda) = a_w(\lambda) + a_{ph}(\lambda) + a_{dg}(\lambda), \quad (8.2)$$

where the subscripts "w", "ph", and "dg" designate sea water, phytoplankton, and the combined contribution of CDOM and detrital material. The spectral absorption coefficient for sea water, $a_w(\lambda)$, is computed for given salinity and temperature based on Pope and Fry (1997) and Pegau *et al.* (1997).

The spectral absorption coefficient of phytoplankton is assumed to be:

$$a_{ph}(\lambda) = a_{ph}(\lambda_0)a^+_{ph}(\lambda), \quad (8.3)$$

where $a^+_{ph}(\lambda)$ is an average of normalized phytoplankton absorption spectra (Roesler and Perry, 1995) and λ_0 is commonly set as 440 nm.

The spectral absorption coefficient of the combined absorption by CDOM and detritus is:

$$a_{dg}(\lambda) = a_{dg}(\lambda_0)\exp(-S(\lambda - \lambda_0)), \tag{8.4}$$

where S is the spectral slope of the combined absorption coefficient. This function has been found to be an adequate representation of measured CDOM and detritus absorption coefficient with S ranging between 0.008 to 0.023 nm^{-1} (e.g., Roesler *et al.*, 1989).

The total backscattering coefficient, $b_b(\lambda)$, is approximated by

$$b_b(\lambda) = b_{bw}(\lambda) + b_{bp}(\lambda). \tag{8.5}$$

The spectral backscattering coefficients of sea water ($b_{bw}(\lambda)$) are computed for a given salinity based on the interpolation of the data of Morel (1974) as in Boss and Pegau (2001).

The spectral particle backscattering coefficient is assumed to obey:

$$b_{bp}(\lambda) = b_{bp}(\lambda_0)(\lambda/\lambda_0)^{-Y}. \tag{8.6}$$

This formulation is consistent with many previous studies, though without in-water validation.

To account for variability in space and time of the spectral shapes of the IOPs we perform the r_{rs} inversion allowing the shape parameters (spectral slope S and spectral slope Y) to vary within most of their observed range of variability ($0.01 \le S \le 0.02$, $0 \le Y \le 2$). For each parameter we use 11 different values with equal intervals between their maximum and minimum, resulting in $11^2 = 121$ different inversion computations for each r_{rs}.

It can be shown that with known spectral shapes, Equation 8.1 can be solved to obtain $b_{bp}(\lambda_0)$, $a_{dg}(\lambda_0)$, and $a_{ph}(\lambda_0)$ using a linear matrix inversion technique (Hoge and Lyon, 1996). When the number of wavelengths exceeds the number of unknowns (3 in our case), this solution is the best solution in a least-square sense (Press, 1992).

From all the solutions to Equation 8.1 we select the solution for which $a_{dg}(440)$ and $a_{ph}(440) > -0.005$ m^{-1} and $b_{bp}(440) > -0.0001$ m^{-1} (slightly negative values are accepted to compensate for finite uncertainties in measurements and calibrations). We further restrict ourselves to the solutions whose reconstructed r_{rs} (calculated by substituting the solutions into Equation 8.1) obeys:

$$|r_{rs,\text{reconstucted}}(\lambda) - r_{rs,\text{known}}(\lambda)|/r_{rs,\text{known}}(\lambda) < 0.1 \text{ or } 0.2 \text{ for every } \lambda.$$

These criteria can result in cases where no solution could be found for a given r_{rs}. The choice of the criteria should be driven by knowledge of uncertainties in observed r_{rs} as well as the assumed spectral shapes (in particular that of phytoplankton).

We thus present the results from the two different solution selection criteria in the tables, but only the criteria of 0.1 in the plots. We provide uncertainties for the solutions on the plots based on the distance between the 84^{th} and 16^{th} percentile of the obtained solutions (∼ ± one standard deviation for a normal distribution).

Given the application to remote sensing we used only the R_{rs} values at 410, 440, 490, 510 and 550 nm (or nearby for the *in situ* data set).

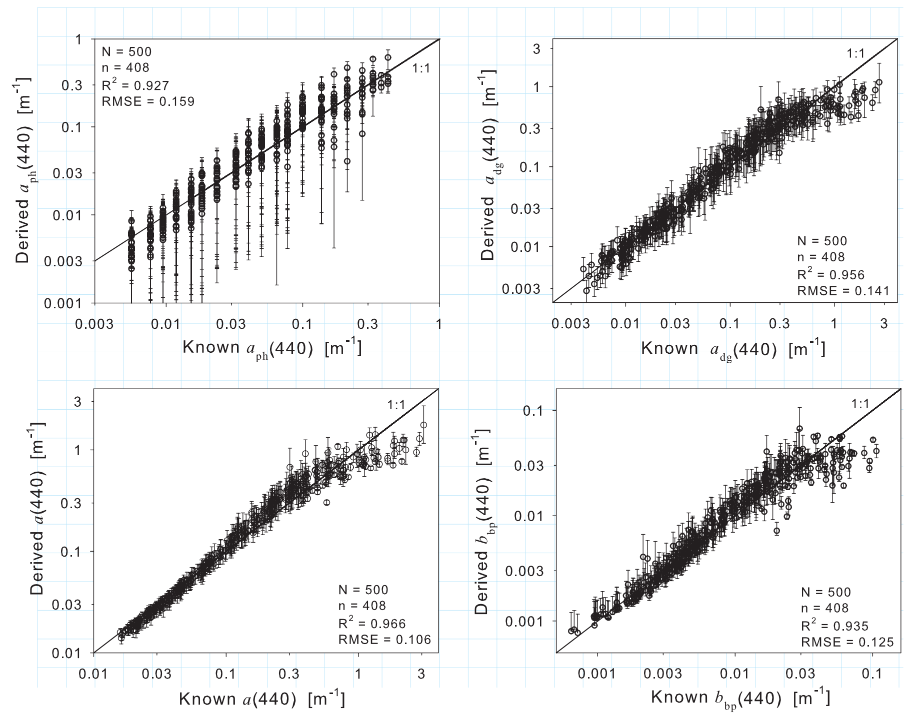

Figure 8.1 Comparison of inverted and the simulated data set (Sun at 30^o from zenith) for $a_{ph}(440)$, $a_{dg}(440)$, $a(440)$, and $b_{bp}(440)$ for the 10% criteria (statistics in Table 8.1). Vertical lines denote the 90% confidence intervals in the solutions. R_{rs} values at 410, 440, 490, 510 and 550 nm were used as inputs for IOP retrieval.

8.2 Results and Discussion with IOCCG Data Sets

8.2.1 Simulated data set

Over the large dynamic range of the data set the inversion fares rather well for both 10 and 20% criteria (Figure 8.1, Tables 8.1 and 8.2). Not surprisingly the stringent criteria provide less but better solutions (in terms of RMSE error and bias). The agreement between derived and known IOPs can be further improved by choosing other wavelengths (*e.g.* 410 nm for a_{dg} and 550 nm for b_{bp}) and

Table 8.1 RMSE and regression (Type II) results for the synthesized data set. Statistics of comparison of the median of all possible inversion solutions with a 10% agreement criterion. R_{rs} values at 410, 440, 490, 510 and 550 nm were used as inputs for IOP retrieval. N is the number of data tested, while n is the number of valid retrieval.

	N	n	intercept	slope	R^2	RMSE	bias
$a(440)$	500	408	-0.001	0.995	0.966	0.106	0.003
$b_{bp}(440)$	500	408	-0.055	0.972	0.935	0.125	0.003
$a_{dg}(440)$	500	408	-0.021	0.982	0.956	0.141	-0.001
$a_{ph}(440)$	500	408	0.160	1.113	0.927	0.159	-0.002

Table 8.2 RMSE and regression (Type II) results for the synthesized data set. Statistics of comparison of the median of all possible inversion solutions with a 20% agreement criterion. R_{rs} values at 410, 440, 490, 510 and 550 nm were used as inputs for IOP retrieval. N is the number of data tested, while n is the number of valid retrieval.

	N	n	intercept	slope	R^2	RMSE	bias
$a(440)$	500	438	-0.074	0.938	0.946	0.145	-0.025
$a_{ph}(440)$	500	438	-0.025	1.014	0.878	0.201	-0.044
$a_{dg}(440)$	500	438	-0.082	0.942	0.944	0.169	-0.023
$b_{bp}(440)$	500	438	-0.186	0.925	0.898	0.168	-0.034

by adding more wavelengths (*e.g.* Wang *et al.* (2005) added a 670 nm channel and the successful retrieval increased from 408 to 472 with the 10% criteria). It is encouraging that the uncertainty estimates for both $a_{dg}(440)$ and $b_{bp}(440)$ intersect the 1:1 line suggesting the constraint criteria is working well.

8.2.2 *In situ* data set

Large uncertainties in inverted parameters (in particular a_{ph}) suggest that some of these data have many possible solutions and thus large uncertainties for a given $R_{rs}(\lambda)$. Some data points are way off the line, possibly due to large sun angles and/or poor measurements (Figure 8.2 and Tables 8.3 and 8.4).

In Wang *et al.* (2005) we used a more complicated phytoplankton absorption formulation which increased the computation by a factor larger than 10. We found that this complexity did not improve the match ups significantly and thus decided here to use a single phytoplankton absorption function. It can easily be demonstrated that a different choice of wavelengths for inversions or a different choice of wavelength for the parameter can significantly improve/degrade the retrieval. Thus, if we are after a_{dg}, inverting a near UV wavelength provides the best inversion; while for b_b, it is in the NIR that the inversion does best; as long as adequate R_{rs} at those wavelengths could be available.

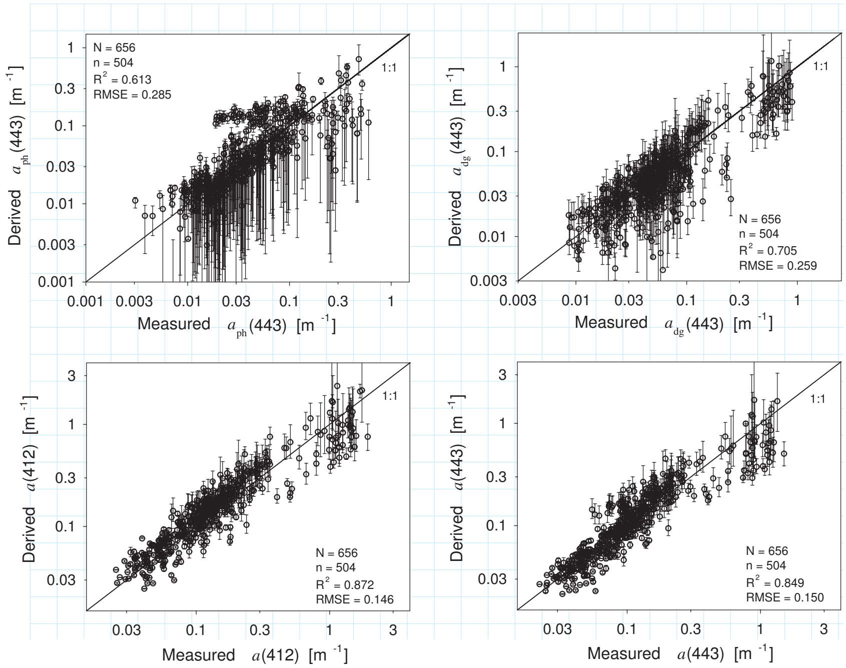

Figure 8.2 Comparison of inverted and the *in situ* data set (Sun at 30° from zenith) for $a_{ph}(440)$, $a_{dg}(440)$, $a(440)$, and $b_{bp}(440)$ for the 10% criteria (statistics in Table 8.3). Vertical lines denote the 90% confidence intervals in the solutions. R_{rs} values at 410, 440, 490, 510 and 555 nm were used as inputs for IOP retrieval.

8.3 Summary

The inversion method presented here was designed to provide uncertainty estimates of inversion products and is dependent on the reality of the assumptions of the model. For example, it is well known that Equation 8.6 is likely not a good representation of particulate spectral backscattering, yet it is the only simple model currently available. Much work is still needed to understand spectral IOPs, and such work will, without a doubt, improve our ability to retrieve in-water parameters from remote sensing.

Table 8.3 RMSE and regression (Type II) results for the *in situ* data set. Statistics of comparison of the median of all possible inversion solutions with a 10% agreement criterion. R_{rs} values at 412, 443, 490, 510 and 555 nm were used as inputs for IOP retrieval. N is the number of data tested, while n is the number of valid retrievals.

	N	n	intercept	slope	R^2	RMSE	bias
$a(412)$	656	504	-0.022	0.942	0.872	0.146	0.029
$a(443)$	656	504	-0.029	0.969	0.849	0.150	0.001
$a_{dg}(443)$	656	504	-0.018	1.043	0.705	0.259	-0.072
$a_{ph}(443)$	656	504	0.068	1.031	0.613	0.285	0.024

Table 8.4 RMSE and regression (Type II) results for the *in situ* data set. Statistics of comparison of the median of all possible inversion solutions with a 20% agreement criterion. R_{rs} values at 412, 443, 490, 510 and 555 nm were used as inputs for IOP retrieval. N is the number of data tested, while n is the number of valid retrievals.

	N	n	intercept	slope	R^2	RMSE	bias
$a(412)$	656	629	-0.036	0.939	0.867	0.157	0.019
$a(443)$	656	629	-0.039	0.977	0.842	0.165	-0.017
$a_{dg}(443)$	656	629	0.086	1.069	0.63	0.298	-0.013
$a_{ph}(443)$	656	629	-0.057	1.014	0.714	0.266	-0.075