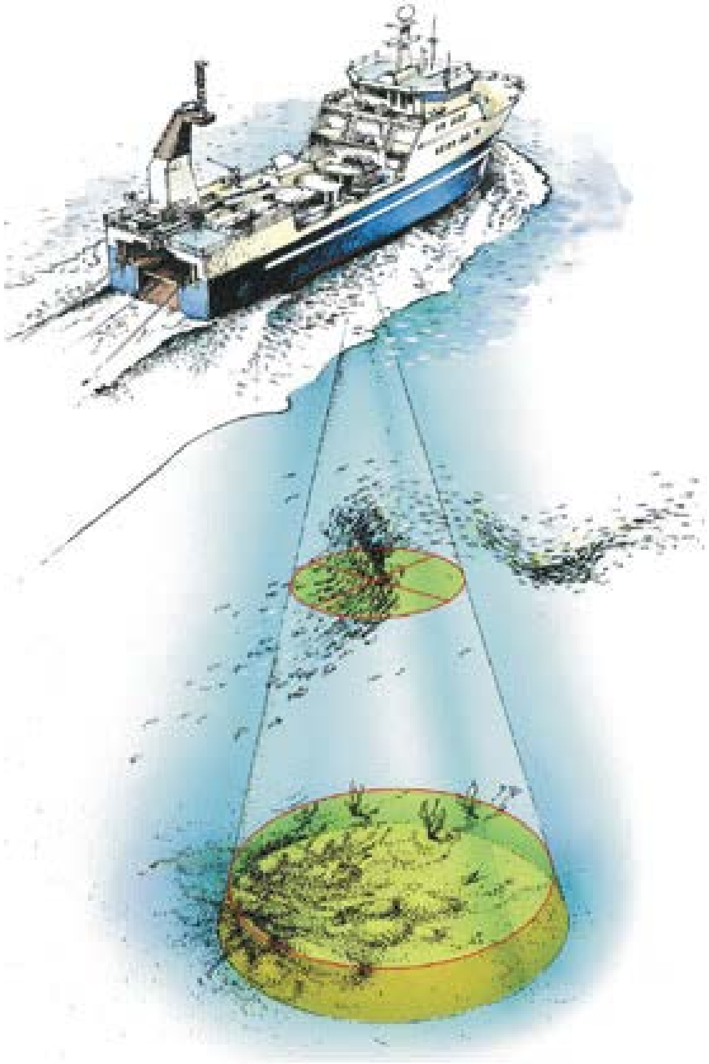
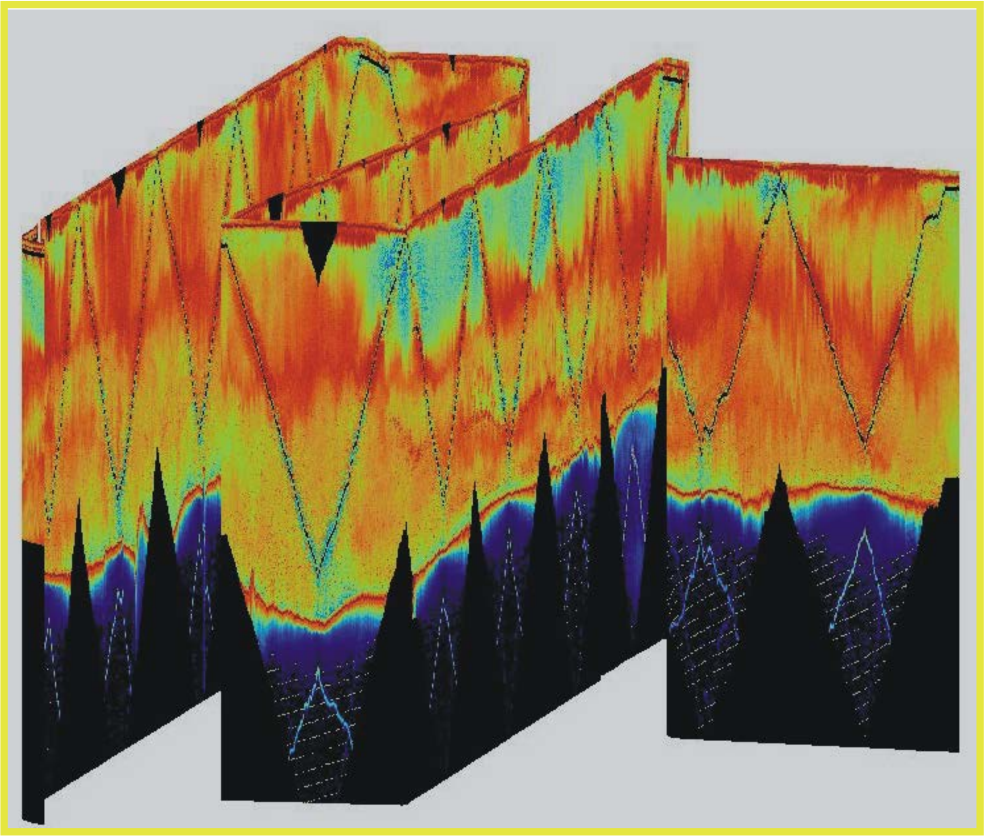
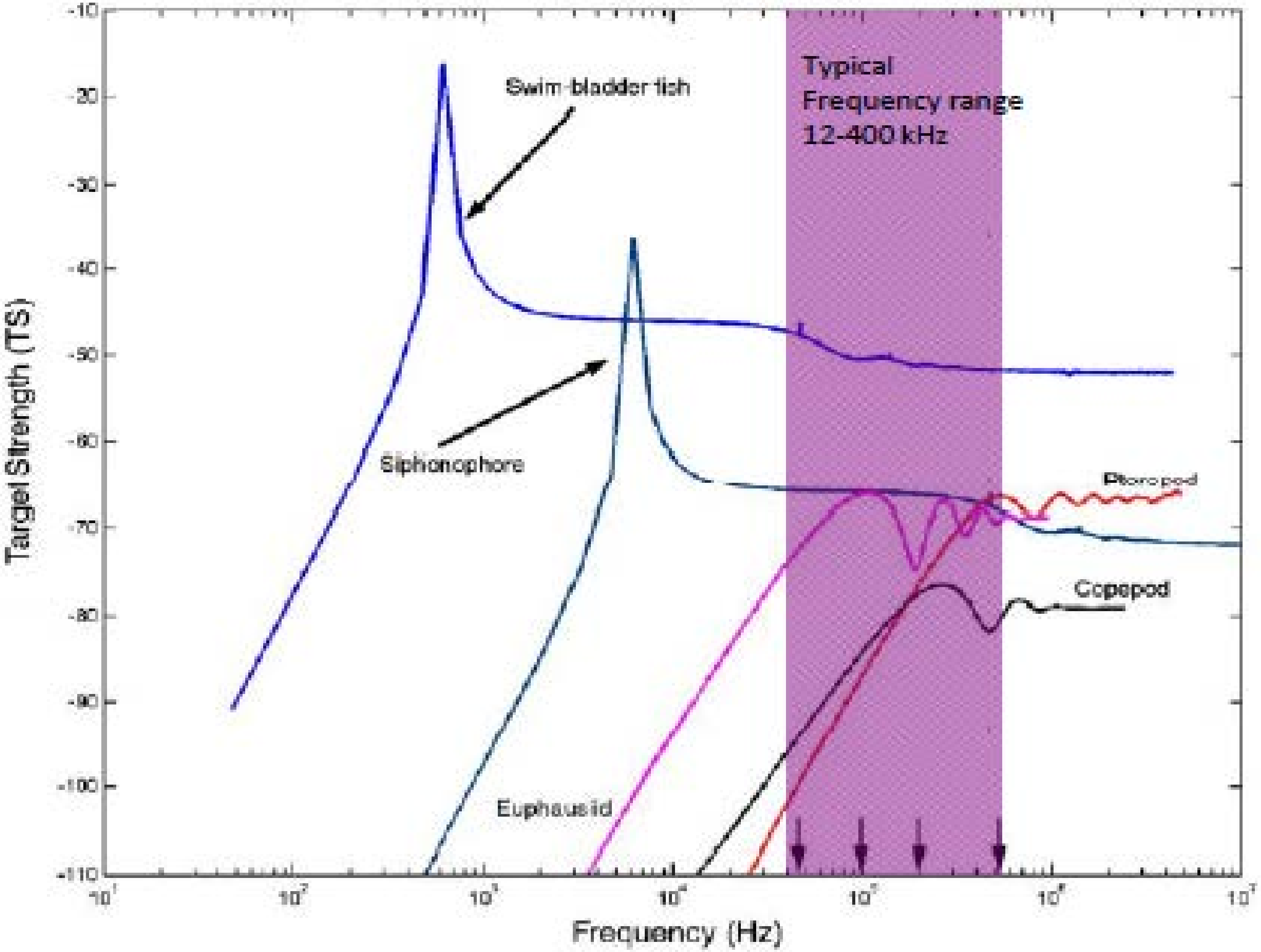
Data Processing & Analysis I





Acoustic Scattering Regions



Lavery et al. 2002

Factors Influencing Acoustic Data Quality

1. Sensor: echosounder, transducer
2. Sampling platform: vessel, transducer mounting, electrical supply, grounding
3. Environment: water structure, surface conditions
4. Targets: behavior, material properties, orientation
5. Operator: parameter settings, biological sampling, recording
6. Analyst: size and density translations, abundance or biomass estimates

Echosounder Outputs: Sample Data

‘Raw’ Data

- sample power (units watts)
- sample angle (units degrees) along: fore, aft; athwart: port, starboard

Processed Data

- area backscatter strength: S_a
- (Mean) volume backscatter strength: S_v
- Nautical Area Backscattering Strength: $S_{A,}$ NASC
- target strength: TS

Processing Steps

1. Pre-Processing: passive noise calculation, calibration tuning, data editing (bottom, noise, empty pings)
2. Processing: gridding, classification,
3. Exports: densities, distributions for maps, demographics (e.g. length frequencies), behaviors (e.g. target tracks)

Partitioning Backscattered Energy aka Scrutinizing

- the division of recorded backscatter into categories
- 3 components: quality control, dividing echo integrals, associating integrals with user-defined categories
- can be subjective (e.g. single frequency) or objective (e.g. multifrequency)
- each transect partitioned by operator or objective rules
- species mixtures complicates the task

Backscatter Categories

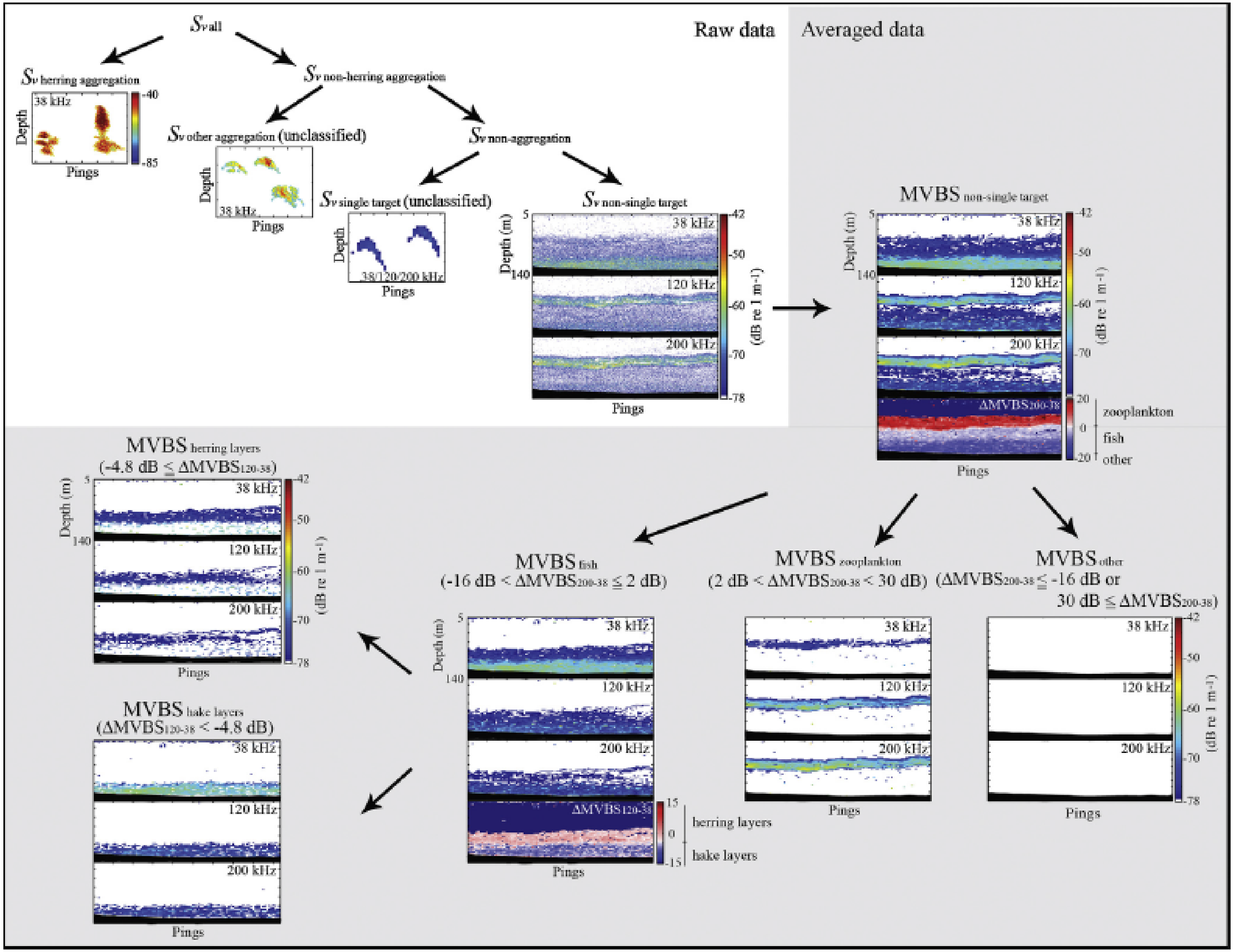
Subjective:

- analyst determined
- species and/or size/length classes
- based on prior knowledge, echogram, direct samples

Objective

- rule determined (*rules set by analyst*)
- frequency- and/or amplitude-dependent backscatter
- trawl catches for species compositions and length frequencies and when constituents unknown
- proportion energy by % species composition or weight

Mean Volume Backscatter Differencing



Sato et al. 2015

Relative Frequency Response

- Choose a baseline frequency (*f* = 38 kHz typical) and compare backscatter integral amplitudes to that frequency

$$r(f) = \frac{S_v(f)}{S_v(f_{ref})} = \frac{S_A(f)}{S_A(f_{ref})}$$

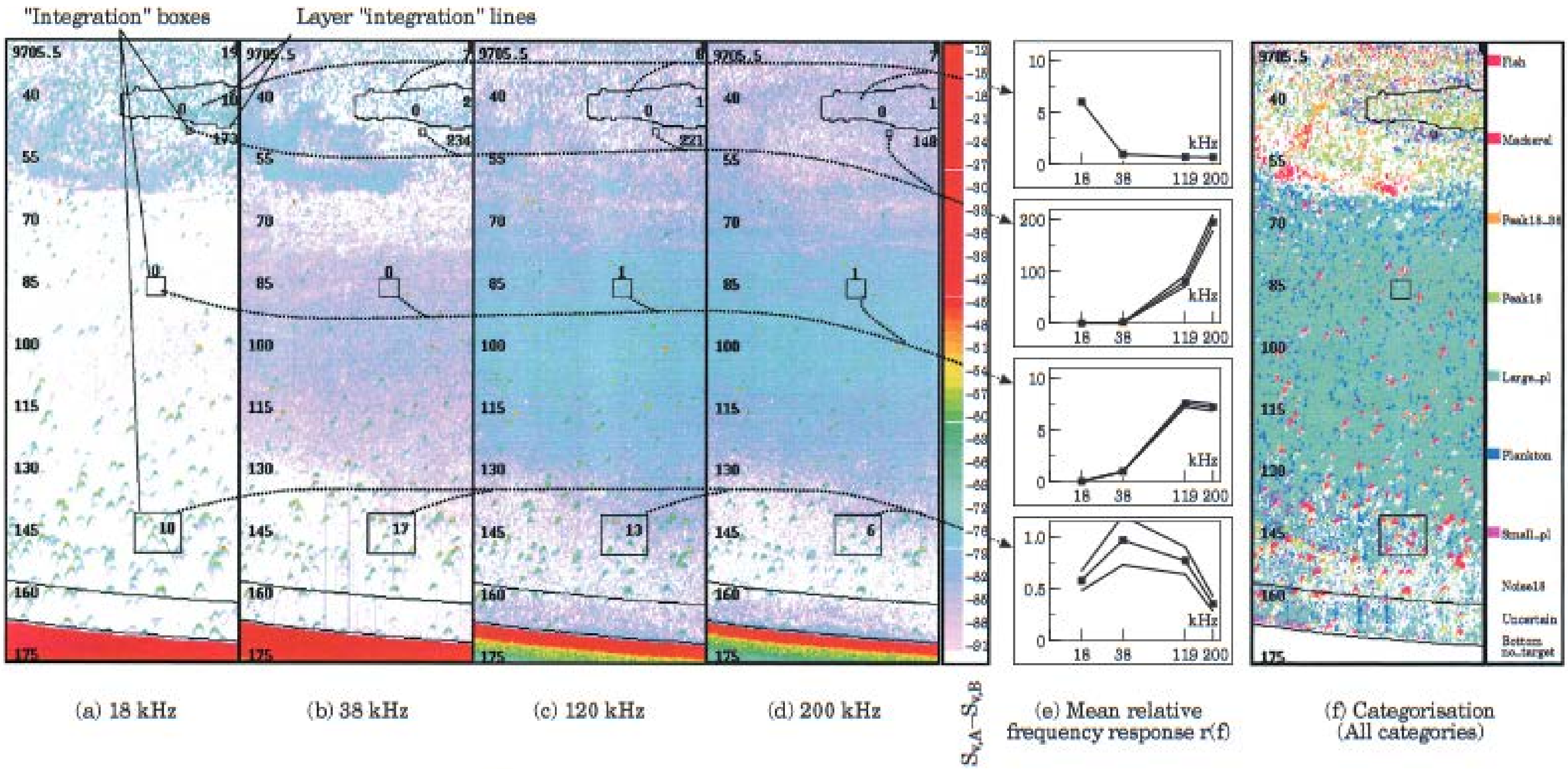
(Korneliussen and Ona 2002)

For Single Targets

$$r_T(f) = \frac{\sigma_{bs,T,i}(f)}{\sigma_{bs,T,i}(f_{ref})}$$

(Pedersen and Korneliussen 2009)

Frequency Response Example



(a) 18 kHz

(b) 38 kHz

(c) 120 kHz

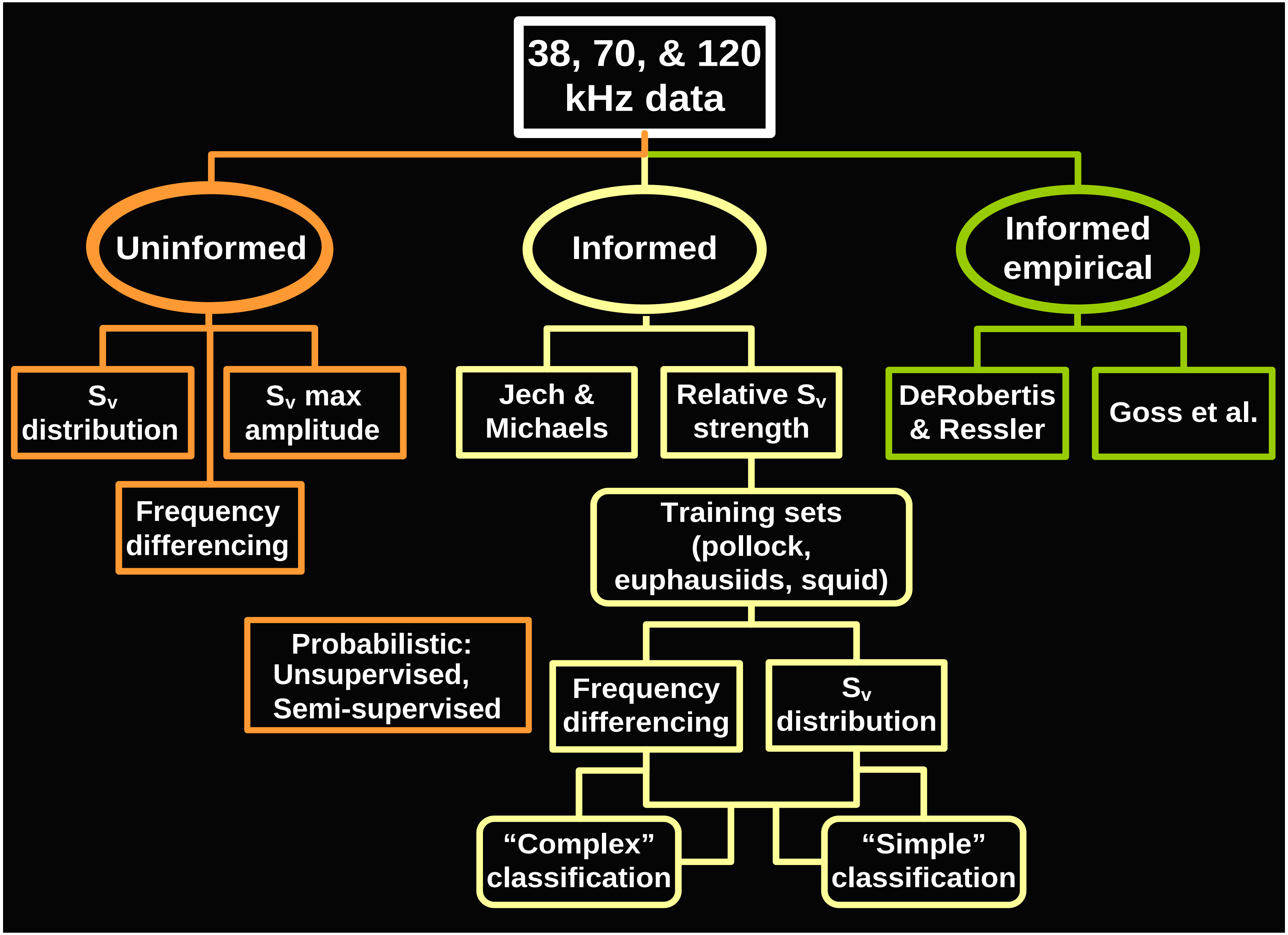
(d) 200 kHz

(e) Mean relative frequency response r(f)

(f) Categorisation (All categories)

Korneliussen and Ona 2002

Multifrequency Classification Approaches



38, 70, & 120 kHz data

Uninformed

Informed

Informed empirical

S_v distribution

S_v max amplitude

Frequency differencing

Probabilistic: Unsupervised, Semi-supervised

Jech & Michaels

Relative S_v strength

Training sets (pollock, euphausiids, squid)

Frequency differencing

S_v distribution

"Complex" classification

"Simple" classification

DeRobertis & Ressler

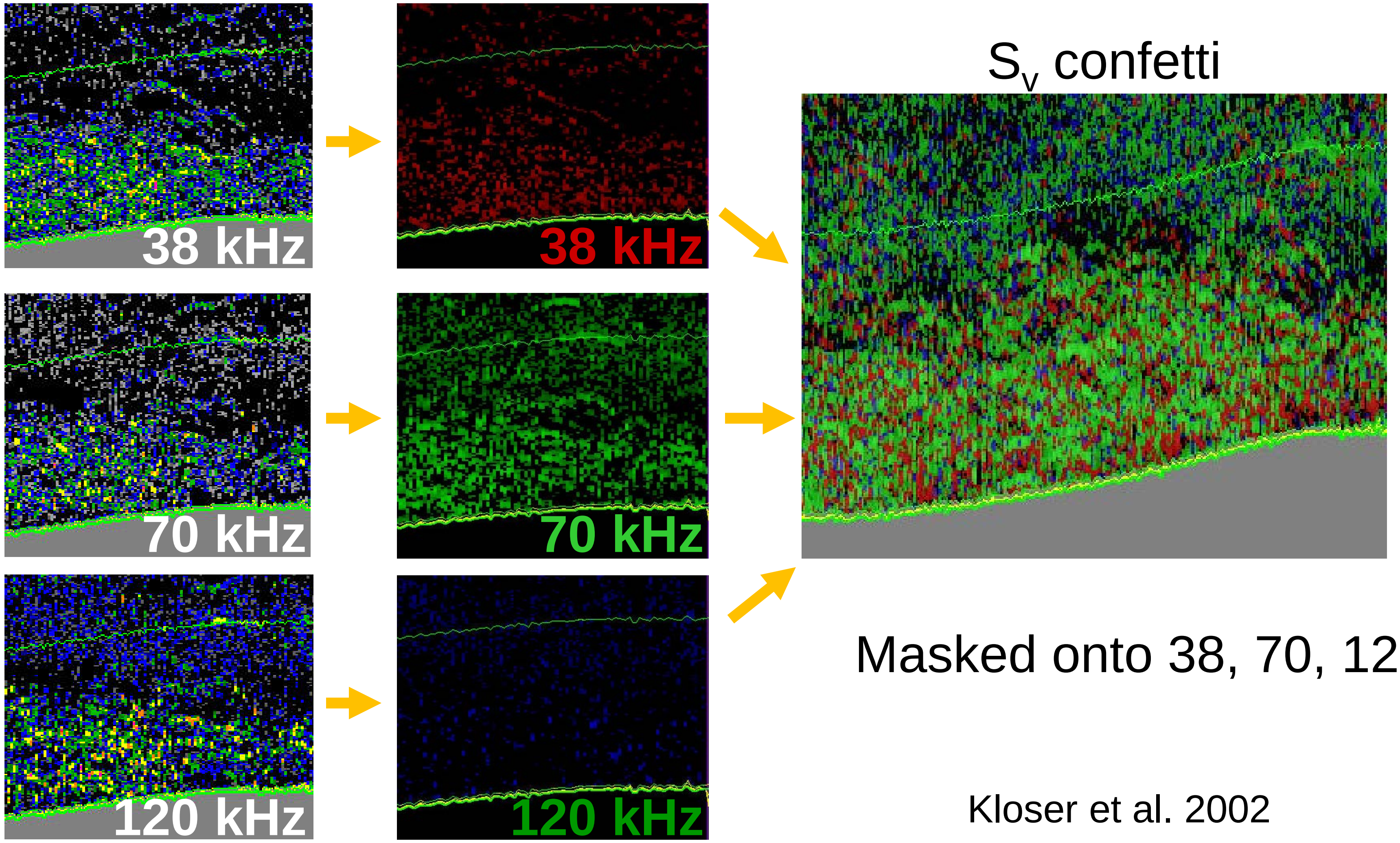
Goss et al.

Unsupervised Classification

S_v maximum amplitude

Original

Maximum



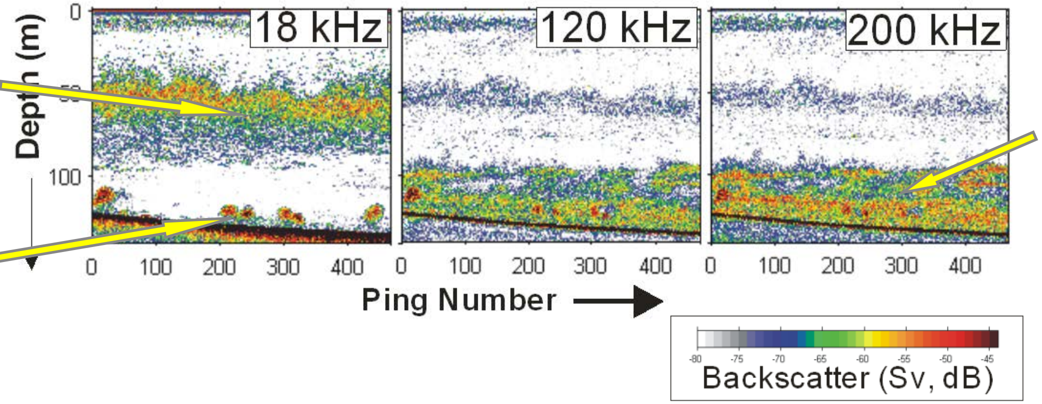
S_v confetti

Masked onto 38, 70, 120

Kloser et al. 2002

Unsupervised Probabilistic Classification

3 Frequency Backscatter:



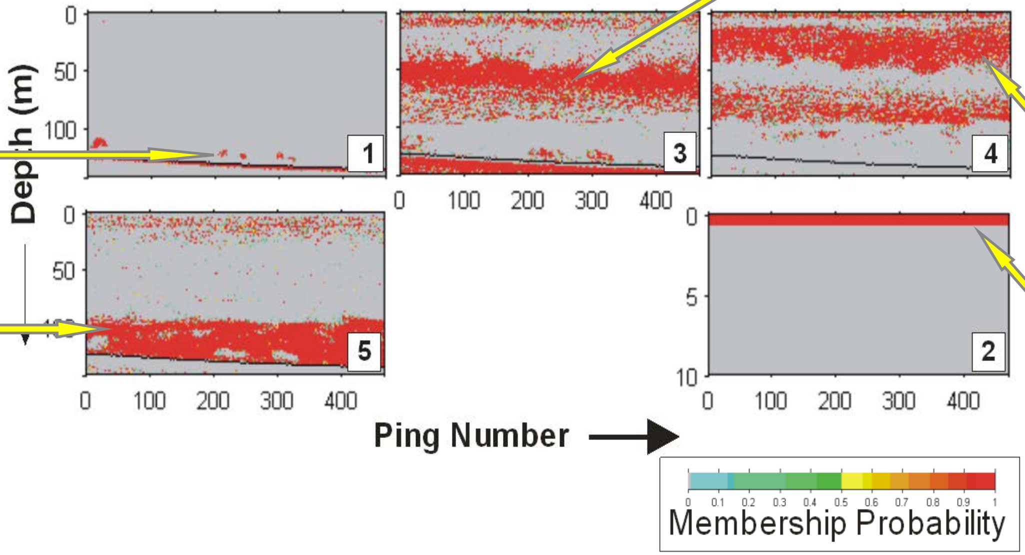
small fish

walleye
pollock

krill

Expectation
maximization of
finite mixture models

5 Cluster Identification:



'fish'

dense
schools

krill

back-
ground

ring
down

Anderson et al. 2007

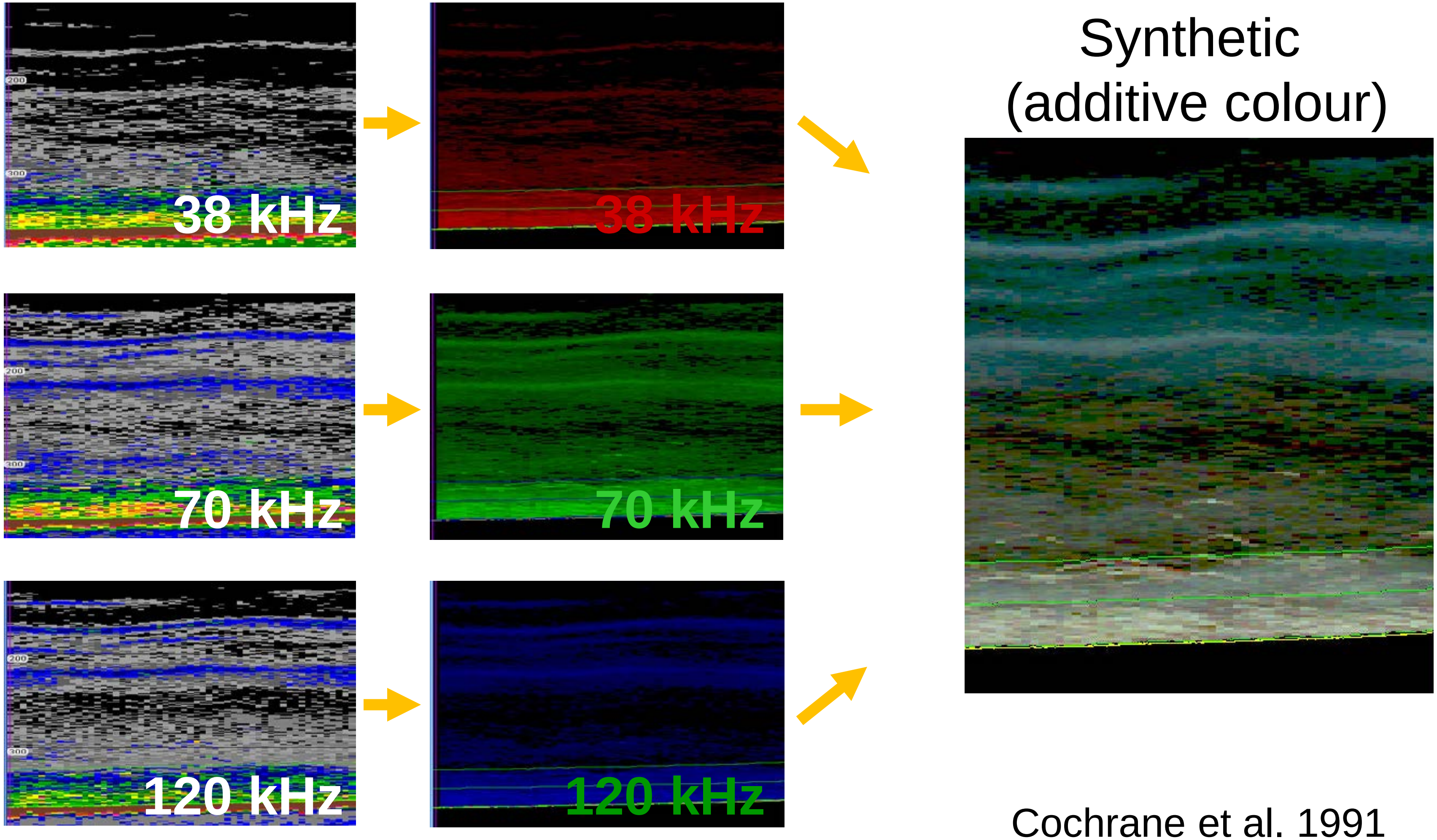
Supervised Classification: Integral Amplitude

S_v relative strength, verified samples

Original

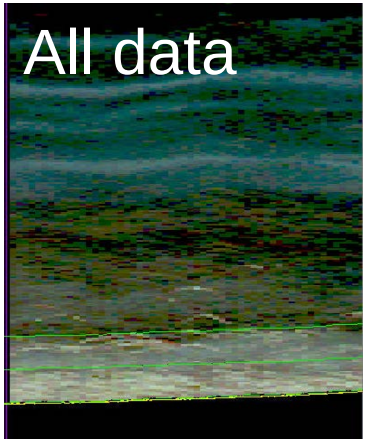
Color scale

Synthetic
(additive colour)



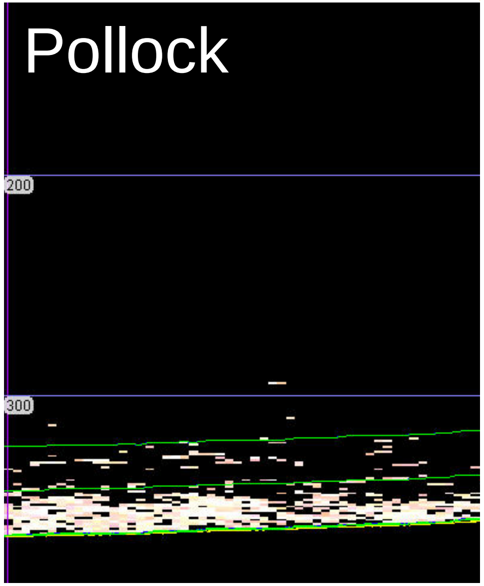
Cochrane et al. 1991

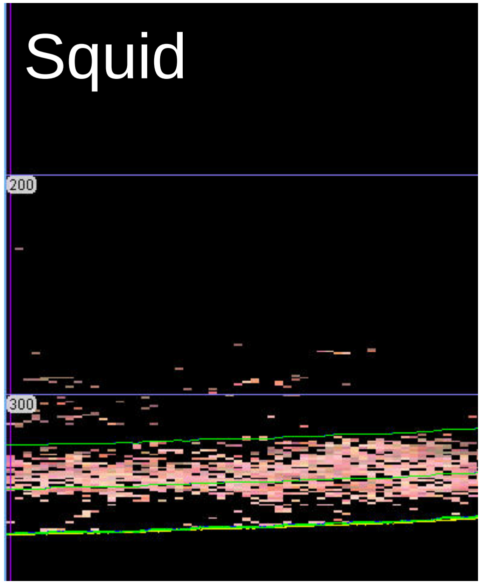
Supervised Classification: Ref Library

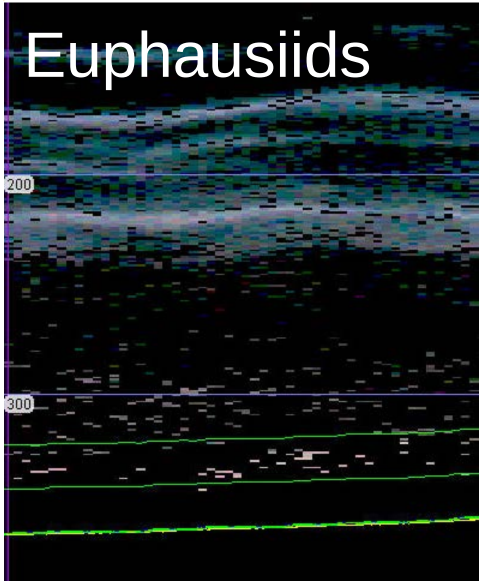


Using verified data as training sets

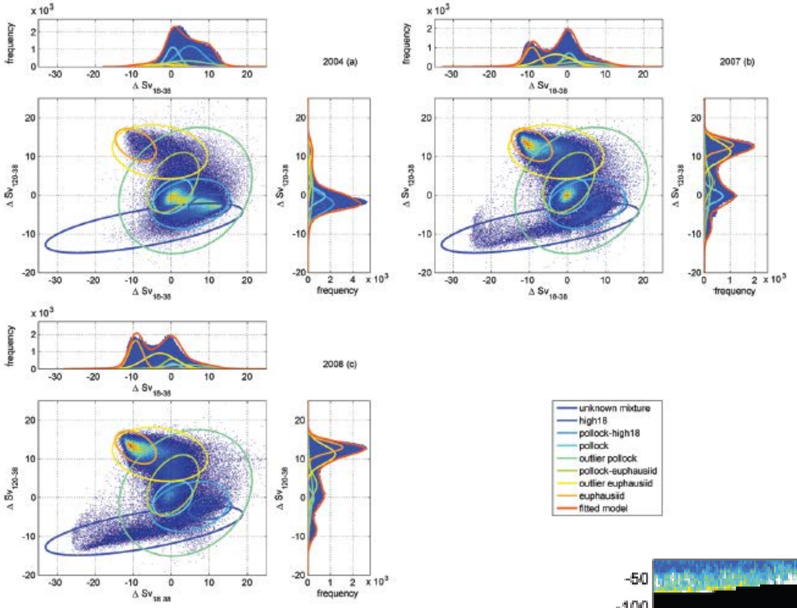
Analyst-assigned categories



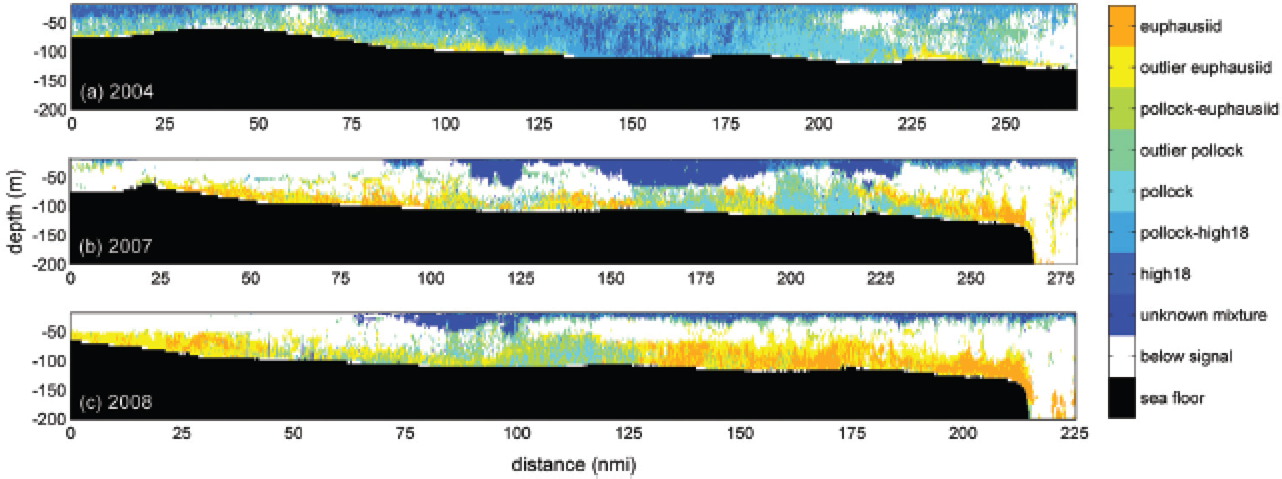




Seimi-supervised Probabilistic Classification



Generalized Gaussian mixture model with class discovery



Woillez et al. 2007

Trawl Classification

Trawl Catches, length-based single or multiple species:

- numbers and weight by species
- lengths by species
- lengths by sex

Proportion (P) of catch by species *i* in length class *j*:

$$P_{ij} = \sum_{k=1}^{M_i} \left(n_{ijk} / N_{ik} \right) / M_i$$

where

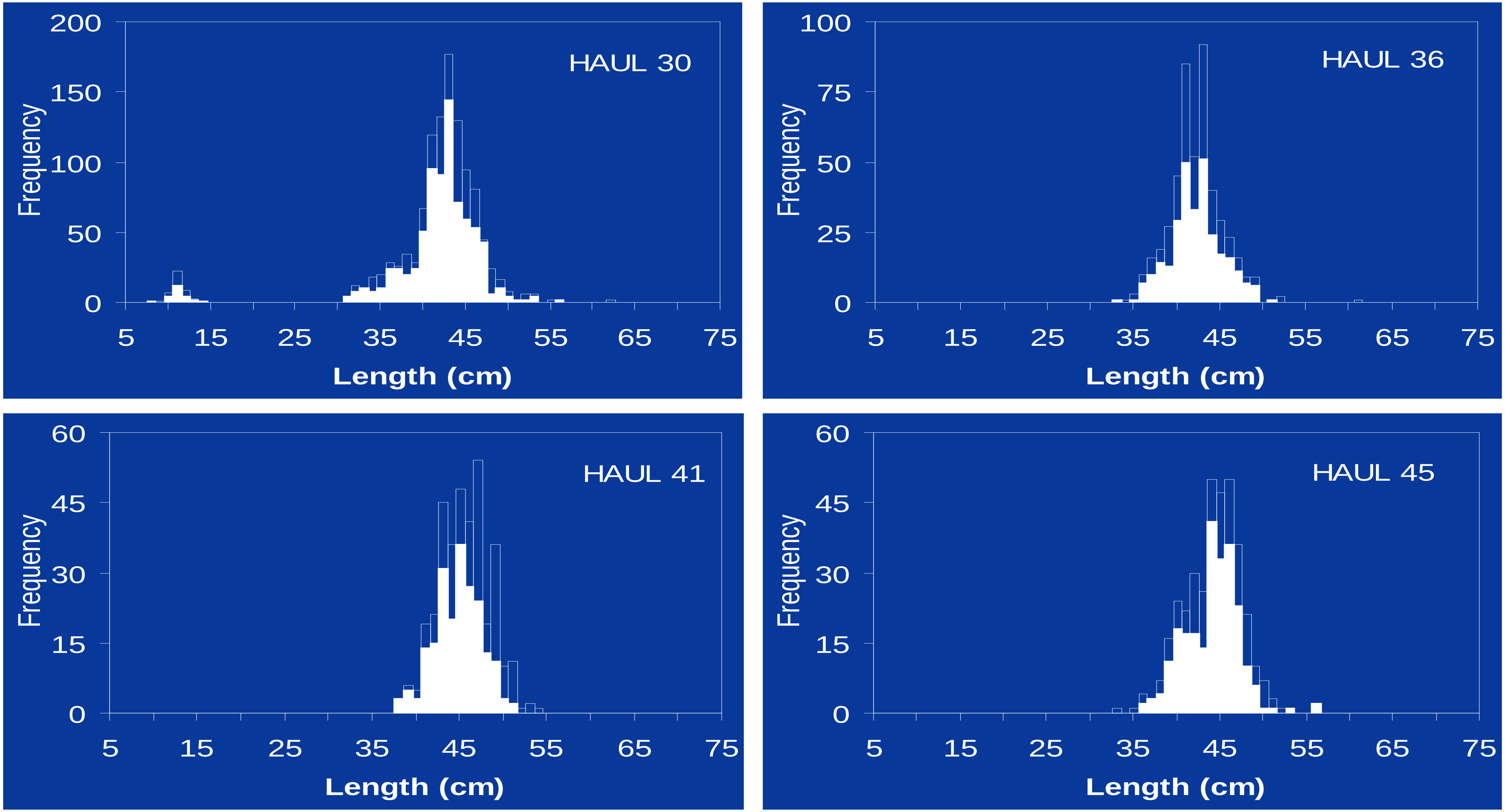
M_i is # hauls, n_{ijk} is number in length class *j*, and N_{ik} is total # species

- equal weight by catch rate of each species
- weight by echo integrals in area

Length Frequency Data

best case scenario

Walleye Pollock, Bering Sea, 2000



Proportions By Species

Trawl Catches, Mixed Species:

Equal weight (w) to each species *i* at station *k*:

$$w_i = \sum_{k=1}^{M_i} \left(q_{ik} / q_k \right) / M$$

where

M_i is # hauls, q_{ik} is quantity of *i* th species at station *k*, and q_k is total catch

- weight by catch rate

- weight by echo integrals in area

Grouping Homogeneous Regions

- once all transect segments are categorized, homogeneous regions are grouped (typically species and/or length structured)

- can use Kolmogorov-Smirnov test to examine differences in distribution of two trawl catches (Campbell 1974):

D statistic: range 0 (identical) to 1 (no similarity)

if $0.1 < p < 0.3$ then samples same

0.1 threshold to incorporate sampling error

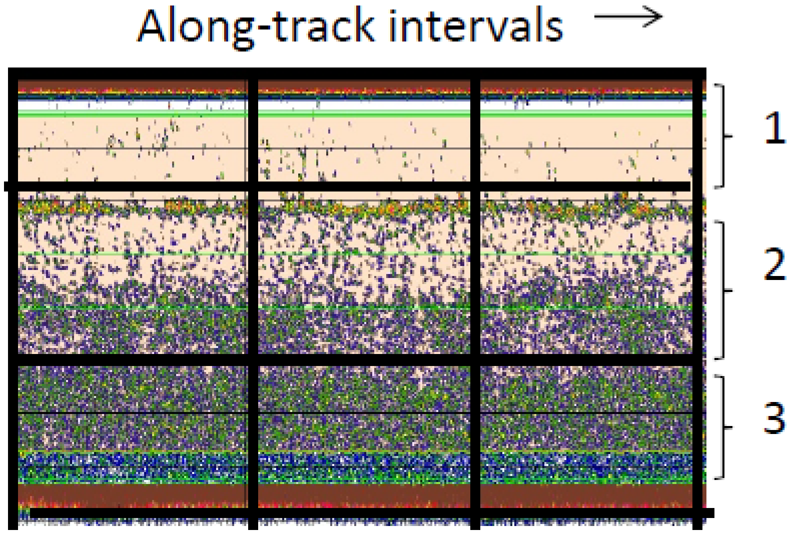
Converting Echo Integrals

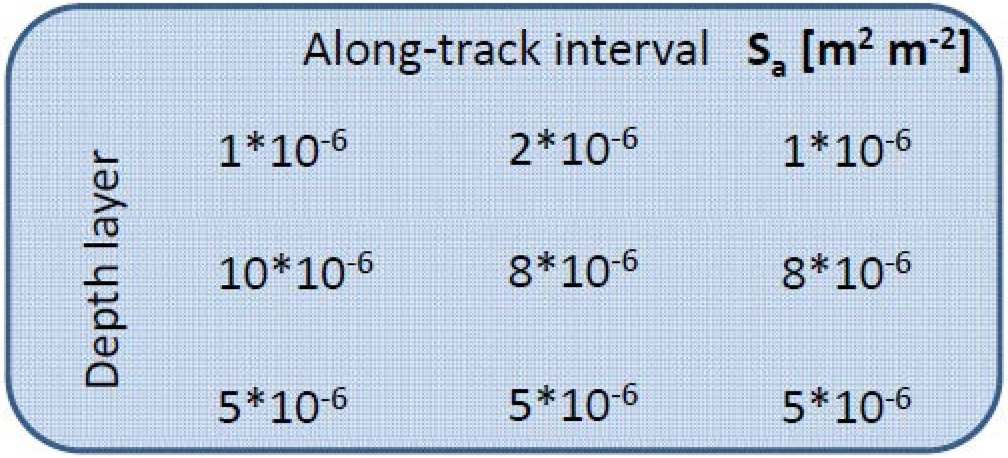
- estimate density of targets ρ_i from the observed echo integrals E_i

$$\rho_i = C_E \frac{E_i}{\langle \sigma_{bsi} \rangle}$$

where C_E is a equipment calibration factor, $<\sigma_{bsi}>$ is the mean backscattering cross section of class or species i

Aggregating Backscatter: SVAH



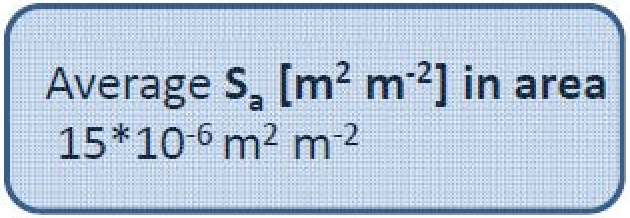


Depth layer	Along-track interval S_a [m^2 m^{-2}]		
	$1*10^{-6}$	$2*10^{-6}$	$1*10^{-6}$
	$10*10^{-6}$	$8*10^{-6}$	$8*10^{-6}$
	$5*10^{-6}$	$5*10^{-6}$	$5*10^{-6}$

Sum Vertically

Along-track interval S_a [m^2 m^{-2}]		
$16*10^{-6}$	$15*10^{-6}$	$14*10^{-6}$

Average Horizontally



Average S_a [m^2 m^{-2}] in area

$15*10^{-6}$ m^2 m^{-2}

Determining <$\sigma_{\mathrm{bs}i}$>

In situ measurements (may or may not be direct samples)

- dual or split beam transducer required

- more common in fresh water surveys than in marine

Acoustic size - target size relationship

$$\left\langle \sigma_{bsi} \right\rangle = \frac{1}{n} \sum_{i} f_{ij} 10^{\left[\left(a_i + b_i \log L\right)/10\right]}$$

$$\left\langle \sigma_{bsi} \right\rangle = \frac{1}{n} \sum \left[r^2 I_{bs}\left(r\right) 10^{\alpha r/10}\right] / I_o$$

Determining <σ_{bs}*i*>

Model estimates (requires acoustic size-animal size (length) relationship):

- numerous models available, most combine anatomical representation with material properties to estimate backscatter

$$\langle \sigma_{bsi} \rangle = \frac{1}{n} \sum RSL * L = \frac{1}{n} \sum \frac{l_{bs}}{L} * L$$

RSL is a non dimensional, linear measure

$$TS = 20 \log(RSL) + 20 \log(L)$$

* log of the means ≠ mean of the logs*

Determining $<\sigma_{bs}i>$

Challenge of Multiple species (need direct sample data):

weight by backscatter proportions using catch, catch-rate of each species, or echo integrals in vicinity of catches

Direct biomass estimate:

derive TS -Weight relationship and adjust parameters accordingly (by size class as L-W curves are non-linear)

Example: Pacific Hake *(Merluccius productus*)

-36 dB kg^{-1} for 50 to 55 cm fish

(Williamson and Traynor 1984)

Abundance Estimates

- calculate densities independently for each EDSU, species or category of target, and depth interval (resolution depends on distributions of animals)
- if randomly distributed then mean density x area of interest
- if contagious then calculate abundance in each region and sum
- partition abundances by length classes
- interpolation among transects: area includes up to ½ the distance between two transects, assume observed density throughout area, sum areal or volumetric estimates
- other methods: contour maps, geostatistical (e.g. kriging)
- extrapolation outside of transects: don’t do it! adjust your survey design and transect layout

Abundance/Biomass Estimates by Length

Assume you have:

σ_{bs} for each length class *i*

frequency *f* of each length class *i*

known distance between transects

defined homogeneous regions *j*

length-weight relationship (*a* & *b* constants) for each length class

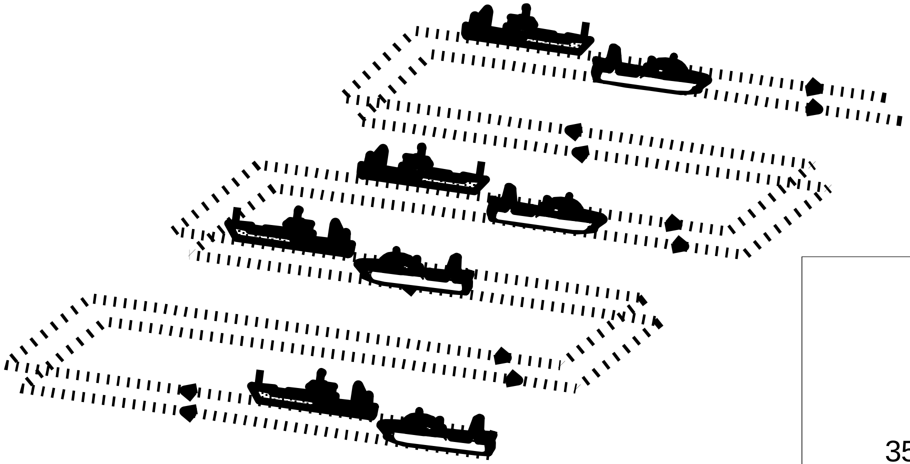
Abundance $$N_{i,j} = S_{a,j} \times area_j \times \frac{1}{\sigma_{bsi}} \times f_{i,j}$$

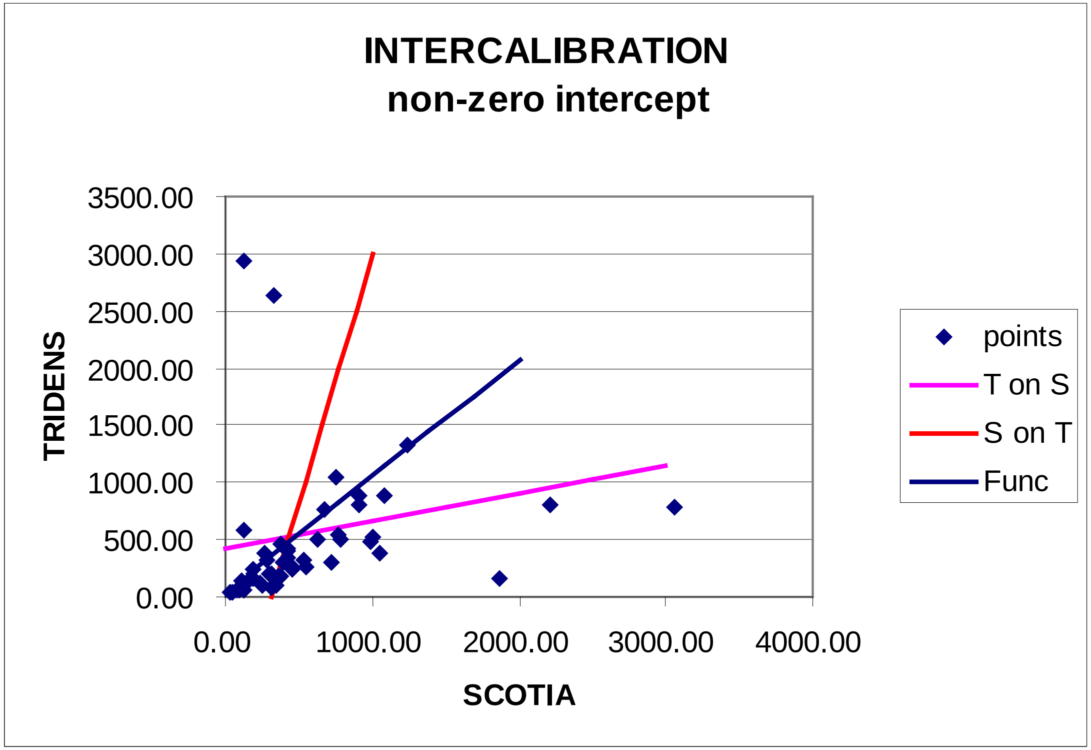
Biomass $$B_{i,j} = N_{i,j} \times a \times \left(L_{i,j}\right)^b$$

$$Total = \sum_i^n \sum_j^m N_{i,j} \; or \; B_{i,j}$$

Intercalibration

- use of multiple vessels to obtain a single estimate
- new vessel replacing an older vessel





- quiet vs non-quiet vessel
- ideally amounts will match

See also: De Robertis et al. 2008, 2010a,b,