Shallow-water Remote Sensing:

Lecture 2B: Database Methods for Spectrum Matching

Curtis Mobley

Vice President for Science and Senior Scientist
Sequoia Scientific, Inc.
Bellevue, WA 98005
curtis.mobley@sequoiasci.com

IOCCG Course
Villefranche-sur-Mer, France
July 2012

Database Spectrum Matching

Mobley et al., 2005. *Applied Optics*, 44(17), 3576-3592

Use a radiative transfer code to create a database of R_{rs} spectra that correspond to all possible combinations of water absorption and scattering properties, bottom depths, and bottom reflectances that might be found in the area being studied.

Each R_{rs} spectrum in the database corresponds to a known set of water properties (a, b and b_b spectra), a bottom reflectance spectrum (bottom type), and a water depth.

Then search the database to find the closest-matching database spectrum to the given image spectrum

The retrieved environmental properties are then whatever values were used to create the closest-matching database spectrum.

CRISTAL

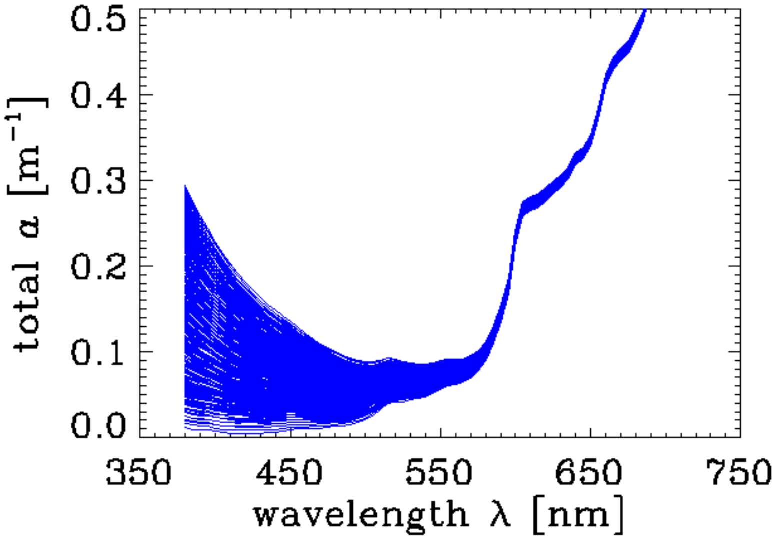
The following results were generated using CRISTAL

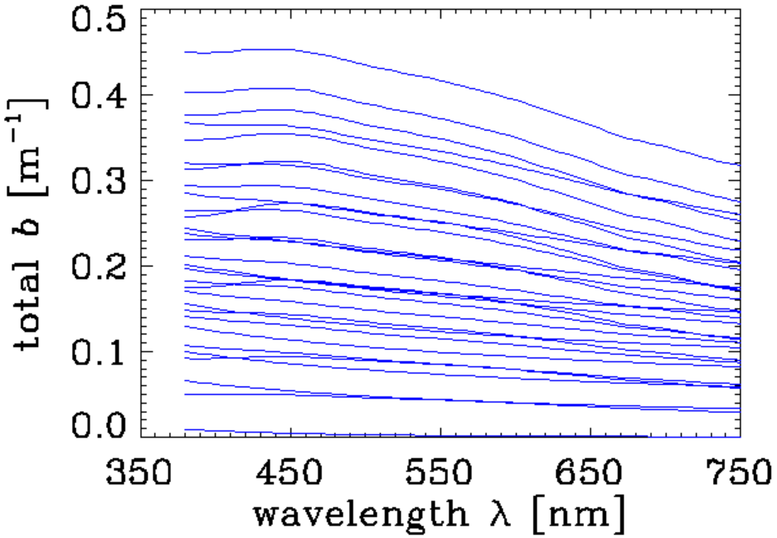
CRISTAL (Comprehensive Reflectance Inversion based on Spectrum matching and TAble Lookup) is a software package developed by me to handle the creation of R_{rs} databases, retrieval of environmental properties (water IOPs, bottom depth, and bottom reflectance or type) from hyperspectral imagery, and display of retrieved results.

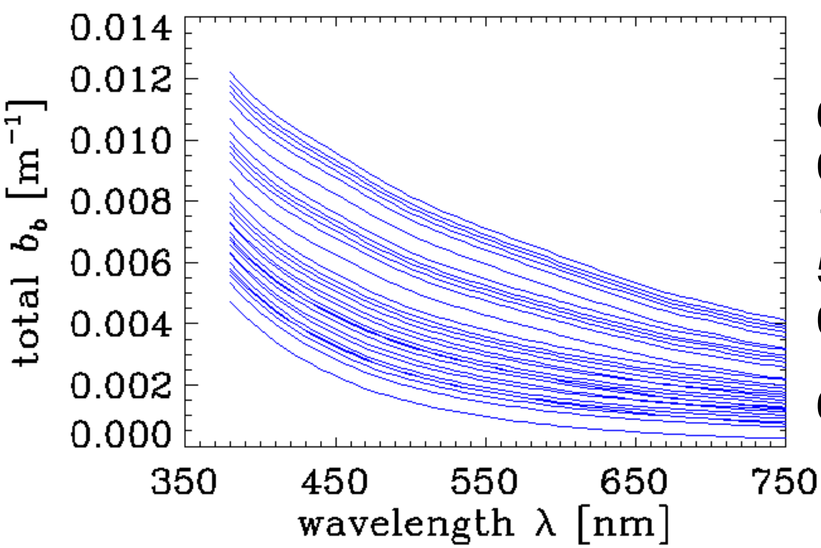
Parts of CRISTAL are coverved by U.S. Patent 7369229

Publications will be submitted asap and the code will eventually be made public.

R_{rs} Database Creation: IOPs







For the Bahamas use

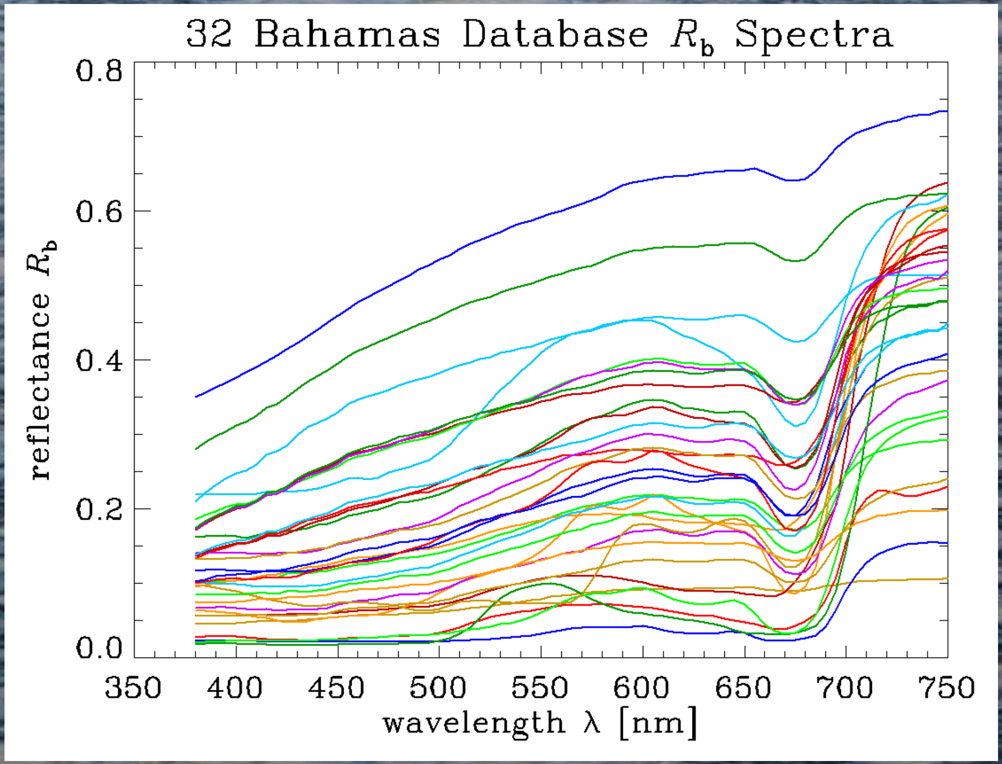
6 Chl values: 0.0, 0.05, 0.10, 0.15, 0.20, 0.30 mg m^{-3}
11 a_{CDOM} values: 0.0 to 0.1 m^{-1} by 0.01 m^{-1}
5 mineral concentrations: 0.0, 0.05, 0.1, 0.2 and 0.3 gm m^{-3}

6x11x5 = 330 IOP sets

R_{rs} Database Creation: Bottom Reflectance

32 different bottom reflectance spectra (pure bottom types and mixtures of bottom types)

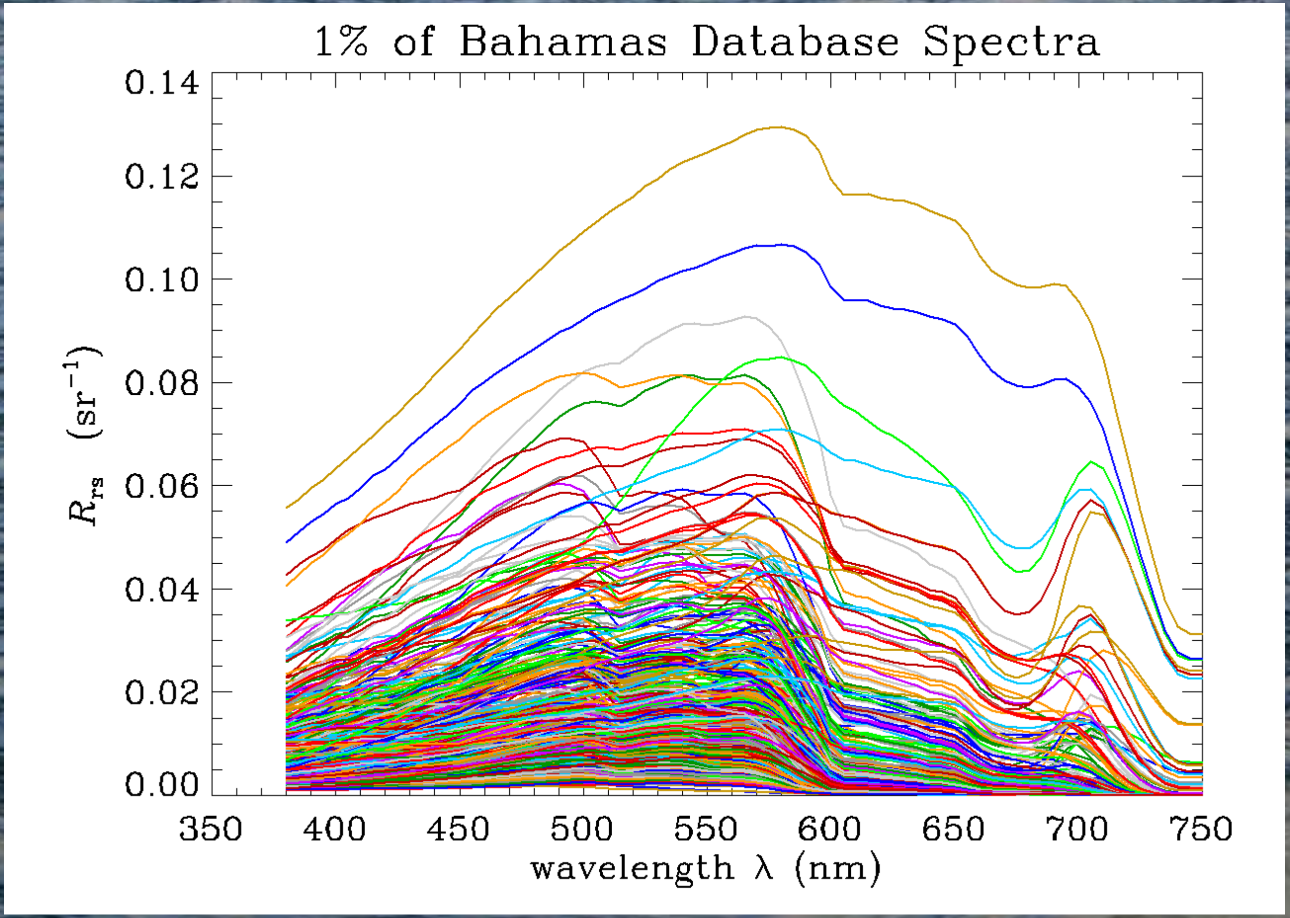
The bottom was placed at 56 depths:
z_b = 0.25, 0.50, 0.75, 1.0, ...,14.75, 15.0, 16.0, ...,19, 20 m, and ∞



The database creation run shown here (for Bahamas waters) used 330 sets of water properties x 32 bottom reflectances x 56 depths, so 330 x (32*55 + 1) ≈ 581,130 RTE solutions to create R_{rs} spectra from 380 to 750 nm by 5 nm (about a week of computer time on a 2 GHz PC).

Database creation is a one-time calculation for a given environment.

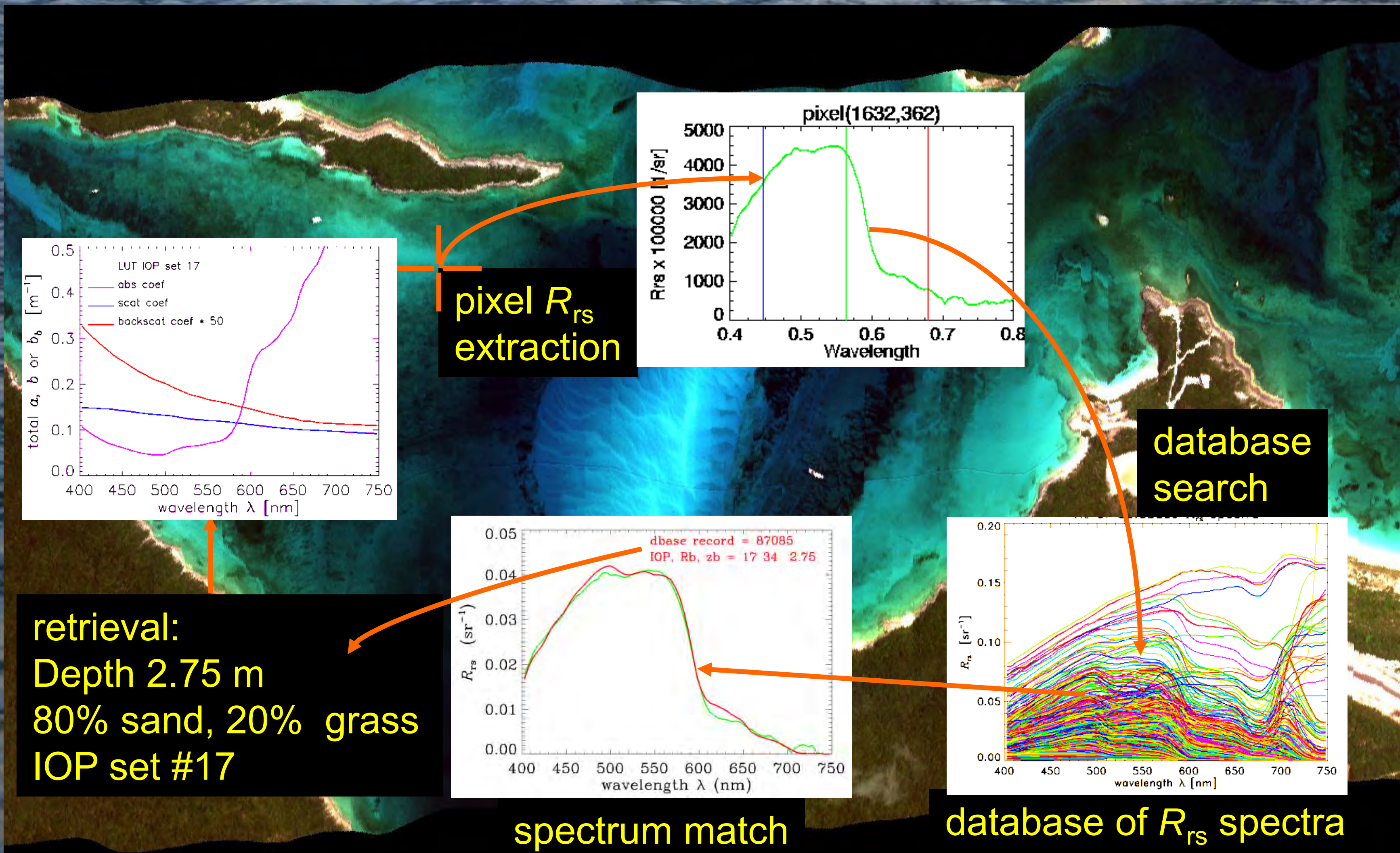
R_{rs} Database Creation



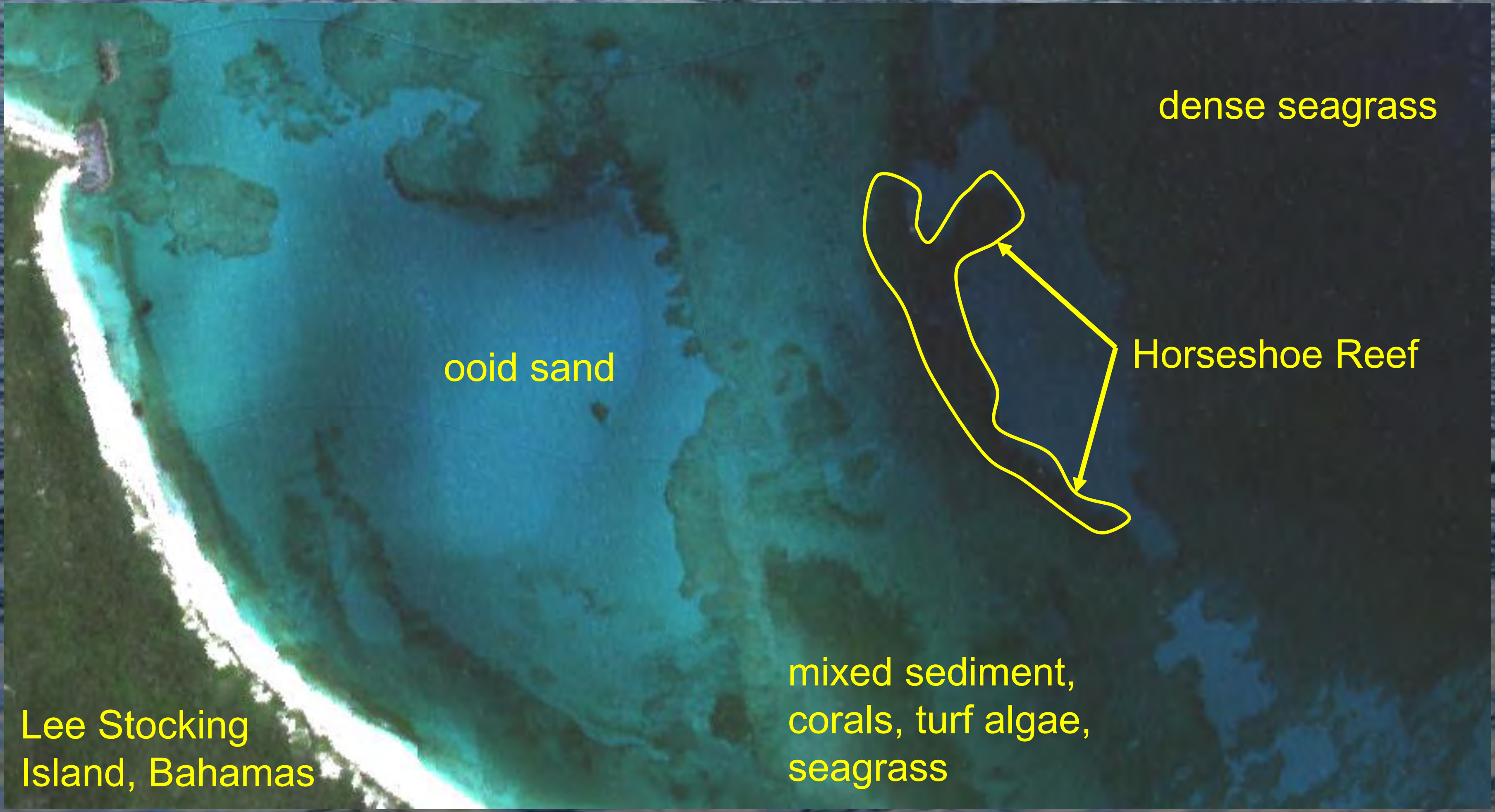
Each R_{rs} spectrum in the database corresponds to a known set of water properties (a, b and b_b spectra), a bottom reflectance spectrum (bottom type), and a water depth.

Image Processing

(after atmospheric correction)

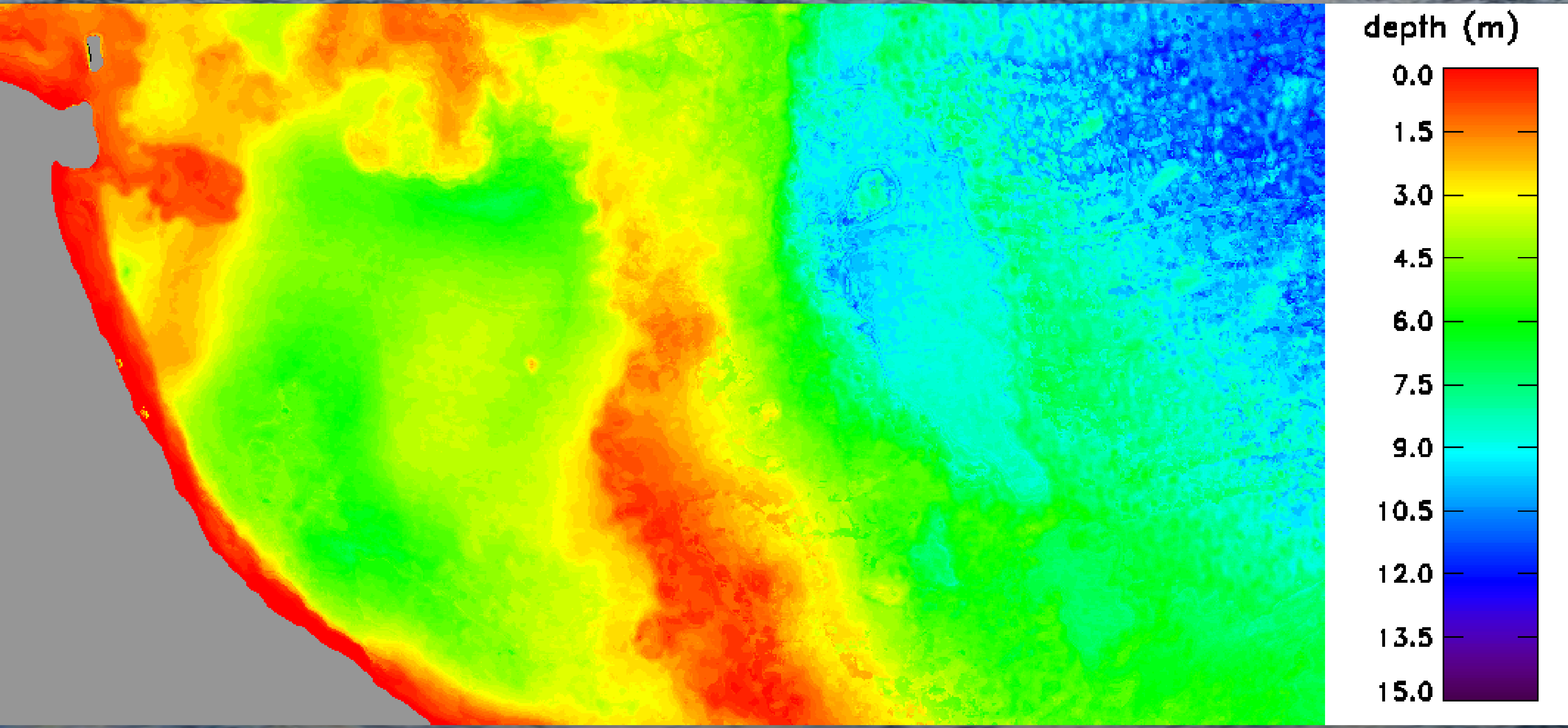


Example: Airborne Hyperspectral Image of Very Clear Water in the Bahamas

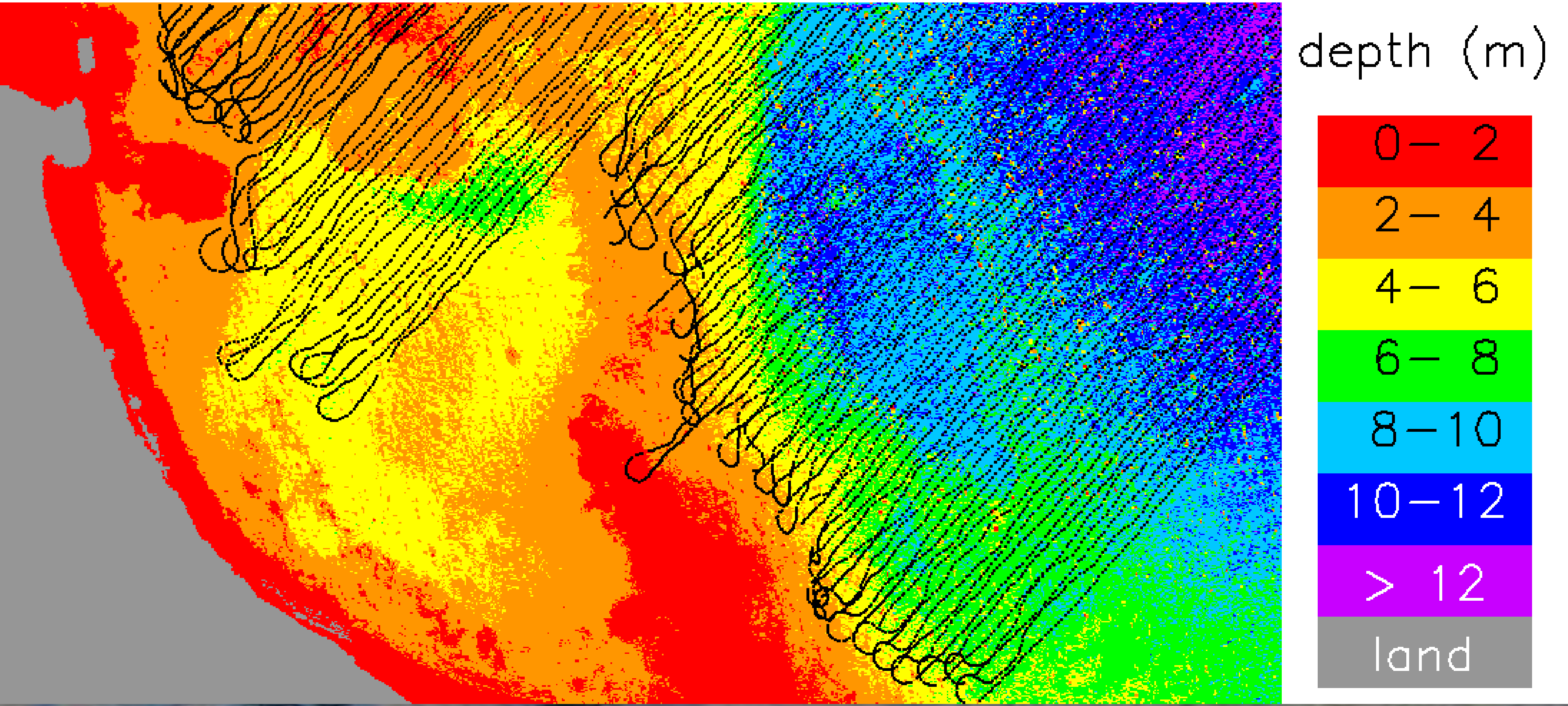


NRL-DC PHILLS image from ONR CoBOP program, May 2000
501x899 pixels at ~1.3 m resolution

Bathymetry Retrieval

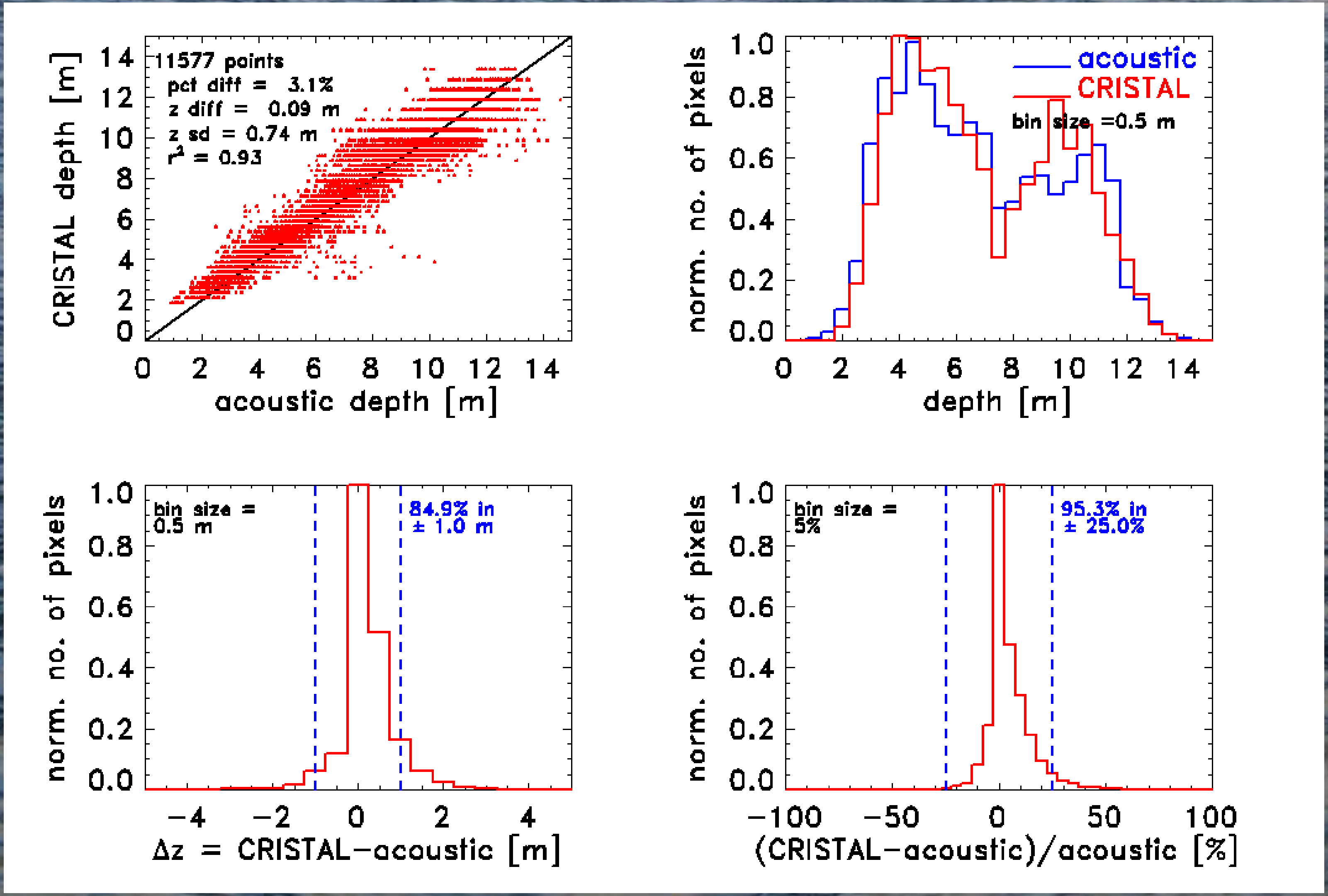


Validation with Acoustic Bathymetry



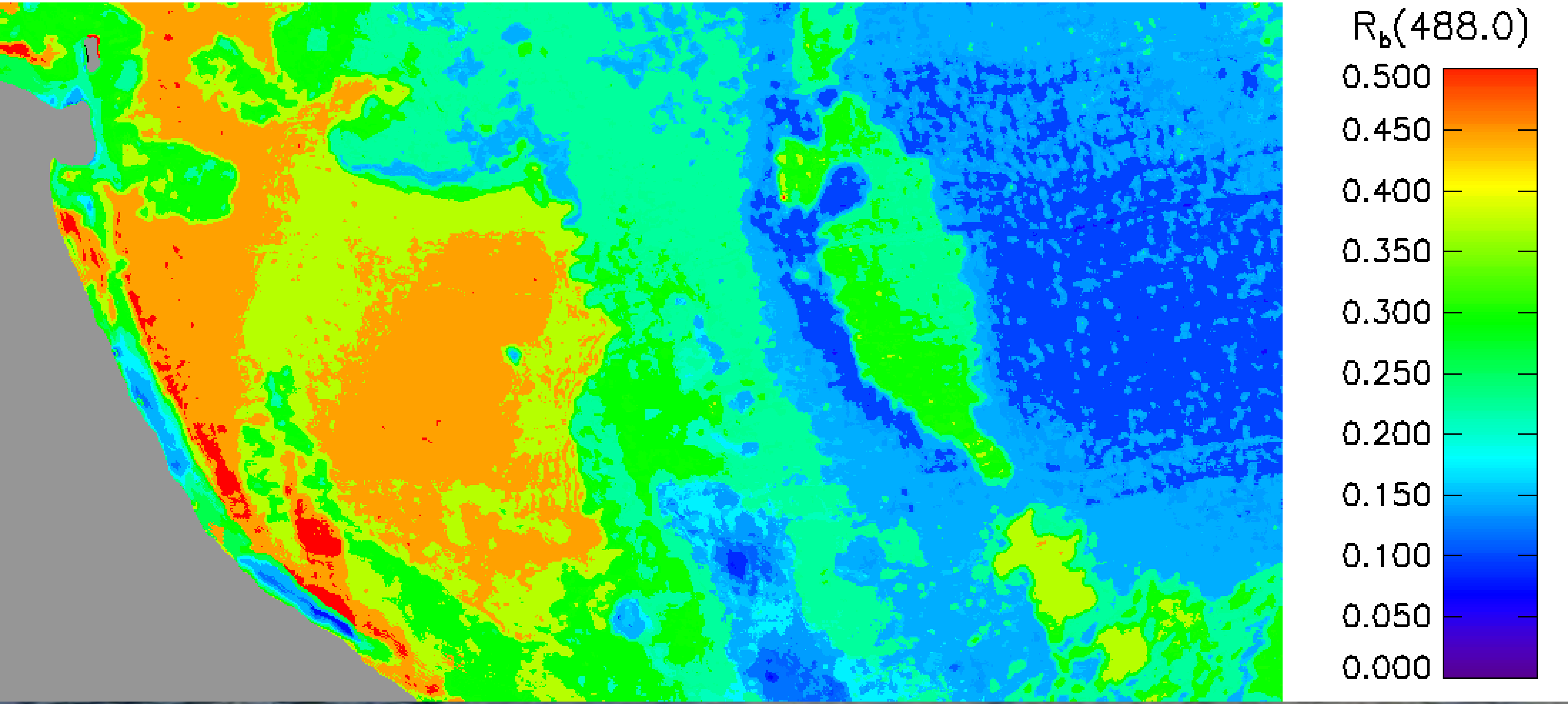
Black: NRL acoustic survey for ONR CoBOP program
Color: CRISTAL depth retrieval

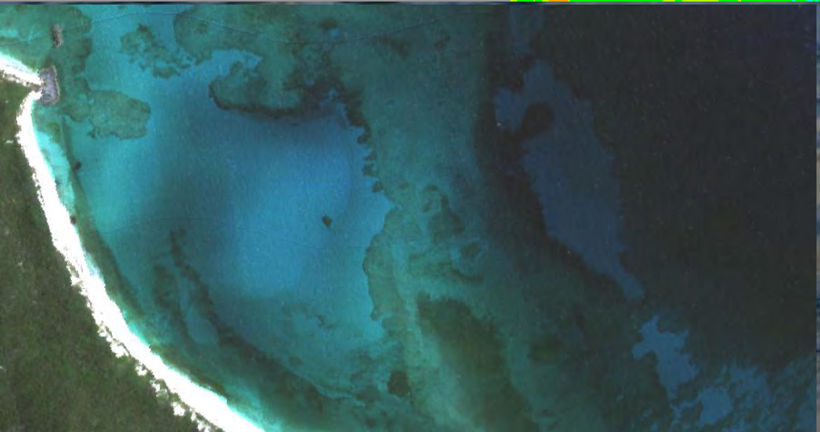
Depth Retrieval vs. Acoustic Bathymetry



These retrieval errors also include errors due to latitude-longitude calculations in mapping acoustic ping locations to image pixels (horizontal errors of several meters or more due to failure of built-in navigation instrument), and due to whitecaps

Bottom Reflectance

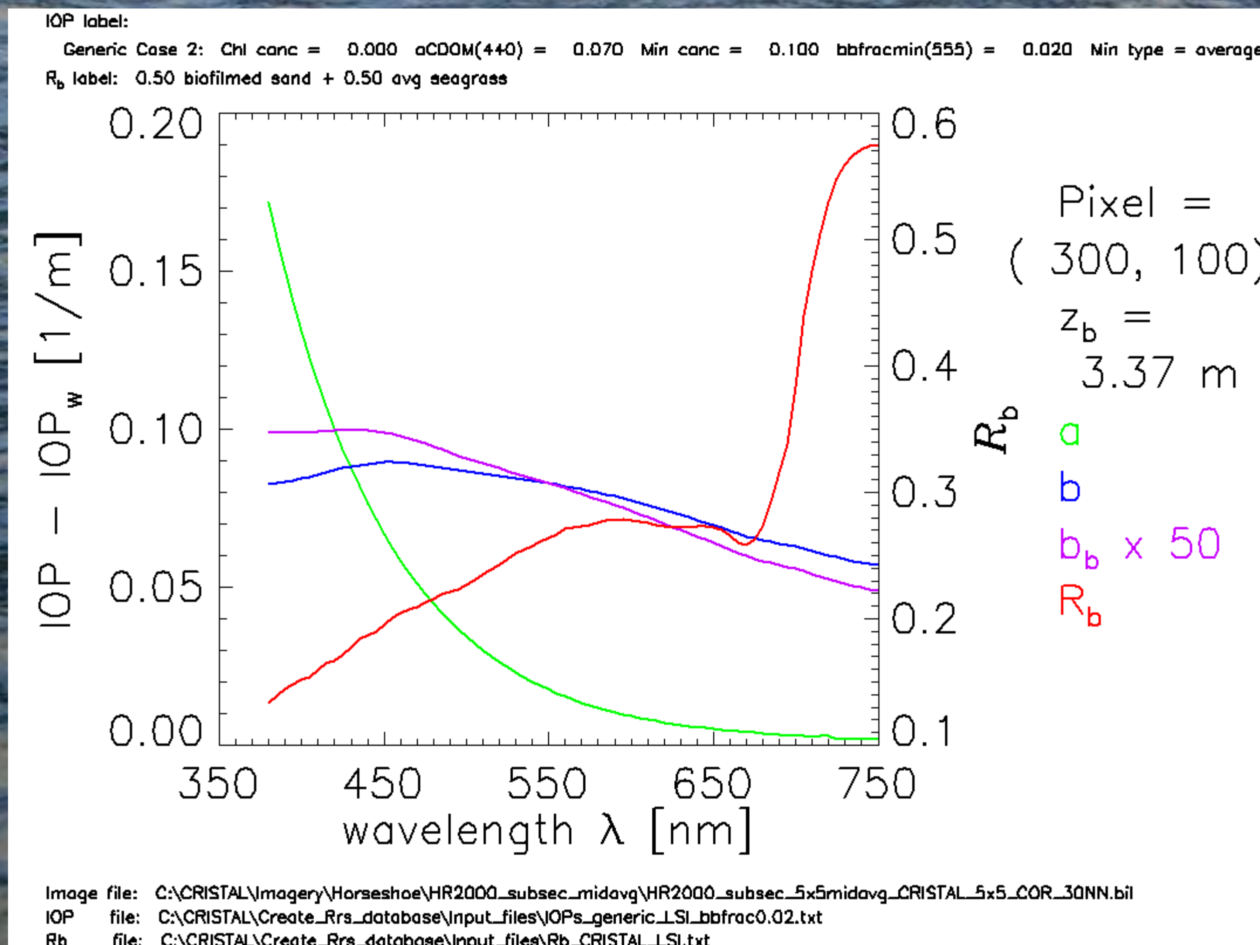




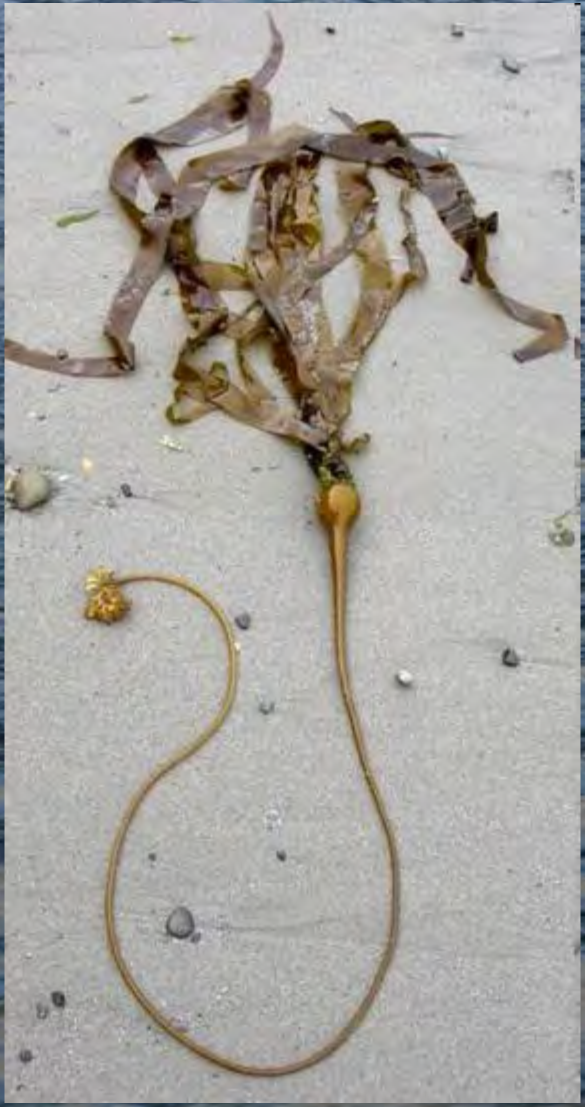
$R_b(488)$ is what you would need for performance evaluation of a 488 nm bathymetric lidar

Retrieval Information

Keep in mind that a database spectrum matching retrieval retrieves full spectral information at each pixel



Kelp Mapping



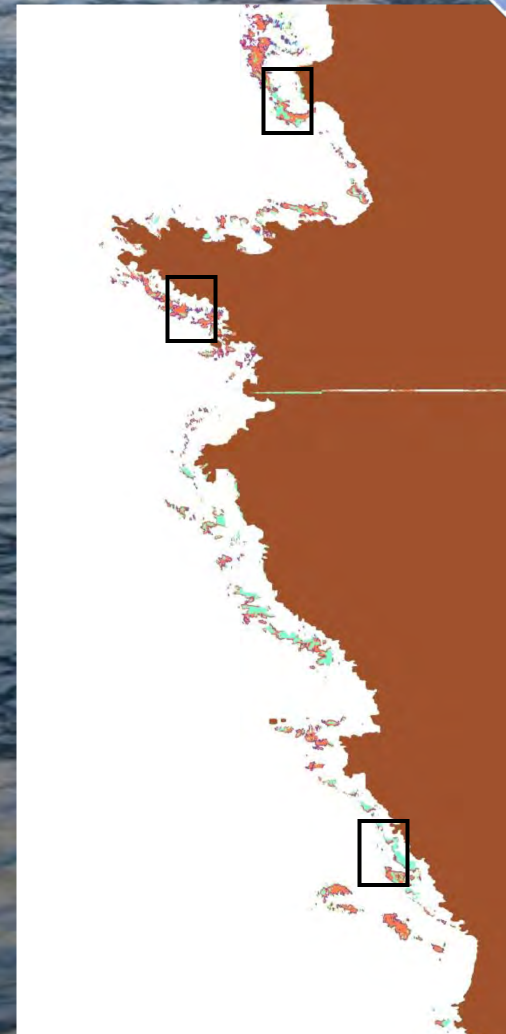
http://www.beachwatchers.wsu.edu/ezidweb/seaweeds/Nereocystis.htm

Bull kelp (*Nereocystis luetkeana*) is very important for food, medicines, sheltering of fish, and recreational diving. Harvesting is strictly managed in the US.

http://www.bestpicturesof.com/misc/pictures%20of%20bull+kelp/?page=2#Google

Mapping of Kelp Coverage
California Coast

2002



2004

DEM Land | 0 % | 10 % | 20 % | 30 % | 40 % | 50 %

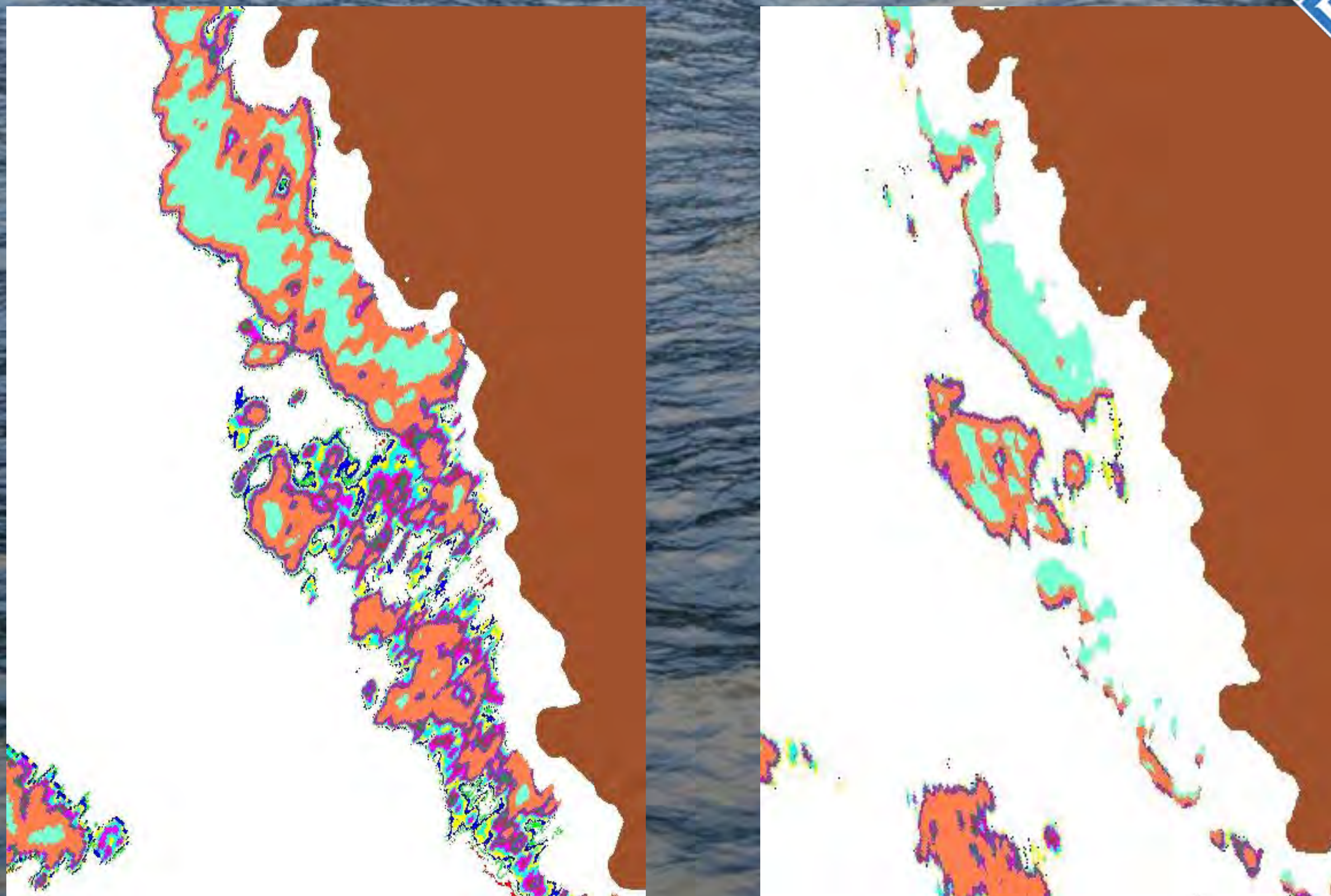
60 % | 70 % | 80 % | 90 % | 100 % | *100 %

images courtesy of Paul Bissett, FERI

Mapping of Kelp Coverage California Coast

CICORE

2002



2004

FERI Florida Environmental Research Institute

DEM Land 0 % 10 % 20 % 30 % 40 % 50 %

60 % 70 % 80 % 90 % 100 % *100 %

images courtesy of Paul Bissett, FERI

Humboldt Bay California
Eel Grass Mapping

Chaeli Judd, MS Thesis, Judd et al., 2006

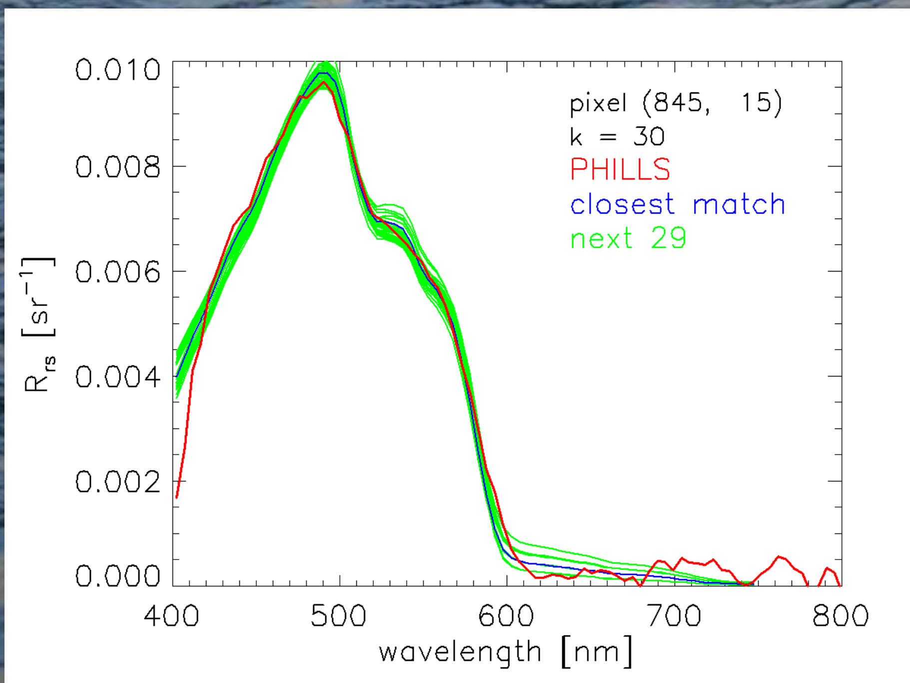
HSI determined eel grass distributions, previously unknown.

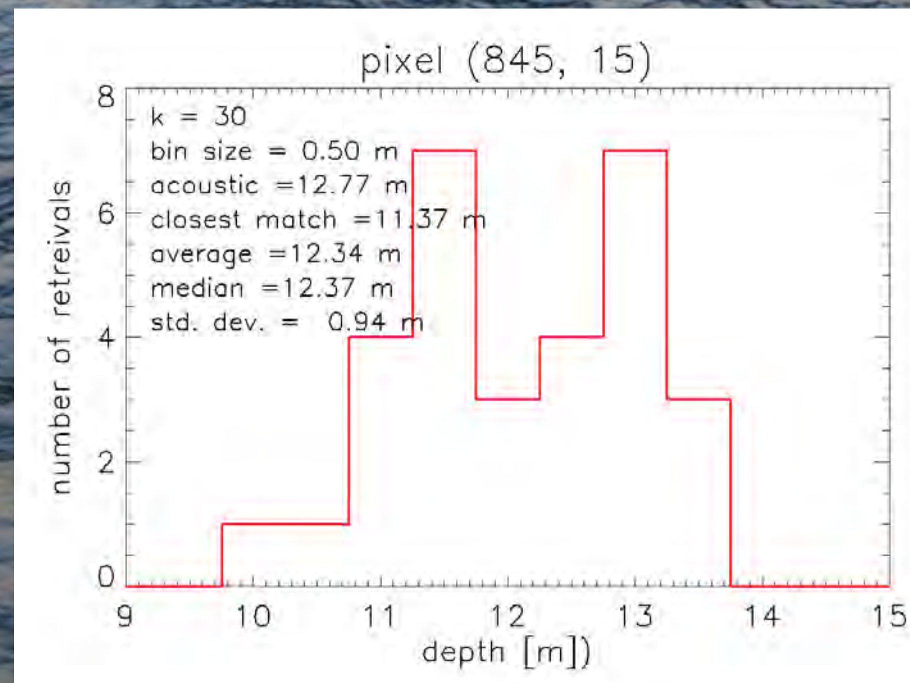
image courtesy of Paul Bissett, FERI

Error Analysis

Being able to place error bars or confidence estimates on retrievals is often as important as the retrieved value itself

Can do this statistically from the distribution of retrieved values for the *k* closest matching spectra (k Nearest Neighbors, or kNN)

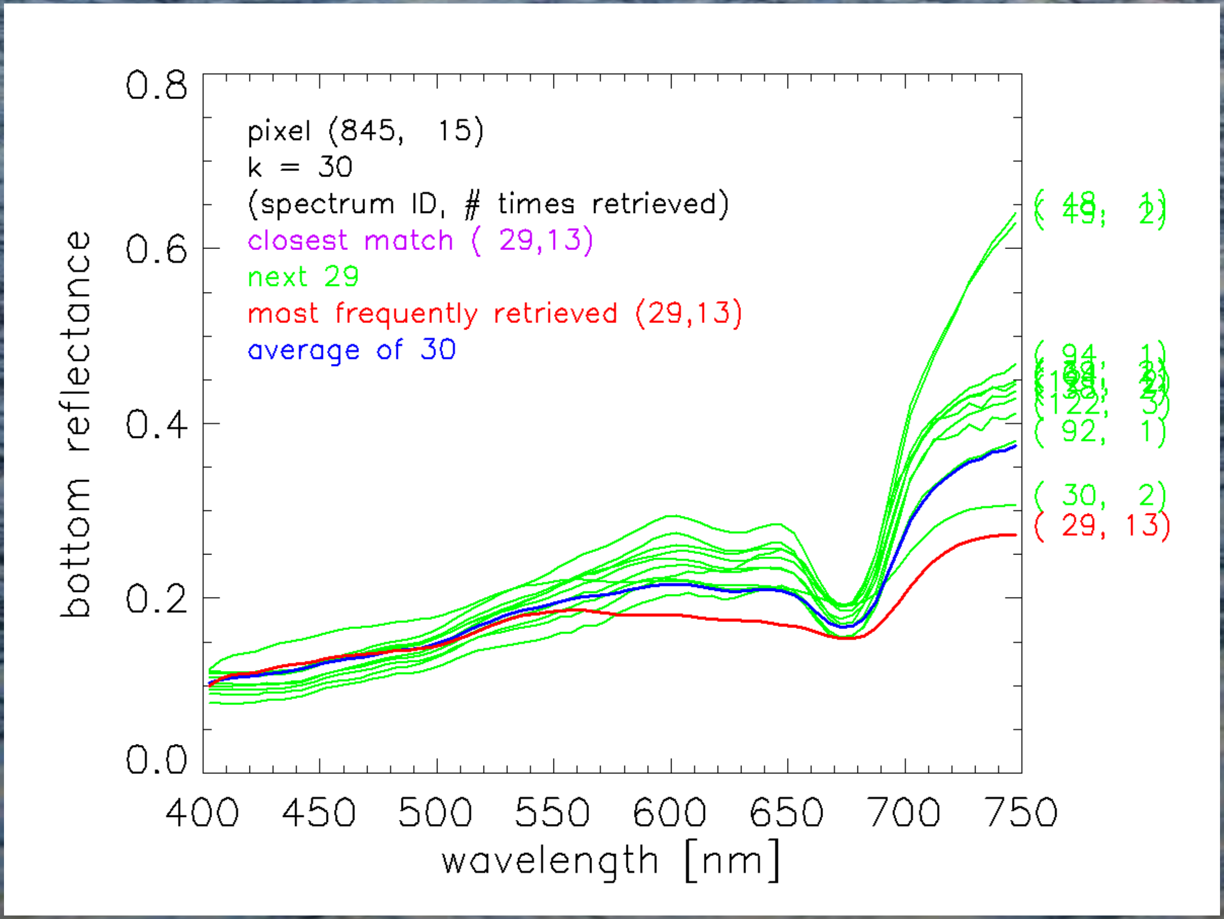




the 30 closest matches give a histogram of retrieved depths

the average or median gives a better estimate of the depth, plus an error estimate

Error Analysis

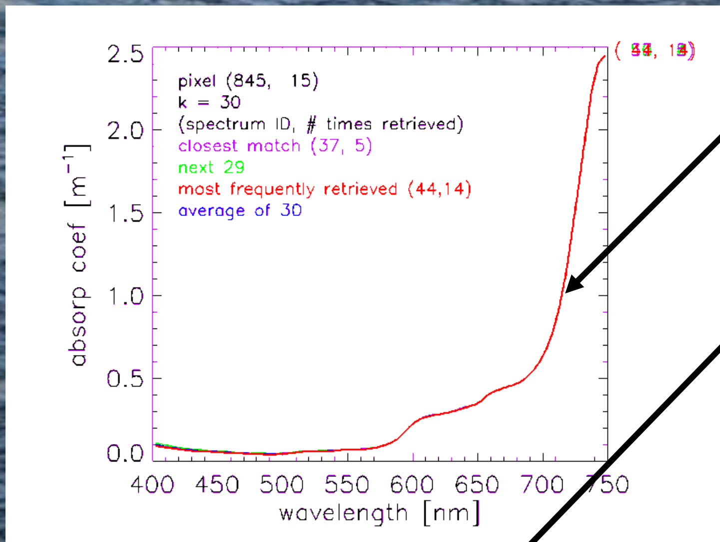


The closest and most frequently retrieved bottom reflectance spectrum was 30% sand and 70% seagrass.

The other bottoms are similar mixtures of sand and grass, sargassum, turf algae, and macrophytes.

So we can be fairly certain that the bottom is dense vegetation, probably sea grass

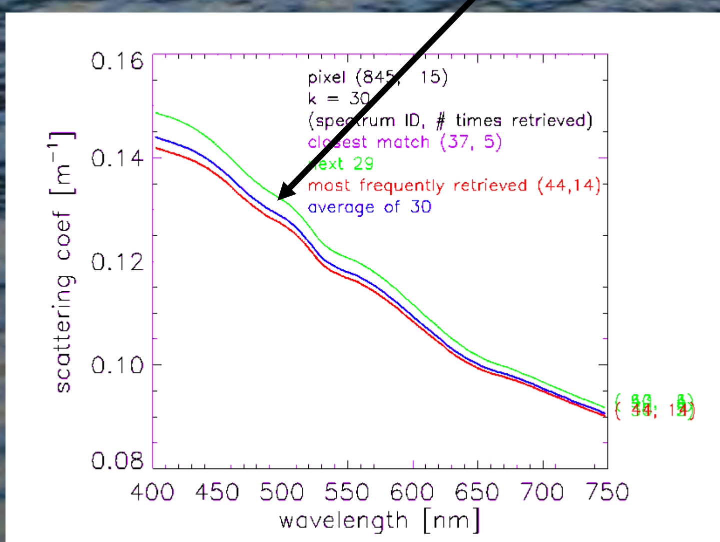
Error Analysis

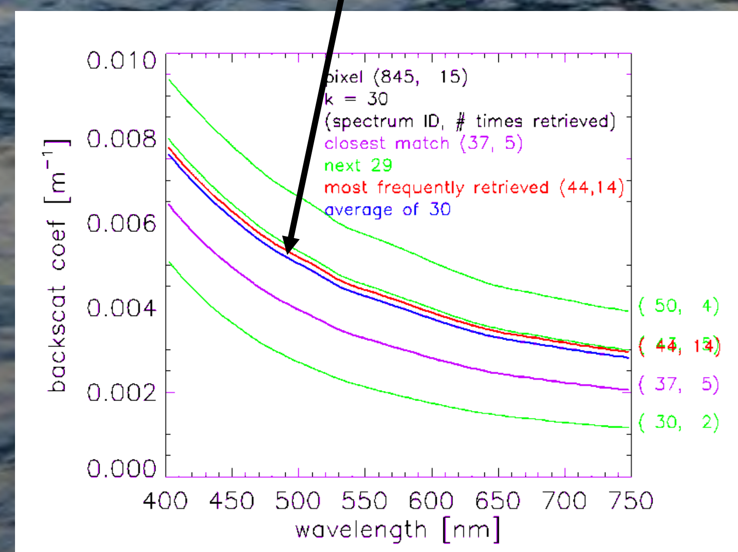


The retrieval is very certain about the absorption coefficient

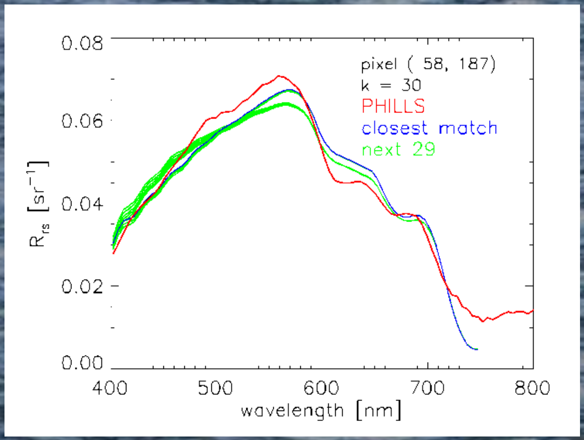
The retrieval is fairly certain about the scattering coefficient

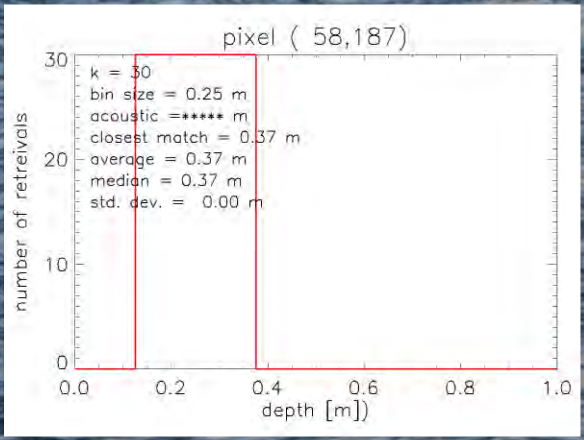
The retrieval is UNcertain about the backscatter coefficient



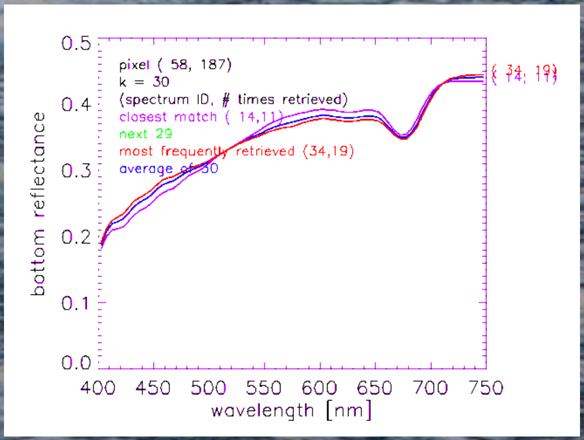


Error Analysis: A Shallow-water Pixel

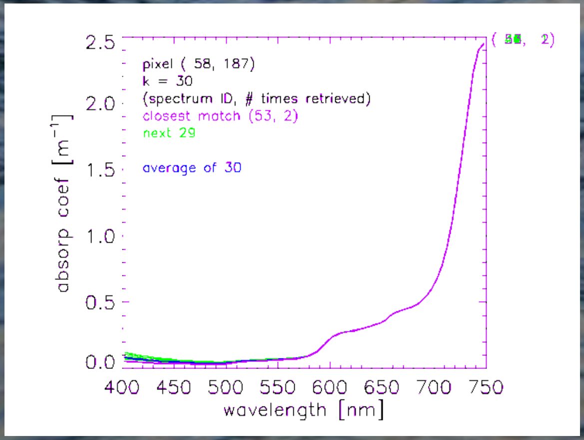




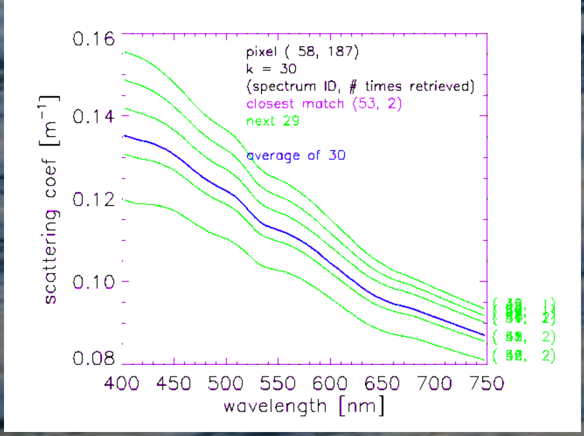
all depths the same; very confident



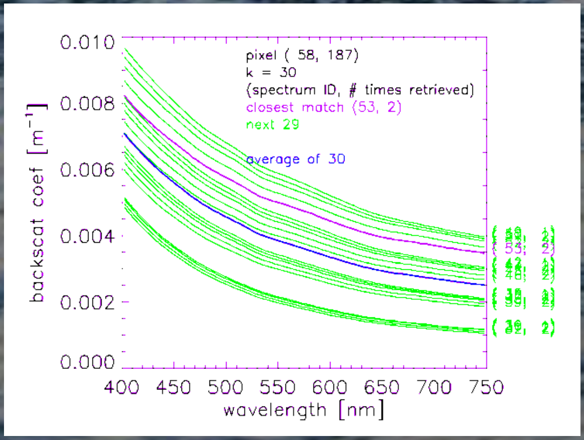
bottoms very similar (sand or grapestone); very confident



absorption; very confident



scattering; uncertain



backscatter; very uncertain

Does This Make Sense?

- In these very clear waters, the water absorption determines how much light gets to the bottom and back to the surface. Water-column scattering and backscatter contribute less to the water-leaving radiance in shallow water than does the bottom reflectance.
- The retrieval was therefore most certain about the absorption coefficient, and least certain about backscatter.
- The bottom reflectances all had similar reflectance spectra because it's the reflectance that is important. The retrieval wasn't able to distinguish between sea grass, turf algae, *sargassum*, and macrophytes, which all have similar reflectances.
- In very shallow (<5 m) clear water, the retrieved bottom reflectance becomes very certain and the water scattering and backscatter very uncertain (i.e., least important in determining R_{rs})

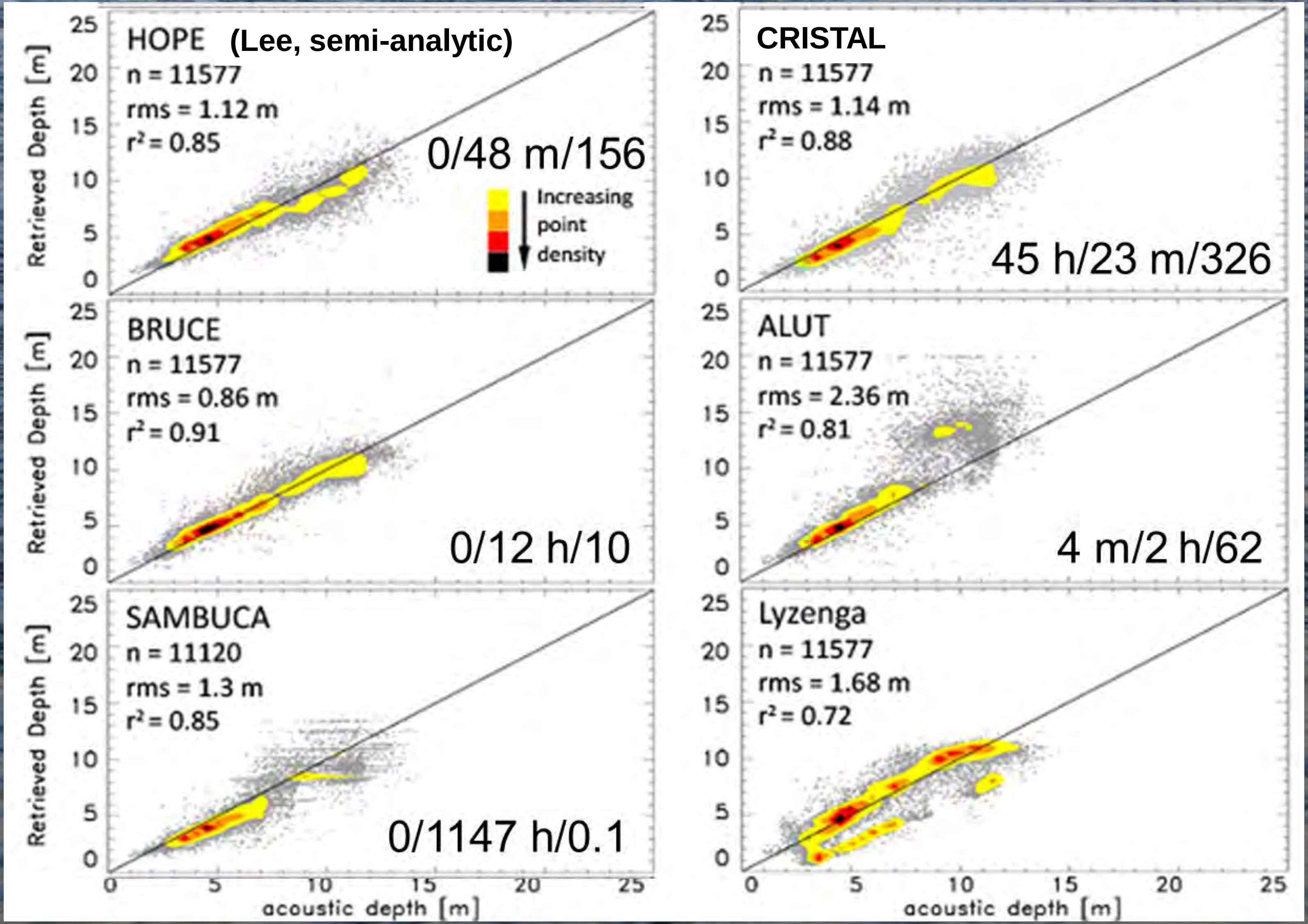
Comparison of Database & SA Algorithms

Feature	CRISTAL	Semianalytical
Algorithm basis	exact solution of the RTE as expressed in the R_{rs} database spectra	approximate solution of the RTE as expressed in the semi-analytical model
Fundamental advantage	accounts for wavelength fine structure of spectra, thus allowing for species-level identification of biota	applicable to any water body without the need for pre-computing underlying databases
Fundamental limitation	Retrievals are good only if the R_{rs} database is representative of the environment.	Retrievals are good only if the semi-analytical model is representative of the environment.
Convergence to a solution	no convergence problems because a closest-matching database spectrum is always found (even if the match is poor because the database is not representative of the environment)	the optimization may not converge or may converge to a local minimum if the initial guess parameter values are not sufficiently accurate or if the model is not representative of the environment
Applicable environment	any water body described by the R_{rs} database	any water body described by the semi-analytical model
Imagery required	R_{rs} spectra must be well calibrated and atmospherically corrected	R_{rs} spectra must be well calibrated and atmospherically corrected
Preprocessing	An R_{rs} database must be pre-computed for the given environment before image processing	No preprocessing is required.
Image processing time	fast when optimized database searches are used	fast or slow, depending on search algorithm and implementation

CRISTAL Users' Guide Table 1.1

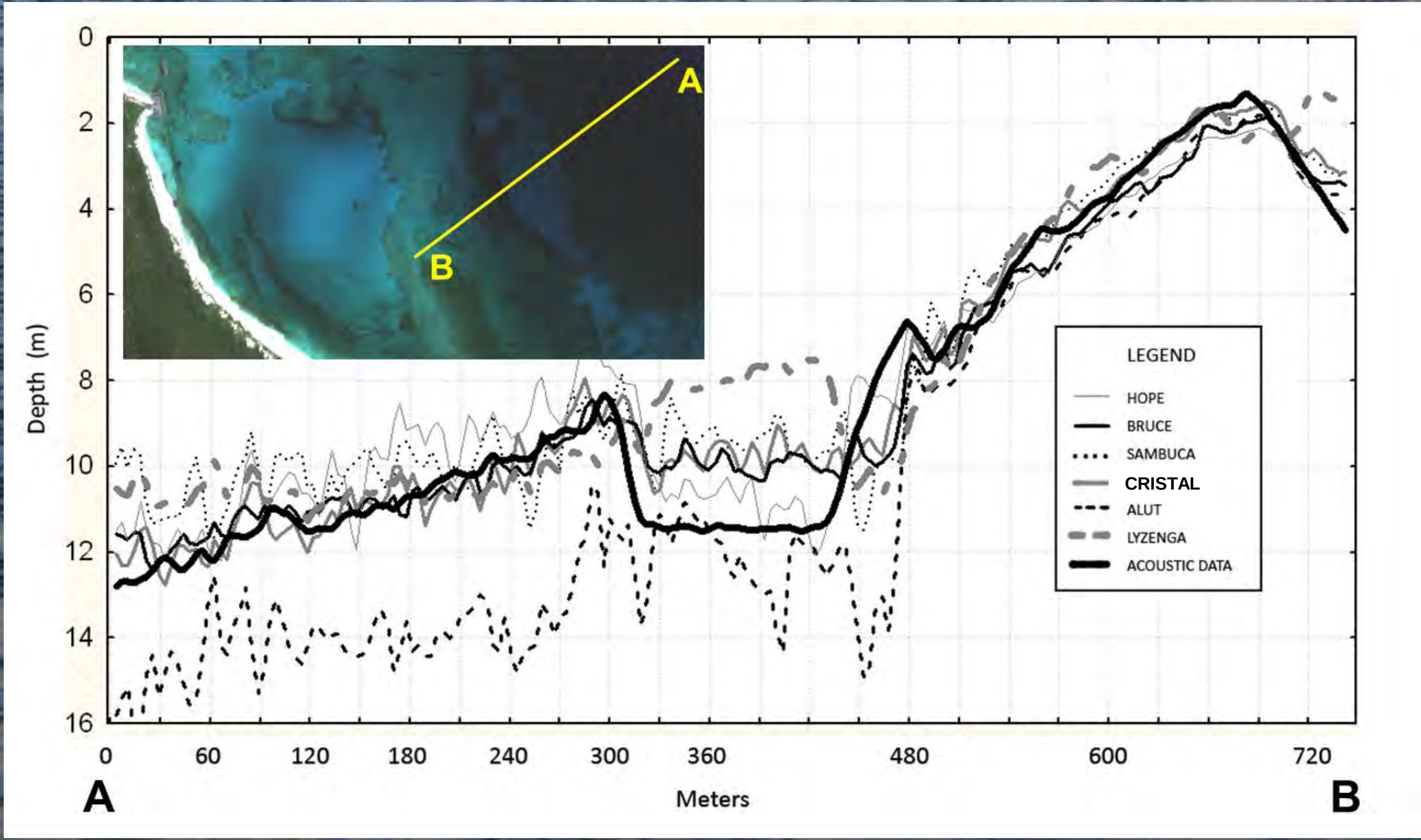
Comparison of Algorithms

preprocessing time / image processing time / pixels per sec



Bahamas Image. From Dekker et al., *Limnol Ocean. Methods*, 2011

Comparison of Algorithms



Bahamas Image. From Dekker et al., *Limnol Ocean. Methods*, 2011

Other Issues

- What is the best metric for spectrum matching?
- What is the best metric for quantifying results?
- How to do glint and whitecap removal?
- Constrained inversions

- How to do atmospheric correction (next lecture)

Computational Issues:
Metrics for Spectrum Matching

Name	Key word	Description	Quantity Computed
Euclidean	EUC	sum of squared differences	$\sum_{k=1}^{N_w} w(\lambda_k)\,[R_{rs}^{im}(\lambda_k) - R_{rs}^{db}(\lambda_k)]^2$
Manhattan	MAN	sum of absolute differences	$\sum_{k=1}^{N_w} w(\lambda_k)\,\lvert R_{rs}^{im}(\lambda_k) - R_{rs}^{db}(\lambda_k)\rvert$
Chebyshev	CHE	largest absolute difference	$\max_k\; w(\lambda_k)\,\lvert R_{rs}^{im}(\lambda_k) - R_{rs}^{db}(\lambda_k)\rvert$
Canberra	CAN	sum of absolute differences divided by sum of values	$\sum_{k=1}^{N_w} w(\lambda_k) \frac{\lvert R_{rs}^{im}(\lambda_k) - R_{rs}^{db}(\lambda_k)\rvert}{[R_{rs}^{im}(\lambda_k) + R_{rs}^{db}(\lambda_k)]}$
Bray-Curtis	BRA	sum of absolute differences divided by sum of absolute values	$\frac{\sum_{k=1}^{N_w} w(\lambda_k)\,\lvert R_{rs}^{im}(\lambda_k) - R_{rs}^{db}(\lambda_k)\rvert}{\sum_{k=1}^{N_w} w(\lambda_k)\,[R_{rs}^{im}(\lambda_k) + R_{rs}^{db}(\lambda_k)]}$
Spectral Angle	COS	cosine of the angle between the spectra	$\frac{\sum_{k=1}^{N_w} R_{rs}^{im}(\lambda_k)\,R_{rs}^{db}(\lambda_k)}{\left\{\sum_{k=1}^{N_w} [R_{rs}^{im}(\lambda_k)]^2 \cdot \sum_{k=1}^{N_w} [R_{rs}^{db}(\lambda_k)]^2\right\}^{1/2}}$
Correlation Coefficient	COR	cosine of the angle between the spectra after the spectra are centered on their means	$\frac{\sum_{k=1}^{N_w} [R_{rs}^{im}(\lambda_k) - \bar{R}_{rs}^{im}]\,[R_{rs}^{db}(\lambda_k) - \bar{R}_{rs}^{db}]}{\left\{\sum_{k=1}^{N_w} [R_{rs}^{im}(\lambda_k) - \bar{R}_{rs}^{im}]^2 \cdot \sum_{k=1}^{N_w} [R_{rs}^{db}(\lambda_k) - \bar{R}_{rs}^{db}]^2\right\}^{1/2}}$

CRISTAL Users' Guide Table 3.1

There is no unique way to say which two spectra are "closest".

The simple Euclidean and Manhattan metrics run the fastest and usually give the best results.

Spectral angle and correlational metrics run slowly and often give poor results because they discard the magnitude information (they compare only the spectral shapes), but are less sensitive to bad atmospheric correction.

Computational Issues:
Metrics for Validation of Retrievals

Name	Description	Quantity Computed
pct diff	average signed relative difference in retrieved vs true depth, in per cent	$\frac{100}{N_t}\sum_i^{N_t}(z_i^r - z_i^t)/z_i^r$
z diff	average signed difference in retrieved vs true depth, in meters	$\bar{z}_{\text{dif}} = \frac{1}{N_t}\sum_i^{N_t}(z_i^r - z_i^t)$
z sd	standard deviation between retrieved and true depths, in meters	$\left[\frac{1}{N_t - 1}\sum_i^{N_t}(z_i^r - z_i^t - \bar{z}_{\text{dif}})^2\right]^{1/2}$
r^2	square of linear correlation coefficient	$\dfrac{\left(N_t\sum_i^{N_t}(z_i^r z_i^t) - (\sum_i^{N_t} z_i^r)(\sum_i^{N_t} z_i^t)\right)^2}{\left(N_t\sum_i^{N_t}(z_i^r)^2 - (\sum_i^{N_t} z_i^r)^2\right)\left(N_t\sum_i^{N_t}(z_i^t)^2 - (\sum_i^{N_t} z_i^t)^2\right)}$
pct ± 1 m	percent of pixels with a retrieved depth within ± 1 m of the true depth	percent of pixels with $\lvert z_i^r - z_i^t\rvert \le 1$ m
pct ± 25 %	percent of pixels with a retrieved depth within ± 25 % of the true depth	percent of pixels with $\lvert(z_i^r - z_i^t)/z_i^t\rvert \le 0.25$

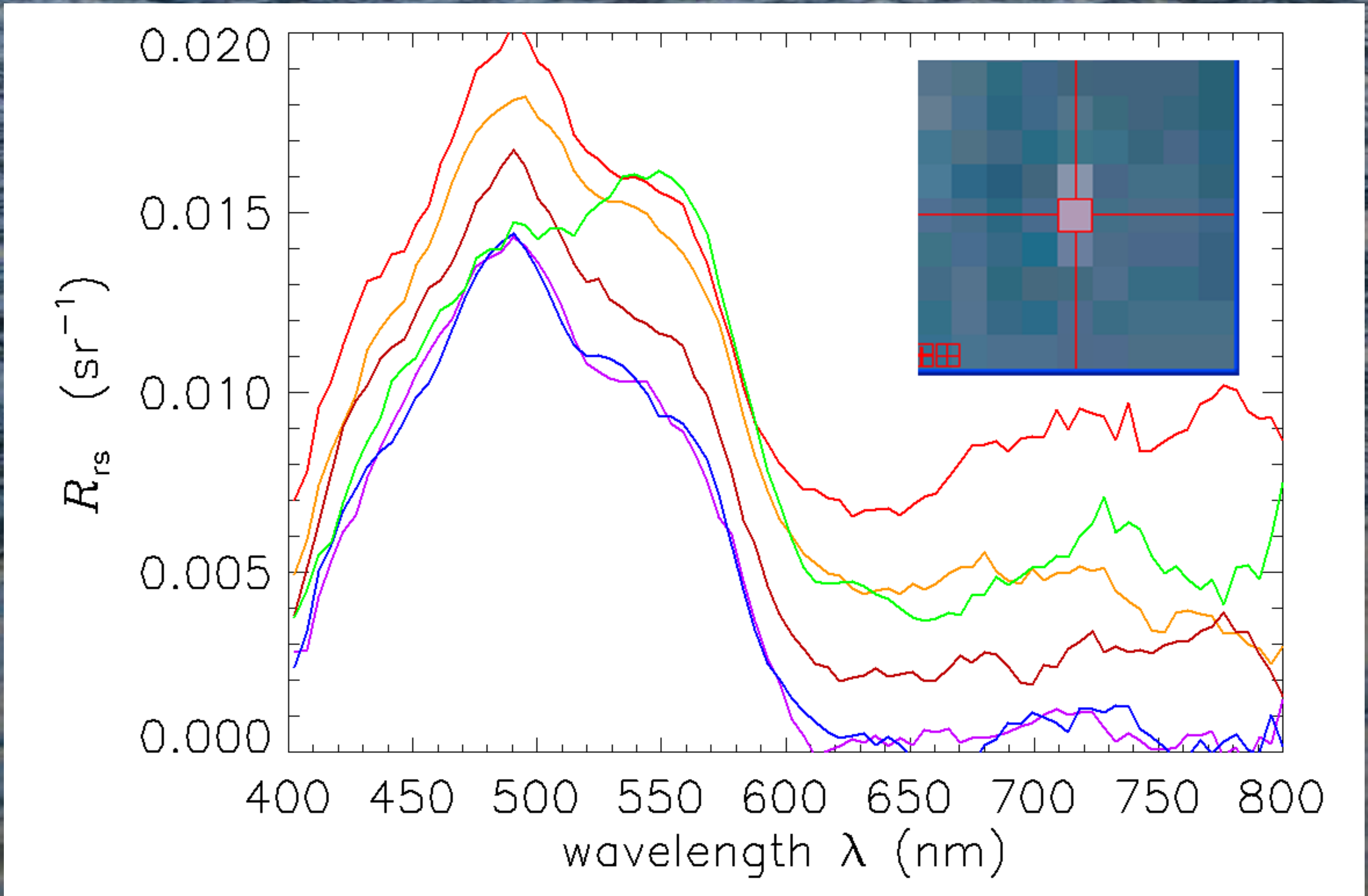
There is no unique way to say which retrieval is "best".

What is "best" often depends on the application.

CRISTAL Users' Guide Table 3.1

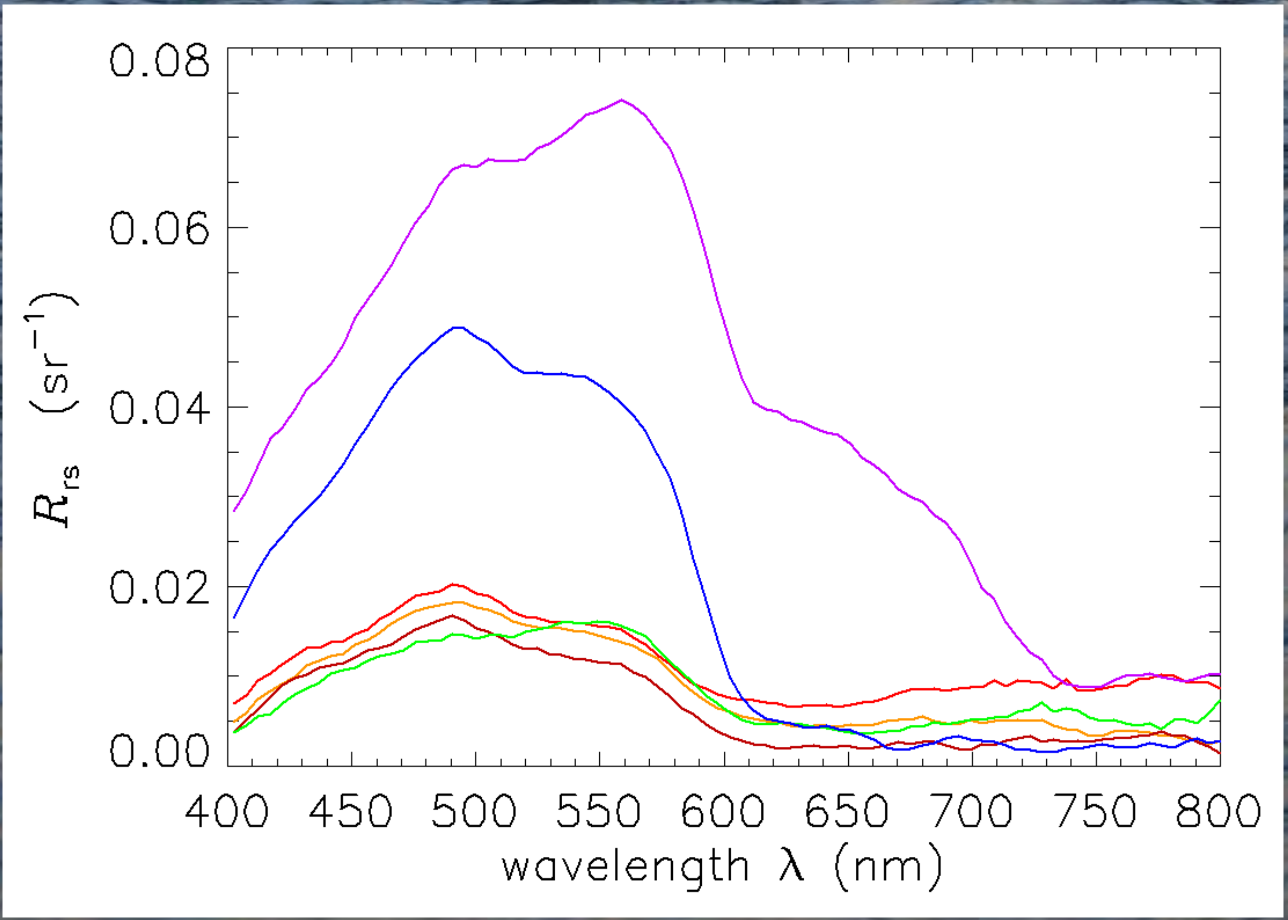
Glint and Whitecap Removal

Sun glint can usually be avoided, but background sky glint is always present. Whitecaps and clouds may be present. All raise the spectrum magnitude at all wavelengths.



Glint and Whitecap Removal

Glint removal algorithms for deep water look at the magnitude of R_{rs} at NIR wavelengths, and flag if too high. However, uncontaminated shallow-water spectra can also be high because of bottom reflectance.



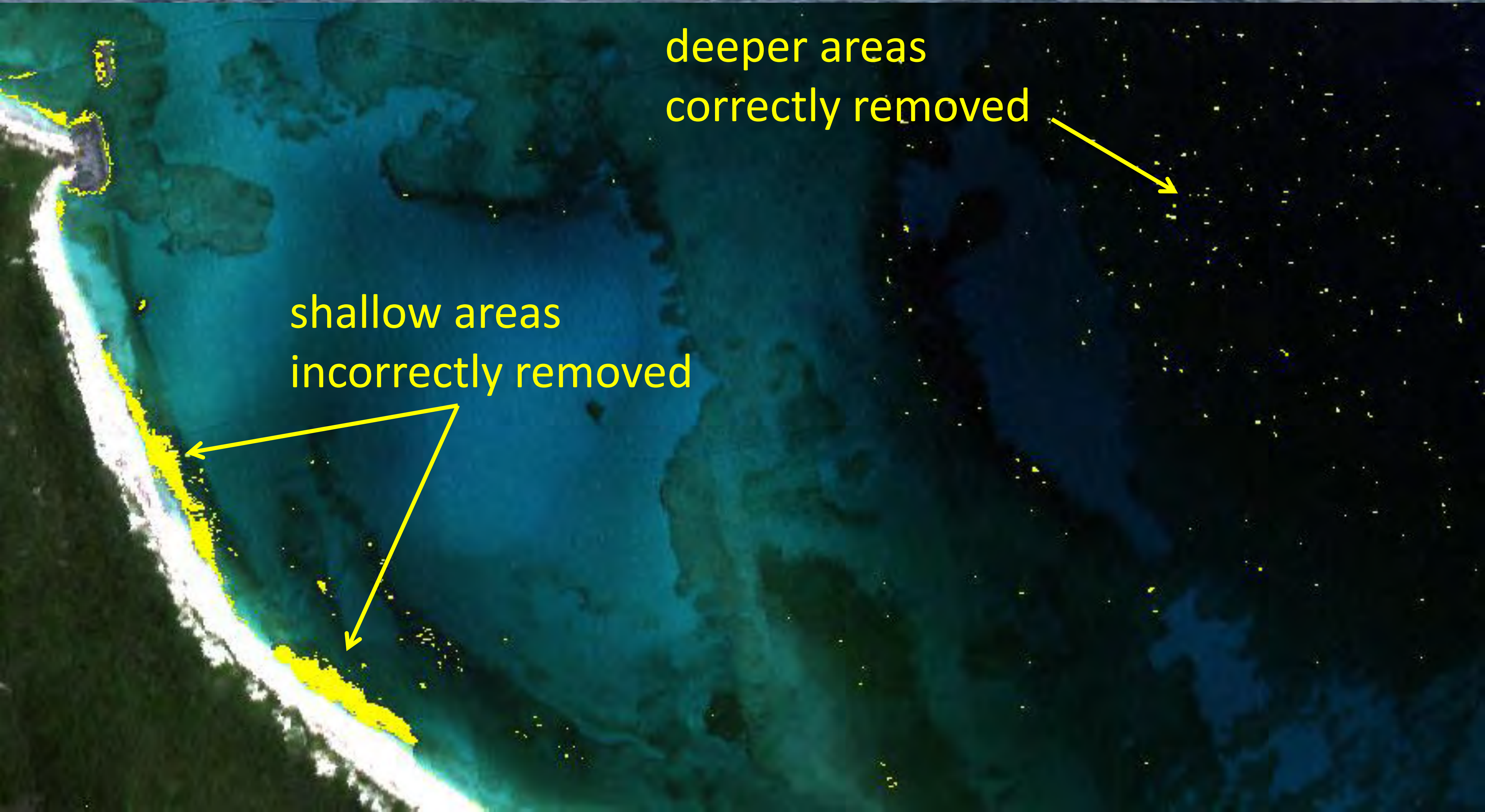
glint-contaminated deep water (red, orange, brown)

uncontaminated shallow-water, bright-bottom (purple, blue)

uncontaminated shallow-water dark bottom (green)

Glint and Whitecap Removal

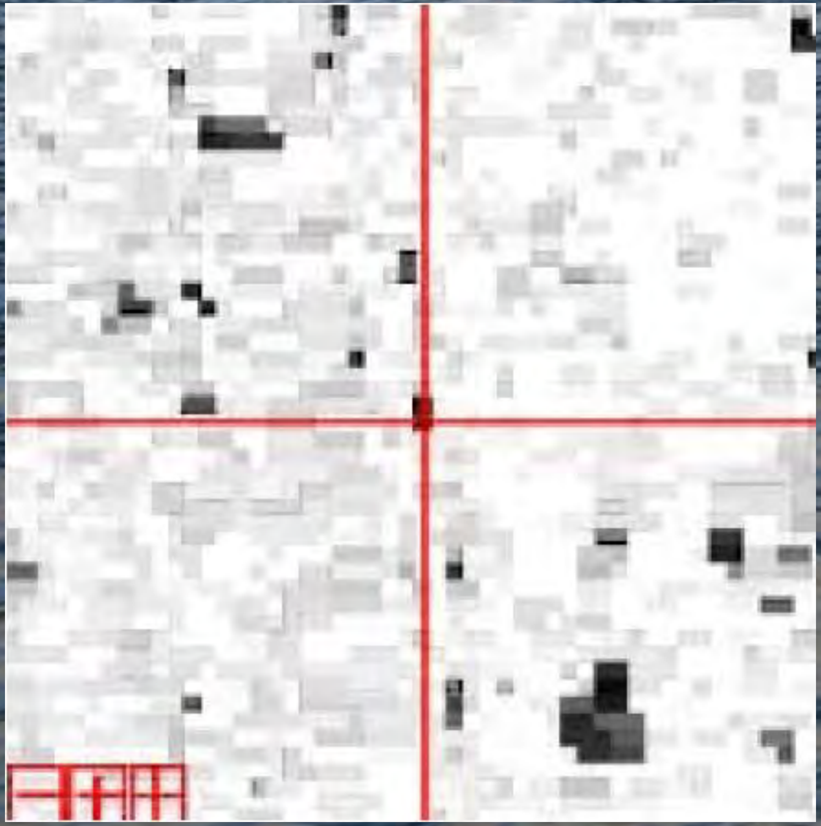
Correct and incorrect glint removal using a single-spectrum NIR threshold algorithm



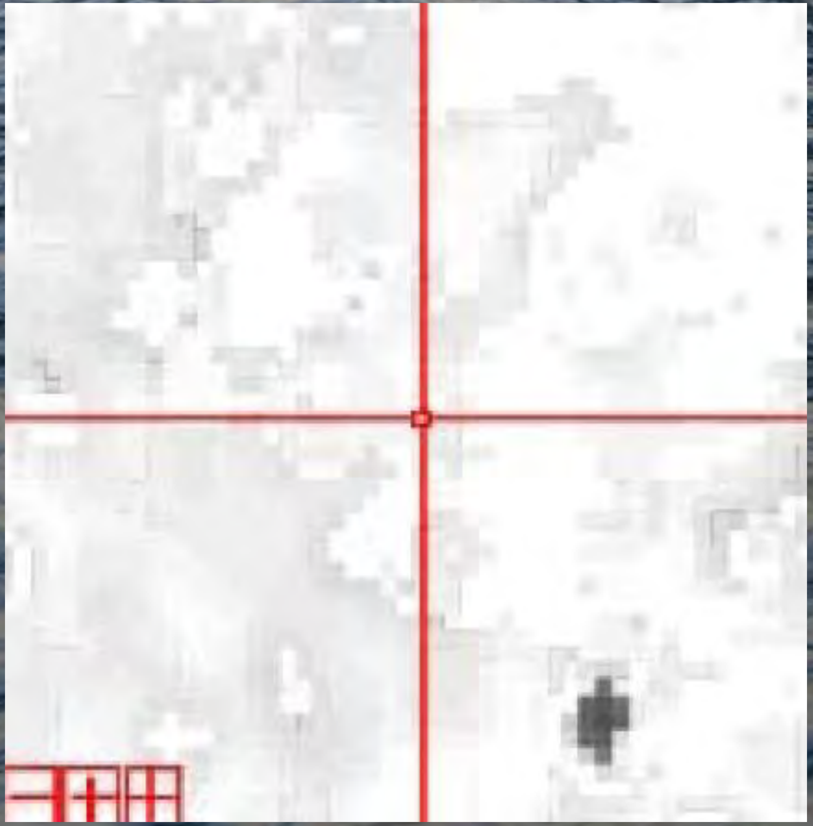
Glint and Whitecap Removal

Can use spatial filtering. Look at bright pixel and surrounding pixels. replace bright pixel with median or average of surrounding dark pixels. Can remove most glint, but degrades spatial resolution.

Original (dark is bright pixels)



Spatially filtered with 5x5 pixel block; discard brightest 2 spectra

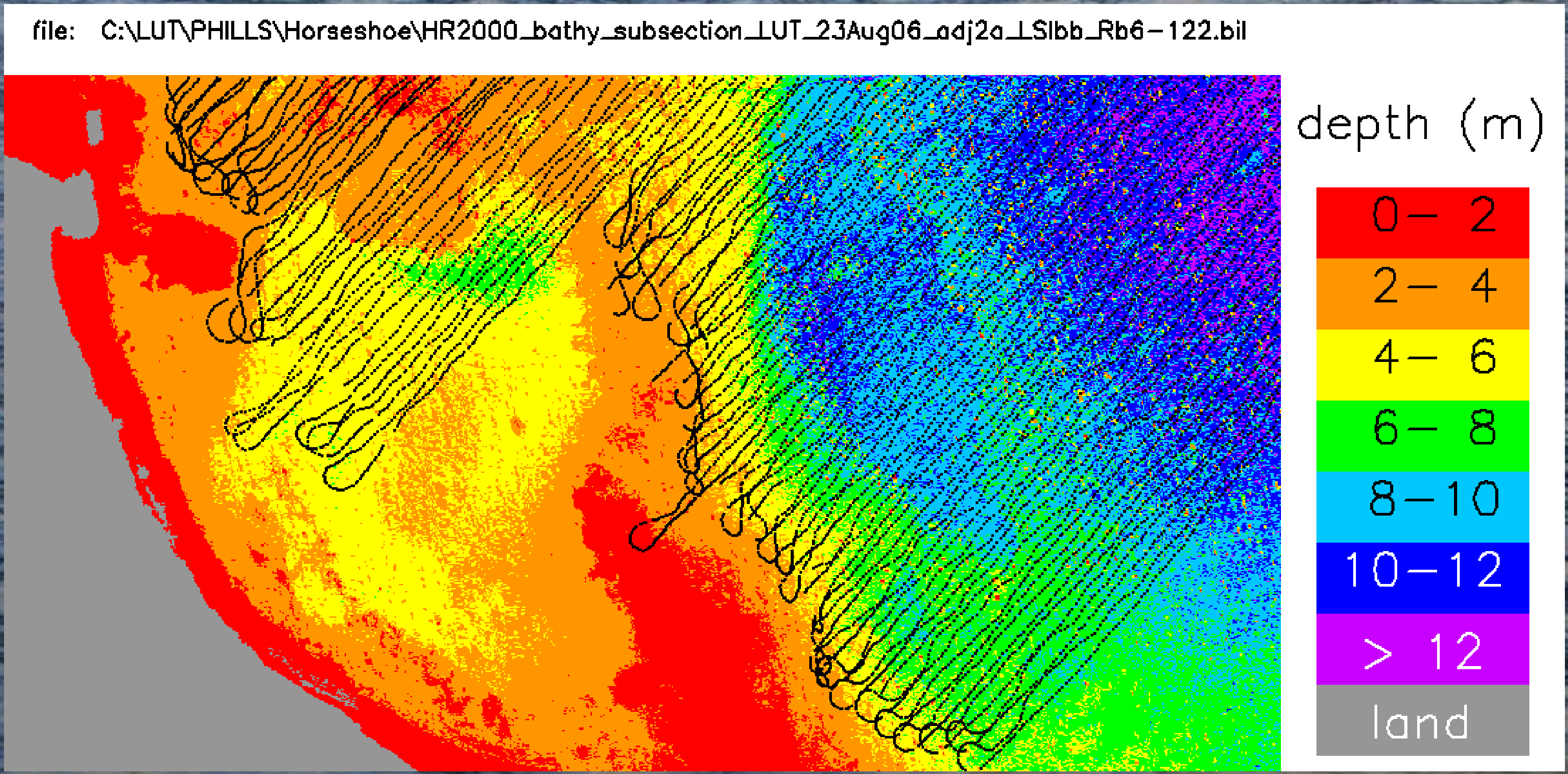


Constrained Inversions

Usually do not know anything about the imaged area, so must do simultaneous retrieval of depth, bottom reflectance, and water IOPs.

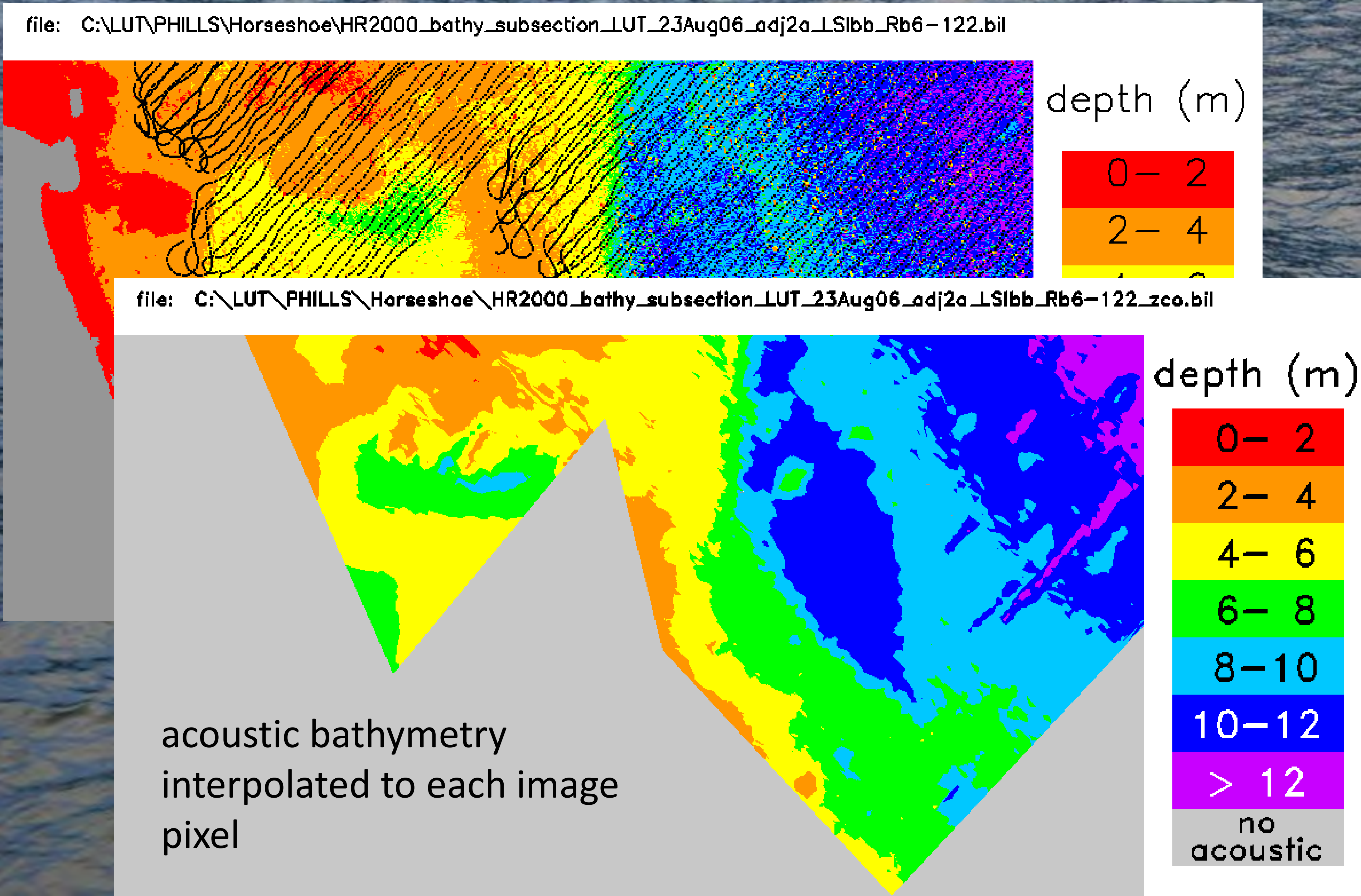
However, if some information is known (e.g., depth from acoustics or a bathymetric lidar, or IOPs from measurement), we can make use of that information and do a constrained inversion. This adds information to the inversion, and should improve the retrievals of the remaining unknowns.

Depth-constrained Inversions



acoustic bathymetry for Bahamas image

Depth-constrained Inversions



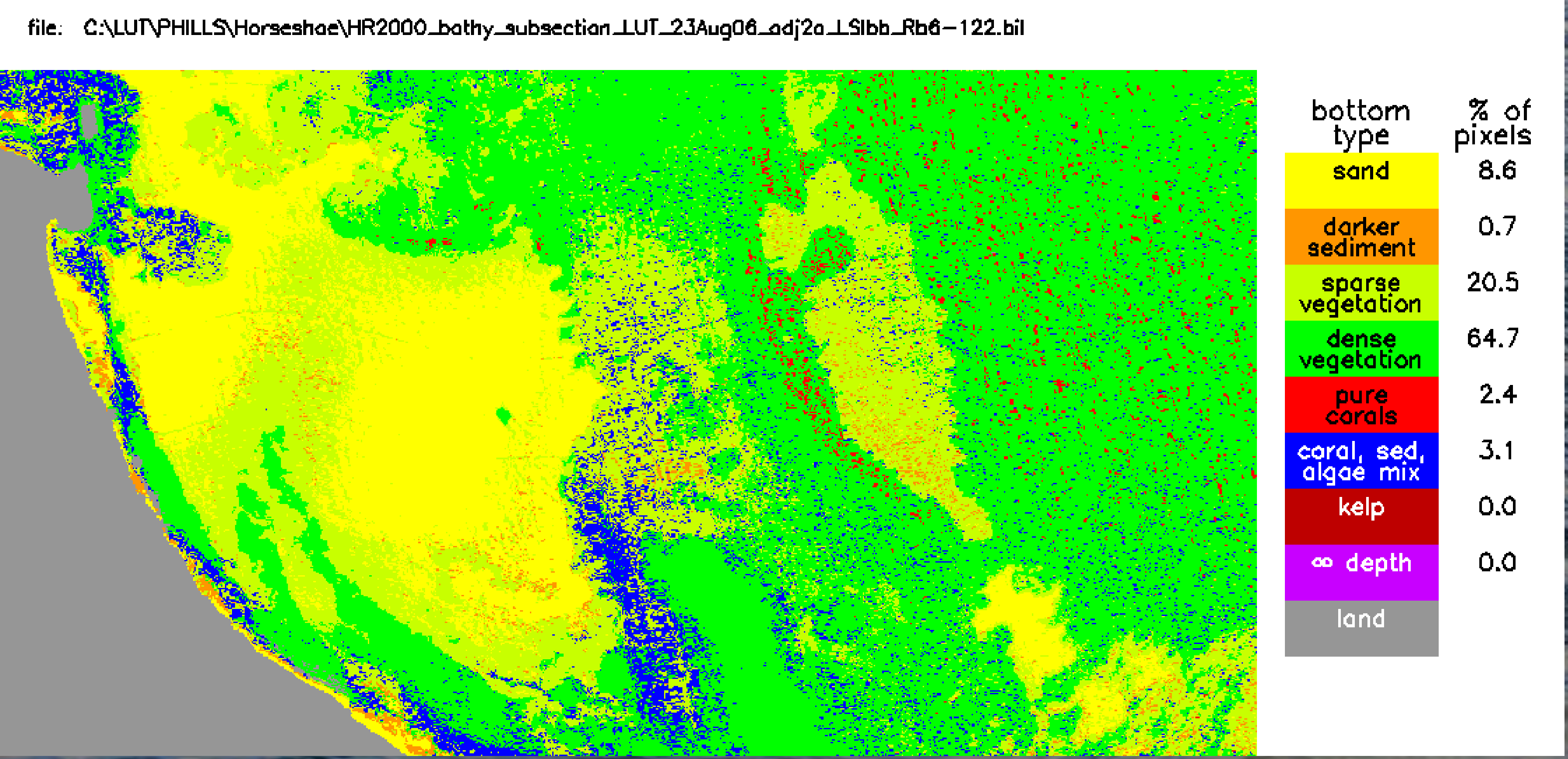
acoustic bathymetry interpolated to each image pixel

Depth-constrained Inversions

Now consider the depth known at each pixel where acoustic info was available for interpolation.

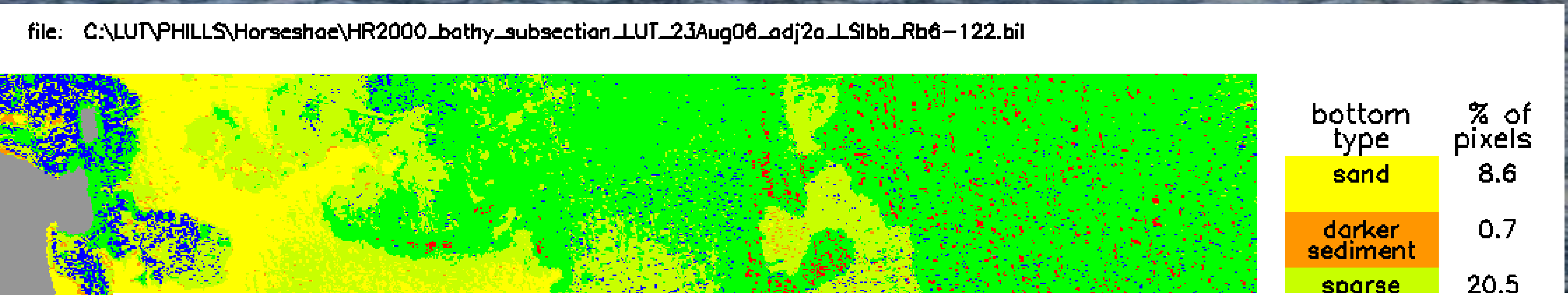
Search the database at each pixel only for spectra that correspond to a depth close to the known depth. Retrieve just bottom reflectance and IOPs.

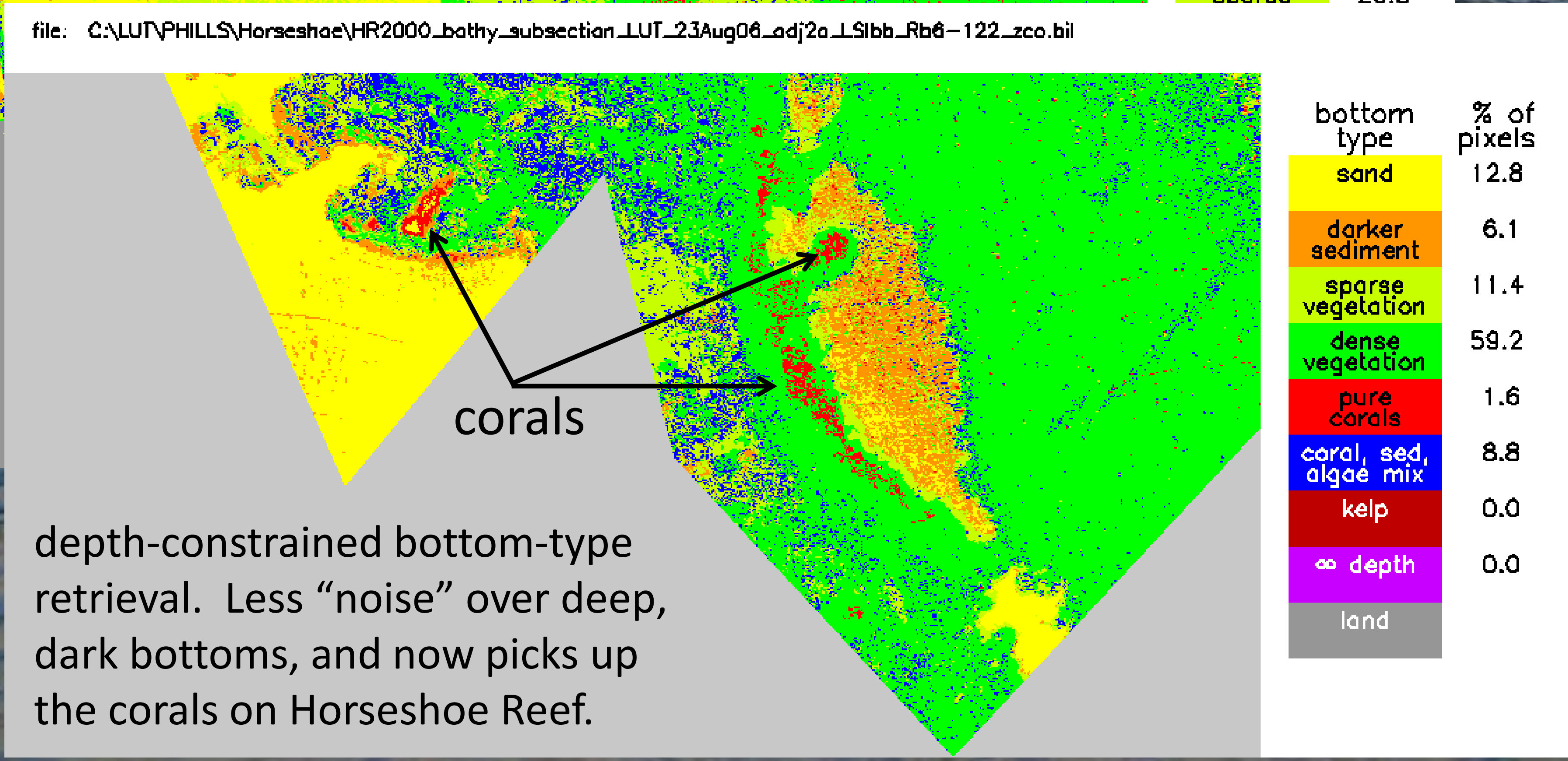
Depth-constrained Inversions



unconstrained bottom-type retrieval. Overall pretty good, but lots of "noise" over deep, dark bottoms, probably due to glint. Not sure what is a coral and what isn't.

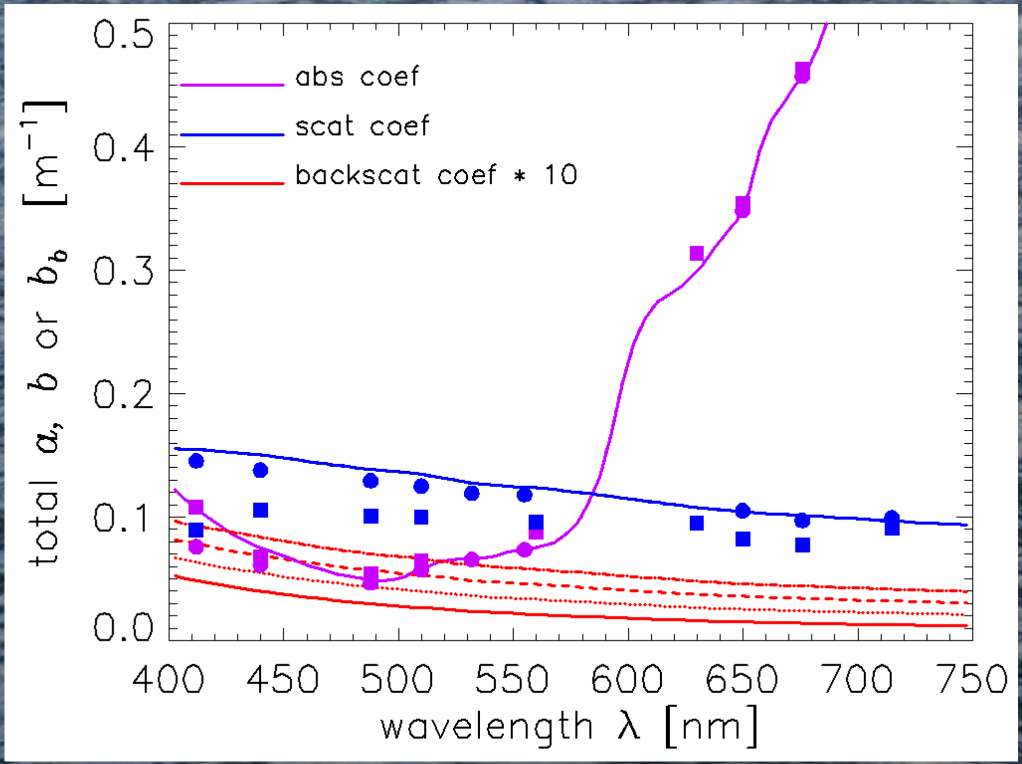
Depth-constrained Inversions





depth-constrained bottom-type retrieval. Less "noise" over deep, dark bottoms, and now picks up the corals on Horseshoe Reef.

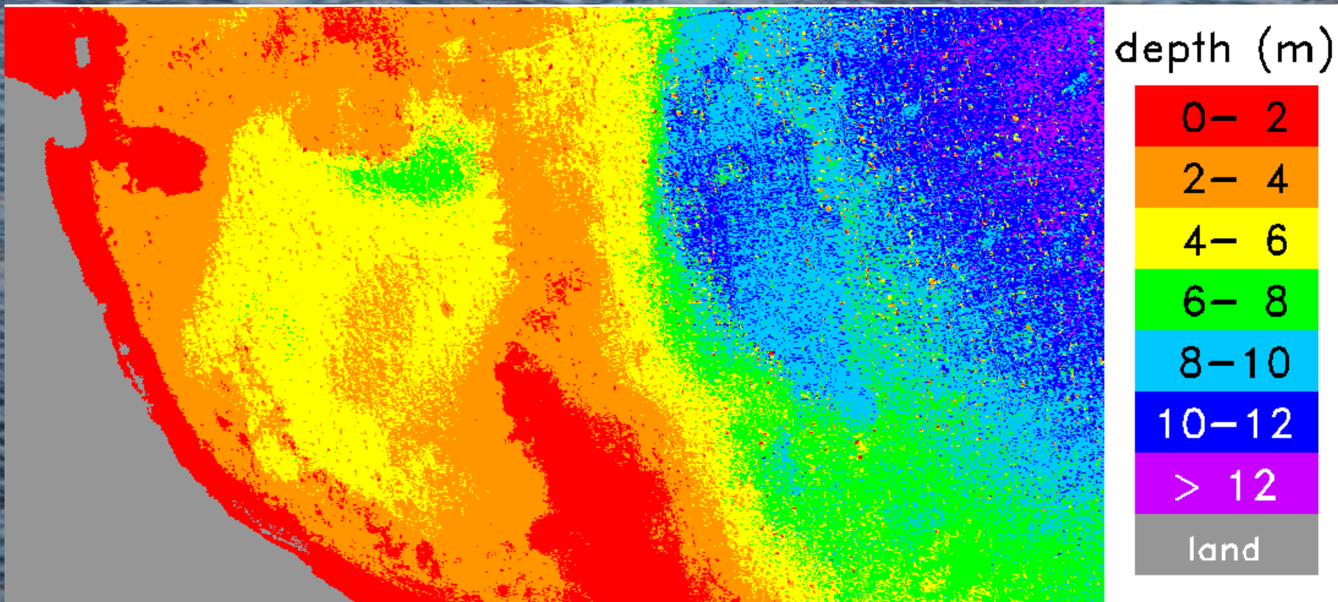
IOP-constrained Inversions



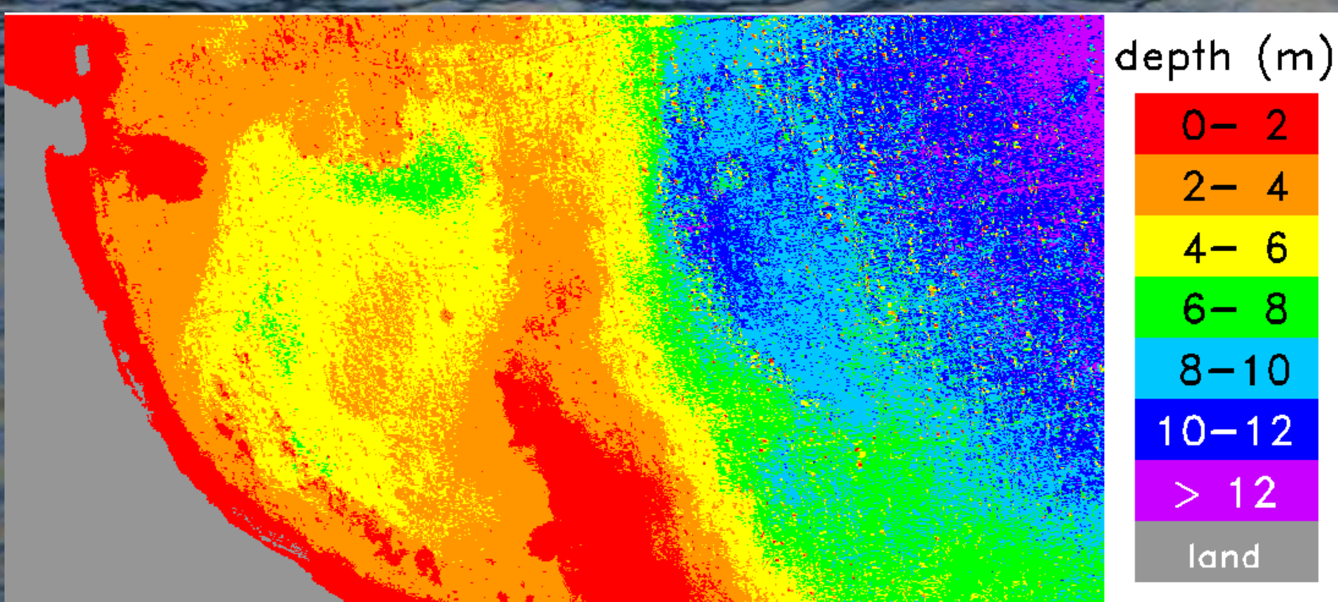
dots and squares: two sets of ac9 data from the Horseshoe Reef area.
lines: similar *a* and *b* from the LUT IOP database; the four backscatter curves have particle backscatter fractions of 0.01, 0.02, 0.03, and 0.04

To constrain the IOPs, assume that *a* and *b* are constant over the image area (probably wrong: CDOM decreases as go off shore, and resuspended sediment likely higher near shore)

IOP-constrained Inversions



Unconstrained inversion for depth



IOP-constrained inversion for depth. Not much different because the unconstrained depth retrieval was already very good.

Computer Processing Times

Even if constrained inversions do not greatly improve the remaining retrievals because the unconstrained inversion were already good, constraining the retrieval does greatly speed up the image processing time because less of the R_{rs} database needs to be searched for each pixel.

For the Horseshoe Reef image (on a 2 GHz PC):

unconstrained inversion:	71 minutes (>10^{10} R_{rs} comparisons)
depth-constrained inversion:	25 min
IOP-constrained inversion:	27 min
depth- and IOP-constrained inversion:	3.5 min

Kayak Camp, Lofoten Islands, Norway, June 2010