Issues & Methods for Analyzing Zooplankton Time Series – Sample Applications of the WG125 Toolkit

Todd O'Brien[1],
Dave Mackas[2],
Mark Ohman[3],
WG125 members,
& 3 'data partners' (G. Mazzocchi, R.Harris, W. Greve)

[1]National Marine Fisheries Service, Silver Spring, MD

[2]Fisheries and Oceans Canada, Institute of Ocean Sciences, Sidney, BC

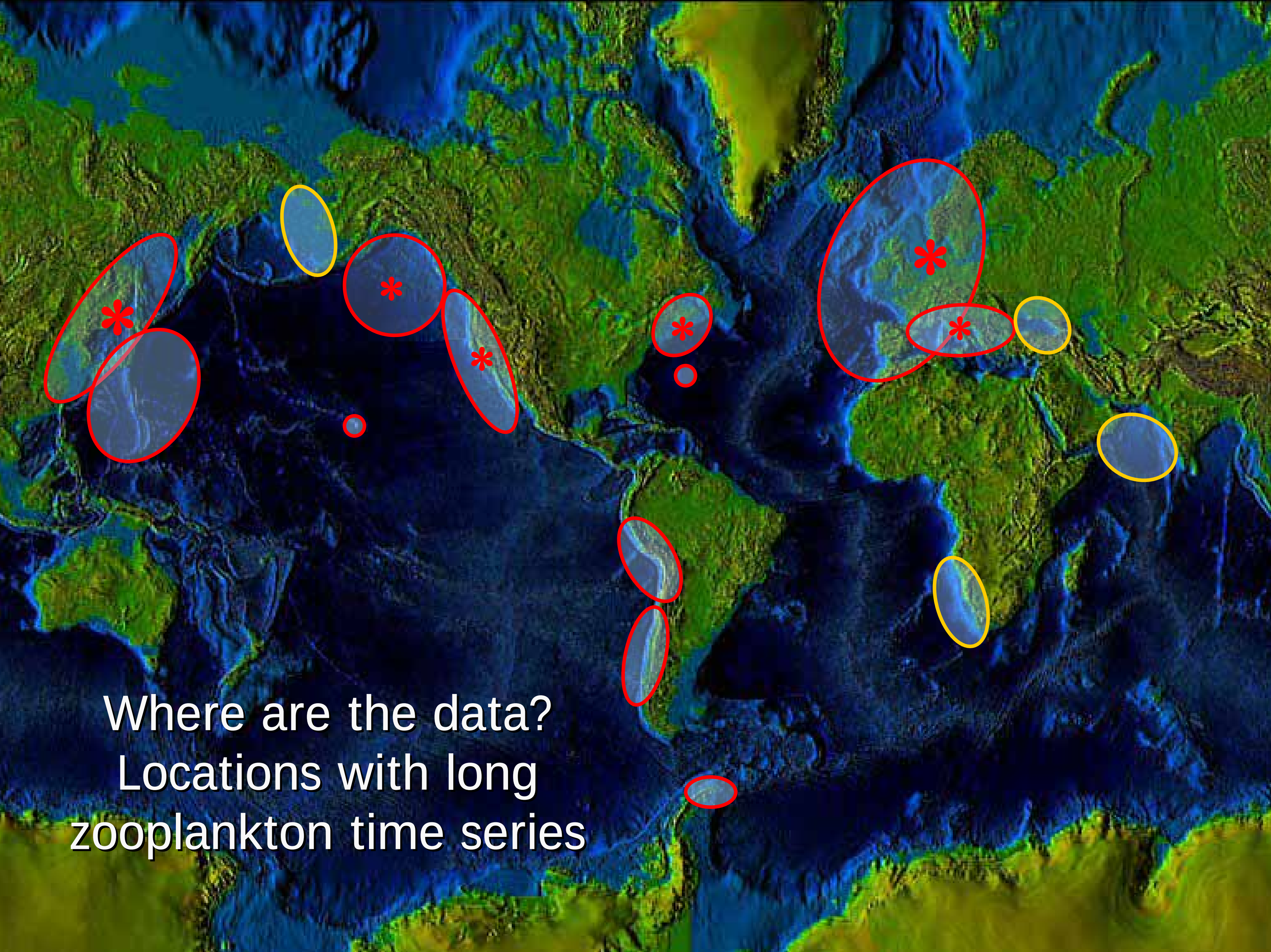
[3]Scripps Institution of Oceanography, La Jolla, CA

Topics & practices:

- Value & availability of zooplankton time series
- Diversity of sampling designs
- Modes and time scales of variability
- Data transformation & averaging
- Removing seasonal cycle -> Climatology
- Deviations from climatology -> Anomalies
  - Dealing with data gaps
  - Comparing years within data sets
  - Comparing taxa within data sets
  - Comparing across data sets
- Visualization tools

Advantages of mesozooplankton for tracking & interpreting ‘Ocean Change’

- Key intermediate step in marine food webs
- Abundant and relatively easy to sample
- Life cycle duration (>>phytoplankton, <<fish) allows good resolution/response at seasonal-to-interannual time scales
- No direct fishery -> reduces ambiguity about cause of observed population changes

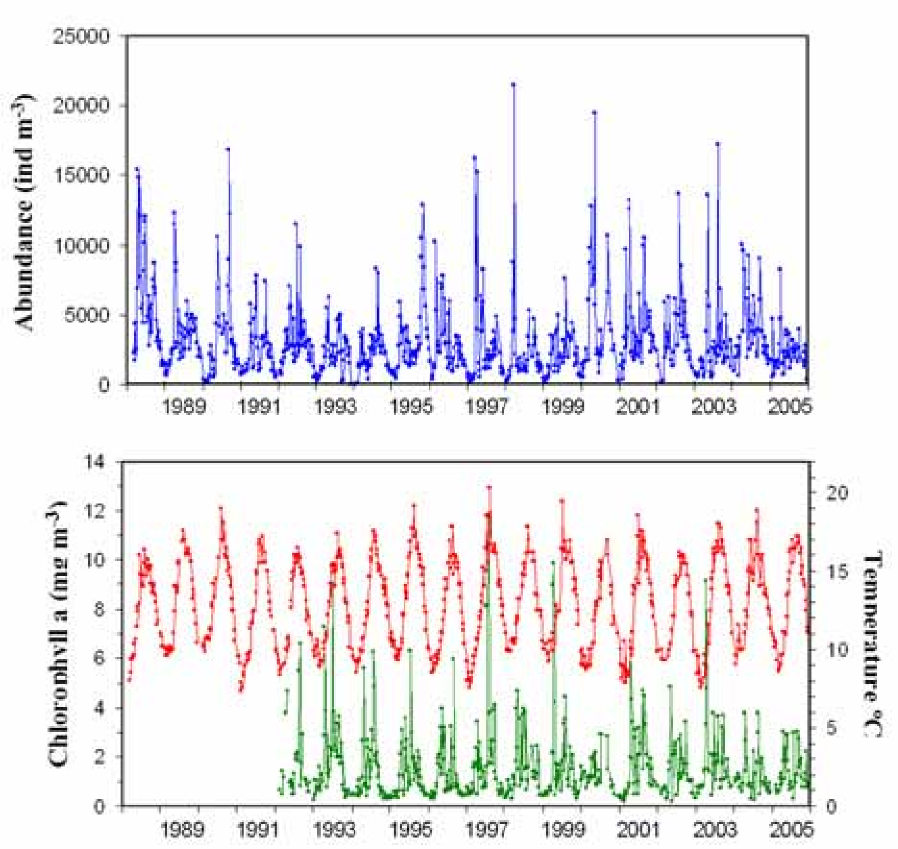

Where are the data?
Locations with long
zooplankton time series

Diverse sampling methods & schedules but most share the following traits:

- Densest coverage along continental margins
- Sampling gear fairly basic
- Best for mesozooplankton (~1mm-1 cm). Smaller, larger, or fragile taxa under-sampled.
- Sampling interval weekly-to-annual (varies with distance to home port)
- Longer time series often contain:
  - Time gaps
  - Changes in sampling grid or method
  - Changes in ‘taxonomy’

Diverse sampling designs: (1) Frequent sampling at a single site

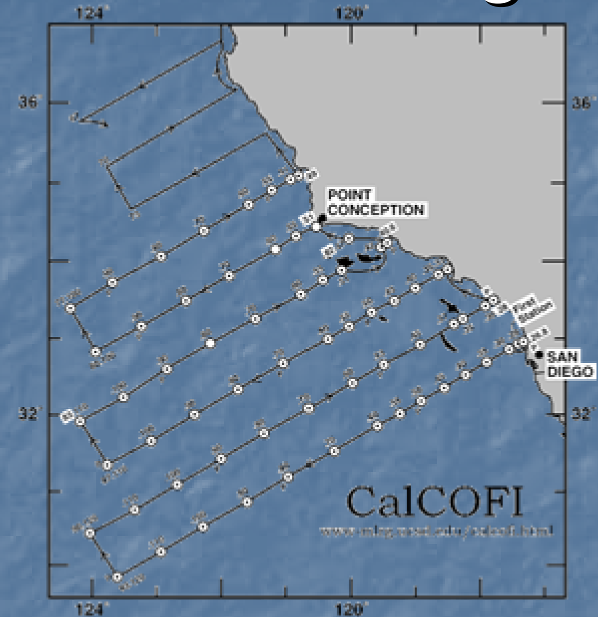
Examples:
- Plymouth L4 (at right)
- Helgoland Roads
- Naples Bay
- Station P
- Hawaii Ocean Time-series (HOT)
- Newport NH5

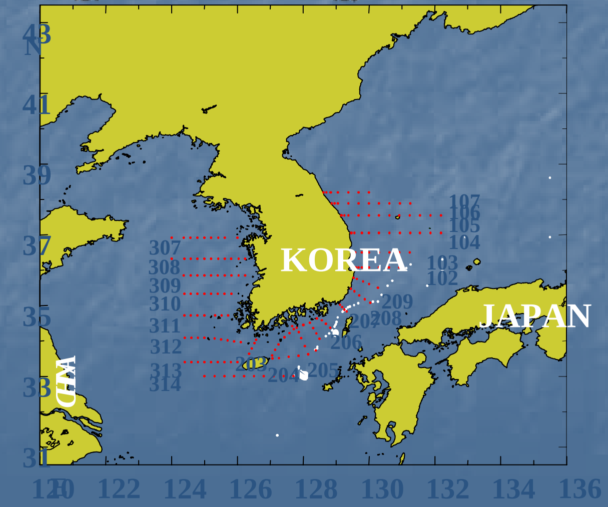


Typical Advantages:
- Frequent & regular
- Year-round, good resolution of seasonal cycle
- Ease of visualization & analysis

Typical Disadvantages:
- Lack within-sampling-period replication
- Often very nearshore

Diversity of sampling designs:
(2) Repeat surveys of ‘standard’ line or grid





Examples:

- CalCOFI (at right top)
- Korean coastal waters (at right bottom)
- Georges Bank GLOBEC
- Vancouver Island margin
- Hokkaido A-line
- Line P

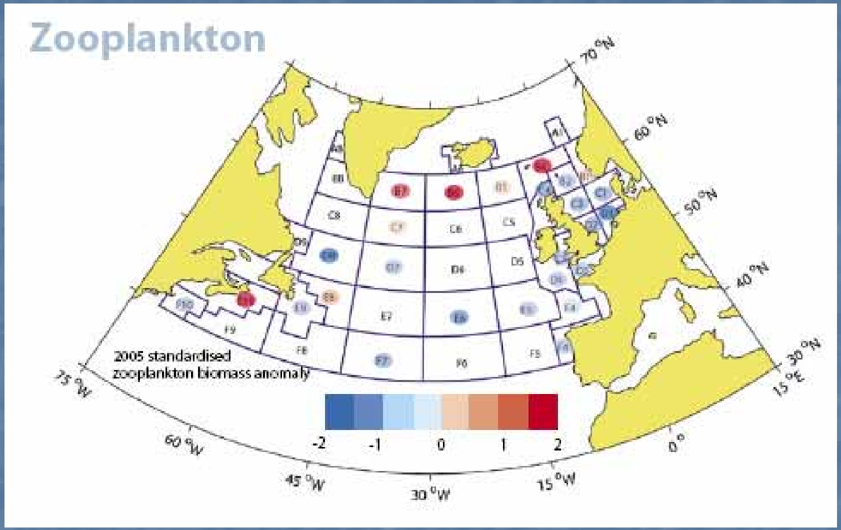
Typical Advantages:

- Several per year – fair-to-good resolution of seasonal cycle
- Can compare 2D patterns (zooplankton, chl a, T, S, currents)
- Can quantify and/or filter spatial patchiness

Typical Disadvantages:

- Mix of day and night samples
- Aliasing of phenologic variability
- May consume lots of dedicated ship time

Diversity of sampling designs:
(3) Variable locations within 'Statistical Areas'



Examples:

- Continuous Plankton Recorder (at left)
- IMARPE (Peru)
- Earlier parts of the ODATE data set (Japan)
- Benguela Current

Typical Advantages:

- Good resolution of seasonal cycle
- Economical (multi-tasked ships)
- Can classify/stratify samples based on water properties (ODATE)

Typical Disadvantages:

- Mix of day and night samples
- Variable sampling-location aliasing? of spatial structure

Diversity of sampling designs: (4) Annual (or less frequent) expeditions

Examples:

- Hokkaido University training-ship cruises
- Icelandic zooplankton monitoring program
- Most Arctic and Southern Ocean surveys

Typical Advantages:

- ‘Distant waters’ & extensive spatial coverage
- Can map changes in large scale zoogeography

Typical Disadvantages:

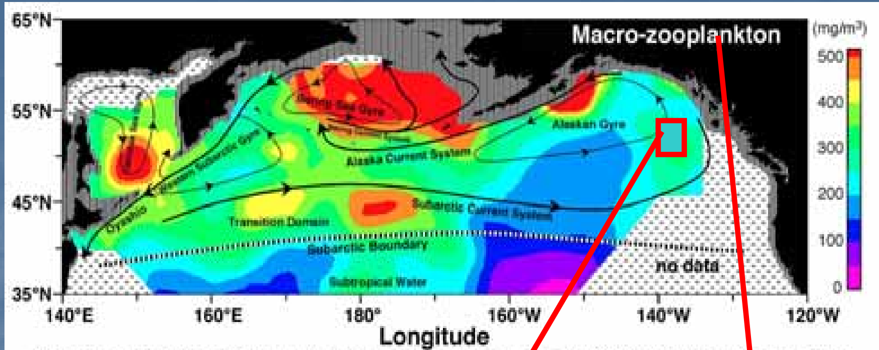
- Poor or no resolution of seasonal cycle
- Significant confounding of variability of sampling date, seasonal cycle, and phenology

All four designs face similar challenges!!

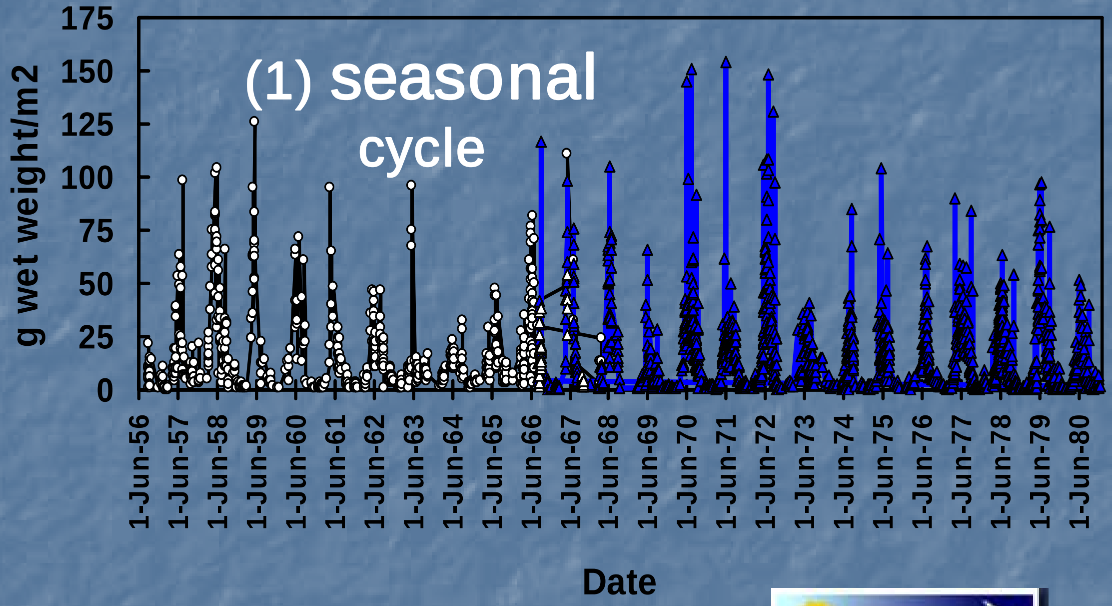
- Separating multi-year change from other large ‘real’ sources of variability (seasonal, spatial, diel, ….)
  [filtering methods, ‘anomalies’ from region climatology]
- Detecting & correcting data biases
  [intercalibration & standardization of gear & taxonomy]
- Statistical consequences of temporal autocorrelation
  [‘effective degrees of freedom’ < *n*, ensemble averaging]
- Sensitivity & consistency of analysis & interpretation
  [‘common currency’, intercalibration/standardization of data processing methods]

Other big components of zooplankton variability:

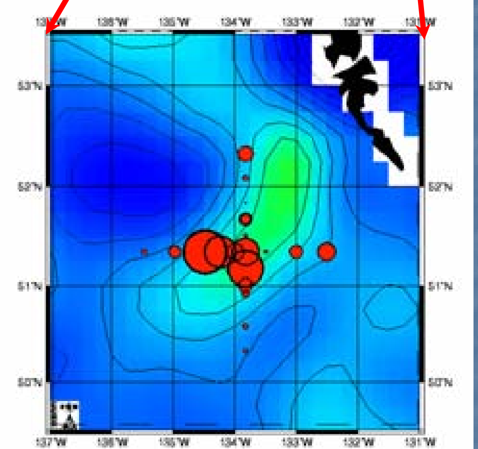
(2) persistent 'regional' patchiness



Mean horizontal distribution of summer zooplankton biomass, 1956-1994 (courtesy K. Tadokoro)



(3) transient 'small-scale' patchiness



Pseudocalanus mimus abundance (red circles) near a coastal-origin eddy (mapped sea surface height). Mackas & Galbraith 2002.

(4) Changes in vertical distribution & catchability

DVM cartoon from website of Marianne Moore, Wellesley Univ.

Selective averaging and differencing are the main strategies for partitioning zooplankton variability

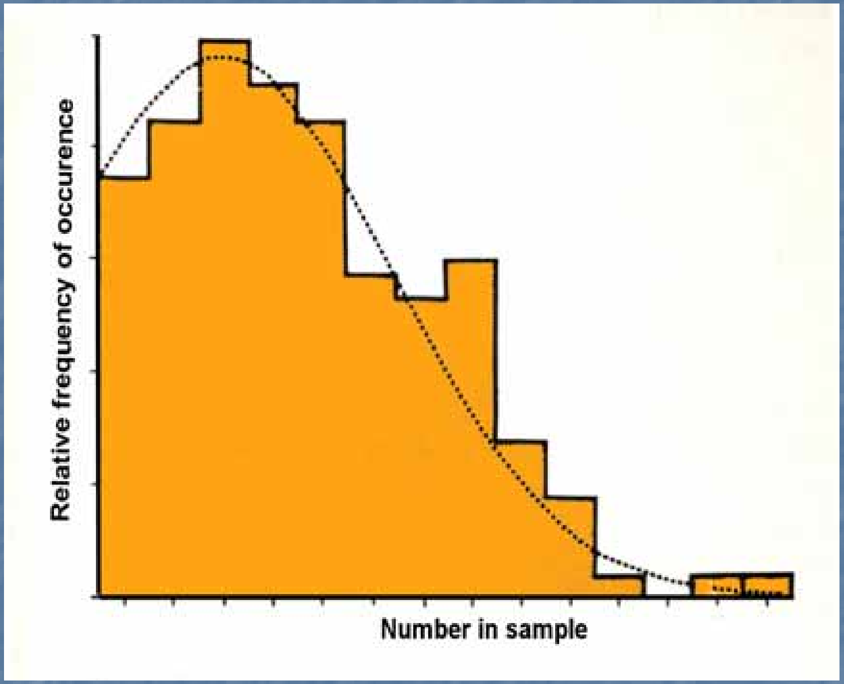
Source/scale of variability	*How dealt with*
Unresolved small-scale patchiness plus sampling error	Minimized by averaging of 'replicates' at all levels
Persistent mesoscale spatial structure	Stratification of samples into spatial averaging units
Annual seasonal cycle	'Climatology': averaging within seasons across years
Interannual & decadal variability	'Anomalies': space-time averaged deviations from climatology
'Global warming' trends	As for decadal

The art of averaging zooplankton data: Non-normal underlying statistical distributions!

- PDFs for single samples are discrete and almost always 'contagious' or 'patchy'

- Observed variance
  $\approx \mu \quad + \quad K\mu^2$ **(Cassie 1963)**
  (Poisson) (~log normal)

- Asymmetric with strong positive skew (not as big a problem as in the past, but still a no-no for many parametric tests & comparisons)



Three common alternatives for averaging (not yet a clear 'best choice' for all cases):

Arithmetic mean
$\Sigma B_t / (n-1)$

Advantages:
- Unbiased
- Easy to calculate
- Uses all data

Disadvantages:
- Noisy
- Asymmetric error bars, unless *n* very large
- Sampling biases multiplicative

Geometric mean
$(\Pi B_t)^{1/n}$
(but usually derived as arithmetic mean of $\log B_t$)

Advantages:
- Narrow & symmetric error bars
- Easy to calculate
- Uses all data
- Sampling biases additive & easy to filter (more later)

Disadvantages:
- Biased low relative to arithmetic mean
- Input data must be >0. May requires an additive data offset (<u>not</u> necessarily +1)

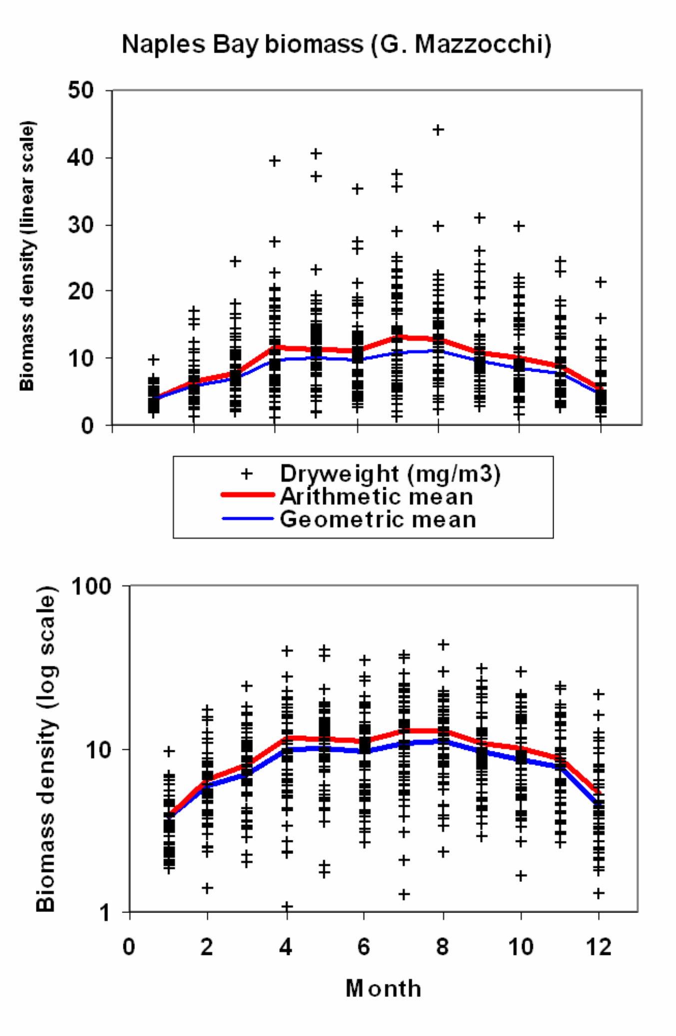
'Trim means'
(discard % from tails of distribution before arithmetic or geometric averaging)

Advantages:
- Often excludes 'bad' data
- Now available in many stat packages

Disadvantages:
- Usually biased low
- Discards good data (a problem if *n* is small)
- Subjective, behavior hard to document

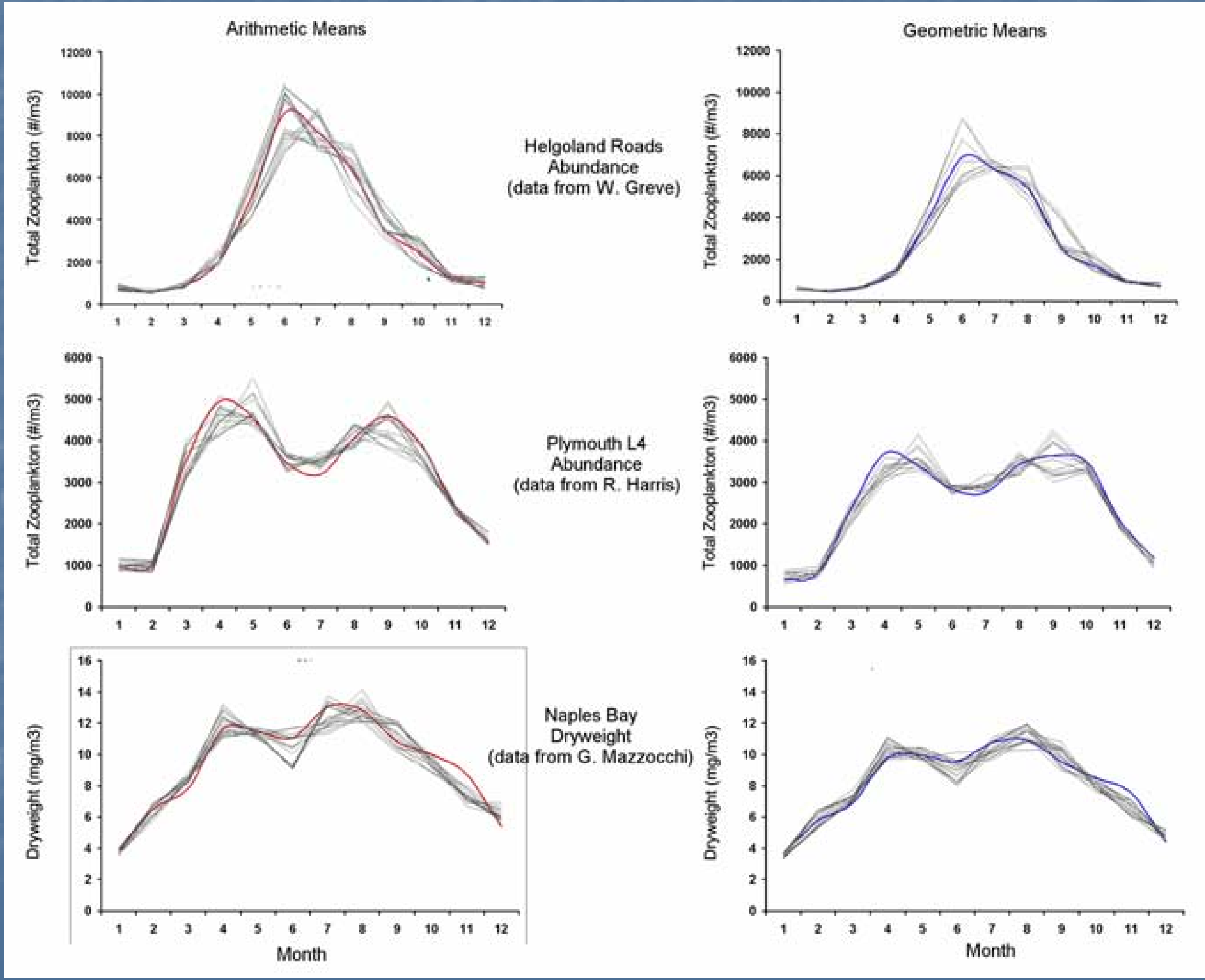
WG125 is developing averaging tools and exploring how method choice affects estimates (e.g. of seasonal cycle)

- The densely-sampled coastal time series (Naples, L4, Helgoland) are especially useful for this.
- If *n* is large:
  - Both arithmetic and geometric means have very similar patterns (compare red & blue lines)
  - Bias of the geometric mean is present but small compared to the overall scatter (black symbols)

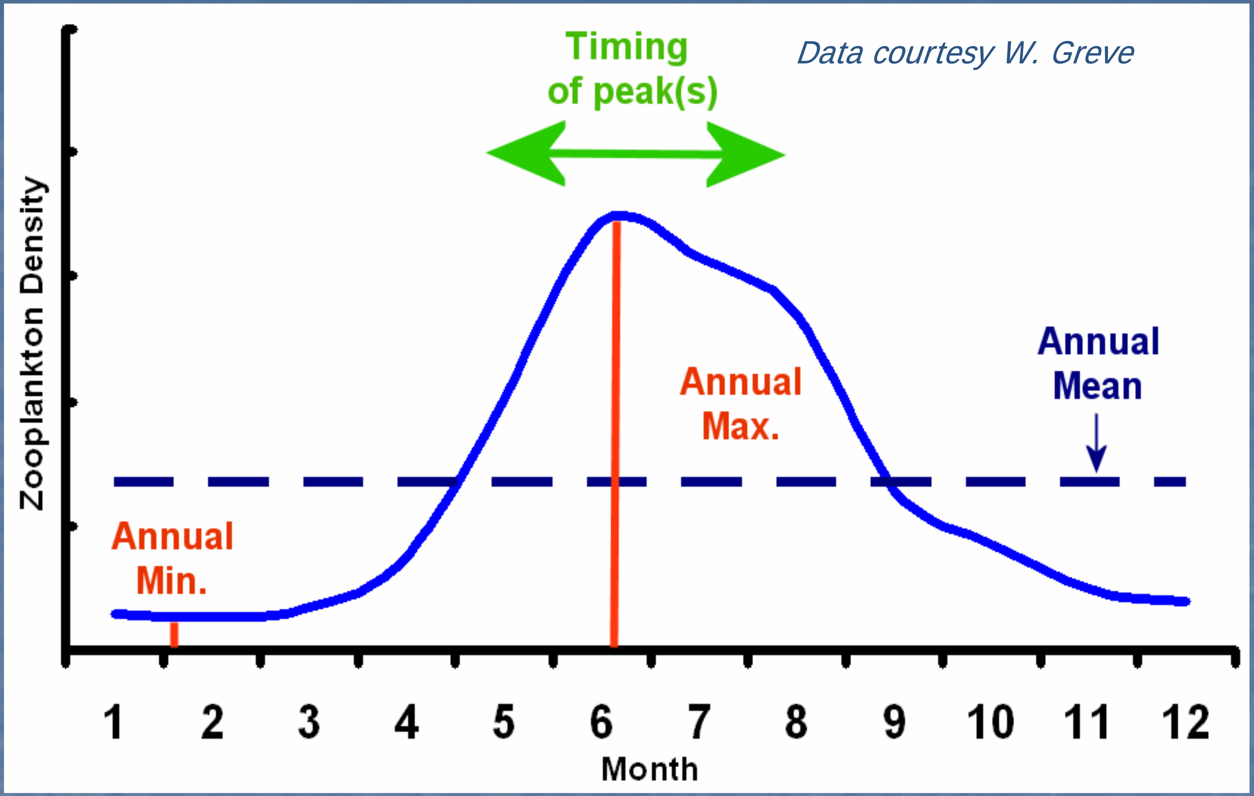


What if the sampling is less dense (*n* small or time coverage gappy)?

- Climatologies calculated from randomized 50% subsample (grey lines) track the '100%' patterns (blue or red lines)
- Peak timing more erratic than level
- Downward bias of geomean is proportional to within-time-period variance



What among-year comparisons might we want?



- In principle, can estimate all of the above from low-order harmonic fits to each year's data (e.g. Dowd et al. 2004)

$$B_t = a + \Sigma Cj^*\sin(jt+\Phi_j)$$

- In practice, data gaps & lack of within-time-period replication limit usefulness of this approach

Alternative for among year comparison: Derive anomalies A_t (differences of data B_t from climatology $\bar{B}$) for each observation period

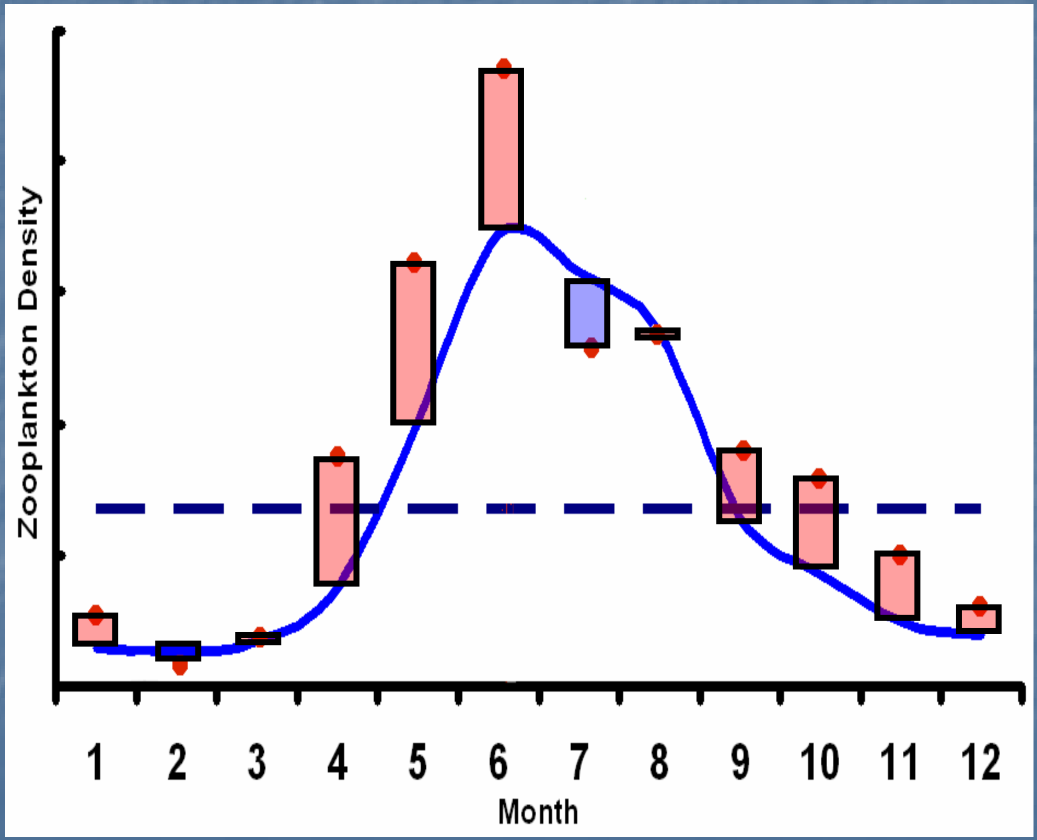
- Anomalies (mostly + here) can be averaged to give an annual anomaly A_y
- Three commonly used equations:

$$A_t = \log(B_t/\bar{B})$$

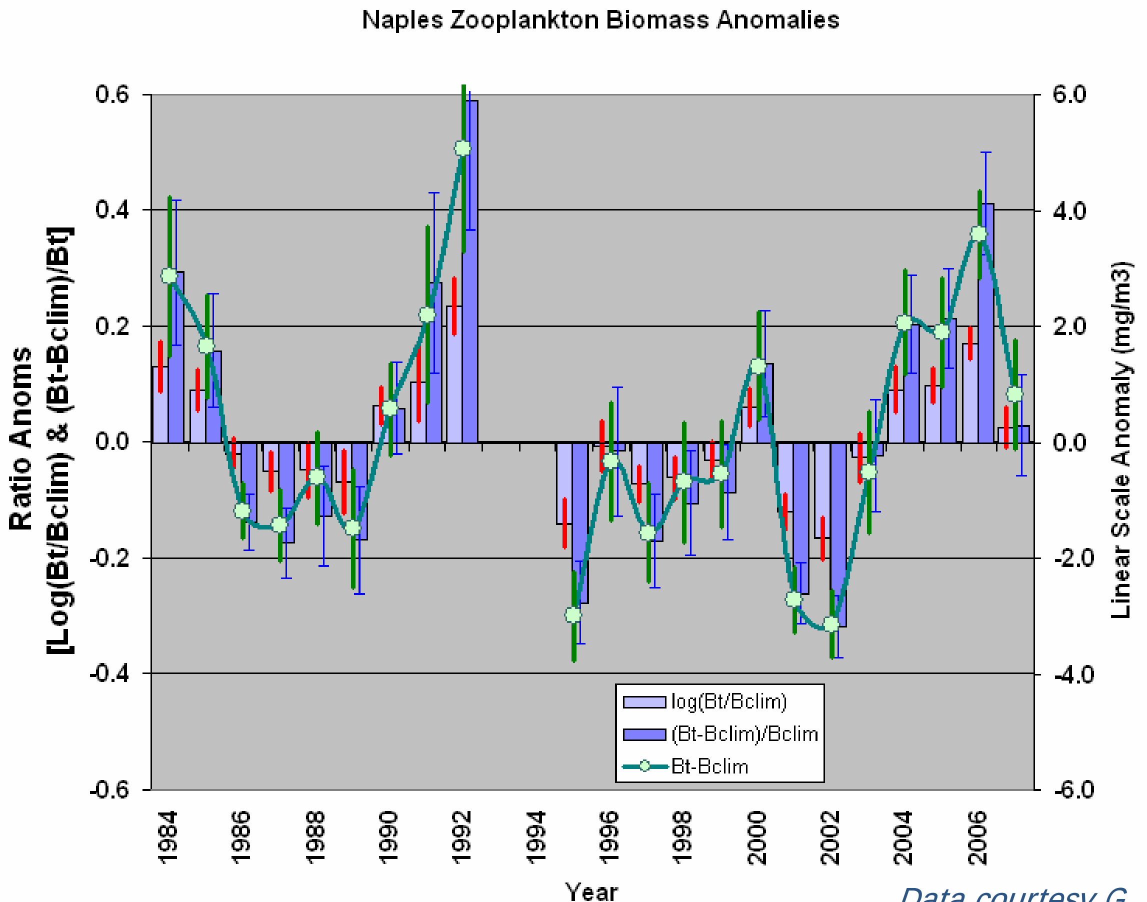
$$A_t = (B_t - \bar{B})/\bar{B}$$

$$A_t = B_t - \bar{B}$$

The first two are unit-free ratios measuring % change (more later on why this is useful). The last has the same units as the raw data.

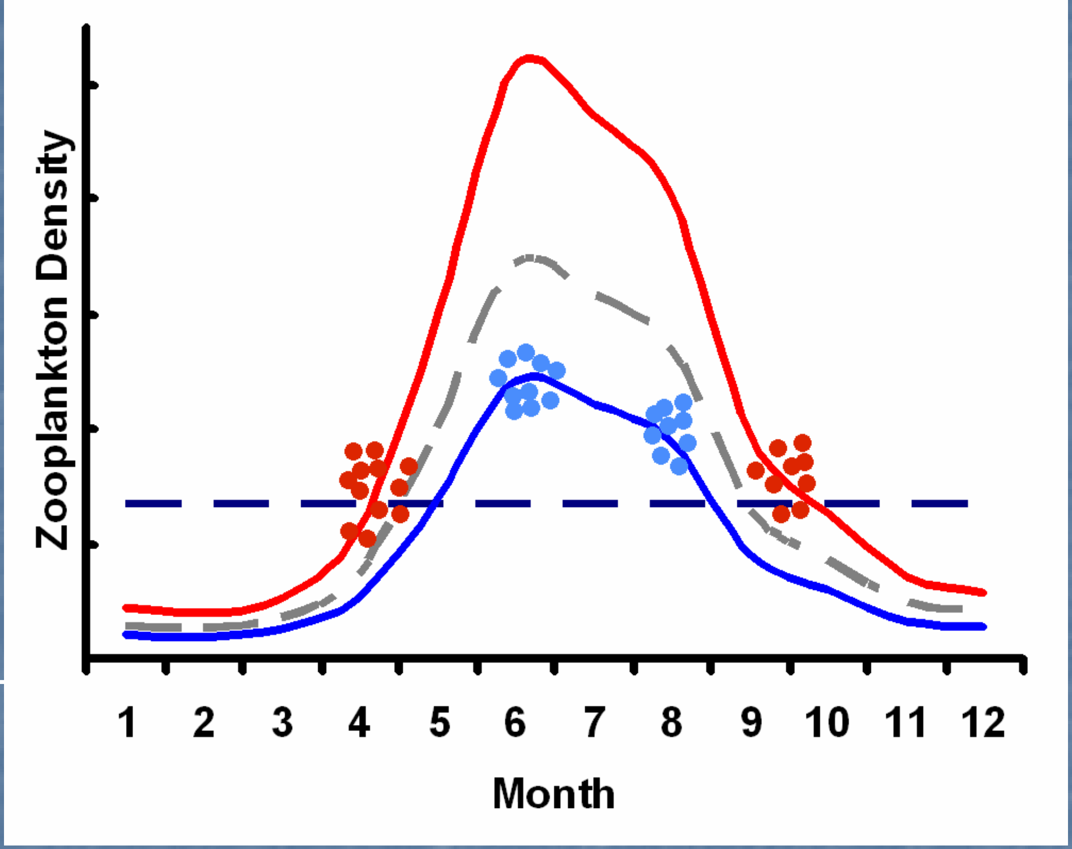


With sufficient data, all three A_t estimates yield similar multi-year time series. The log scale index is usually less noisy



Data courtesy G. Mazzocchi

Beware of using overall annual mean for $\bar{B}$ (especially if seasonality is strong and sampling is infrequent)!!!



Measurements obtained before or after the annual peak in a "good year" (red) will often be lower then measurements made closer to the peak in a "bad year" (blue)

Comparisons between zooplankton time series

- Regional time series differ in sampling methodology, depth range, gear bias, and/or units of measurement.
- Resistant to conversion to externally-imposed 'standard methods' because:
  - Method changes risk loss of time continuity (a serious concern both internally and externally)
  - Local methods are usually optimized for local conditions (Have you ever processed 100µ-mesh net samples from an upwelling region??)
  - Costs of change are local, benefits largely external
  - Biological oceanographers are rabid individualists: "I'd sooner use his toothbrush than his #*!&@ method".

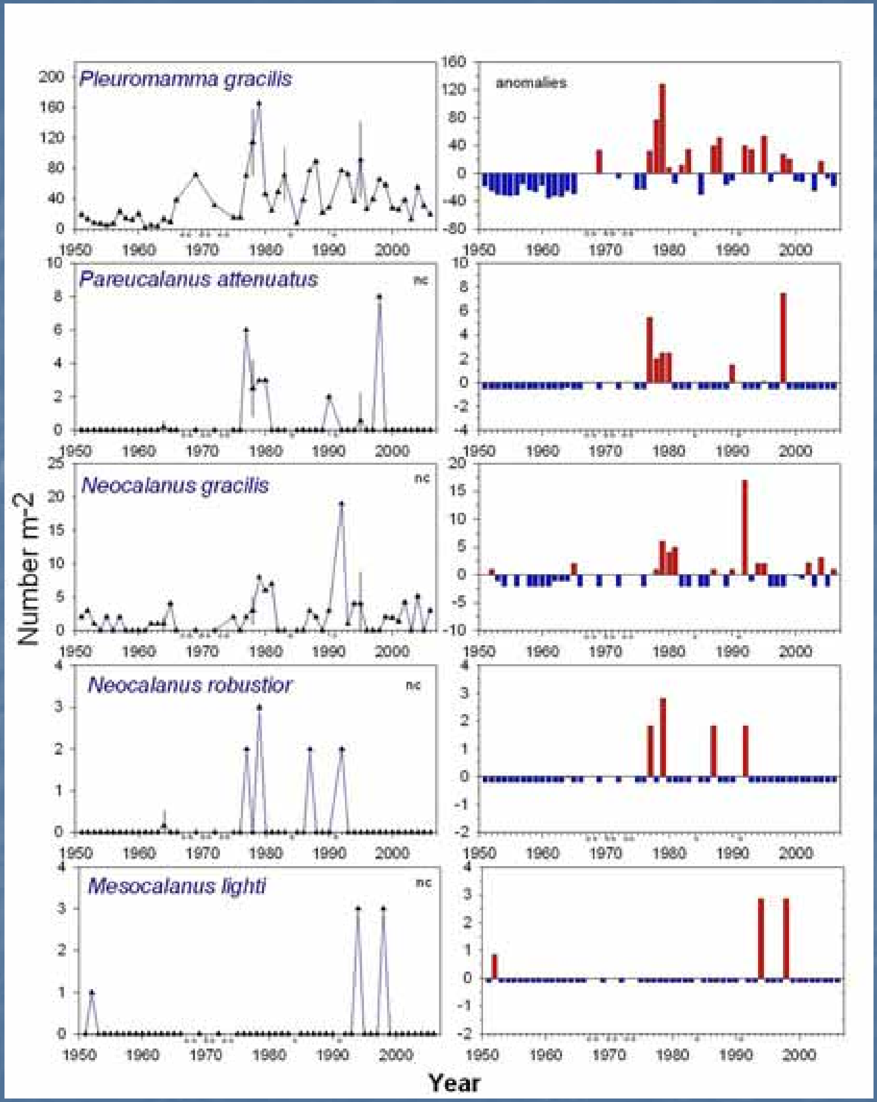
Nevertheless, the dimensionless anomalies A_t provide a basis for quantitative comparison!

- Any and all programs can compare ratios of local within-survey means (of one-to-many samples) to local climatologies (based on many samples)
- With sufficient averaging, the outputs reliably show changes in relative amount (2x more, 3x less, ...) AND
- Filter out seasonal cycle, spatial patchiness, AND
- Filter out any persistent local gear bias.
  This is because the bias c is in both the local data [numerator] and the local climatology [denominator]

$A_t = c(B_t - \bar{B}) / c\bar{B}$ or $A_t = \log (cB_t / c\bar{B})$

Comparison/compositing of species time series:

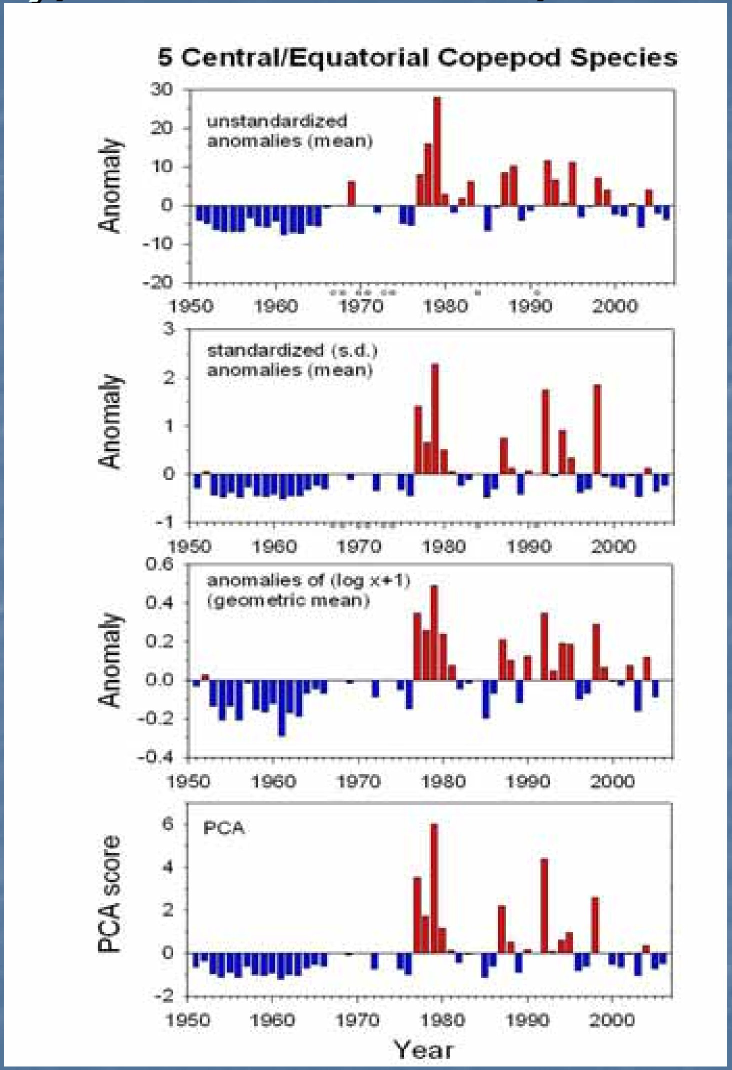
- Taxa sharing ecological niche or zoogeographic affinity often have similar time series, but differ greatly in mean and peak abundance
- As abundance gets lower, 'noise' increases (variance progressively more dominated by Poisson-distributed sampling error)
- How can we best identify 'shared pattern'?



Abundance & anomaly time series of 5 "central/equatorial Pacific" species in the CalCOFI region

Choices for averaging/merging of single species anomalies: Linear, Ratio (linear or log-scale), 'standardization', PCA

- Linear & unstandardized – outcome weighted by mean abundance (therefore strongly dominated by abundant taxa)
- Standardization – all taxa (and their associated noise) have identical weight
- 'Ratio' – average weighted by % change for each species
- Principal Components – large covariance (+ or -) emphasized, 'species-specific' components are isolated (& usually ignored)



Outcome of manipulations: all derived time series are strongly correlated
'Ratio' is the smoothest.
PCA? or 'Ratio' ? best captures the core shared pattern

WG125 Data Visualization Tools

A posteriori inference is very prone to misuse!!

Despite the above:

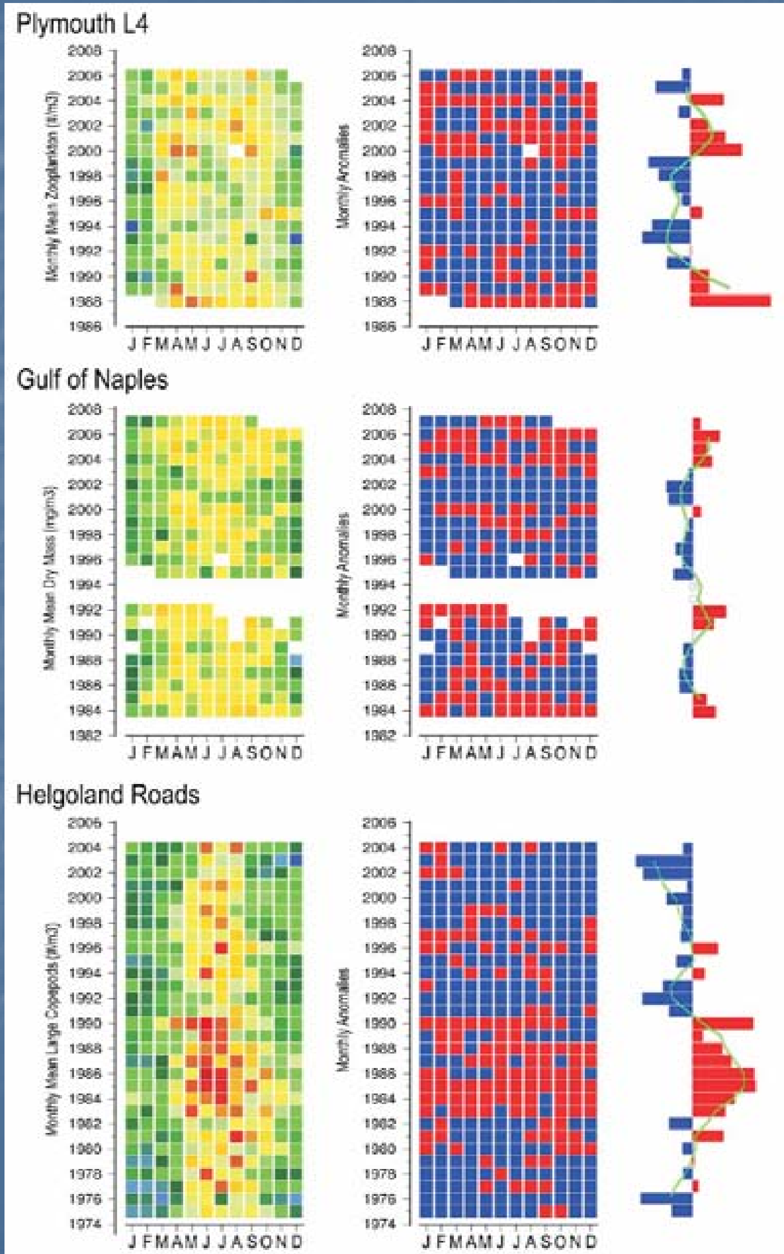
- "You learn a lot by looking" (Berra)
- 'WYSIWYG' = 'What you see is what you get' (Apple/MacIntosh?)
- 'WYDSYWNC'= " What I don't see somehow, I will never consider" (DLM + many others?)

A prioiri data exploration is risky but also very useful. Let's us compare across years and regions:

- Raw time series
- Within-year-anomalies,
- Annual anomalies

Lesson learned:

- Anomalies are often truly 'annual' to 'interannual' in duration



Planned additions to the WG125 'Tool Kit'

- More visualization tools
- Methods for comparing 'synchrony' and 'abruptness' of changes
- 'Free-ware' routines (probably mostly in 'R')
- Consequences of spatial and temporal autocorrelation
- More on covariance among species and various measures of 'condition'

If you have a zooplankton time series you are willing to include in the comparison, please contact us.

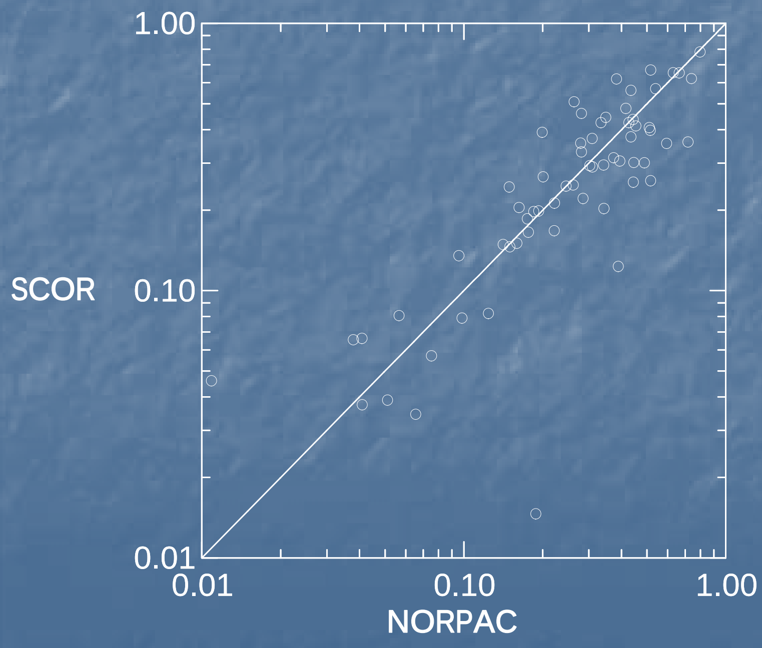
Spares for questions

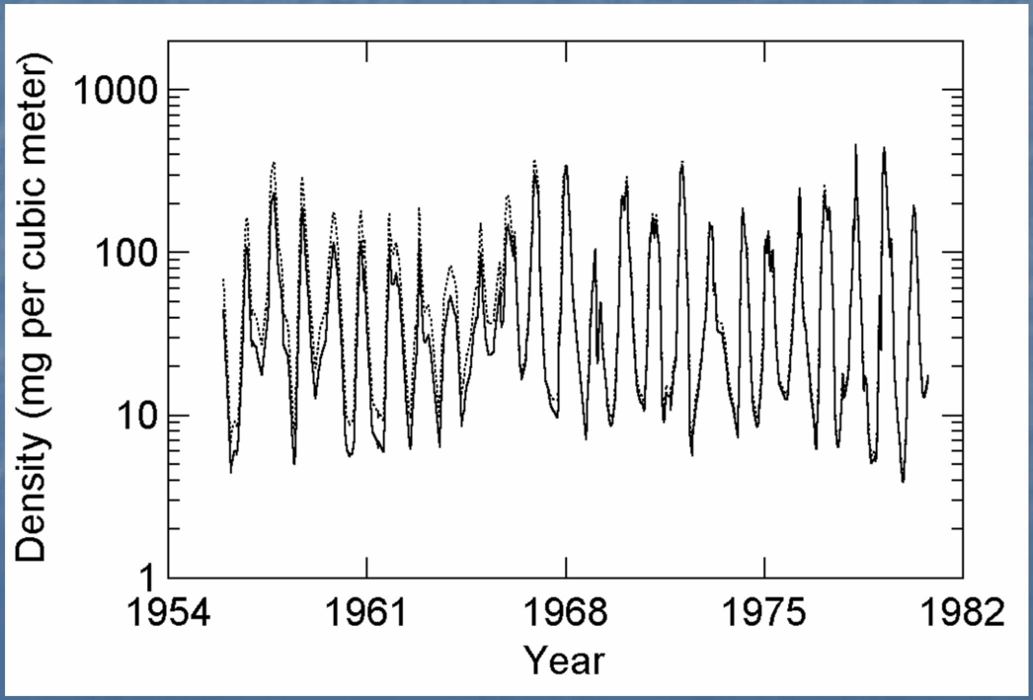
Standardization & intercalibration of sampling methods

Method changes risk loss of time continuity (a serious concern)

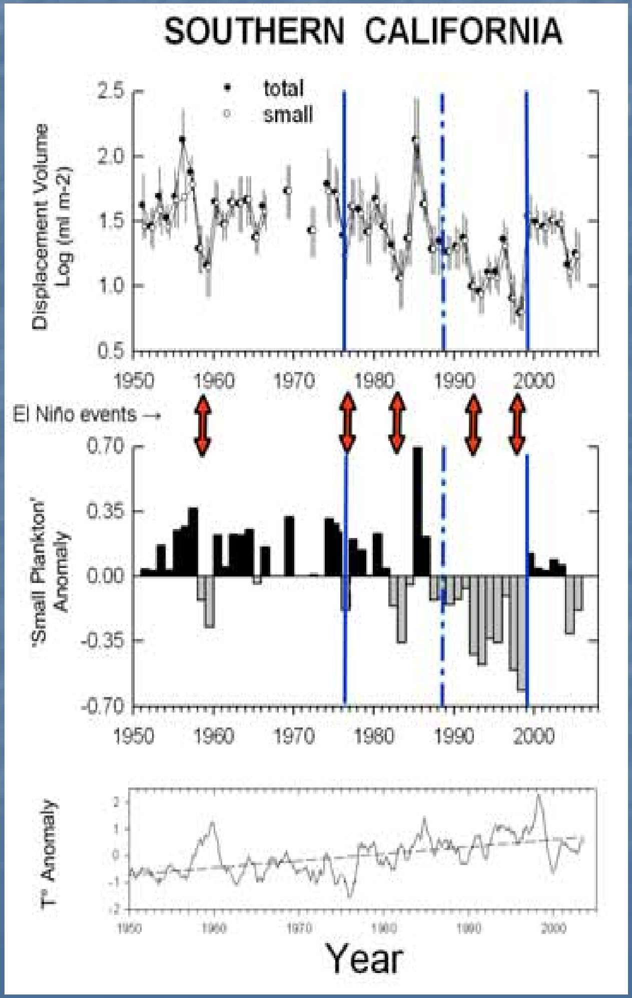
Net intercalibrations are feasible (e.g. McKinnell & Mackas 2003):

- Mesh size, flow metering, tow depth all matter, <u>but</u>
- Modern nets (if flow-metered) perform similarly
- Correction factors often small compared to real signal

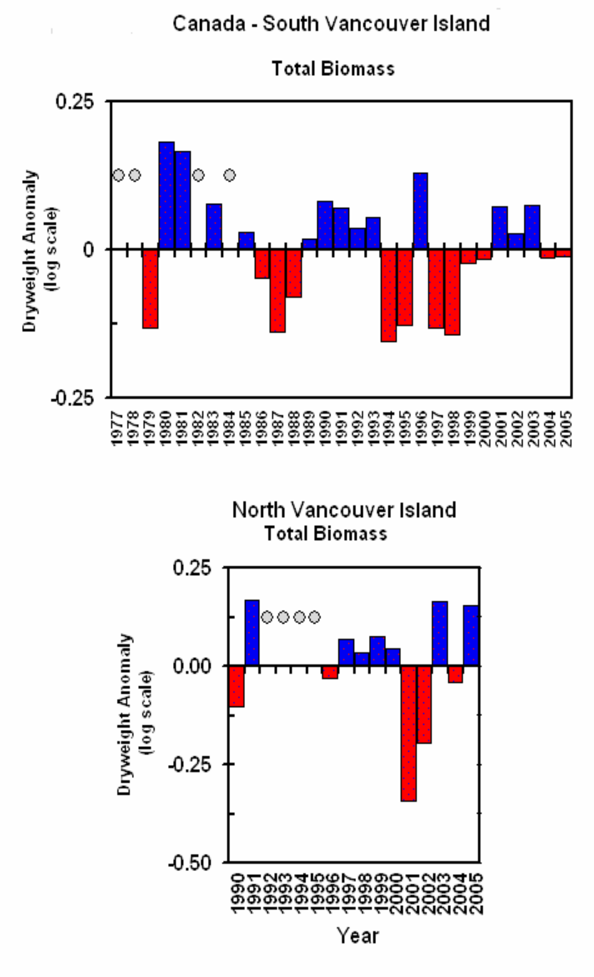




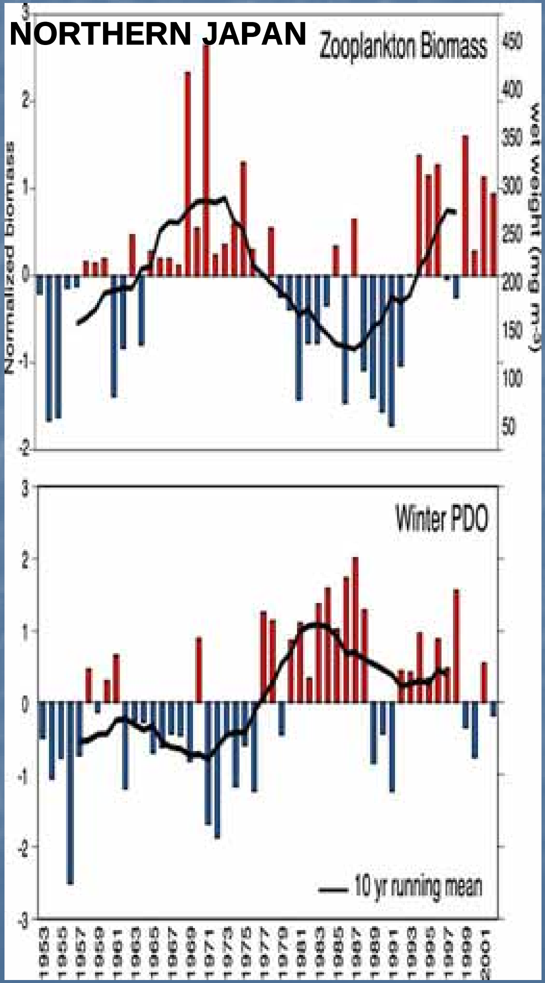
Many North Pacific examples of interannual change in total zooplankton biomass: (continental margins)



Ohman, updated from Lavaniegos & Ohman 03



Mackas et al 06



Chiba et al 05

Summary of zooplankton biomass variability:

Amplitude - 3 to 5 fold range

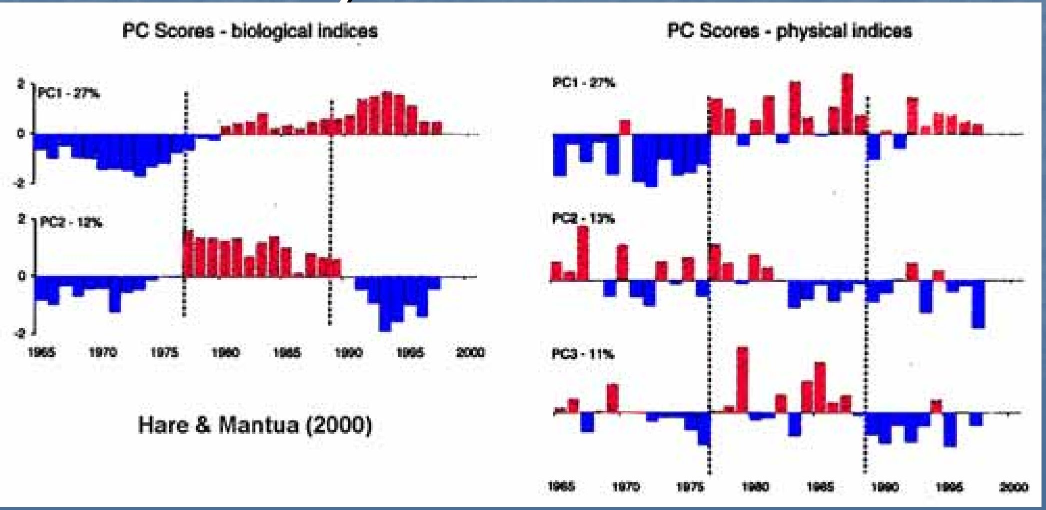
Important time scales (~same as physical environment):

interannual = 1-3 year duration

decadal 'regimes' = 5-20 years, abrupt transitions?

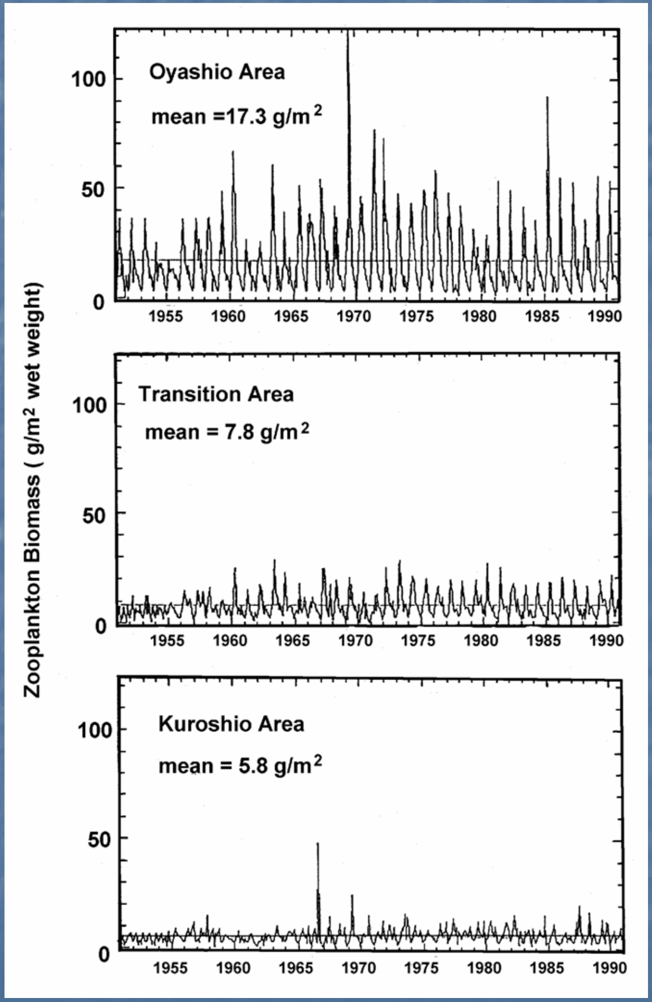
'climate trend' = 50-100 years and longer

Covariance: With ocean climate (stratification, winds & currents) and with various fishery indices (catch, survival, recruitment)

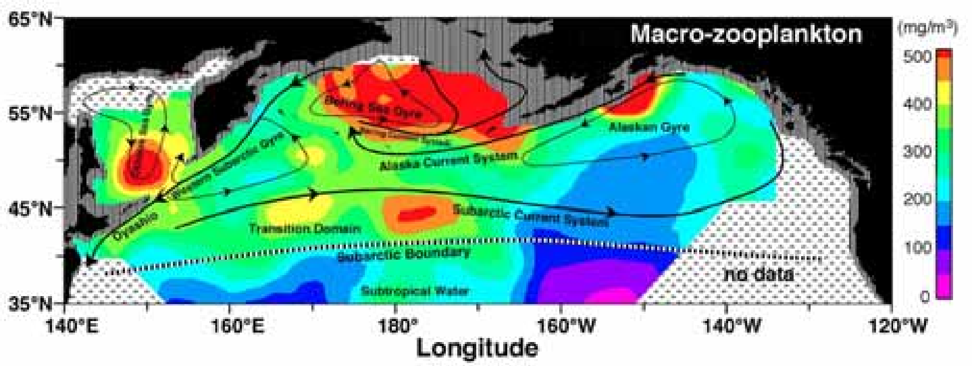


(1) & (2) can be reduced/eliminated by estimating and subtracting a baseline climatology (log transformation often useful)

(1) seasonal cycle



(from K. Odate, 1994)



Mean horizontal distribution of summer zooplankton biomass, 1956-1994 (courtesy K. Tadokoro)

(2) persistent 'regional' spatial pattern

Standardization & intercalibration of sample processing and data analysis:

Continuity/comparability of ‘taxonomy’ and ‘archival categories’:

- Rarely addressed
- Worst dangers come when data user $\neq$ data originator
- Safer to analyze at less than full taxonomic resolution??

Standardization/intercomparison of data analysis

- Also needs more work (or play)