2015 Summer Course

on Optical Oceanography and Ocean Color Remote Sensing
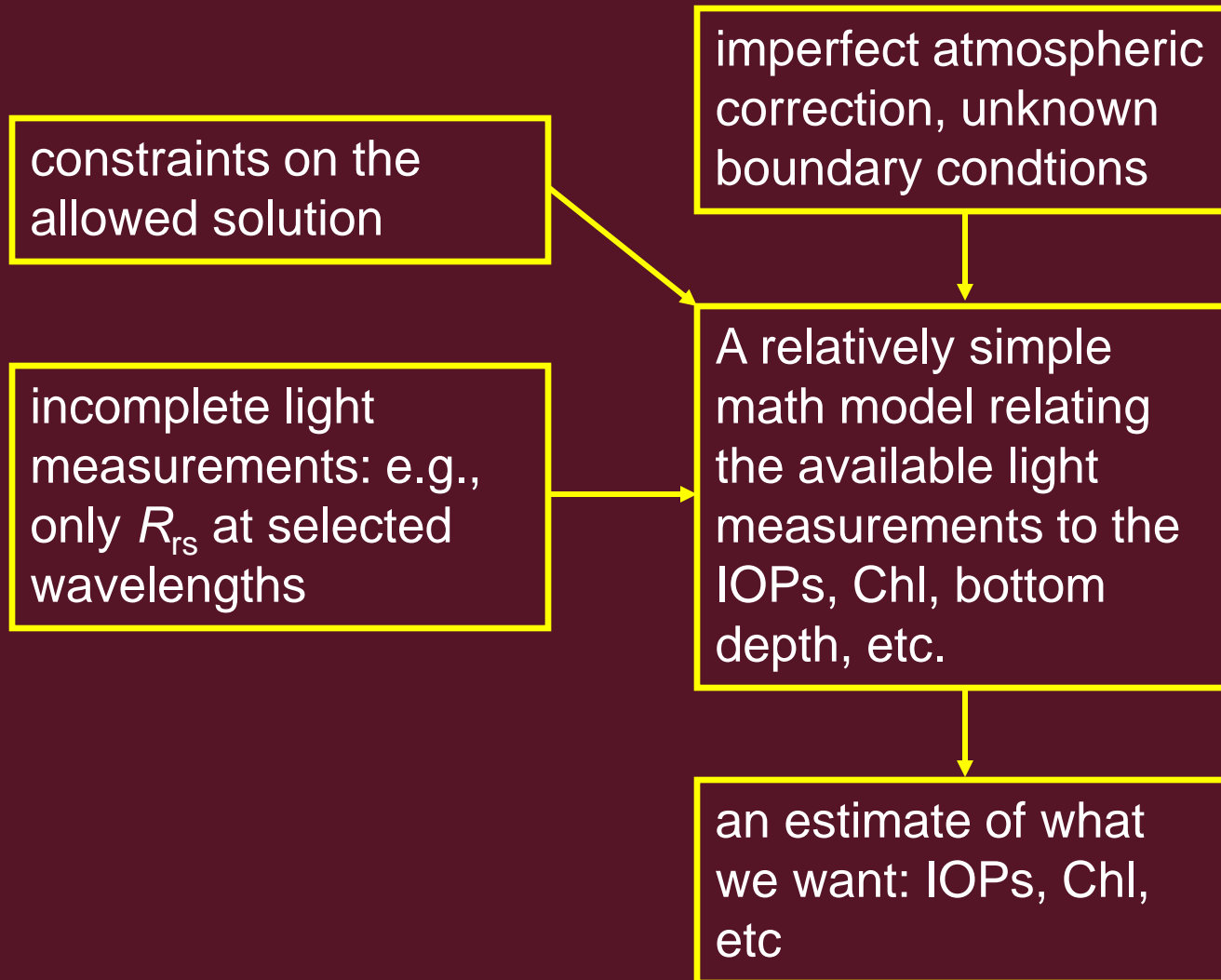
Curtis Mobley

Statistical Methods for Remote Sensing

Delivered at the Darling Marine Center,
University of Maine
July 2015

Recall the Remote-Sensing Inverse Problem

constraints on the allowed solution

imperfect atmospheric correction, unknown boundary condtions

incomplete light measurements: e.g., only R_{rs} at selected wavelengths

A relatively simple math model relating the available light measurements to the IOPs, Chl, bottom depth, etc.

an estimate of what we want: IOPs, Chl, etc

Statistical Inverse Models

One family of simple math models relating the available measurements to what we want is *statistical* models.

These models are essentially just correlational models obtained from inspection of data sets containing both the inputs inputs (R_{rs}) and outputs (Chl, water depth, etc). The models are not necessarily based on any physical insight as to why the correlation exists.

The general forms of the models contain unknown parameters (proportionality constants, weighting functions, fitting coefficients). The parameter values are determined by *forcing the model to fit data containing both the inputs and outputs*. That is, the parameter values give the statistical best-fit of the model to the data, hence the name "statistical" or "empirical" models.

After the parameters have been determined using known inputs and outputs, the model with the same parameter values can be applied to new input data, to obtain new outputs.
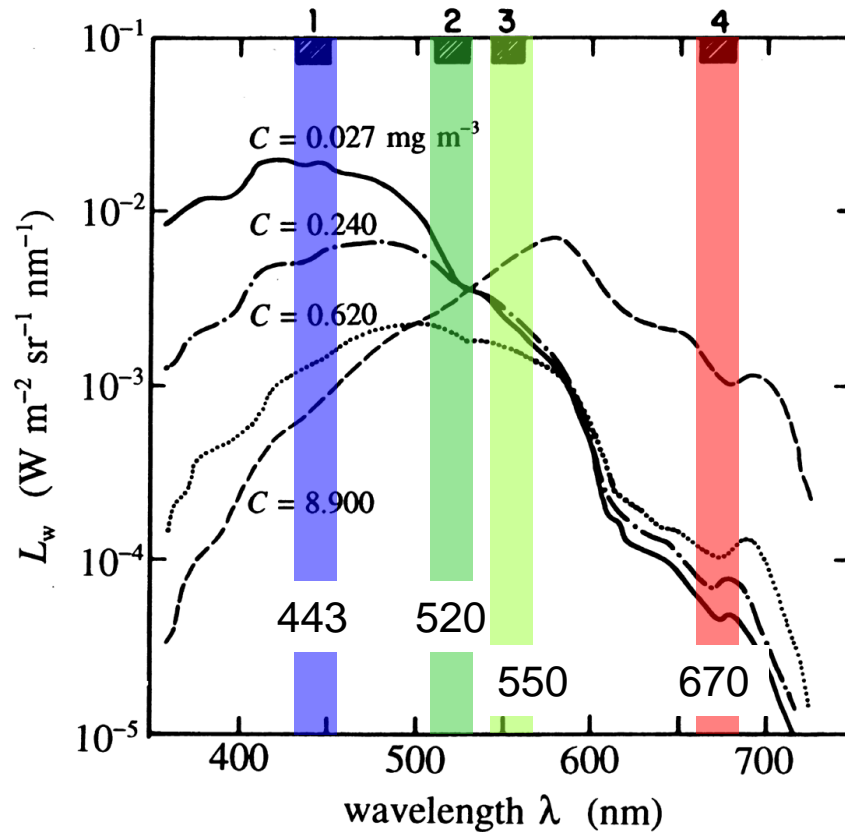
Statistical methods are how ocean color remote sensing got started 40 years ago
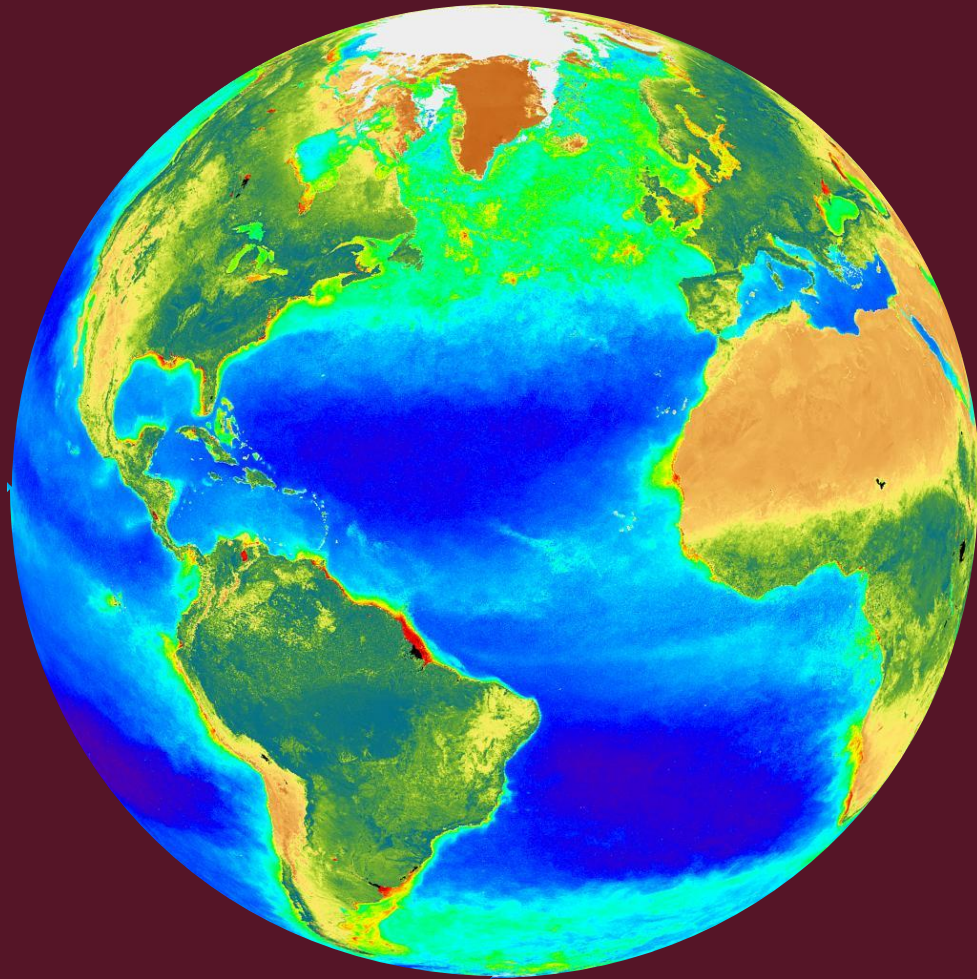
Two examples:
- band-ratio algorithms
- neural networks

Where It All Started

Fig. 10.1. Water-leaving radiances L_w as a function of wavelength for four chlorophyll concentrations C, in case 1 waters. The shaded regions labeled 1-4 indicate the detector bandwidths of the CZCS sensor. [redrawn from Gordon, *et al.*, (1985), by permission]

The seminal idea of ocean color remote sensing: *Chl* concentration and water-leaving radiance are correlated.

$R(1,3) = L_w(\lambda_1=443)/L_w(\lambda_3=550)$ vs *Chl*
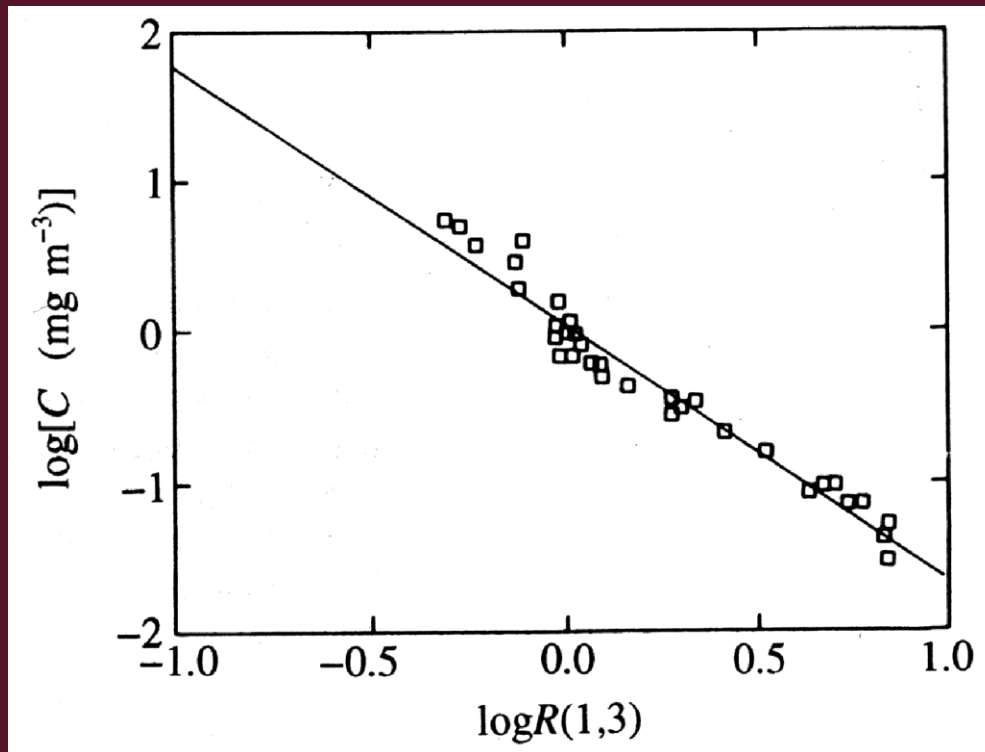
Note: only 33 data points were initially available!

This suggests the band-ratio model:

$\log_{10}(Chl) = C_1 + C_2\log_{10}[L_w(443)/L_w(550)]$

C_1 and C_2 are the unknown model parameters whose values are determined by a best fit to the data

CZCS Image
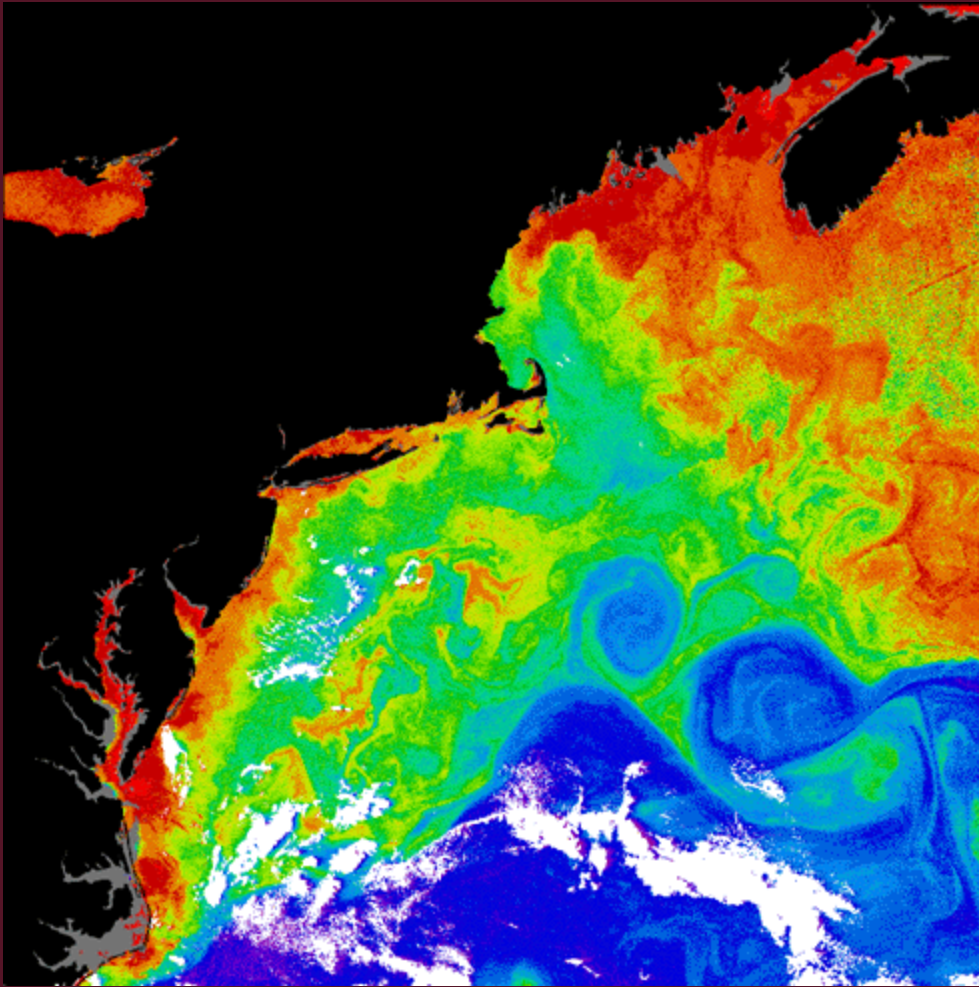
Coastal Zone Color Scanner (CZCS)

1978-1986

4 visible, 2 IR bands

66,000 images

revolutionized oceanography with very simple band ratio algorithms

Chl = 0.2 in blue to 30 in red

Examples of Recent Band-Ratio Algorithms

SeaWiFS OC4v4 for Chl:

$X = \log_{10}\{\max[R_{rs}(443)/R_{rs}(555), R_{rs}(490)/R_{rs}(555), R_{rs}(510)/R_{rs}(555)]\}$

$Chl = 10^{\wedge}(0.366 - 3.067X + 1.930X^2 + 0.649X^3 - 1.532X^4)$

MODIS for $K_d(490)$:

$X = L_w(488)/L_w(551)$

$K_d(490) = 0.016 + 0.156445X^{\wedge}(-1.5401)$

MODIS for $a_{CDOM}(400)$ and $a_{phy}(675)$:

$r_{15} = \log_{10}[R_{rs}(412)/R_{rs}(551)]$

$r_{25} = \log_{10}[R_{rs}(443)/R_{rs}(551)]$

$r_{35} = \log_{10}[R_{rs}(488)/R_{rs}(551)]$

$a_{CDOM}(400) = 1.5*10^{\wedge}(-1.147 + 1.963r_{15} - 1.01r_{15}^2 - 0.856r_{25} + 1.02r_{25}^2)$

$a_{phy}(675) = 0.328 [10^{\wedge}(-0.919 + 1.037r_{25} - 0.407r_{25}^2 -$
$3.531r_{35} + 1.702r_{35}^2 - 0.008)]$
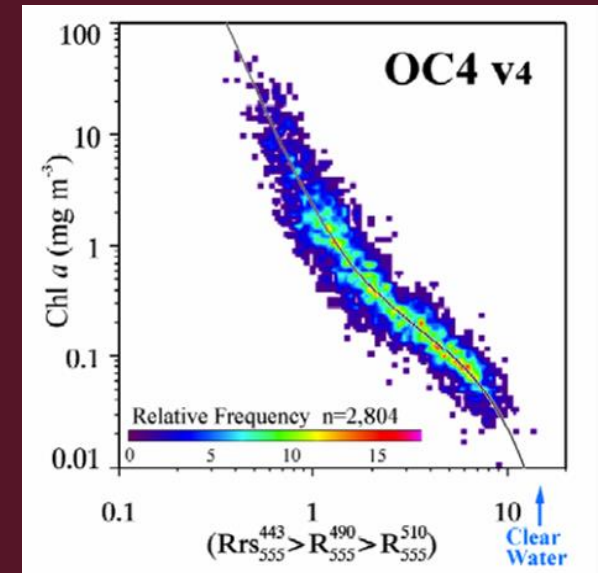
and so on, for dozens more….

A Fun Project

Use HydroLight to generate some R_{rs} spectra for various case 1 and case 2 IOPs. Then run these R_{rs} through various band-ratio algorithms to see how the retrieved values compare with each other and with what went into HydroLight. You can find more on the www.

Darieki and Stramski, *RSE*, 2004

MODIS Chl Case 1

MODIS K_d(490)

MODIS Chl Case 2

MODIS Chl Case 2

MODIS a_ϕ(675) & a_{CDOM}(400)

CZCS Chl

SeaWiFS Chl

Atmospheric Correction Effects

Band-ratio algorithms can be less sensitive to bad atmospheric correction than some other techniques such as spectrum matching
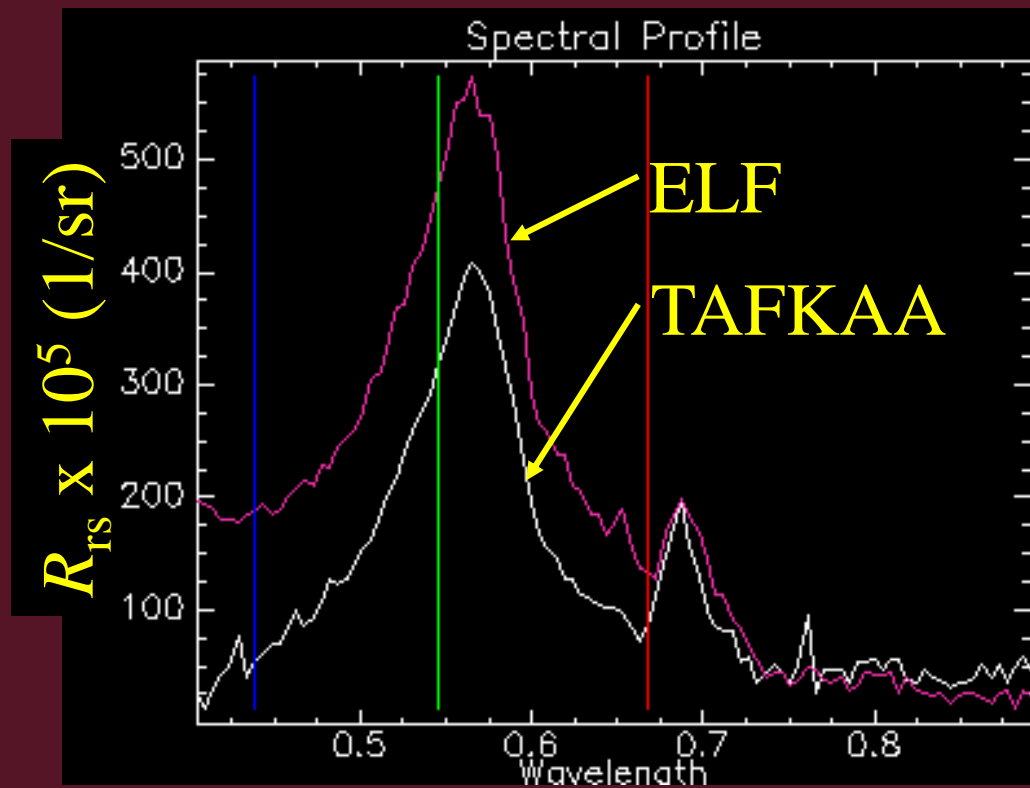
Nonuniqueness

Band-ratio algorithms are vulnerable to non-uniqueness problems because the R_{rs} ratioing throws out magnitude information that makes spectra unique. Every unique spectrum below has $R_{rs}(490)/R_{rs}(555) = 1.71 \pm 0.01$, which gives $Chl = 0.59$ mg/m^3 by the SeaWiFS OC2 algorithm; all of these spectra had Chl < 0.2 mg/m^3 (the spectra are influenced by bottom reflectance).
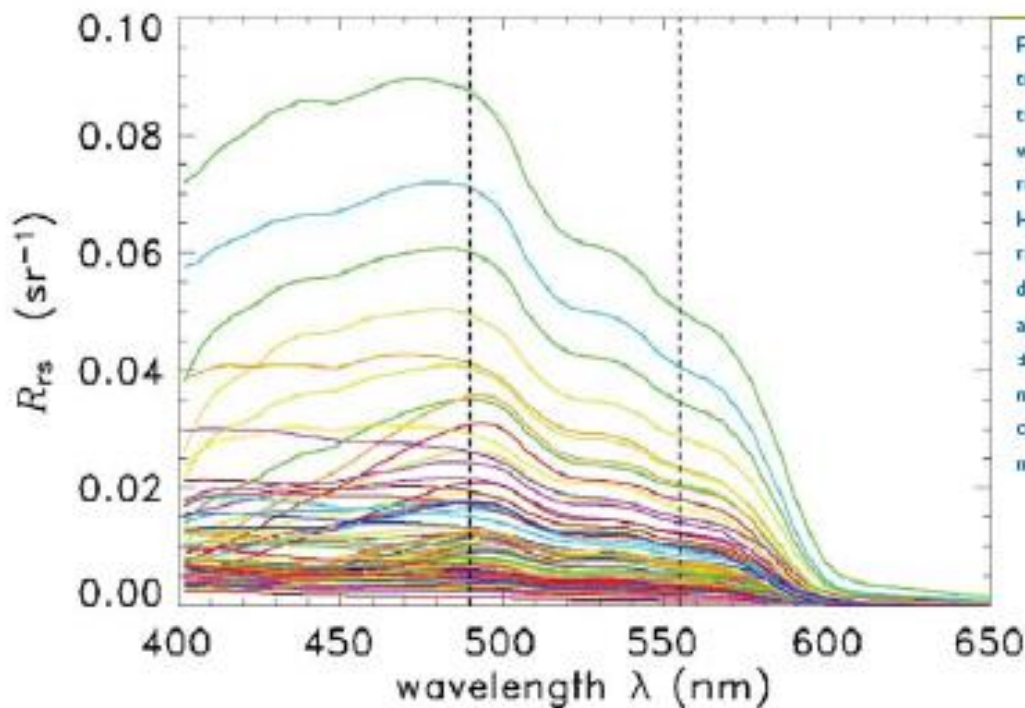
Figure 3. Chlorophyll concentration algorithms designed for multispectral instrumentation may not be useful for shallow, optically clear waters. Shown here are one hundred twenty two Hydrolight-generated remote sensing reflectance (R_{rs}) spectra for Bahamian waters using various combinations of nine different sets of IOPs, 32 different bottom reflectances, and 22 depths between 5.5 and 50 m. These spectra are clearly unique. However, every spectrum has nearly the same remote sensing reflectance wavelength ratio: $R_{rs}(490)/R_{rs}(555) = 1.71 \pm 0.01$ (490 and 555 nm are indicated by the vertical black dashed lines). If this ratio were applied to the commonly used SeaWiFS band-ratio algorithm (OC2; O'Reilly et al., 1998), it would give a chlorophyll concentration of 0.59 ± 0.01 mg Chl m^{-3}. In other words, the same chlorophyll concentration would be determined for all 122 spectra despite the fact that these simulated water bodies have IOPs corresponding to chlorophyll concentrations between 0.0 (pure water) and 0.2 mg Chl m^{-3}. The OC2 algorithm fails here because of bottom effects in optically clear waters.

Nonuniqueness

Dierssen et al. (*Limnol. Oceanogr*. 41(1), 444-455, 2003) developed a band-ratio algorithm for bottom depth in clear Bahamas waters:
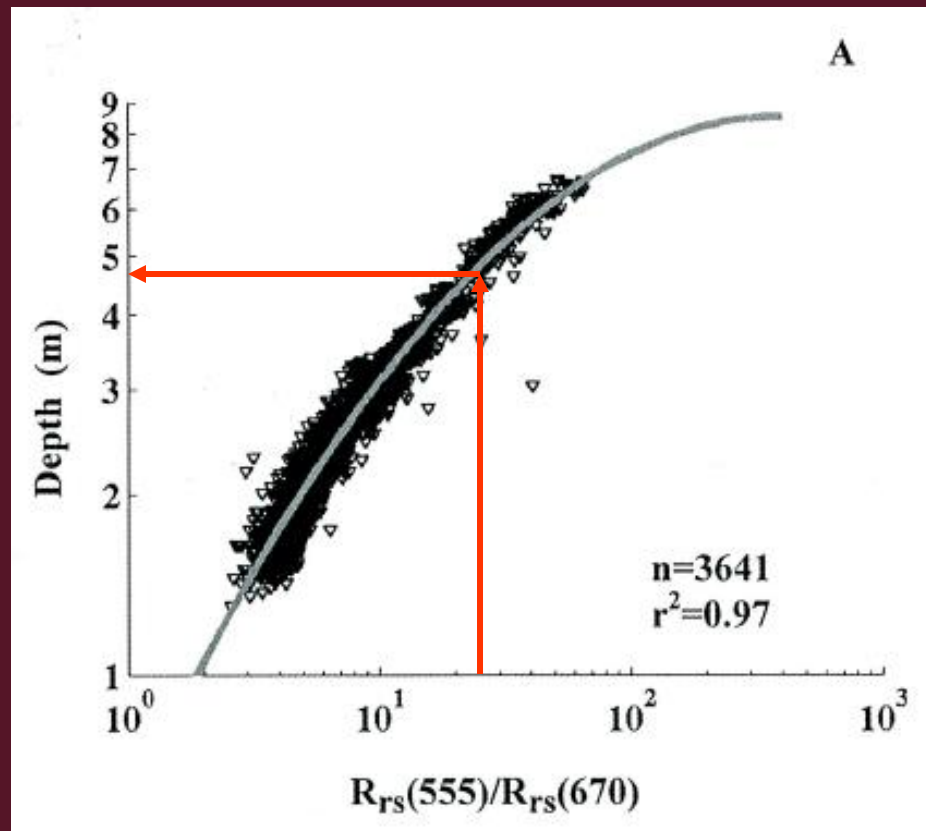
$$x = \log_{10}[R_{rs}(555)/R_{rs}(670)].$$

$$\log_{10}(z_b) = -0.1706\,x^2 + 0.8913\,x - 0.2316.$$

Nonuniqueness

sand   grass

line 220
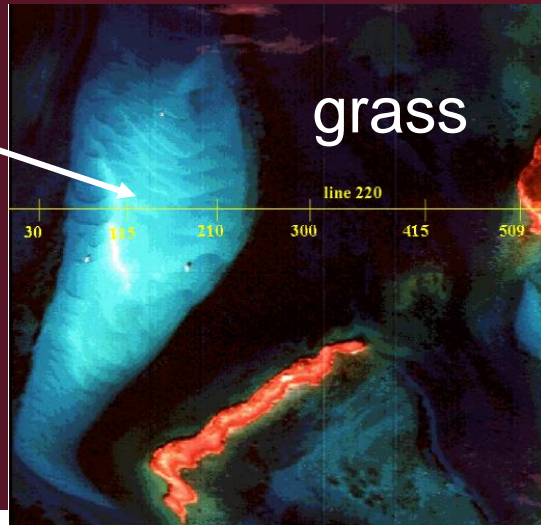
The Dierssen algorithm did OK over shallow sand bottoms, but totally failed over deeper sea grass bottoms. Why?

Dierssen et al zb model

OK   wrong

Algo1_RGS_5750_w1_UnS

z_b (m)

0 – 2
2 – 4
4 – 6
6 – 8
8 – 10
10 – 15
land
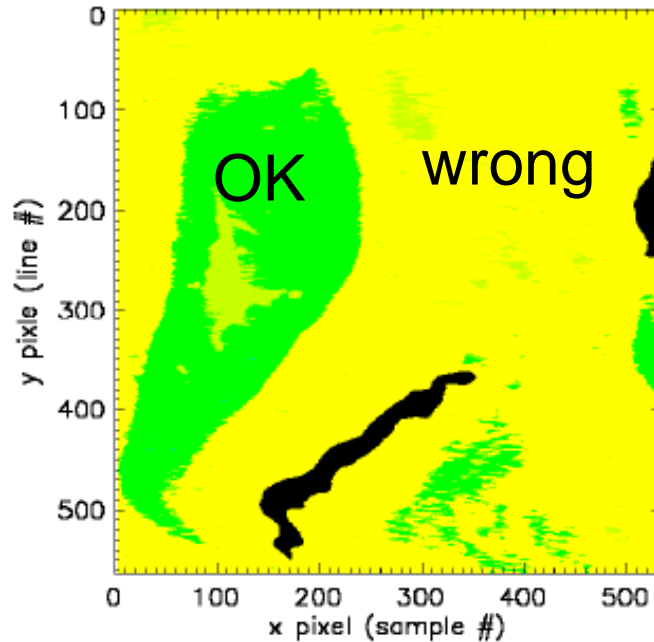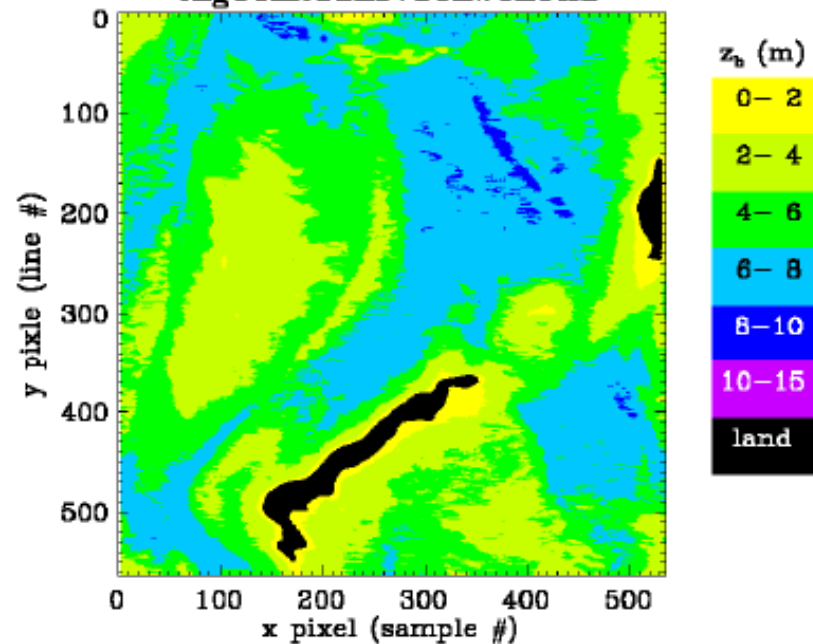
Nonuniqueness

HydroLight simulations of $R_{rs}(555)/R_{rs}(670)$ for two sets of IOPs and two different bottoms (sand and grass), as a function of bottom depth. Nonuniqueness for $z_b > 5$ m and grass bottom.
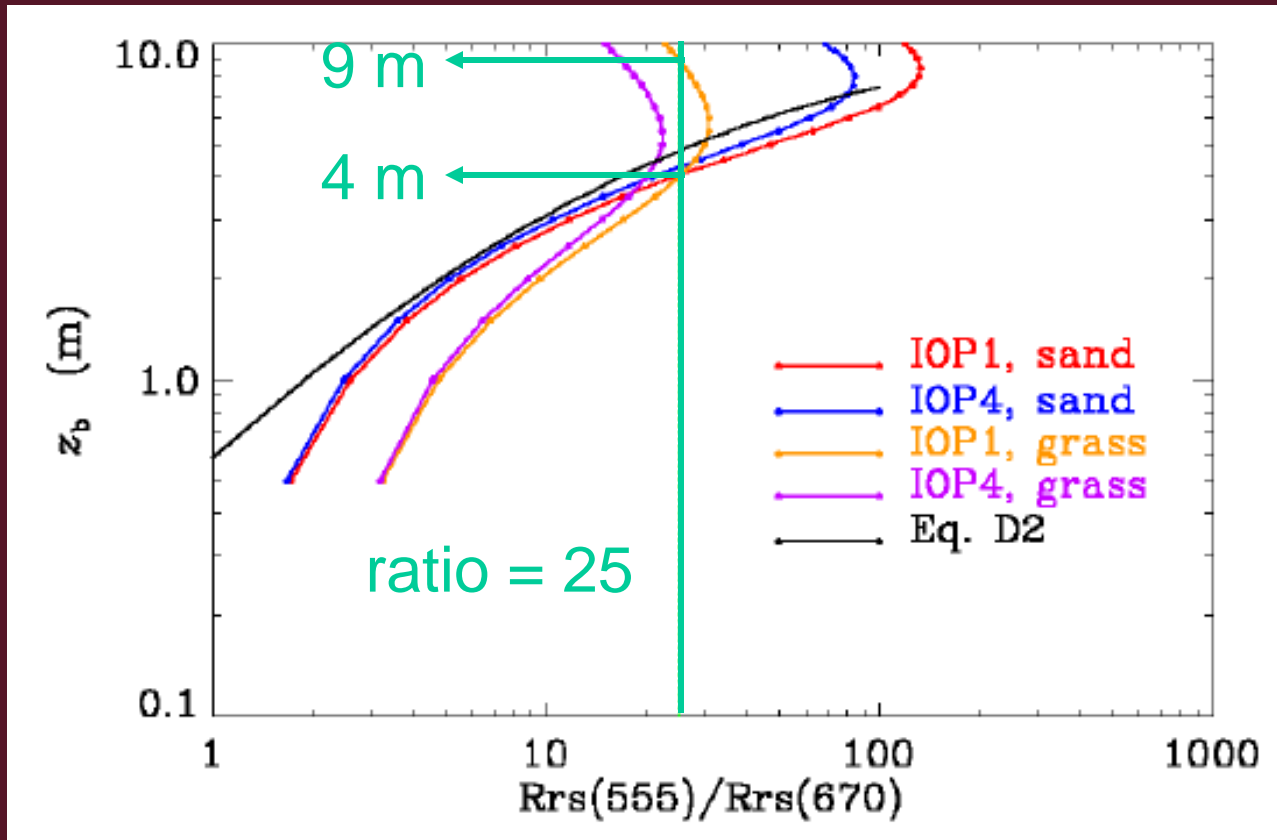
Nonuniqueness

The R_{rs} spectra for z_b = 4 and 9 m depth, grass bottom. Both spectra have $R_{rs}(555)/R_{rs}(670)$ = 25 ± 0.1. The Dierssen model gives z_b = 4.8 m.
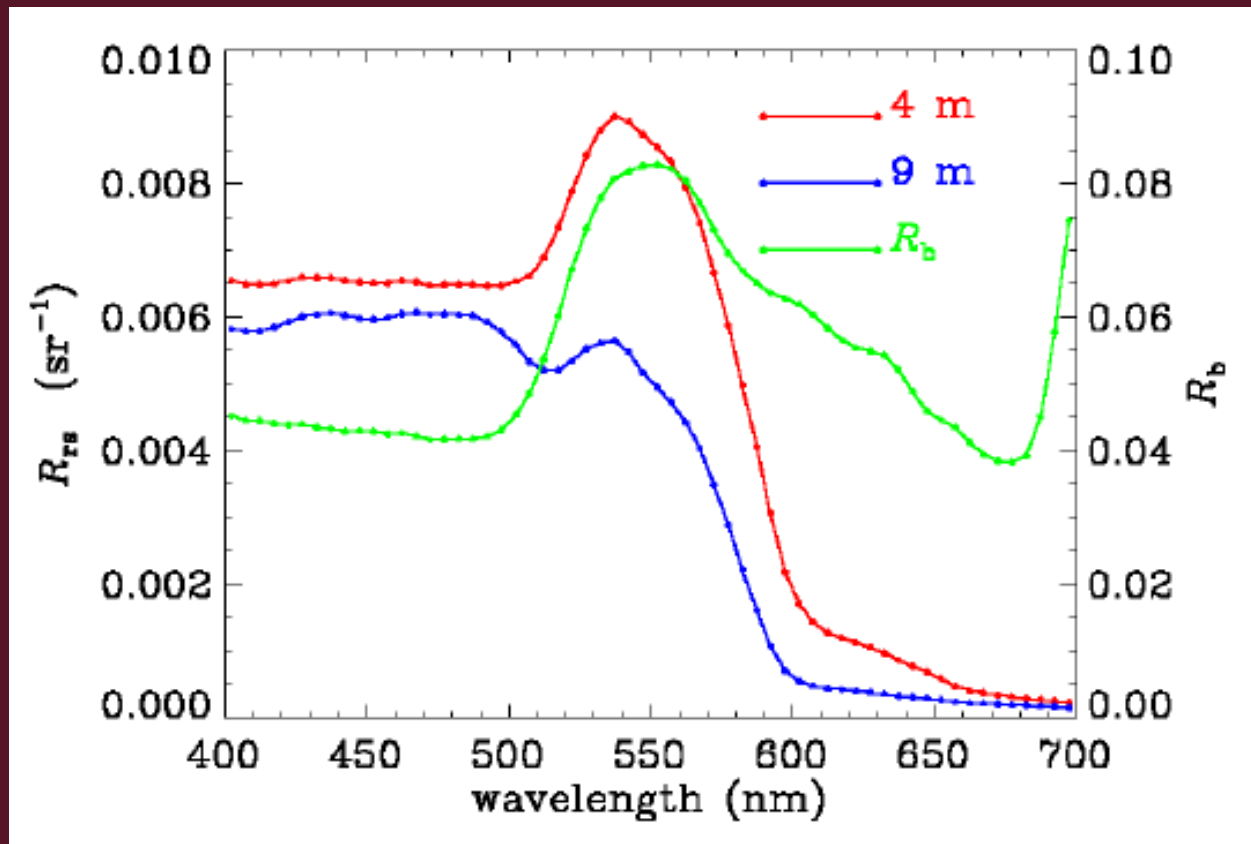
Model Selection

In some situations, you can figure out (from intuition, theoretical guidance, or data analysis) the general mathematical form of the model that links the input and output (e.g., the polynomial functions that relate the band ratios to Chl). You can then use the available data (e.g., simultaneous measurements of $R_{rs}(\lambda)$ and Chl) to get best-fit coefficients in the model via least-squares fitting.
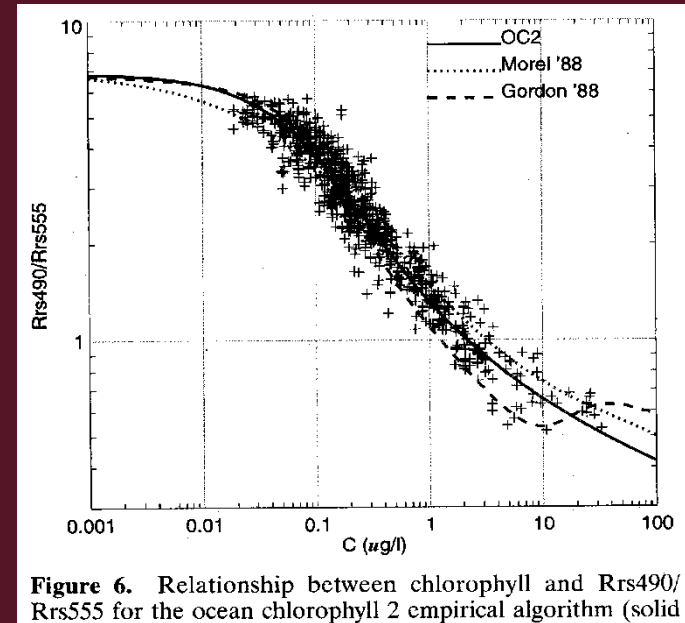
Figure 6. Relationship between chlorophyll and Rrs490/Rrs555 for the ocean chlorophyll 2 empirical algorithm (solid

O'Reilly et al., JGR, 1998

But what if you don't have any idea what the mathematical form of the model is?

Neural Networks

Neural networks are a form of multiprocessor computation, *based on the parallel architecture of animal brains,* with

- simple processing elements

- a high degree of connection between elements

- simple input and output (real numbers)

- adaptive interaction between elements

Neural networks are useful

- where we don't know the mathematical form of the model linking the input and output

- where we have lots of examples of the behavior we require (lots of data to "train" the NN)

- where we need to determine the model structure from the existing data

Biological Neural Networks

inputs

Axon

processing

outputs

Cell body

Synapse

Dendrite
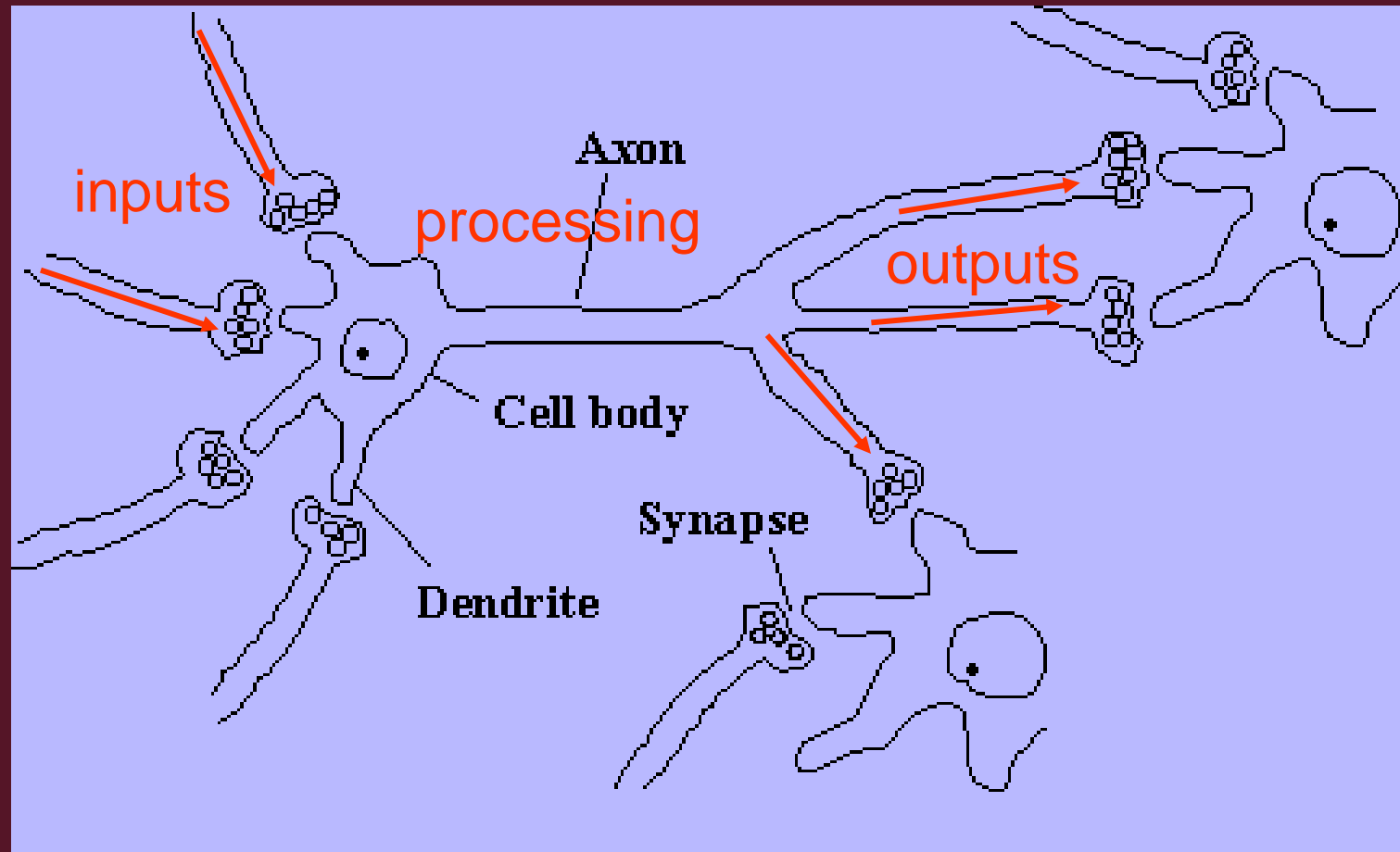
from www.qub.ac.uk/mgt/intsys/nnbiol.html

A Simple Artificial Neural Network

input synaptic   hidden layer   output
layer weights   (neurons)   layer

x_1

w_1

w_2

x_2

If
$x_1 w_1 + x_2 w_2 + b < t$
output = 0, else
output = 1
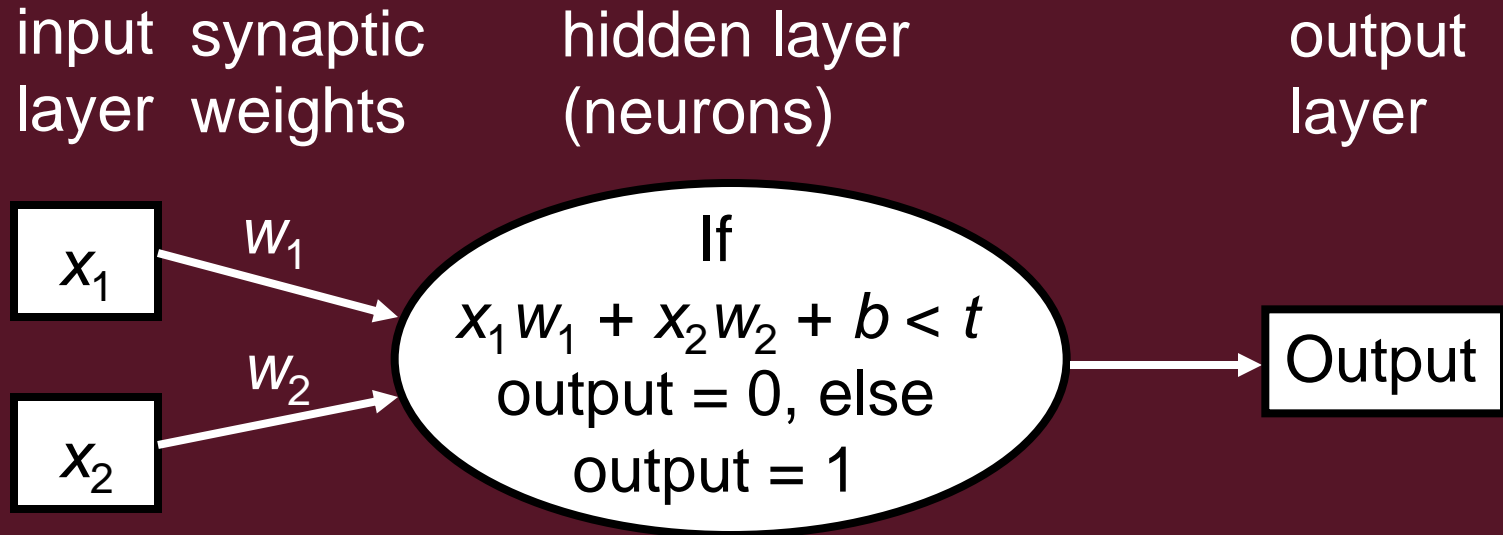
Output

In the neuron, b is the bias, t is the threshhold value

The neuron (processor) does two simple things:
(1) it sums the weighted inputs
(2) compares the biased sum to a threshhold value to
    determine its output

Training the Neural Network (1)

The essence of a neural network is that it can "learn" from available data. This is called *training* the NN. The NN has to *learn* what weighting functions will generate the desired output from the input.

Training can be done by *backpropagation of errors* when known inputs are compared with known outputs. We feed the NN various inputs along with the correct outputs, and let the NN objectively adjust its weights until it can reproduce the desired outputs.

The Java applet at www.qub.ac.uk/mgt/intsys/perceptr.html illustrates how a simple NN is trained by backpropagation.

run the NN applet

Things to Note

The NN was able to use the training data to determine a set of weights so that the given input produced the desired output. After training, we hope (in more complex networks) that new inputs (not in the training data set) will also produce correct outputs.

The "knowledge" or "memory" of a neural network is contained in the weights.

In a more complicated situation, you must balance having enough neurons to capture the science, but not so many that the network learns the noise in the training data.

Training the Neural Network (2)

Another way to train a NN is to view the NN as a complicated mathematical model that connects the inputs and outputs via equations whose coefficients (the weights) are unknown.

Then use a non-linear least squares fitting/search algorithm (e.g., Levenberg-Marquardt) to find the "best fit" set of weights for the given inputs and outputs (the training data).

This makes it clear that NNs are just fancy regression models whose coefficients/weights are determined by fancy curve fitting to the available data (not a criticism!)

An Example NN

From Ressom, H., R. L. Miller, P. Natarajan, and W. H. Slade, 1995. *Computational Intelligence and its Application in Remote Sensing*, in *Remote Sensing of Coastal Aquatic Environments*, R.L. Miller, C.E. Del Castillo, B.A. McKee, Eds.

• Assembled 1104 sets of corresponding R_{rs} spectra and *Chl* values from the SeaBAM, SeaBASS, and SIMBIOS databases.

• Construced a NN with 5 inputs (R_{rs} at 5 wavelengths) and two hidden layers of 6 neurons each, and one output (*Chl*).

• Partitioned the 1104 data points into 663 for training, 221 for validation, and 221 for testing the trained NN.

• The NN predictions of *Chl* using the testing data were compared with the corresponding *Chl* predictions made by the SeaWiFS OC4v4 band-ratio algorithm.

The Ressom et al. NN

input layer   30 weights   hidden layer 1   36 weights   hidden layer 2   6 weights   output layer

$R_{rs}(410)$

$R_{rs}(443)$

$R_{rs}(490)$

$R_{rs}(510)$

$R_{rs}(555)$

n_1 n_2 n_3 n_4 n_5 n_6

m_1 m_2 m_3 m_4 m_5 m_6

output   →   Chl

N.B. not all connections are shown; all neurons in a layer are connected to all in the preceeding and following layers
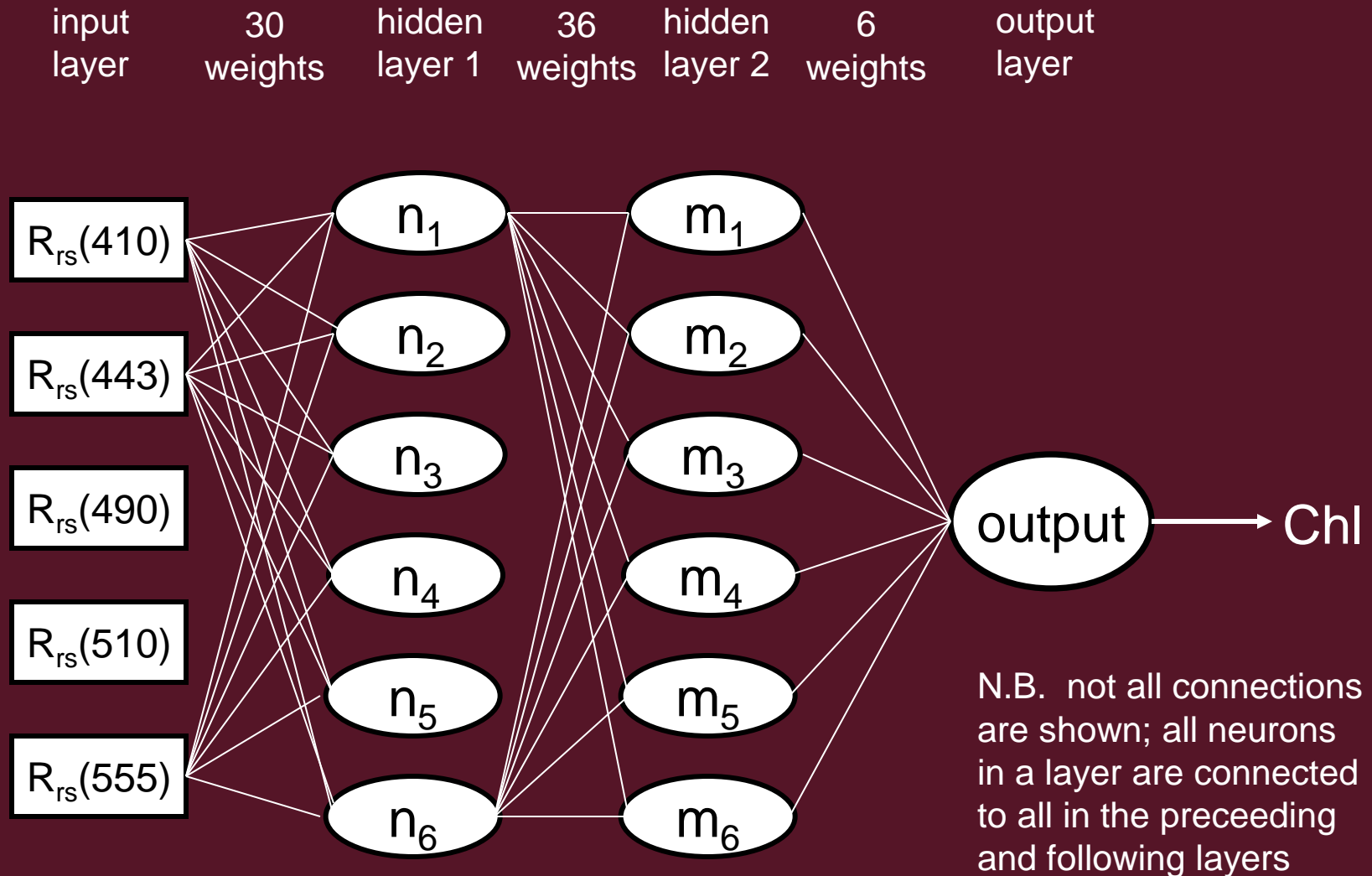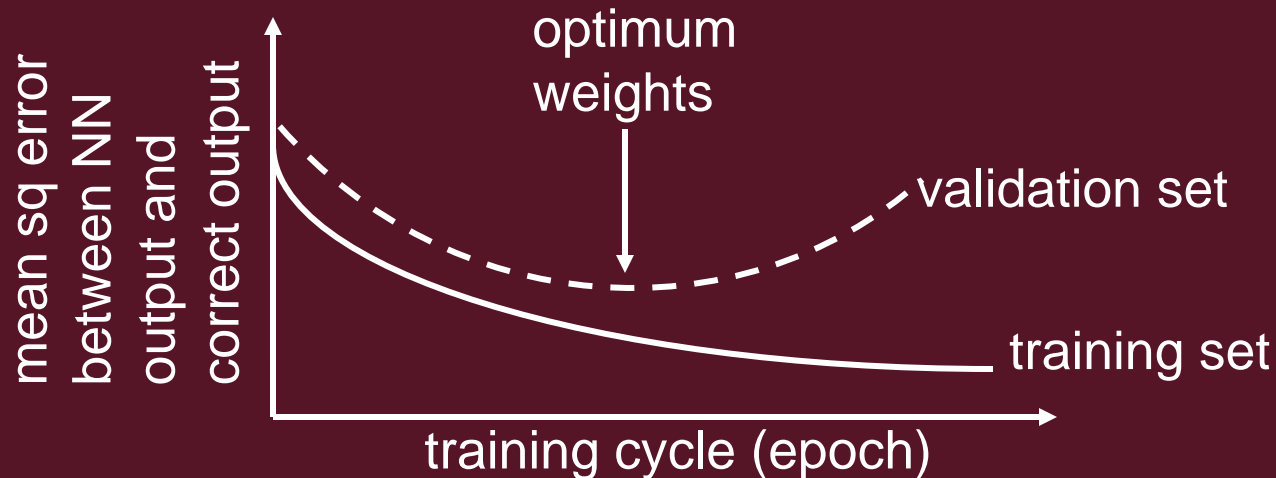
The Ressom et al. NN

Used two layers of 6 neurons, rather than one layer of 12, (for example), so that neurons can talk to each other (gives greater generality to the NN).
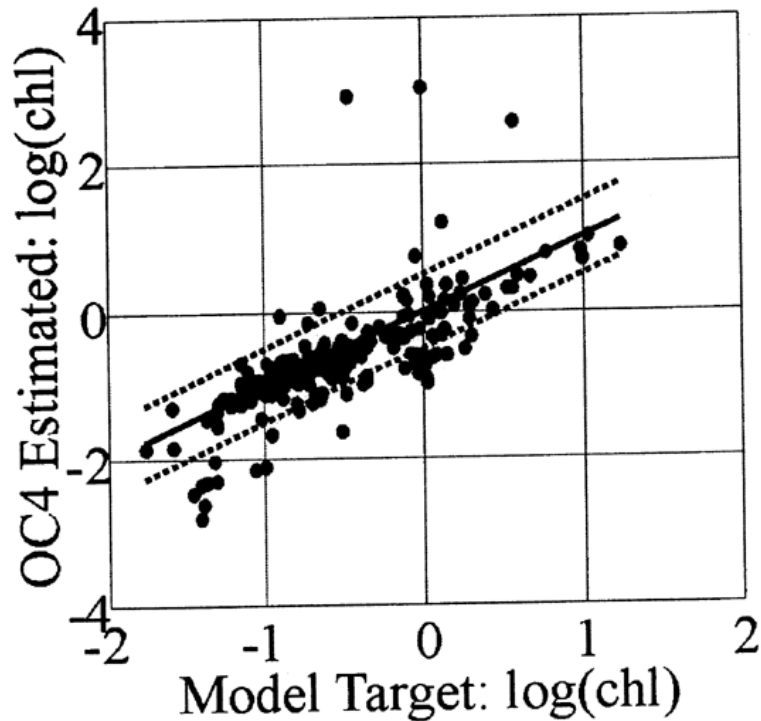
Training uses the training set for weigh adjustments, and the validation set to decide when to stop adjusting the weights.

NN vs. OC4v4 Performance

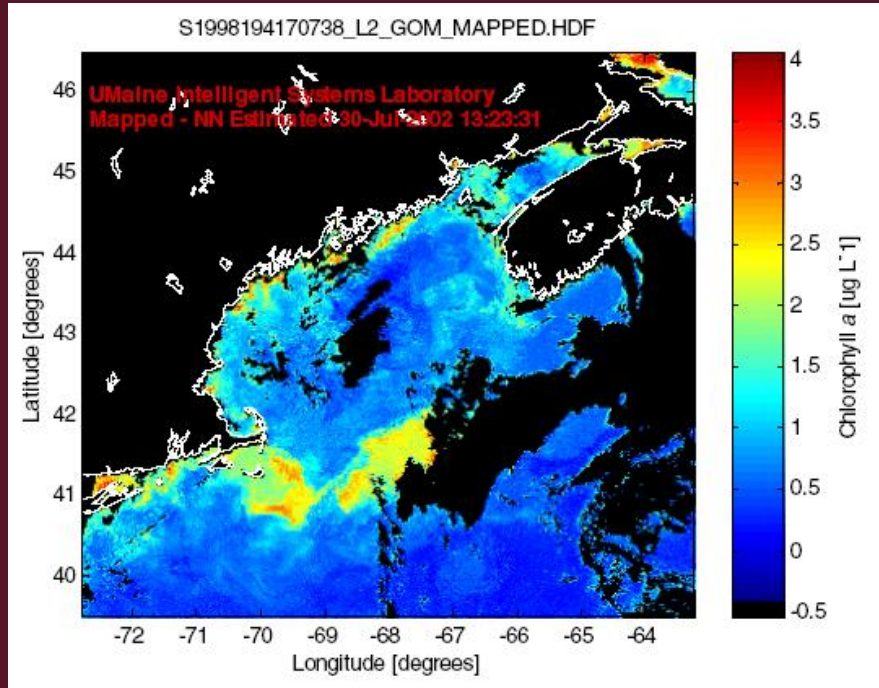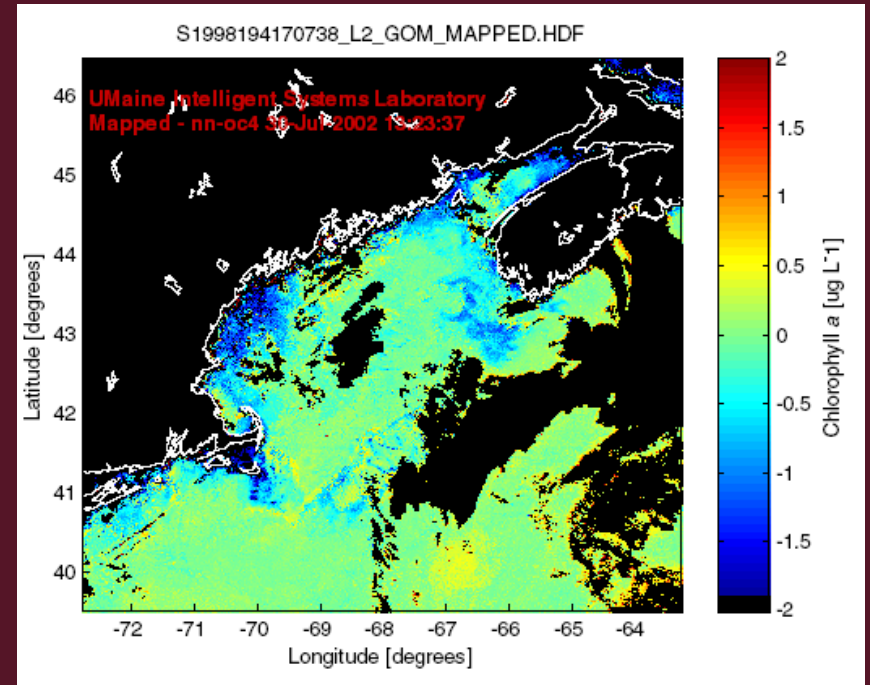|	Training Data (n=662)		Validation Data (n=221)		Testing Data (n=221)	
	r^2	RMS E	r^2	RMS E	r^2	RMS E
OC4	0.651	0.484	0.677	0.450	0.556	0.503
NN5	0.921	0.164	0.837	0.241	0.866	0.199

OC4v4 Results

NN5 Model Testing Results
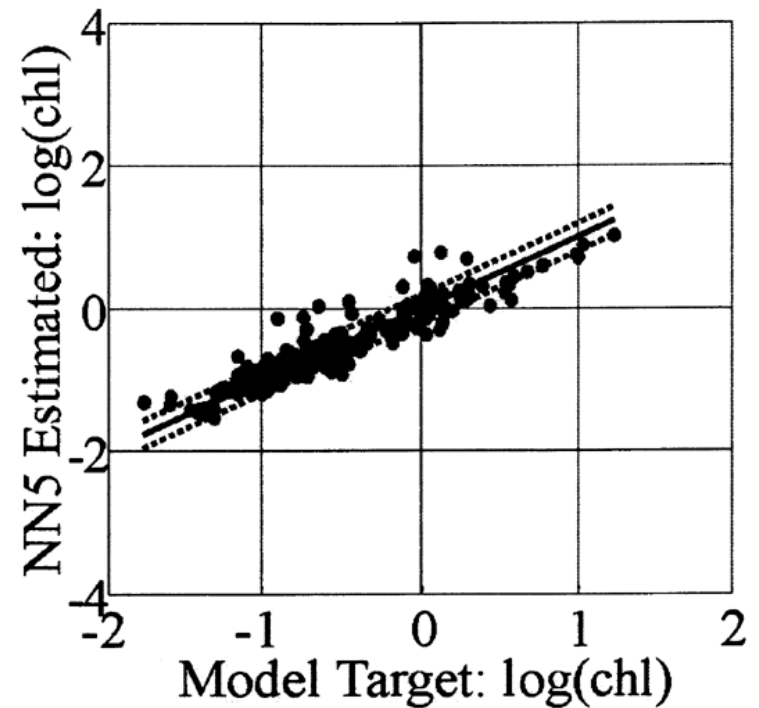
NN vs. OC4v4 Performance

ChI in the Gulf of Maine generated by applying a NN to SeaWiFS data

Difference in the NN and OC4 ChI values (NN-OC4)

from Slade, et al. Ocean Optics XVI

Takehome Messages

Statistical methods for retrieving environmental information from remotely sensed data have been highly successful and are widely used, but...

• An empirical algorithm is only as good as the underlying data used to determine its parameters.

• This often ties the algorithm to a specific time and place. An algorithm tuned with data from the North Atlantic probably won't work well in Antarctic waters because of differences in the phytoplankton, and an algorithm that works for the Yellow Sea in summer may not work there in winter.

• The statistical nature of the algorithms often obscures the underlying biology or physics.

Takehome Messages

Band-ratio algorithms remain operationally useful, but they have been milked for about all they are worth (IMHO). Note that band ratio algorithms throw away magnitude information in the R_{rs} spectra, and they may not use information at all available wavelengths.

Newer statistical techniques such as neural networks are proving to be very powerful, as are other techniques such as spectrum matching and semi-analytical techniques.

Limestone (early-mid Cambrian, 505-525 Myr old) boulder with fossil algal mats, Grand Canyon, photo by Curt Mobley