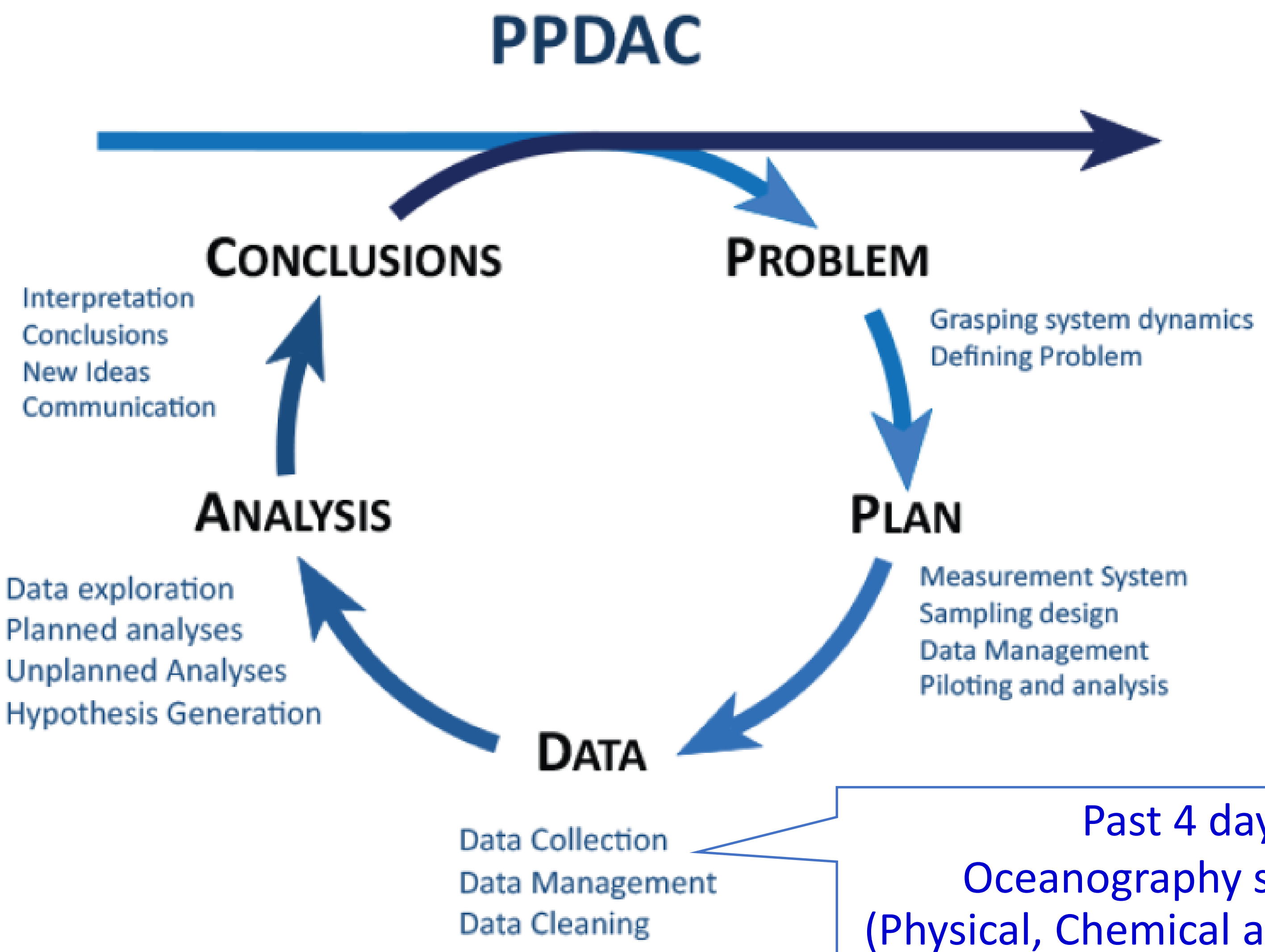
Statistic Test on the Relationship Between Oceanic Environment and Fishery Abundance

Online Training-course

Regional Online Training Course on the Relationship Between Ocean Environment Variability and Marine Resource Abundance and Oceanographic Sampling

26 November 2021

The process of abstracting and solving a statistical problem

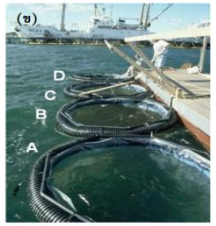
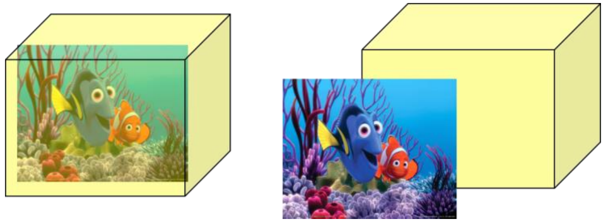


Today

Past 4 days;
Oceanography sampling
(Physical, Chemical and Biological)

Wild C. 2018. The place of data analysis in problem solving. The University of Auckland. 3 p.

Source of data for the study of relationship between environment and fishes

		Control	Replication	Realism	*In Situ* (Y/N)
Laboratory					N
Microcosm					Y or N
Mesocosm					Y or N
Enclosure / Exclosure					Y
Whole-system					Y

Caution on pseudoreplication

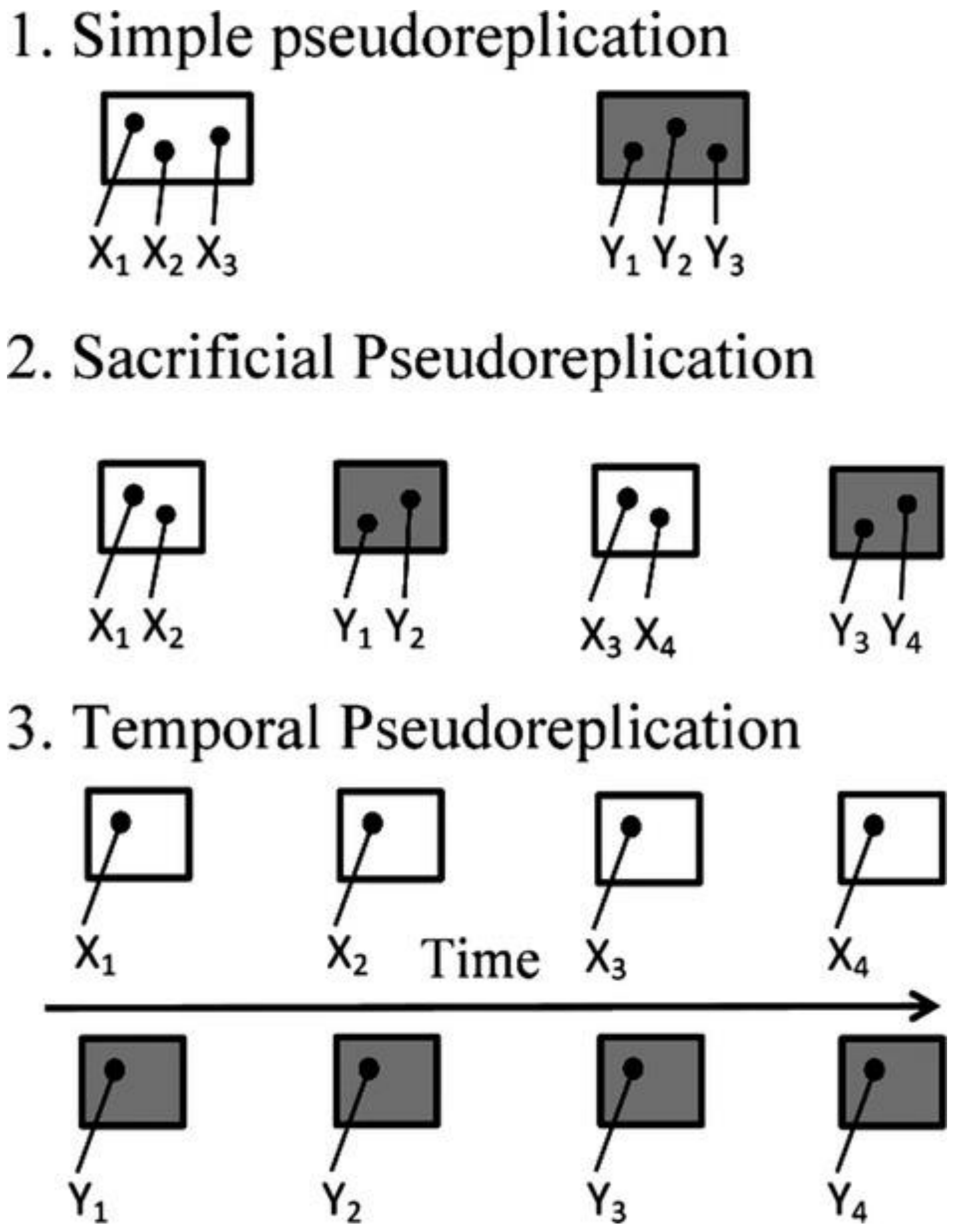
Ecological Monographs, 54(2), 1984, pp. 187–211

PSEUDOREPLICATION AND THE DESIGN OF ECOLOGICAL FIELD EXPERIMENTS[1]

STUART H. HURLBERT
Department of Biology, San Diego State University, San Diego, California 92182 USA

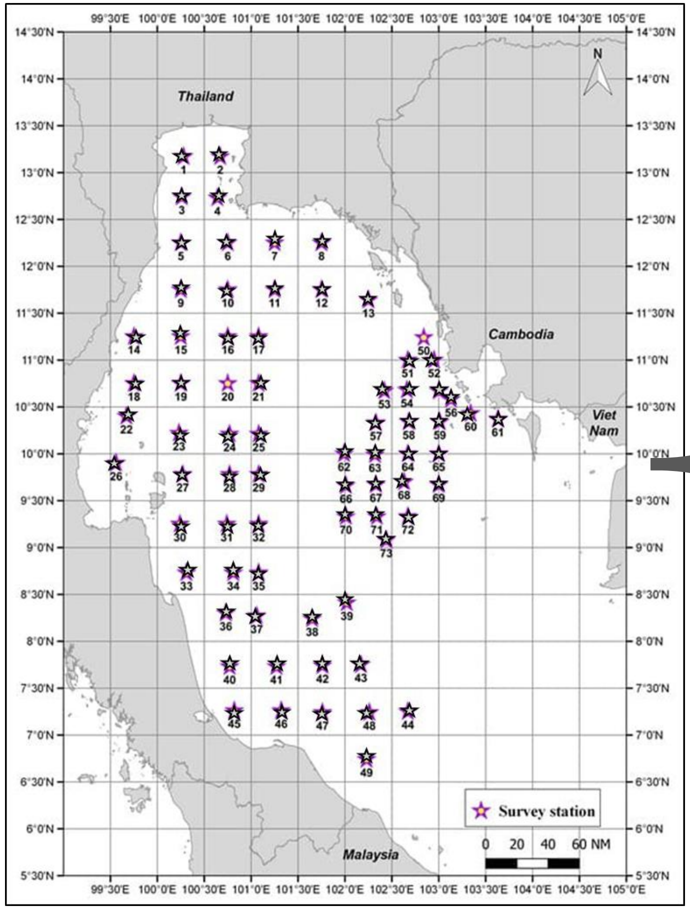
Abstract. Pseudoreplication is defined as the use of inferential statistics to test for treatment effects with data from experiments where either treatments are not replicated (though samples may be) or replicates are not statistically independent. In ANOVA terminology, it is the testing for treatment effects with an error term inappropriate to the hypothesis being considered. Scrutiny of 176 experimental studies published between 1960 and the present revealed that pseudoreplication occurred in 27% of them, or 48% of all such studies that applied inferential statistics. The incidence of pseudoreplication is especially high in studies of marine benthos and small mammals. The critical features of controlled experimentation are reviewed. Nondemonic intrusion is defined as the impingement of chance events on an experiment in progress. As a safeguard against both it and preexisting gradients, interspersion of treatments is argued to be an obligatory feature of good design. Especially in small experiments, adequate interspersion can sometimes be assured only by dispensing with strict randomization procedures. Comprehension of this conflict between interspersion and randomization is aided by distinguishing pre-layout (or conventional) and layout-specific alpha (probability of type I error). Suggestions are offered to statisticians and editors of ecological journals as to how ecologists' understanding of experimental design and statistics might be improved.

Key words: experimental design; chi-square; R. A. Fisher; W. S. Gossett; interspersion of treatments; nondemonic intrusion; randomization; replicability; type I error.



https://ecologycriticalthinking.wordpress.com/2015/10/23/pseudo-replication-and-experimental-design/

General data matrix

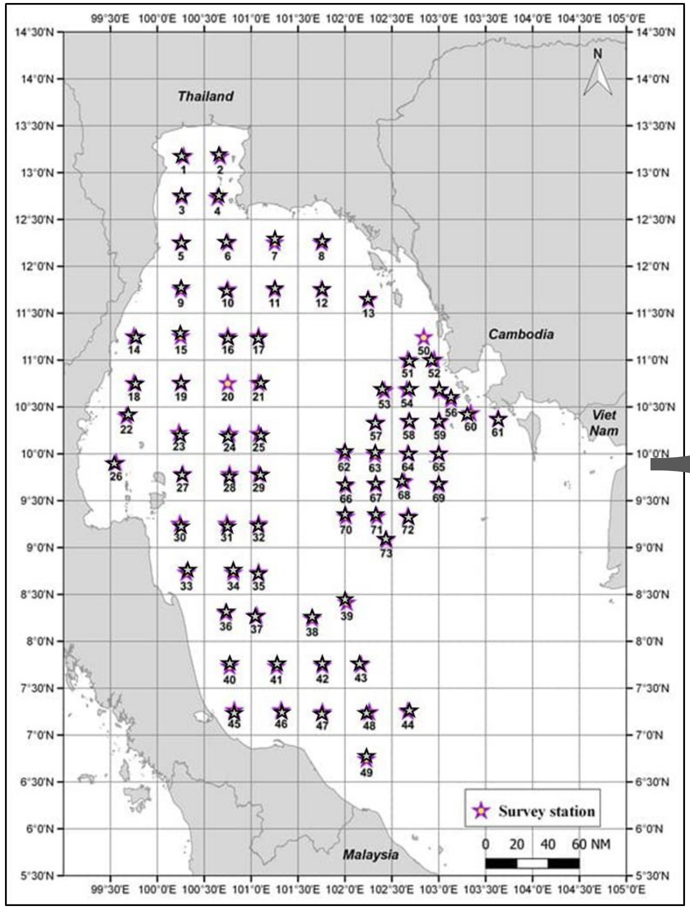


Oceanographic data (physical & chemical variables)

Data of fisheries resources (numbers, weight, CpUE, Present'& Absence)

	V1	V2	V3	V4	V5	sp1	sp2	sp3	sp4	sp5	...	sp20
site1	0.1	18	3	312	0.3	0	1	0	1	0	...	1
site2	0.2	15	4	526	0.1	1	0	0	0	1	...	0
site3	0.6	14	1	489	0.2	0	1	1	1	1	...	0
site4	0.05	19	2	523	0.5	0	0	0	0	1	...	1
site5	0.4	13	3	214	0.6	1	1	1	1	1	...	0
site6	0.8	12	4	265	0.4	1	1	0	1	0	...	1
site7	0.2	15	1	236	0.8	0	0	1	0	0	...	0
site8	0.1	17	3	541	0.1	0	1	1	1	0	...	1
site9	0.6	16	2	845	0.3	0	0	1	1	0	...	0
site10	0.1	18	1	126	0.1	1	1	0	0	1	...	0
...	...	...	...	...	...	...	...	...	...	...	...	...
site100	0.04	15	3	235	0.6	1	1	1	1	0	...	0

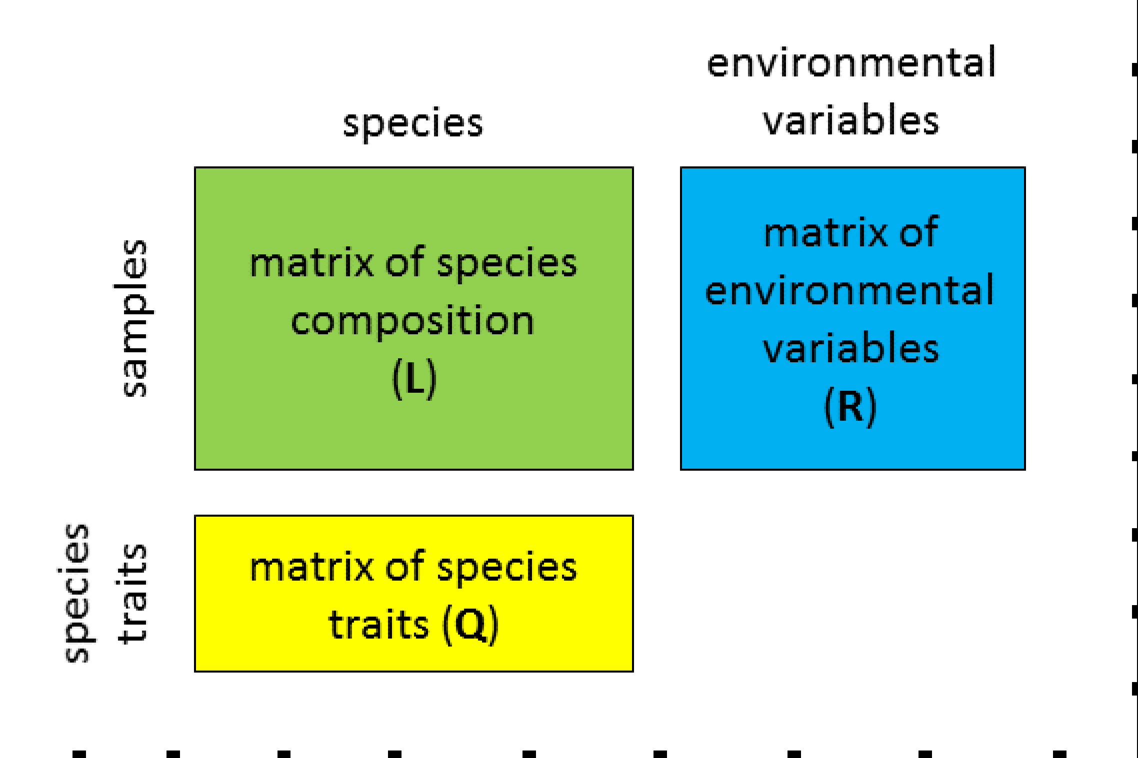
General data matrix



Oceanographic data (physical & chemical variables)

Data of fisheries resources (numbers, weight, CpUE, Present'& Absence)

	V1	V2	V3	V4	V5	*sp1*	*sp2*	*sp3*	*sp4*	*sp5*	...	*sp20*
site1	0.1	18	3	312	0.3	0	1	0	1	0	...	1
site2	0.2											0
site3	0.6											0
site4	0.05											1
site5	0.4											0
site6	0.8											1
site7	0.2											0
site8	0.1											1
site9	0.6											0
site10	0.1											0
...	...											...
site100	0.04	15	3	235	0.6	1	1	1	1	0	...	0



Analytical approaches

	V1	V2	V3	V4	V5	sp1	sp2	sp3	sp4	sp5	...	sp20
site1	0.1	18	3	312	0.3	0	1	0	1	0	...	1
site2	0.2	15	4	526	0.1	1	0	0	0	1	...	0
site3	0.6	14	1	489	0.2	0	1	1	1	1	...	0
site4	0.05	19	2	523	0.5	0	0	0	0	1	...	1
site5	0.4	13	3	214	0.6	1	1	1	1	1	...	0
site6	0.8	12	4	265	0.4	1	1	0	1	0	...	1
site7	0.2	15	1	236	0.8	0	0	1	0	0	...	0
site8	0.1	17	3	541	0.1	0	1	1	1	0	...	1
site9	0.6	16	2	845	0.3	0	0	1	1	0	...	0
site10	0.1	18	1	126	0.1	1	1	0	0	1	...	0
...	...	...	...	...	...	...	...	...	...	...	...	...
site100	0.04	15	3	235	0.6	1	1	1	1	0	...	0

Explanatory variables

Variables to be explained (Responses)

Variables → MODELs

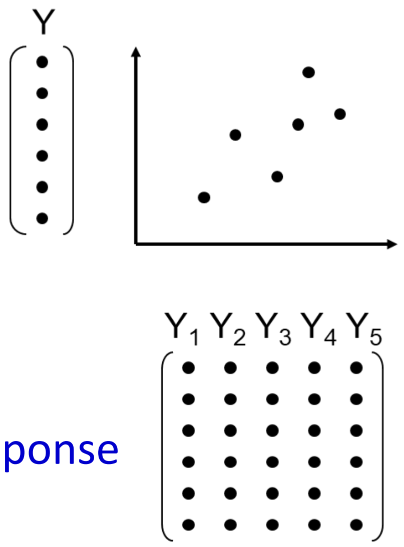
Unsupervised models

Explanatory variables (*independents*) → MODELs → Variables to explain or Responses (*dependents*)

Supervised models

One response

More than 1 response



Principle analysis on fishery resources

Diversity indices

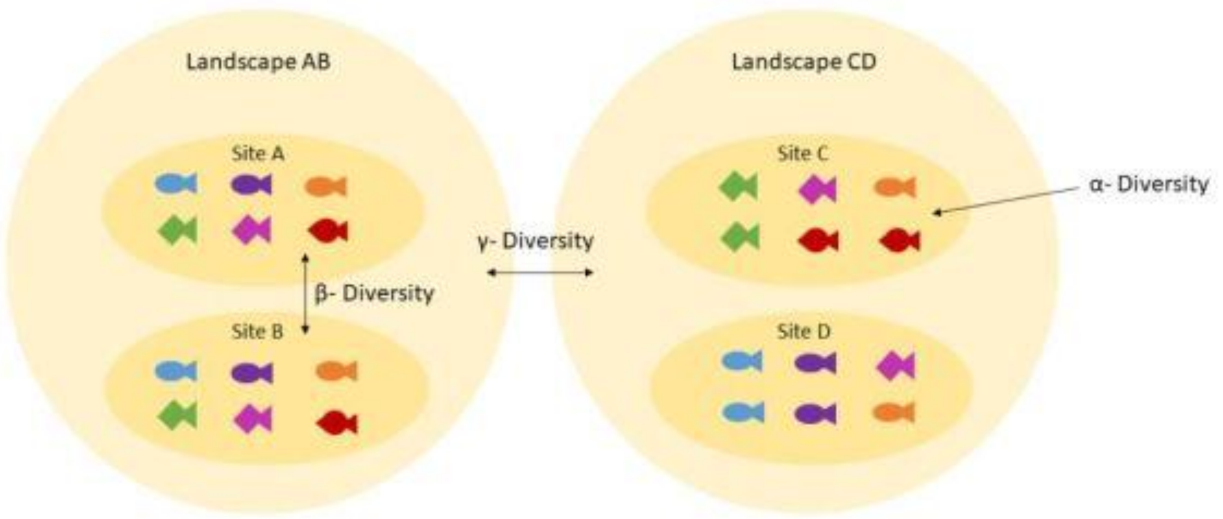
- Richness: Simply the number of species found in each site
- Shannon index: $H' = -\sum_{i=1}^{R} p_i \ln p_i$, where p_i is the proportion of characters belonging to the i^{th} type of letter in the string of interest.
- Simpson index: $\lambda = \sum_{i=1}^{R} p_i^2$ where R is richness, i.e. the total number of species in the dataset, and this index can be presented as "Inverse Simpson index": $\frac{1}{\lambda} = \frac{1}{\sum_{i=1}^{R} p_i^2}$
- Eveness: $J' = \frac{H'}{H'_{\max}}$, where H' is the number derived from the Shannon diversity index and H'_{max} is the maximum possible value of H' and $H'_{\max} = -\sum_{i=1}^{S} \frac{1}{S} \ln \frac{1}{S} = \ln S$., where S is number of species

Principle analysis on fishery resources

Diversity indices

- *Alpha diversity*: the species diversity within a community at a small scale or local scale, generally the size of one ecosystem.
- *Beta diversity*: the species diversity between two communities or ecosystems, i.e. comparative approach.
- *Gamma diversity*: species diversity is compared among many ecosystems.

	Response 1	Response 2	Response 3	…	Response n
Sample 1					
Sample 2					
Sample 3					
…					
Sample n					



https://www.differencebetween.com/what-is-the-difference-between-alpha-beta-and-gamma-diversity/

Principle analysis on fishery resources

Diversity indices

RAFFLES BULLETIN OF ZOOLOGY Supplement No. **32**: 85–94
Date of publication: 6 May 2016
http://zoobank.org/urn:lsid:zoobank.org:pub:28B921A9-5C59-46D8-83DF-2A3483BDF1EF

Assemblages and diversity of fishes in Singapore's marinas

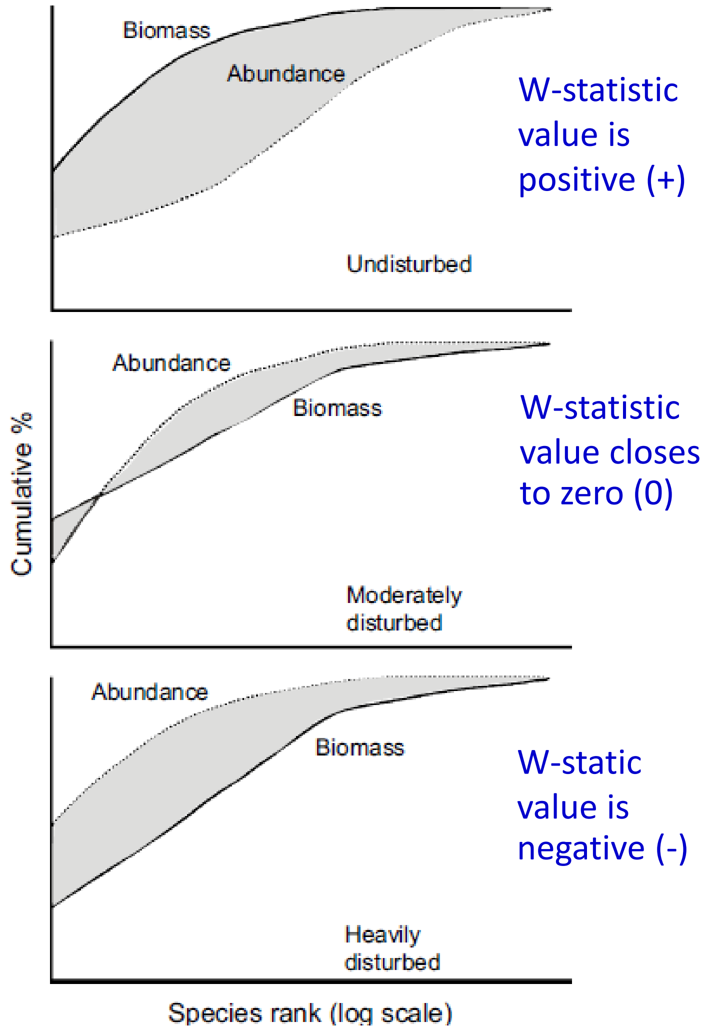
Kok Ben Toh[1,2], Chin Soon Lionel Ng[1,2], Wai-Kit Gavan Leong[1], Zeehan Jaafar[1,3], Loke Ming Chou[2]

Table 1. Average number of fishes caught per trap set, number of species (*S*), number of families (F), Margalef diversity index (D), Shannon's diversity index (H') and Pielou's evenness Index (J') of fishes in ONE°15 Marina Club (OMC), Marina at Keppel Bay (MKB), Raffles Marina (RM).

Site	Average Catch (± s.e.)	S	F	D	H'	J'
OMC	11.2 ± 6.7	29	18	8.15	1.99	0.59
MKB	9.4 ± 8.9	26	16	7.21	2.37	0.73
RM	28.4 ± 10.0	26	19	8.65	1.63	0.50

Principle analysis on fishery resources

abundance/biomass comparison (ABC)



- Rank the fisheries resources according to the abundance
- calculate the cumulative % abundance and % biomass as abundance rank
- The disturbance level of the community was evaluated both by the ABC curve and W-statistic value, in which

$$W = \sum_{i=1}^{S} \frac{(B_i - A_i)}{[50(S-1)]}$$

- where S is the number of species, Ai is abundance value of the each of species rank i, and Bi is biomass value of each species rank, i.

Principle analysis on fishery resources

abundance/biomass comparison (ABC)

Yemane, D., Field, J. G., and Leslie, R. W. 2005. Exploring the effects of fishing on fish assemblages using Abundance Biomass Comparison (ABC) curves. – ICES Journal of Marine Science, 62: 374–379.

The possible effect of fishing on dominance patterns in the South African south coast demersal trawl fishery is assessed using Abundance Biomass Comparison (ABC) curves for the period 1986–2003. The ABC method compares the ranked distribution of abundance among species against the similar distribution of biomass among species. The temporal pattern in the ABC curves and the W-statistic for two depth groups (<100 m and 101–200 m), and for the whole area combined, shows a gradient of change in the demersal assemblages from neutral (W $\geq$ 0) towards negative (W < 0), suggesting a disturbed or stressed condition. This corresponds to the onset of longline fishing effort in 1994, still ongoing in 2003, superimposed upon declining trawl effort in the same region. The ABC method shows promise as a guide for assessing the effects of fishing on fish communities, being based on established r- and k-selection theory. More modelling and comparative work is needed to establish acceptable ranges for the W-statistic, and their application in an ecosystem approach to fisheries management.

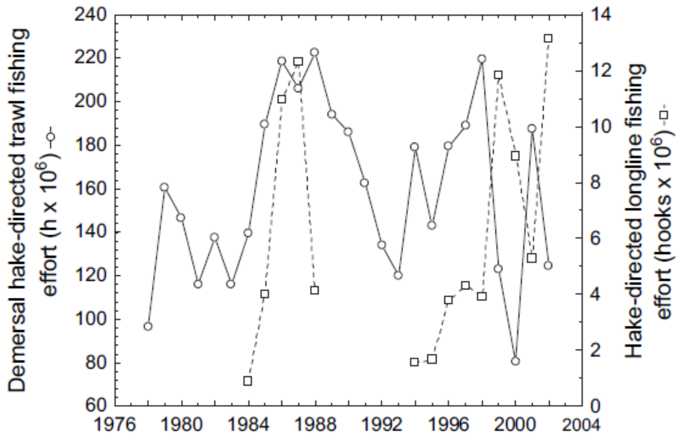
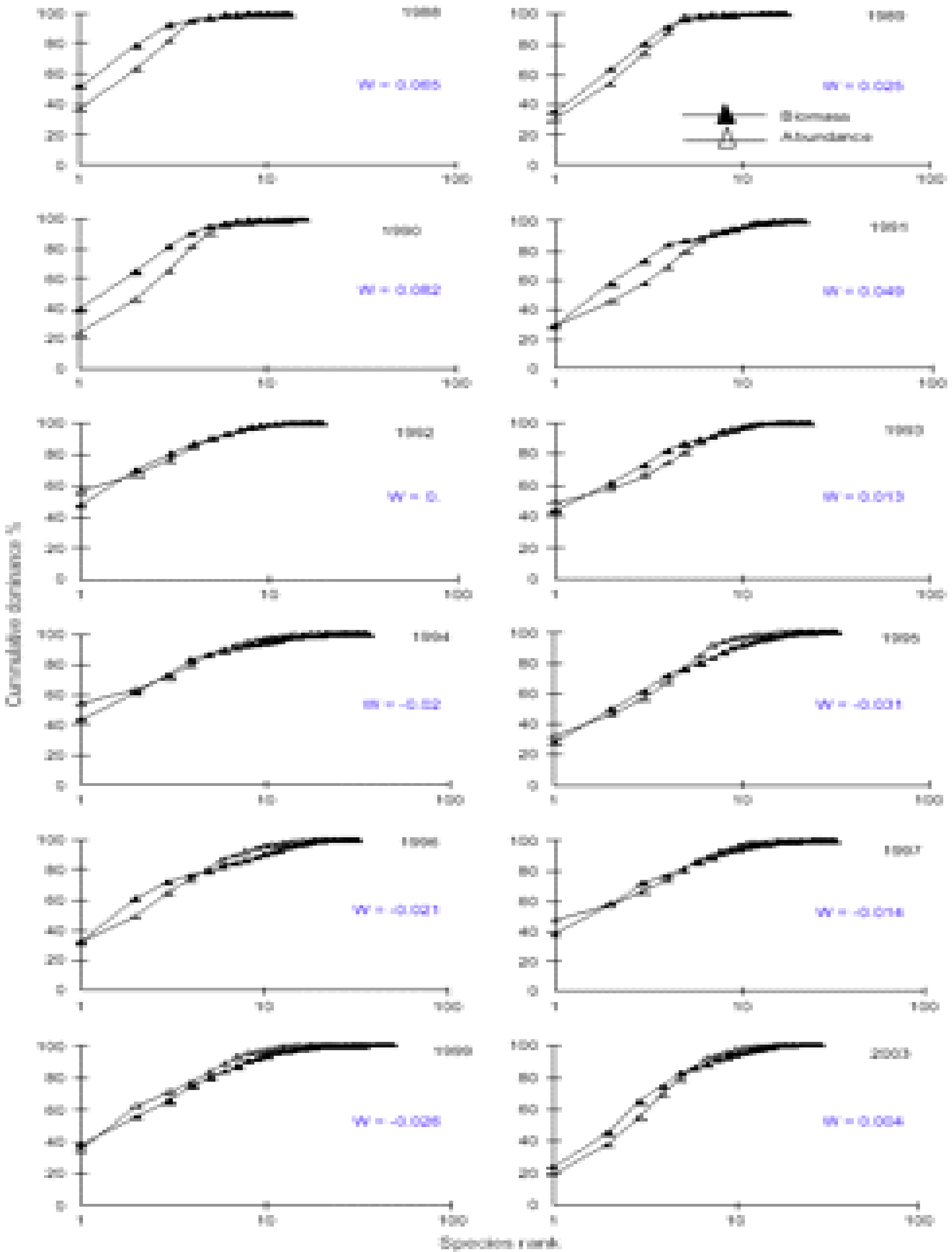
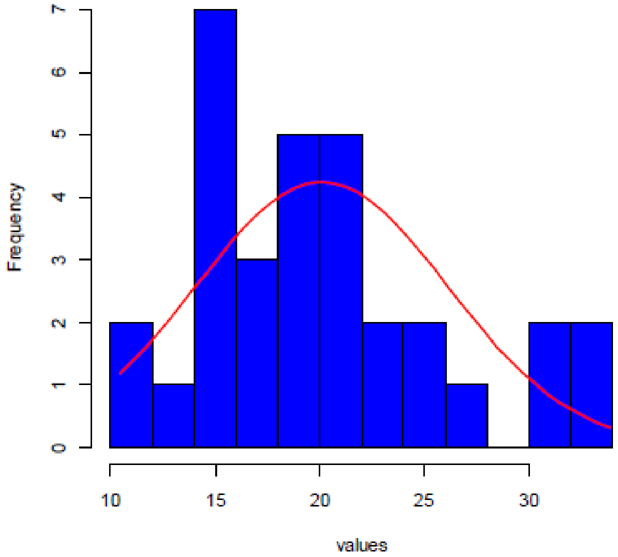


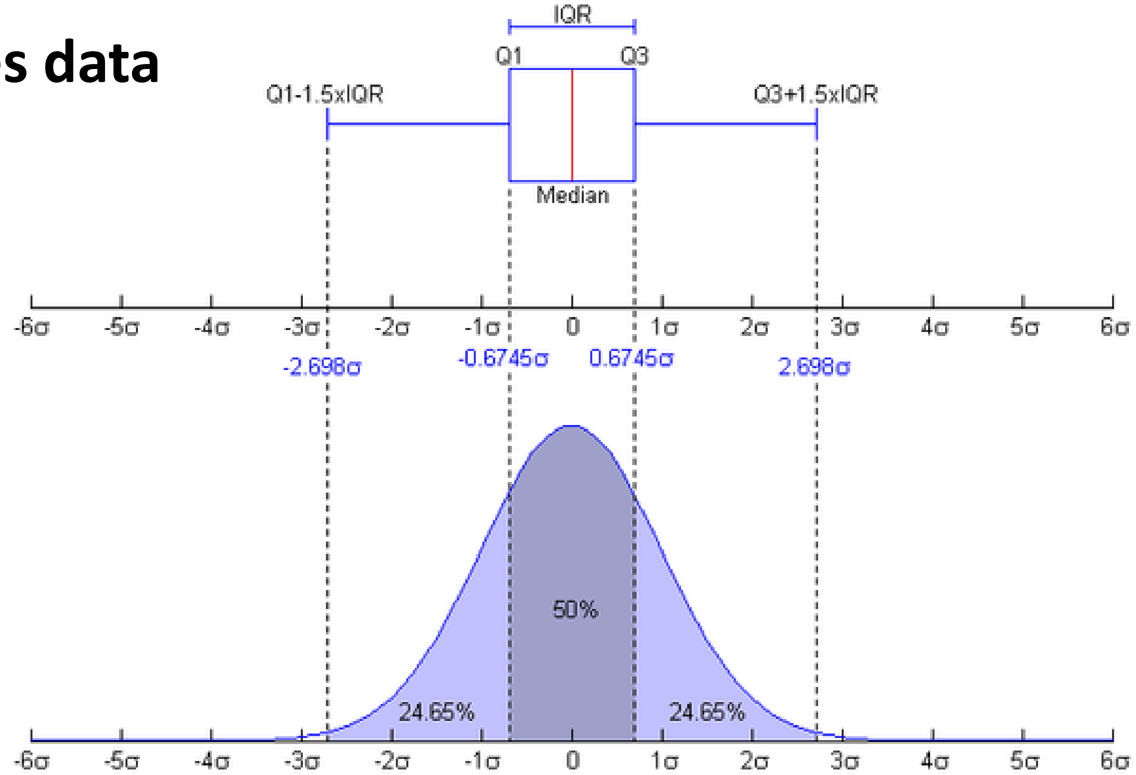
Figure 2. Long-term trend in the total fishing effort of the demersal trawl fishery and longline fishery (from MCM data, and Japp, 1989).

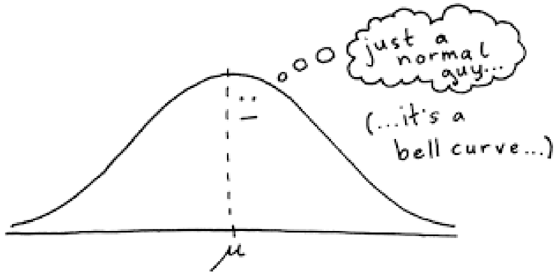


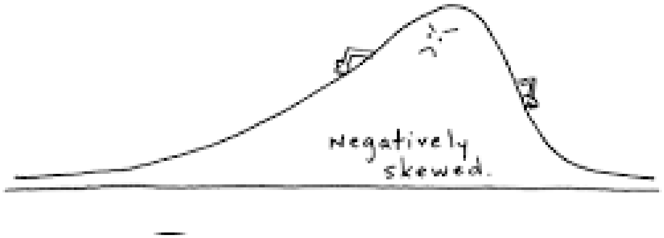
Hypothesis testing on oceanic environment and fisheries data

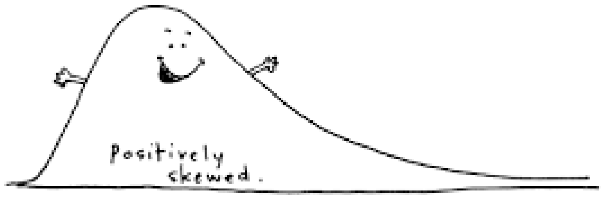
Normality of oceanic environment and fisheries data











The 3 commons

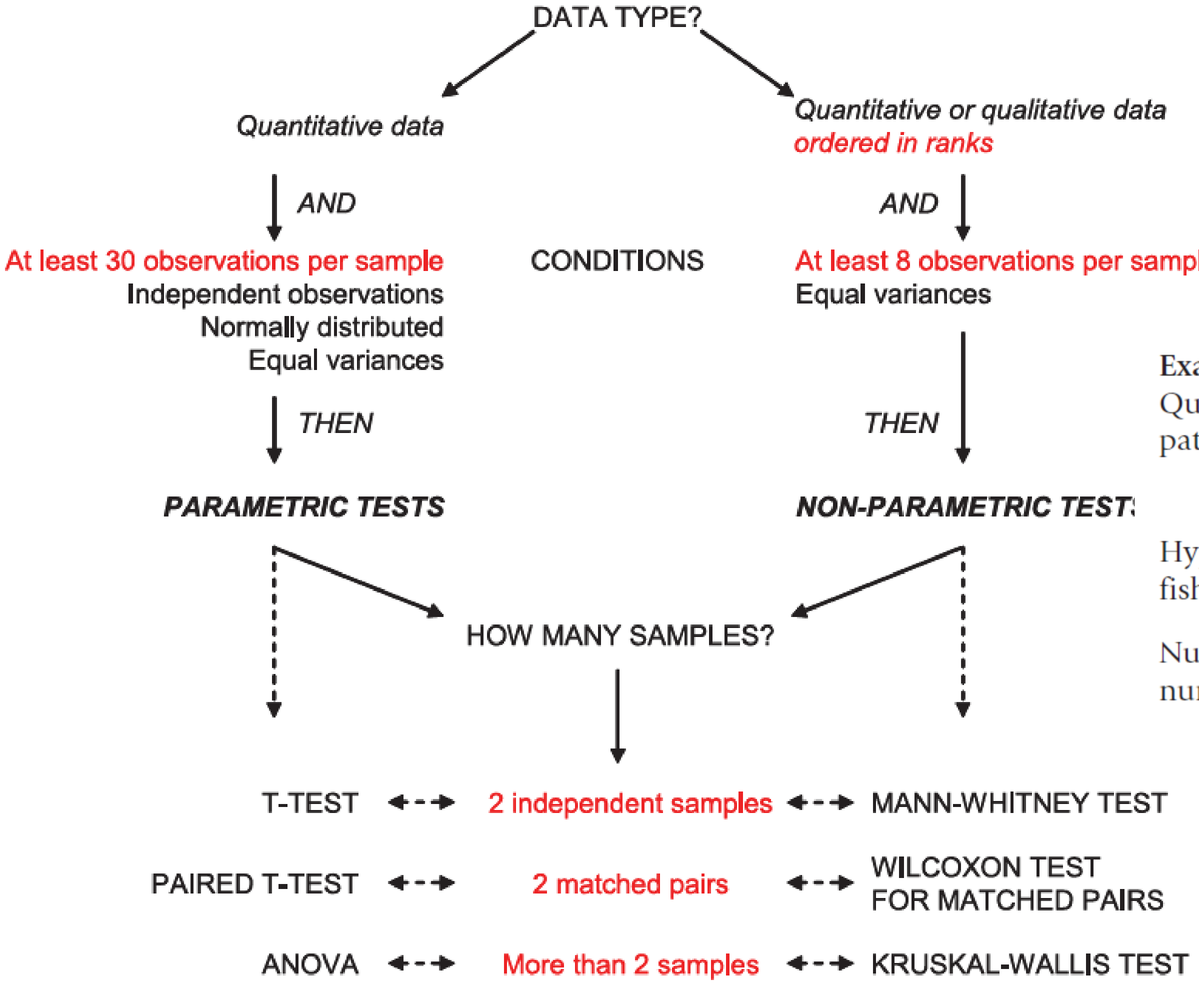
$$y' = \log_{10}(y+1)$$

$$y' = \sqrt{y}$$

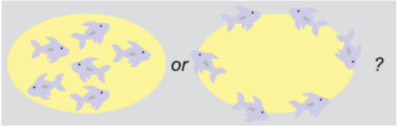
$$y' = \arcsin\sqrt{y}$$

Hypothesis testing on oceanic environment and fisheries data

Hypothesis testing on oceanic environment and fisheries data



Example:
Question: is there the same density of fish inside habitat patches and at the edge of these habitats?



Hypothesis: The number of fish at the edge of habitat patches is different to the number of fish in the interior of habitat patches.

Null hypothesis: The number of fish at the edge of habitat patches is not different to the number of fish in the interior of habitat patches.

Post test, if significant difference is found

Baran E., Warry F. 2008 Simple data analysis for biologists. WorldFish Center and the Fisheries Administration. Phnom Penh, Cambodia. 67 p.

Hypothesis testing on oceanic environment and fisheries data

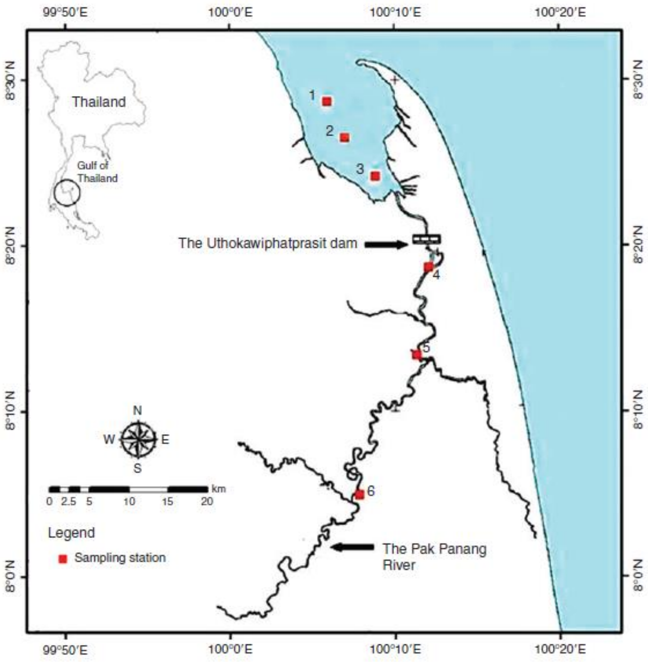
Hypothesis testing on oceanic environment and fisheries data

CSIRO PUBLISHING

Marine and Freshwater Research, 2010, **61**, 288–301 www.publish.csiro.au/journals/mfr

Effects of an anti-salt intrusion dam on tropical fish assemblages

Tuantong Jutagate[A,F], *Amonsak Sawusdee*[B], *Thanitha Thapanand-Chaidee*[C], *Sovan Lek*[D], *Gaël Grenouillet*[D], *Sutheera Thongkhoa*[B] *and Piyapong Chotipuntu*[E]



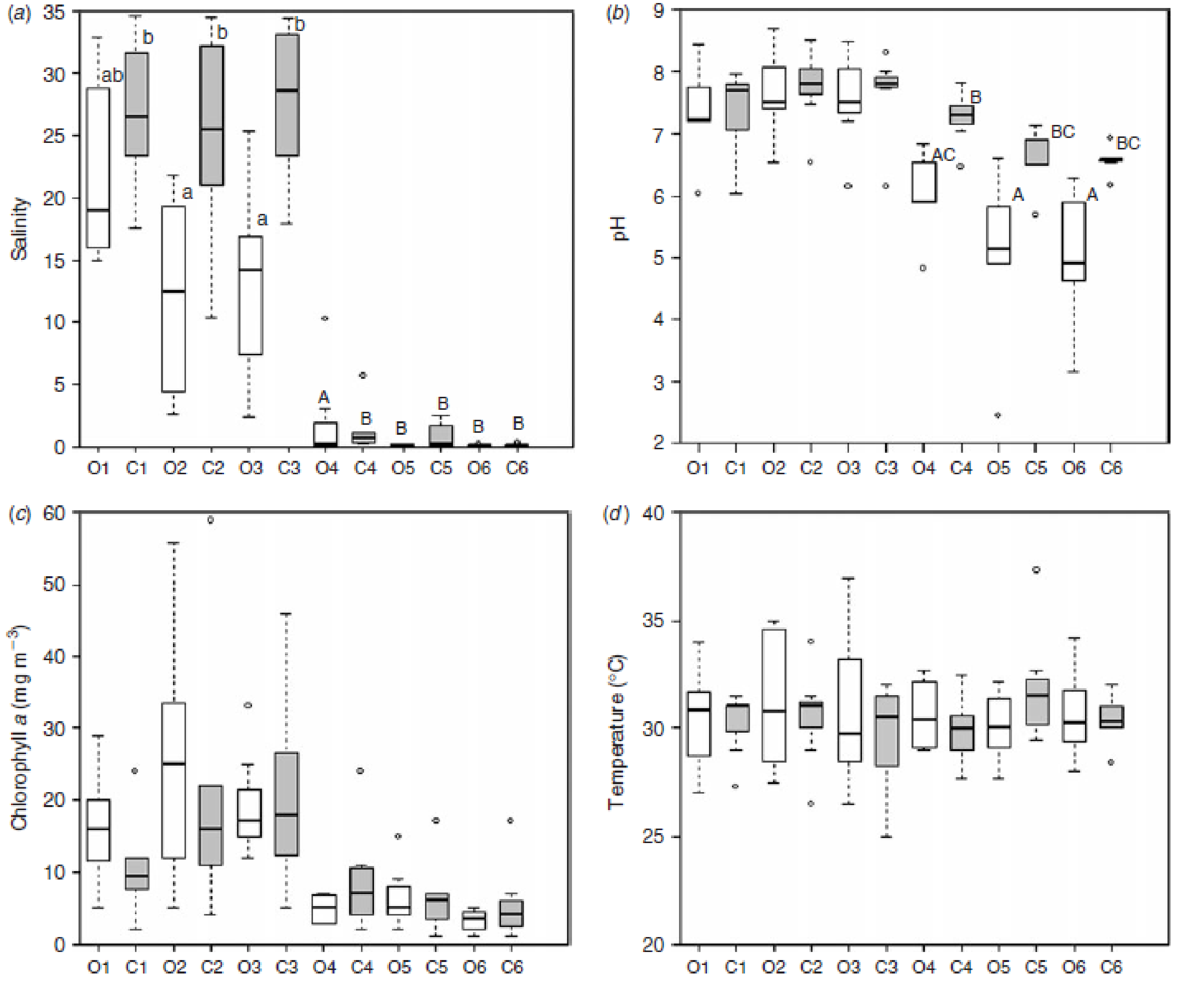
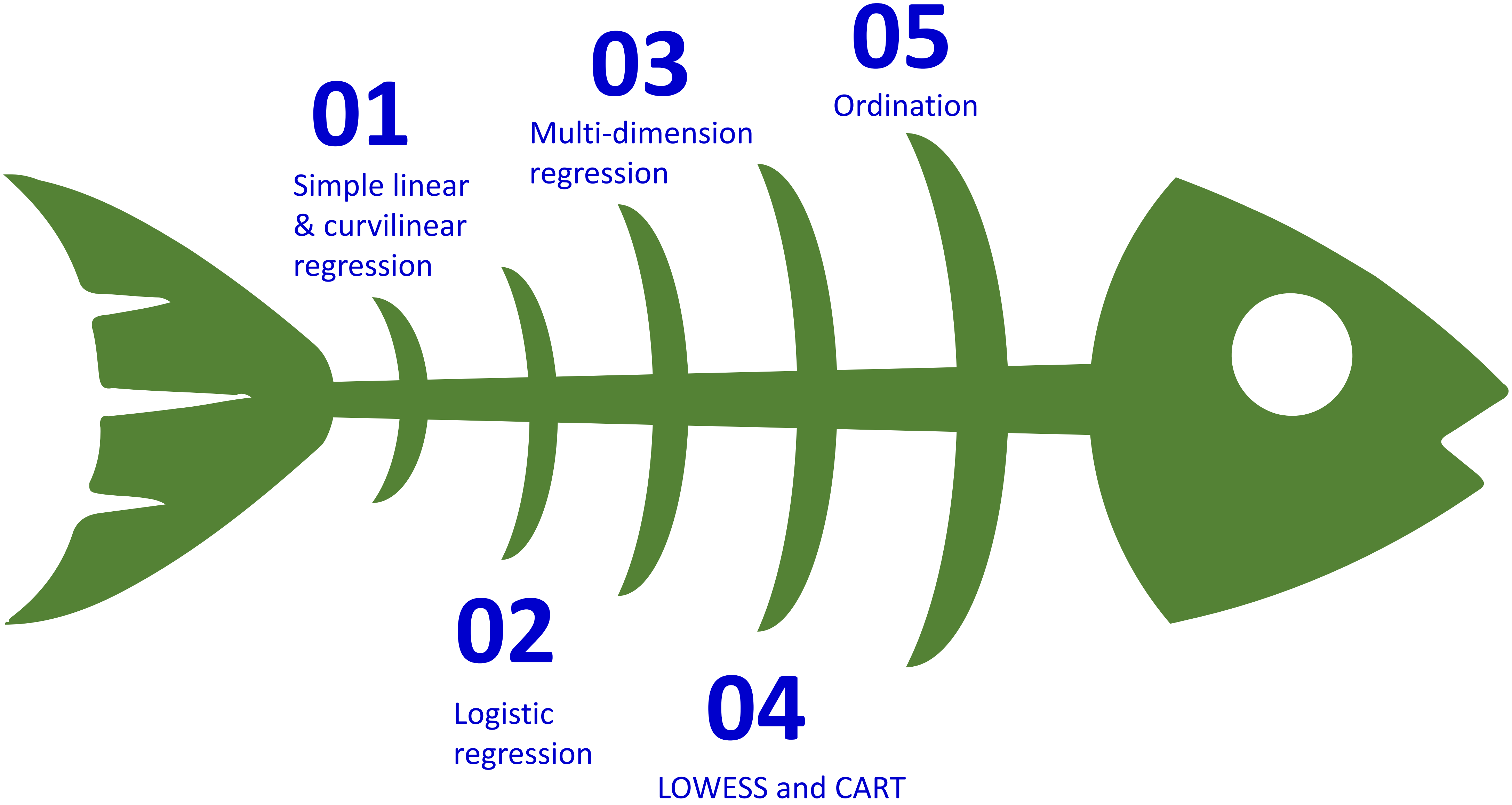
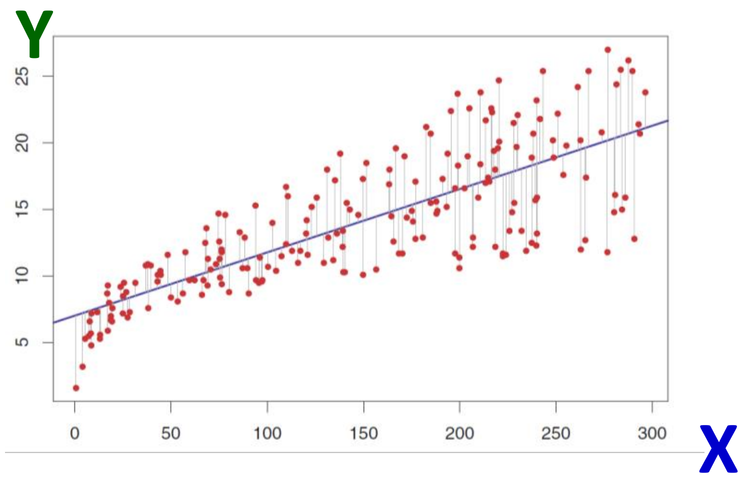


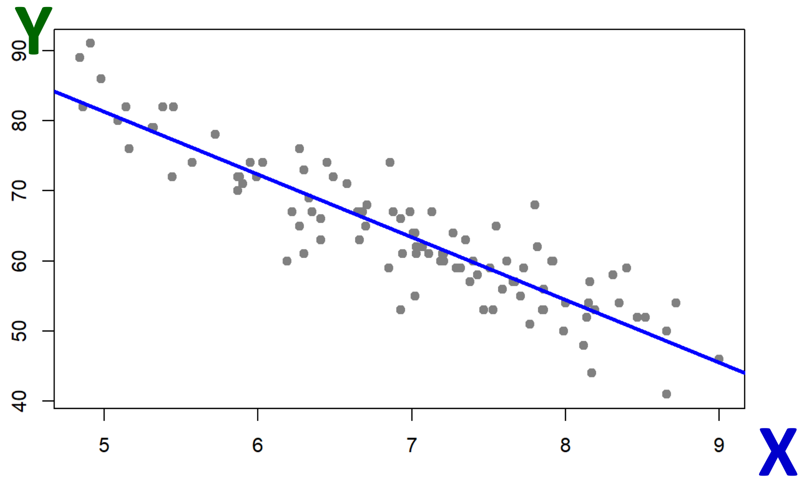
Fig. 2. Environmental attributes of the sampling stations comparing between sluice closing (C) and opening (O) periods based on mean values among all measured water parameters (numbers indicate the sampling stations). The same letter above a box indicates that the values are not statistically different (Dunn's post-hoc tests; $\alpha = 0.05$) when Kruskal–Wallis (H) $P < 0.05$. Lowercase and uppercase letters are for the estuary and river areas respectively.

Principle statistical tests on the relationship between variables



01 Simple linear & curvilinear regression





Common relationships (y ~ f(x)):

- (**Length**) TL ~ f(FL): TL = a + b(FL)
- (**Length**) (dL/dt) ~ f(FL): (dL/dt) = a - b(L)
- (**Reproduction**): Fe ~ f(L): Fe = a + b(L)
- (**Catch**) CpUE ~ f(effort): CpUE = a – b(effort)
- (**Mortality**) Z ~ f(effort): Z = a + b(effort)
- And so on….

General equation for simple linear regression

$$y = a \pm b(x)$$

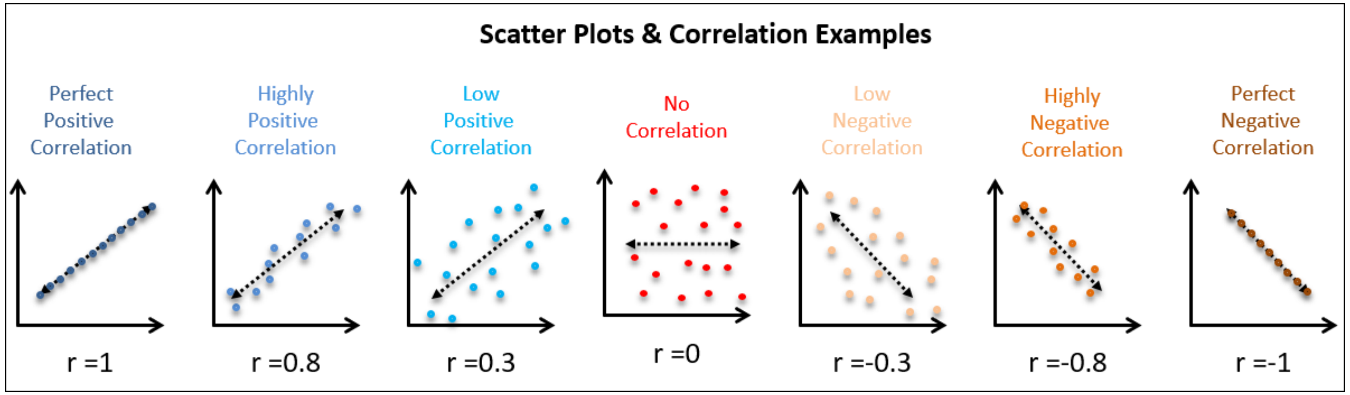
y = Response variable

x = Predictor or explanatory variable

a = y-intercept

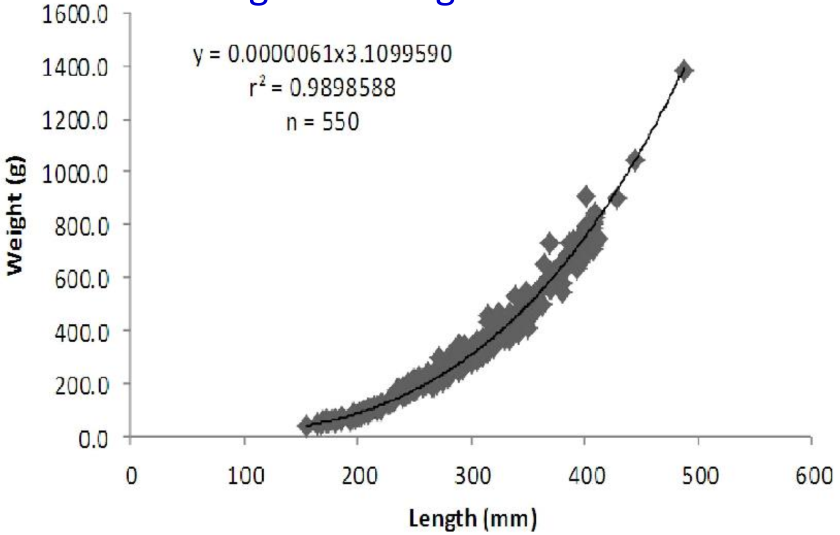
b = Slope

Correlation coefficient (r) and Coefficient of determination (r^2)

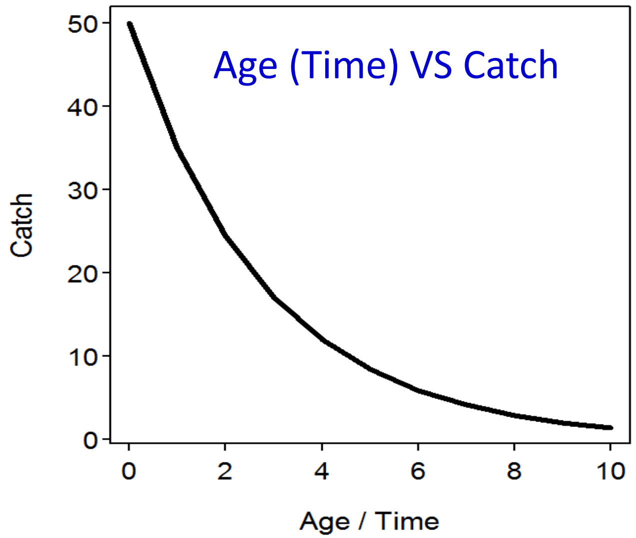


01 Simple linear & curvilinear regression

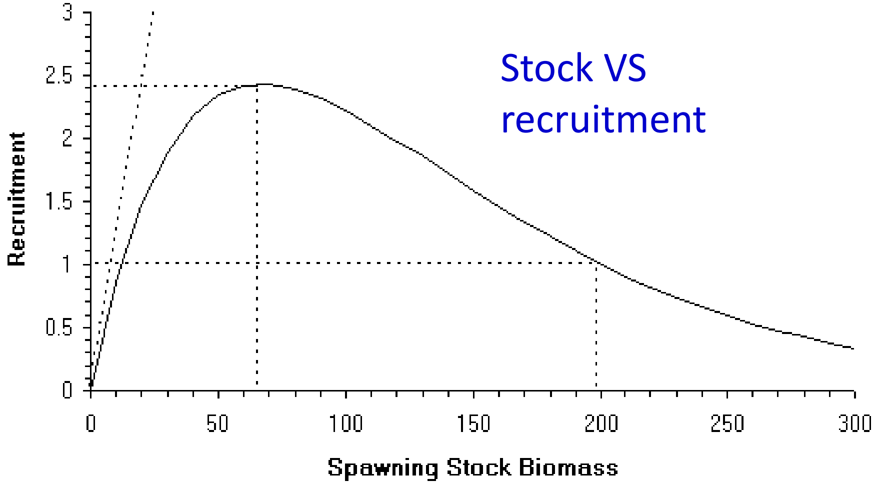
Length VS Weight



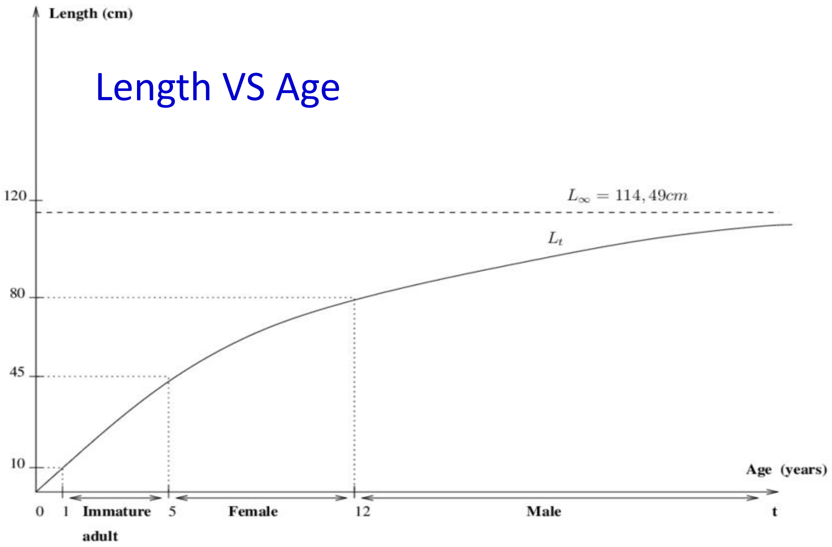
Age (Time) VS Catch



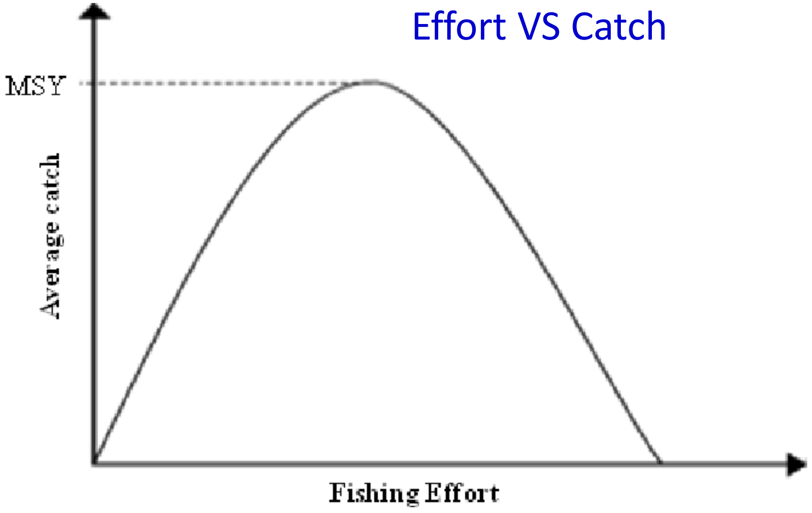
Stock VS recruitment



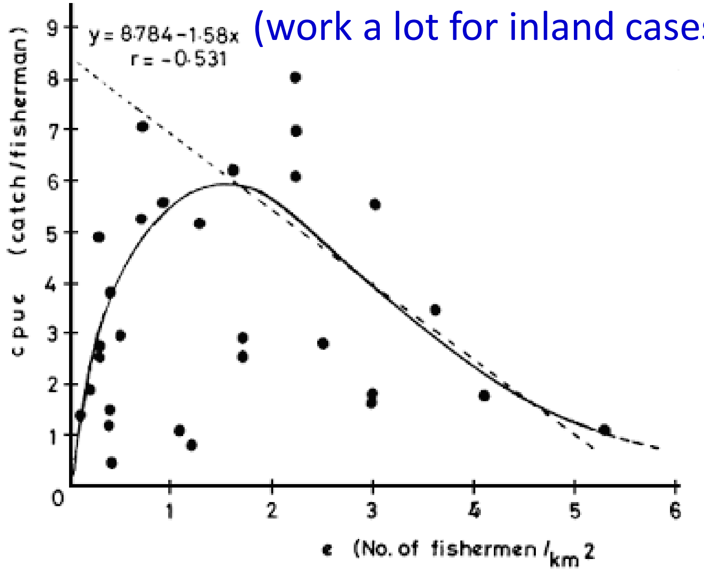
Length VS Age



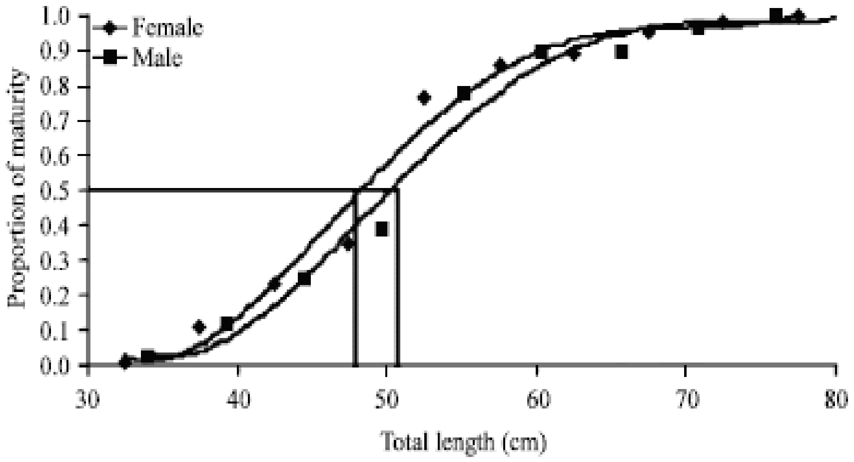
Effort VS Catch

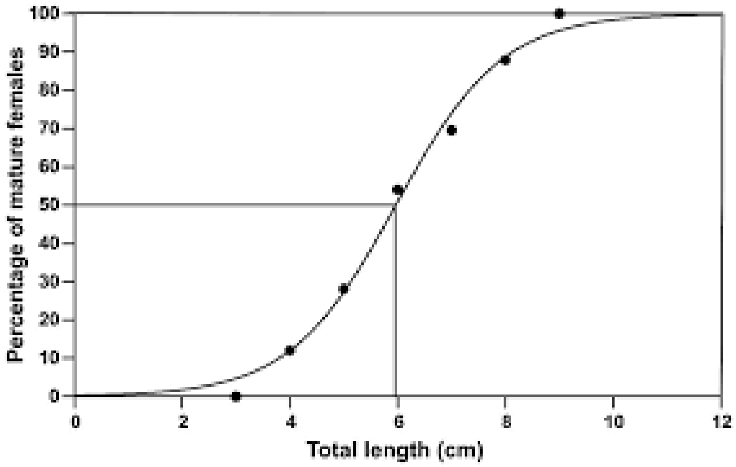


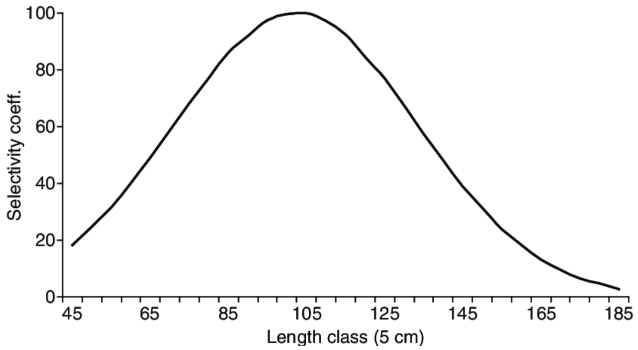
Density of fishers VS CpUE (work a lot for inland cases)



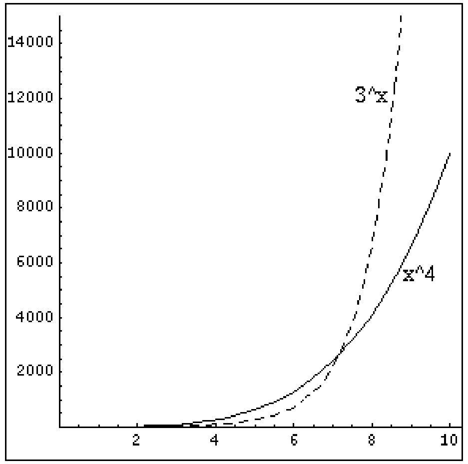
01 Simple linear & curvilinear regression







Rule of thumb for Curvilinear regression



- Check if the relation between predictor and response is not "linear", via "scatter pot
- Check the possible form(s) of relationship
- Analysis could be done through "transformed" values, i.e. making the linear relationship, OR analysis through "fit" the relationship by using "Non-linear least square methods." or "Maximum likelihood methods" or "Bayesian estimation"

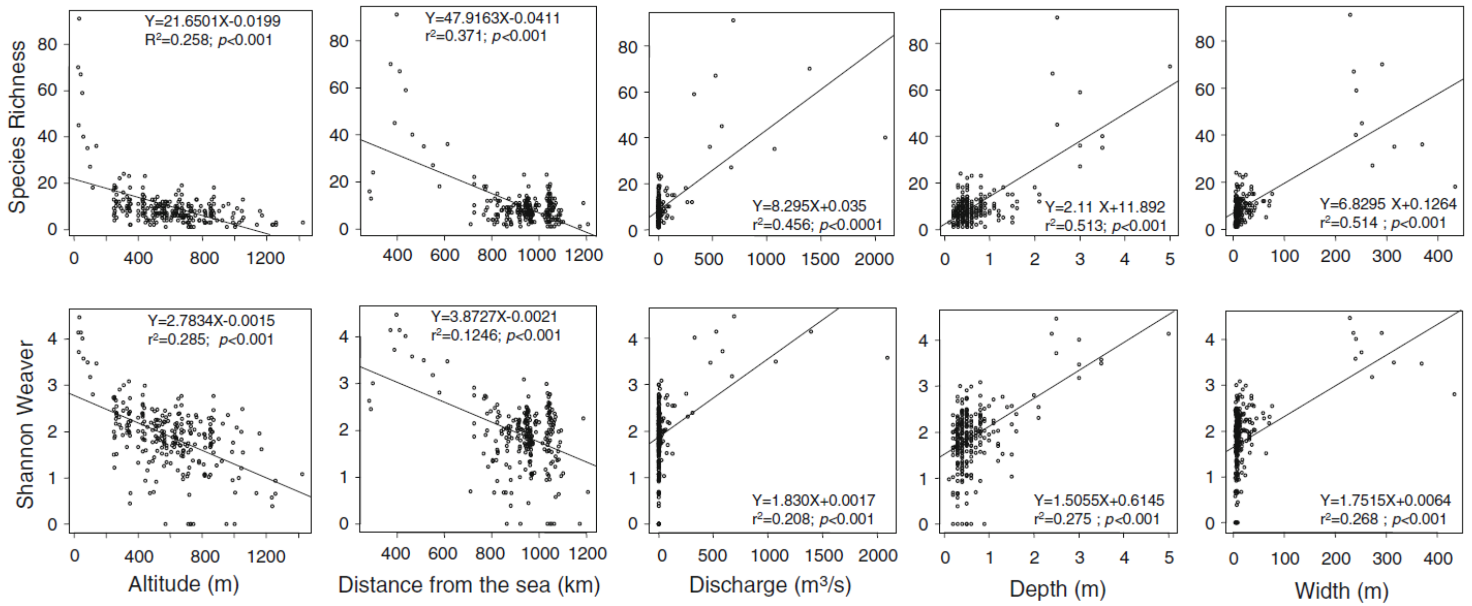
01 Simple linear & curvilinear regression

Hydrobiologia (2012) 694:153–169
DOI 10.1007/s10750-012-1139-4

PRIMARY RESEARCH PAPER

Fish diversity and assemblage patterns along the longitudinal gradient of a tropical river in the Indo-Burma hotspot region (Ping-Wang River Basin, Thailand)

Apinun Suvarnaraksha · Sovan Lek · Sithan Lek-Ang · Tuantong Jutagate



ELSEVIER

Available online at www.sciencedirect.com
ScienceDirect
Fisheries Research 92 (2008) 148–161

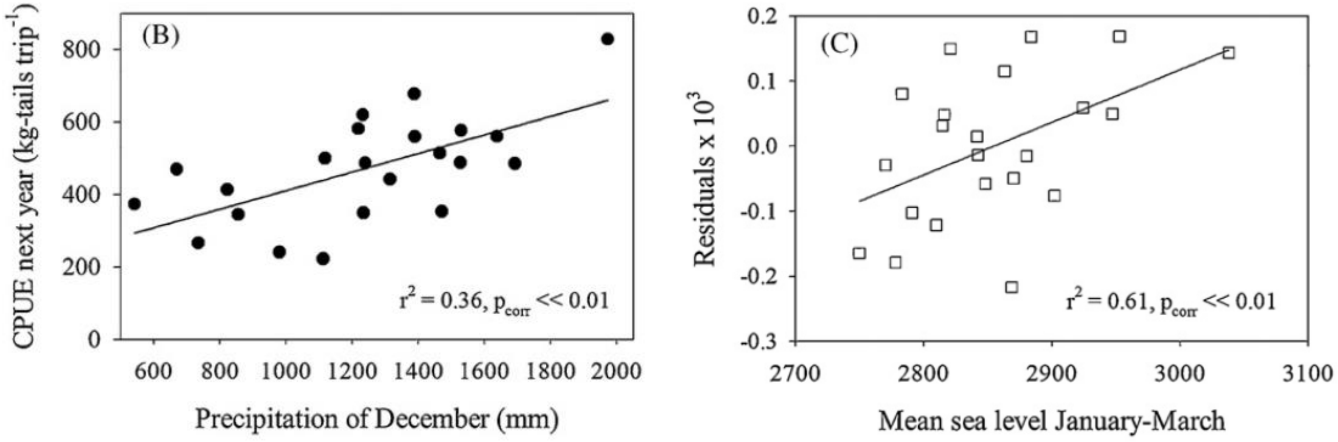
FISHERIES RESEARCH
www.elsevier.com/locate/fishres

Relationship of precipitation, freshwater input, and sea level height with the abundance of the white shrimp (*Litopenaeus occidentalis*; Street, 1871) off Buenaventura, eastern tropical Pacific

Javier A. Díaz-Ochoa [a,b,*], Renato A. Quiñones [a,b]

[a] *Programa de Postgrado en Oceanografía, Departamento de Oceanografía, Universidad de Concepción, Casilla 160-C, Concepción, Chile*
[b] *Centro para la Investigación Oceanográfica en el Pacífico Sur Oriental (FONDAP-COPAS), Universidad de Concepción, Chile*

Received 13 March 2007; received in revised form 21 December 2007; accepted 18 January 2008

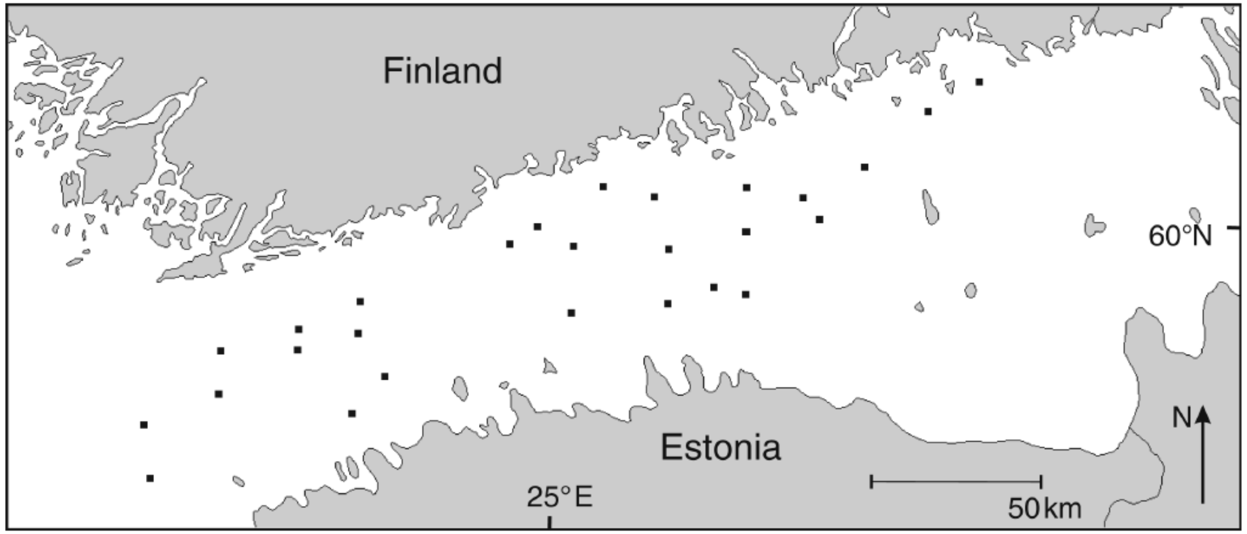


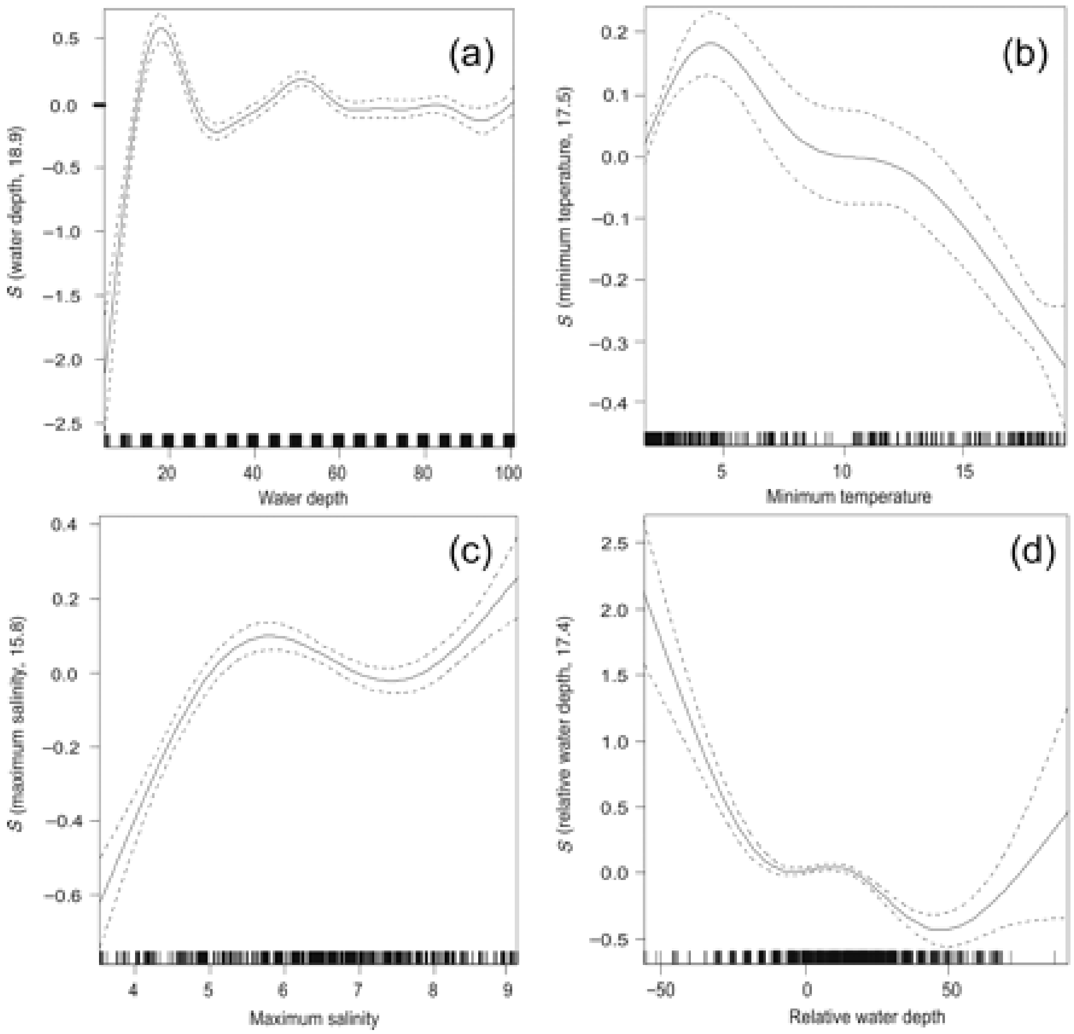
01 Simple linear & curvilinear regression

Pelagic fish abundance in relation to regional environmental variation in the Gulf of Finland, northern Baltic Sea

Heikki Peltonen, Miska Luoto, Jari-Pekka Pääkkönen, Miina Karjalainen, Antti Tuomaala, Jukka Pönni, and Markku Viitasalo

Peltonen, H., Luoto, M., Pääkkönen, J.-P., Karjalainen, M., Tuomaala, A., Pönni, J., and Viitasalo, M. 2007. Pelagic fish abundance in relation to regional environmental variation in the Gulf of Finland, northern Baltic Sea. – ICES Journal of Marine Science, 64: 487–495.



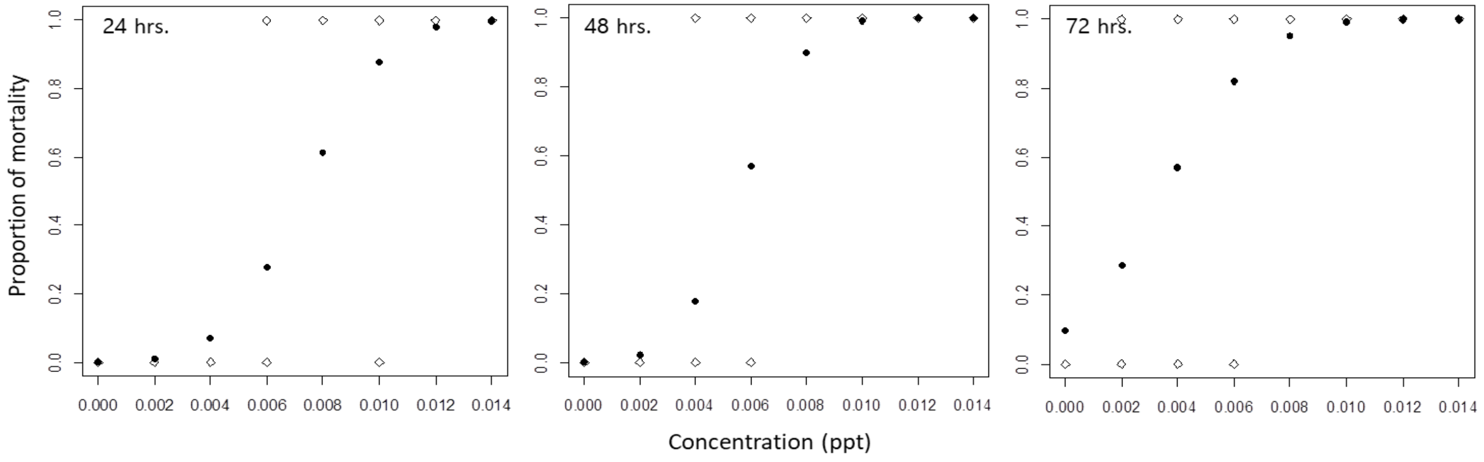


02 Logistic regression

- Sometimes, you fisheries abundance data, i.e. response, is nominal and/or ordinal scales
- Response could be Binary (presence/ absence, Y/ N, Dead/ Not-dead) OR Nominal scale (species composition) OR Ordinal scale (Likert score)
- No assumptions on Normality and Homoscedasticity

$$\ell = \log_b \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

$$g(x) = a \pm b_1\ (x_1) \pm b_2\ (x_2) \pm b_3\ (x_3) \pm \ldots\ldots \pm b_k\ \ (x_k)$$



02 Logistic regression

Environmental and Ecological Statistics **12**, 45–54, 2005

Modelling skewed data with many zeros: A simple approach combining ordinary and logistic regression

DAVID FLETCHER,[1,2,*] DARRYL MACKENZIE[2] and EDUARDO VILLOUTA[3]

[1]*Department of Mathematics and Statistics, University of Otago, P.O. Box 56, Dunedin, New Zealand*
E-mail: dfletcher@maths.otago.ac.nz
[2]*Proteus Wildlife Research Consultants, P.O. Box 5193, Dunedin, New Zealand*
[3]*Department of Conservation, Wellington, New Zealand*

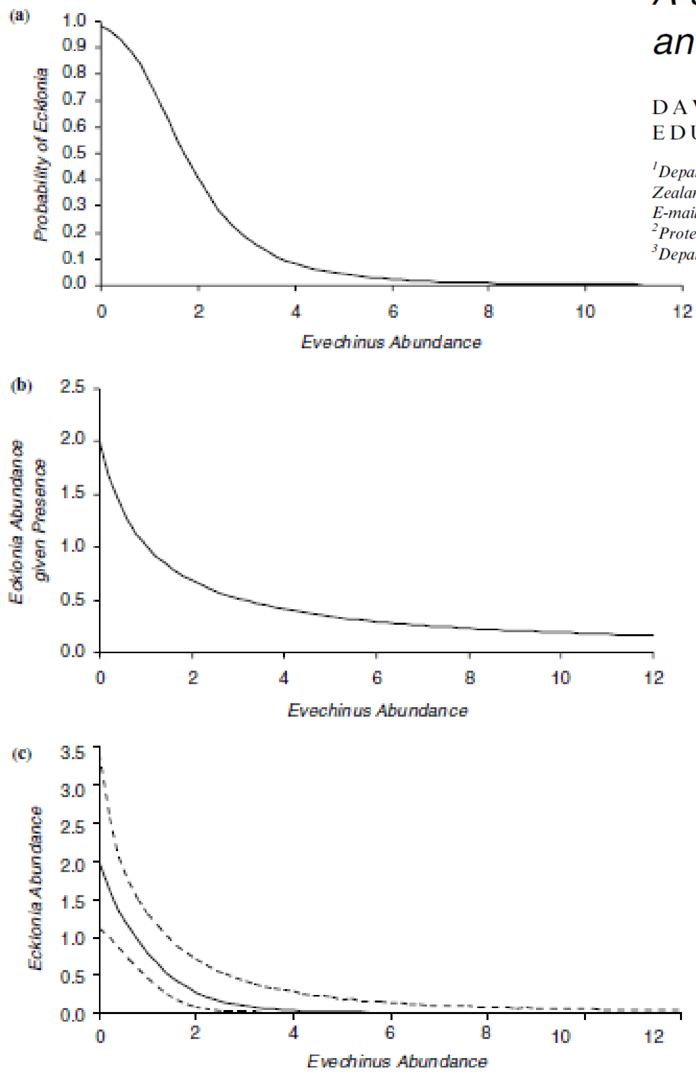


Figure 1. Estimates of (a) probability of presence, (b) expected abundance given presence and (c) expected abundance of *Ecklonia* (dashed lines are 95% confidence limits), plotted against abundance of *Evechinus*. The predictions are for an average site (M, A and F set to 9.22 km, 0.72 km and 32.68 km, respectively). Abundance is measured in individuals per m^2.

Vol. 336: 249–265, 2007 | MARINE ECOLOGY PROGRESS SERIES Mar Ecol Prog Ser | Published April 27

Modeling the pelagic habitat of salmon off the Pacific Northwest (USA) coast using logistic regression

Hongsheng Bi[1,*], Rachel E. Ruppel[1], William T. Peterson[2]

[1]Cooperative Institute for Marine Resources Studies, Oregon State University, and [2]National Marine Fisheries Service, Northwest Fisheries Science Center, Hatfield Marine Science Center, 2030 Marine Science Drive, Newport, Oregon 97365, USA

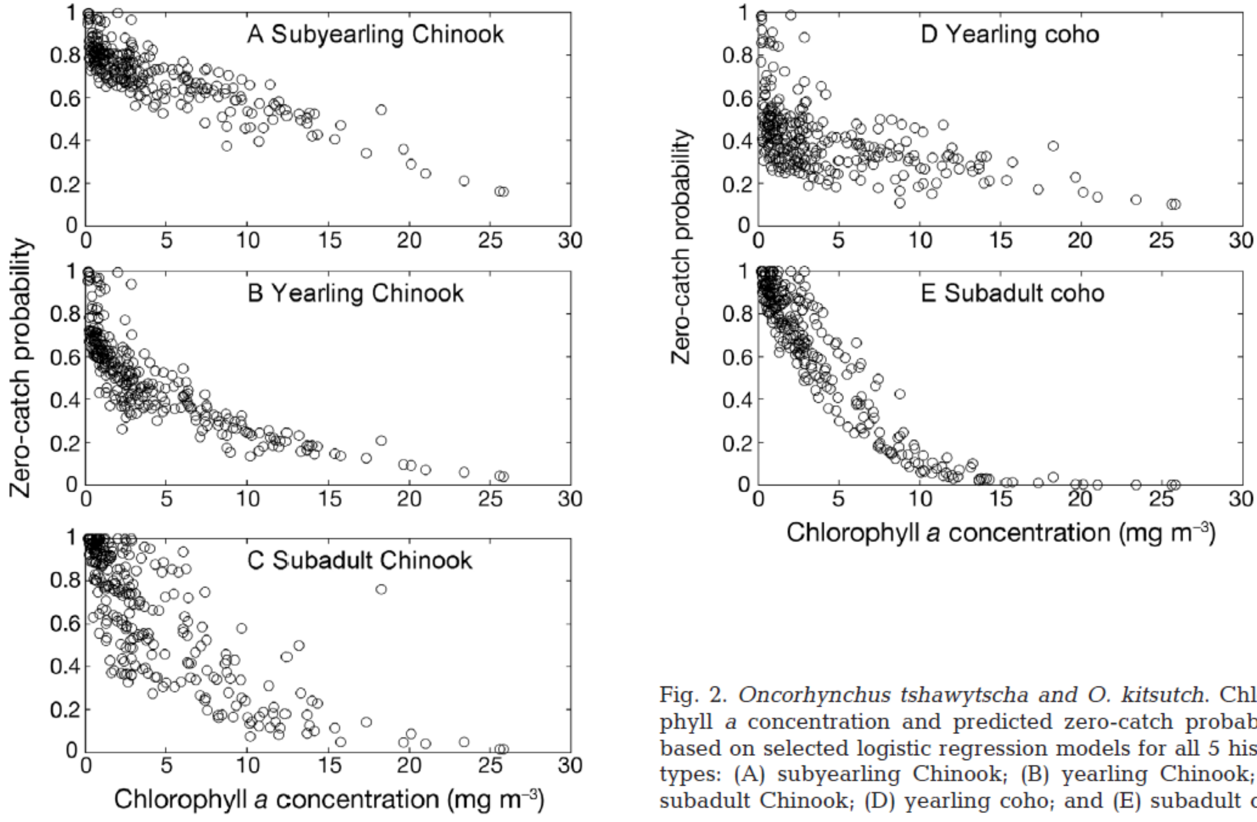
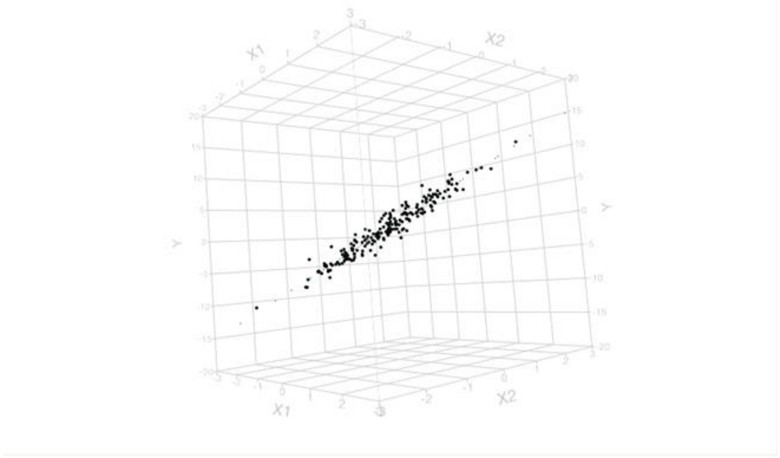


Fig. 2. *Oncorhynchus tshawytscha and O. kitsutch.* Chlorophyll *a* concentration and predicted zero-catch probability based on selected logistic regression models for all 5 history types: (A) subyearling Chinook; (B) yearling Chinook; (C) subadult Chinook; (D) yearling coho; and (E) subadult coho

03 Multiple regression

General equation for simple linear regression

$$y = a \pm b_1(x_1) \pm b_1(x_1) \pm b_2(x_2) \ldots \pm b_n(x_n)$$



y = Response variable); x = Predictor variables;

a = y-intercept and b = regression coefficients

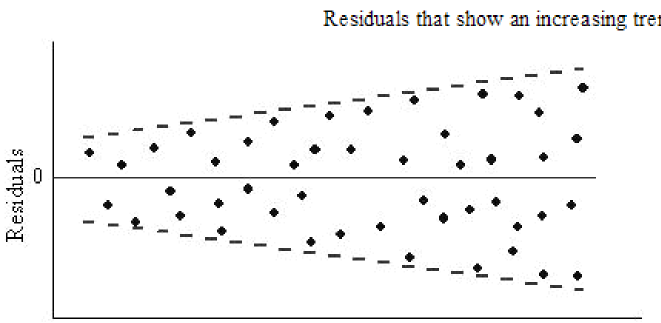
Example in Fisheries Biology (y ~ f(x1, X2, X3)):

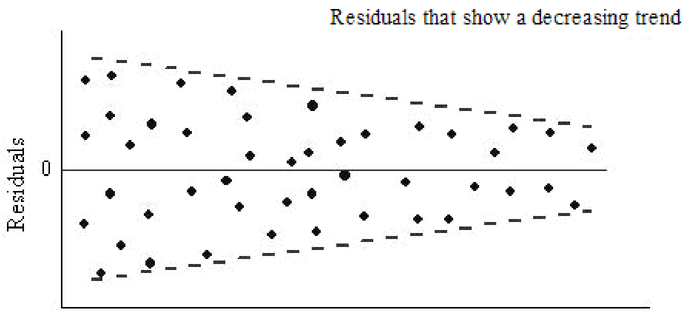
- (**Mortality**) M ~ f(L_∞, K, temp.): $\ln M = - a - b_1 \ln L_\infty + b_2 \ln K + b_3 \ln(temp)$
- (**Consumption**) Q/B ~ f(W_∞, temp., Aspect ratio, Food): $Q/B = - a - b_1 \ln W_\infty + b_2 \ln(temp) + b_3 \ln A + b_4 F$
- (**Swimming speed**) S_r ~ f(SL, Aspect ratio, Mode of swim): $\log_{10} S_r = a - b_1 \log_{10} SL + b_2 \log_{10} A + b_3 \log_{10} M$
- and so on……..

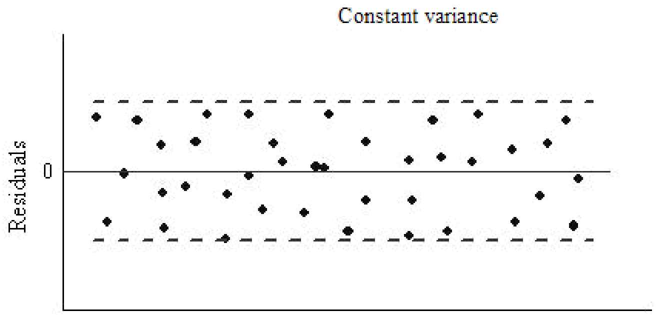
03 Multiple regression

Rule of thumb for Multiple regression

- Check if there are any confound among the predictors, i.e. each predictor variable should be independent to each other.
- Number of the sample should be at least 10 times เท่าของจำนวนตัวแปรที่เป็น Predictors
- Try first the Multiple Linear Model (MLR), thus transformation of some Predictors may be necessary.
- Suggesting Backward stepwise fitting
- If the Response variables are binary or nominal or rank, use (Multiple) Logistic Regression







03 Multiple regression

Vol. 381: 119–127, 2009
doi: 10.3354/meps07969
MARINE ECOLOGY PROGRESS SERIES
Mar Ecol Prog Ser
Published April 17

Influence of surface oceanographic variability on abundance of the western winter-spring cohort of neon flying squid *Ommastrephes bartramii* in the NW Pacific Ocean

Jie Cao[1], Xinjun Chen[1, 2, 3,*], Yong Chen[1, 2, 3, 4]

Table 1. *Ommastrephes bartramii*. Regression model between proportion of favourable-sea surface temperature areas (PFSSTA) and catch per unit effort (CPUE)

Model	95% CI	p
$CPUE = \alpha_0 + \alpha_1 P_1 - \alpha_2 P_2$		
$\alpha_0 = -9.5350$	−18.8267 to −0.2433	0.045
$\alpha_1 = 20.0663$	5.2710 to 34.8616	0.014
$\alpha_2 = -12.7812$	−24.8188 to −0.7436	0.040
P_1 = PFSSTA of February on spawning ground		
P_2 = 4th root of continued product of PFSSTA from Aug–Nov		
Multiple = 0.77, R^2 = 0.60, Residual SD = 4.3426, F = 5.1627, Significance F = 0.0419		

AINLEY ET AL.: DISTRIBUTION OF YOUNG ROCKFISH
CalCOFI Rep., Vol. 34, 1993

OCEANIC FACTORS INFLUENCING DISTRIBUTION OF YOUNG ROCKFISH (*SEBASTES*) IN CENTRAL CALIFORNIA: A PREDATOR'S PERSPECTIVE

DAVID G. AINLEY,
WILLIAM J. SYDEMAN
PRBO International Biological Research
4990 Shoreline Highway
Stinson Beach, California 94970

RICHARD H. PARRISH
Pacific Fisheries Environmental Group
NOAA, National Marine Fisheries Service
P.O. Box 831
Monterey, California 93942

WILLIAM H. LENARZ
Tiburon Laboratory
NOAA, National Marine Fisheries Service
3150 Paradise Drive
Tiburon, California 94920

TABLE 2

Backward Regression of January + February Upwelling Index (UI), February Sea Level (SL), and March Sea-Surface Temperature (SST) with Rockfish Prevalence in the Murre Diet

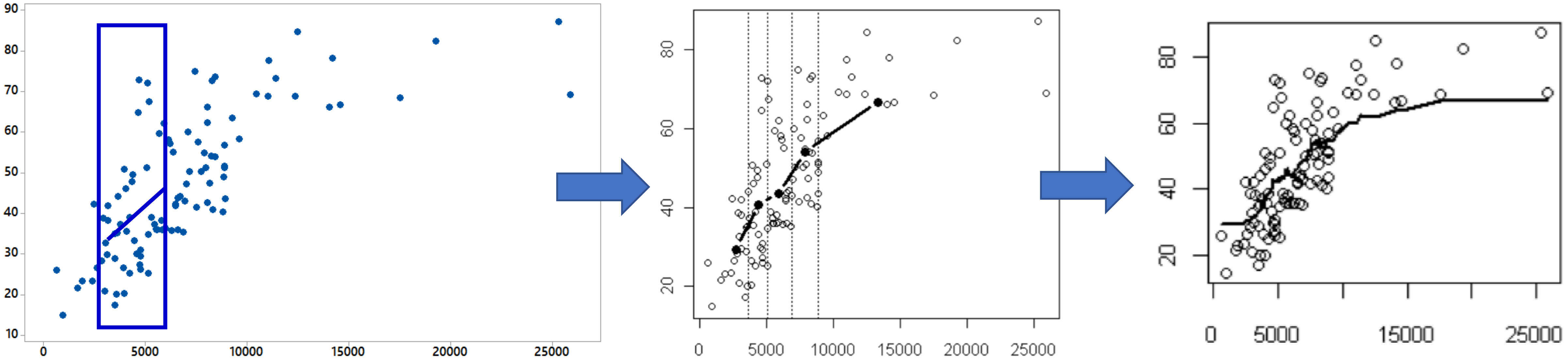
Model	R^2	F	P
UI, UI², SST, SL	0.58	6.81	0.0035
UI, UI², SST	0.59	9.17	0.0013
UI, UI², SL	0.60	9.50	0.0011
UI, UI²	0.61	14.03	0.0004
SL, SST	0.23	3.55	0.0547
UI	0.11	3.11	0.0971
SL	0.28	7.54	0.0144
SST	0.11	3.08	0.0986

R^2 values were adjusted by the degrees of freedom and sample size (n = 18 years).

04 LOESS (LOWESS)

LOESS (LOWESS)

- LOESS (locally estimated scatterplot smoothing) AND LOWESS (locally weighted scatterplot smoothing) are known as moving regressions.
- Non-parametric regression; LOESS uses Linear polynomial, meanwhile LOWESS uses Quadratic polynomial
- Fitting simple models to localized subsets of the data to build up a function that describes the deterministic part of the variation in the data, point by point.

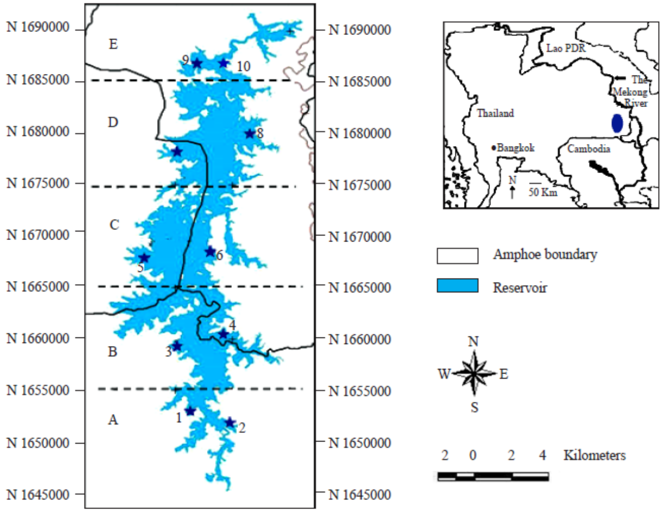


04 LOESS (LOWESS)

Indian J. Fish., 63(3): 11-23, 2016
DOI: 10.21077/ijf.2016.63.3.54491-02

Spatio-temporal variations in abundance and assemblage patterns of fish larvae and their relationships to environmental variables in Sirindhron Reservoir of the Lower Mekong Basin, Thailand

TUANTONG JUTAGATE[1], ACHARA RATTANACHAI[1], SURIYA UDDUANG[2], SITHAN LEK-ANG[3] AND SOVAN LEK[3]



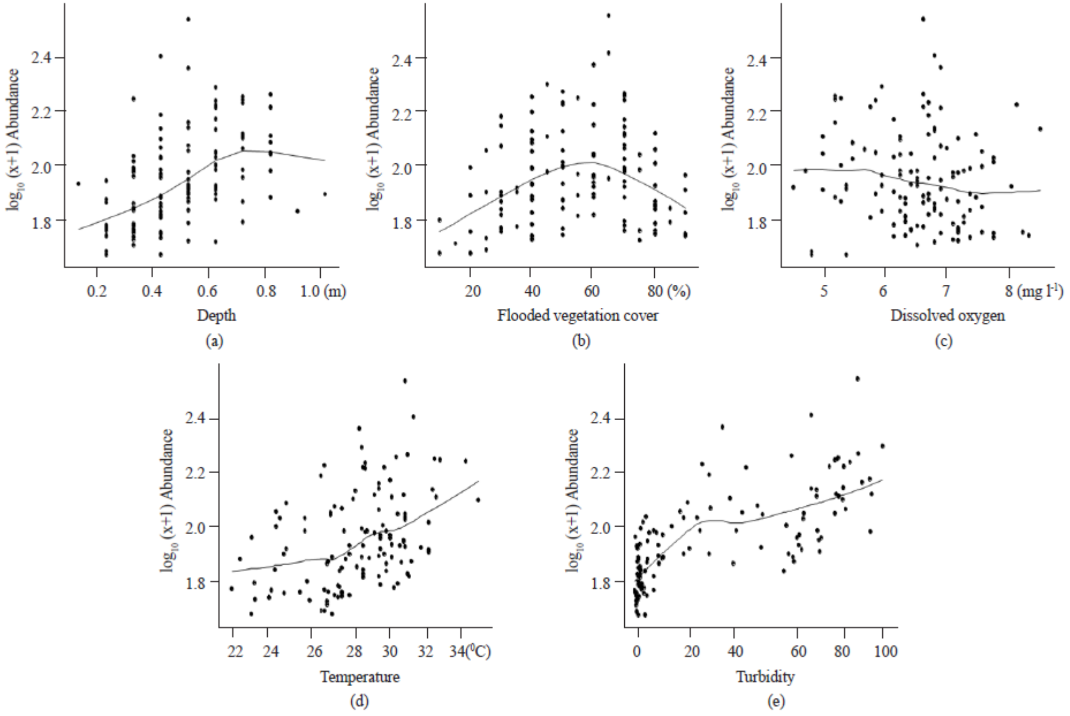


Fig. 5. Bivariate plots of the five environmental variables against fish larval abundance in Sirindhorn Reservoir. The *Lowess* curves (solid lines) were used to fit the data

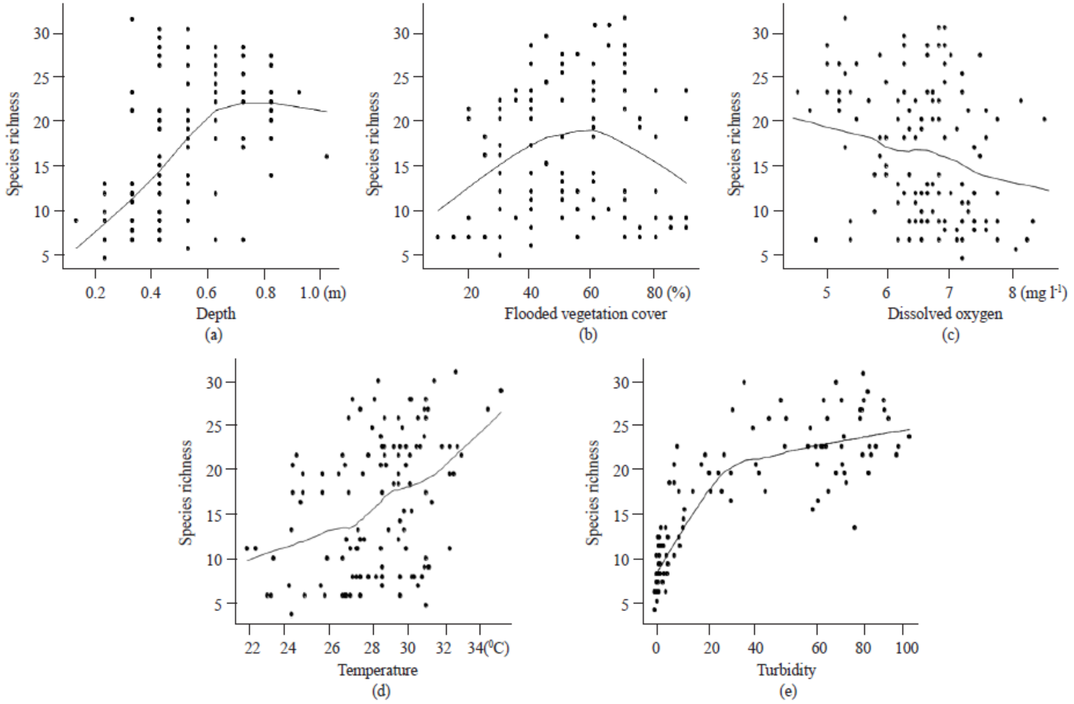
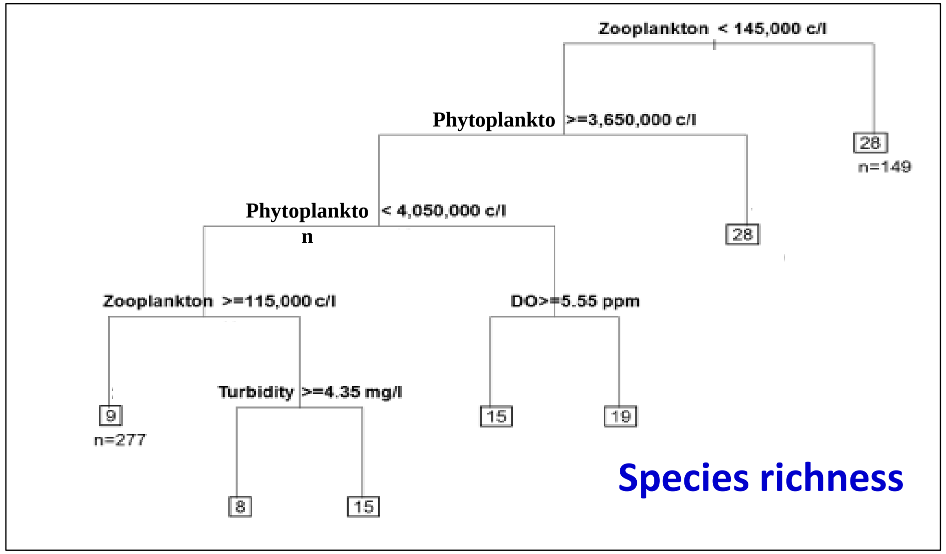


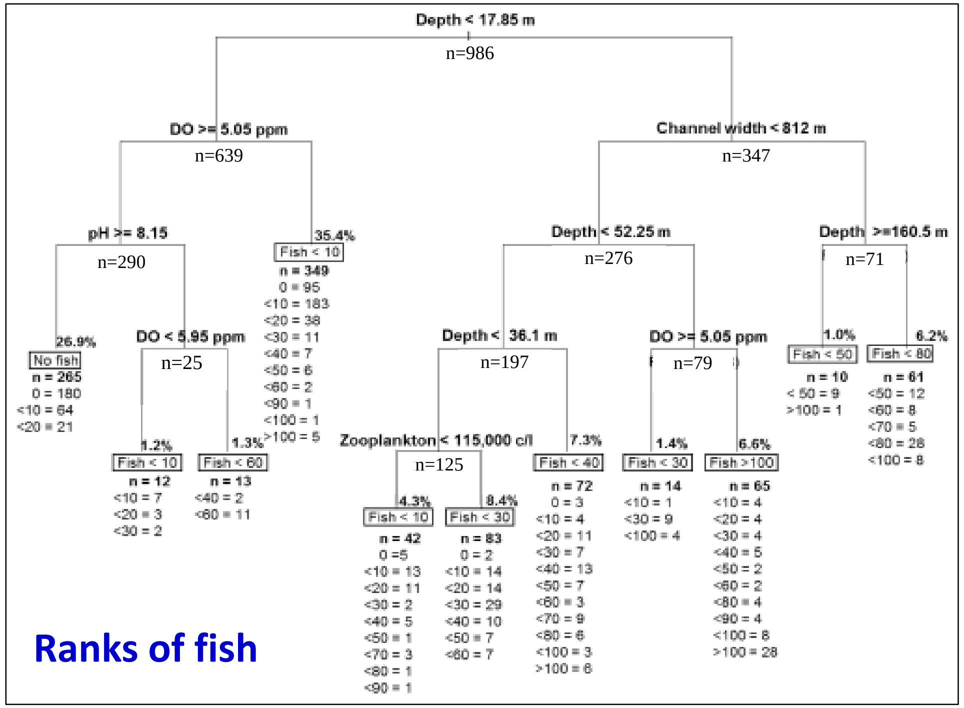
Fig. 6. Bivariate plots of the six environmental variables against species richness of fish larvae in Sirindhorn Reservoir. The *Lowess* curves (solid lines) were used to fit the data

04.1 CART (Classification and Regression Tree)

CART (Classification And Regression Tree)

- Non-parametric machine learning technique for regression and classification problems., which are easy to be interpreted.
- Regression Tree for “Quantitative” response AND Classification Tree for “Qualitative” response
- The resulting tree is composed of decision nodes, branches and leaf nodes. The tree is placed from upside to down, so the root is at the top and leaves indicating the outcome is put at the bottom.
- Each decision node corresponds to a single input predictor variable and a split cutoff on that variable. The leaf nodes of the tree are the outcome variable which is used to make predictions.
- The tree grows from the top (root), at each node the algorithm decides the best split cutoff that results to the greatest purity (or homogeneity) in each subpartition.
- Beware on overfitting / high variance / low bias-tree





04.1 CART (Classification and Regression Tree)

Limnologica 42 (2012) 127–136

Contents lists available at SciVerse ScienceDirect

Limnologica

journal homepage: www.elsevier.de/limno

ELSEVIER

Predicting fish assemblages and diversity in shallow lakes in the Yangtze River basin

Lin Cheng[a,b,c], Sovan Lek[c], Sithan Lek-Ang[c], Zhongjie Li[a,*]

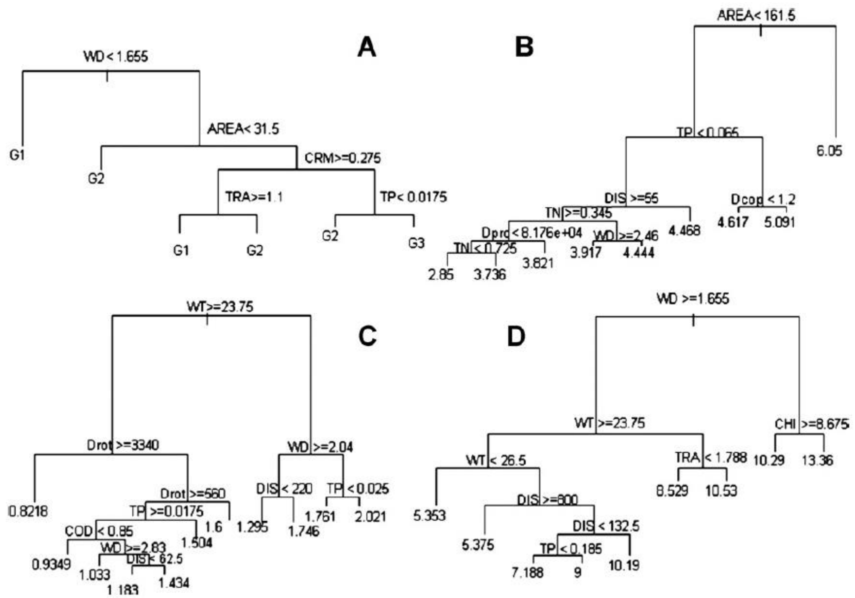


Fig. 6. CART model predicting fish community pattern and richness in studied lakes. Biotic and abiotic variable identifications refer to Section "Materials and methods". (A) Predicting fish assemblages. (B) Predicting relative abundance. (C) Predicting Shannon index. (D) Predicting species richness. The overall percentage of successful prediction is 60.7%, 63.9%, 59.3% and 56.5%, respectively.

Indian Journal of Geo Marine Sciences
Vol. 45 (12), December 2016, pp. 1677-1687

Use of different approaches to model catch per unit effort (CPUE) abundance of fish

*Vinod K. Yadav[1,2], Shrinivas Jahageerdar.[2], Ramasubramanian V.[2], Vidya S. Bharti[2] & Adinarayana J.[1]

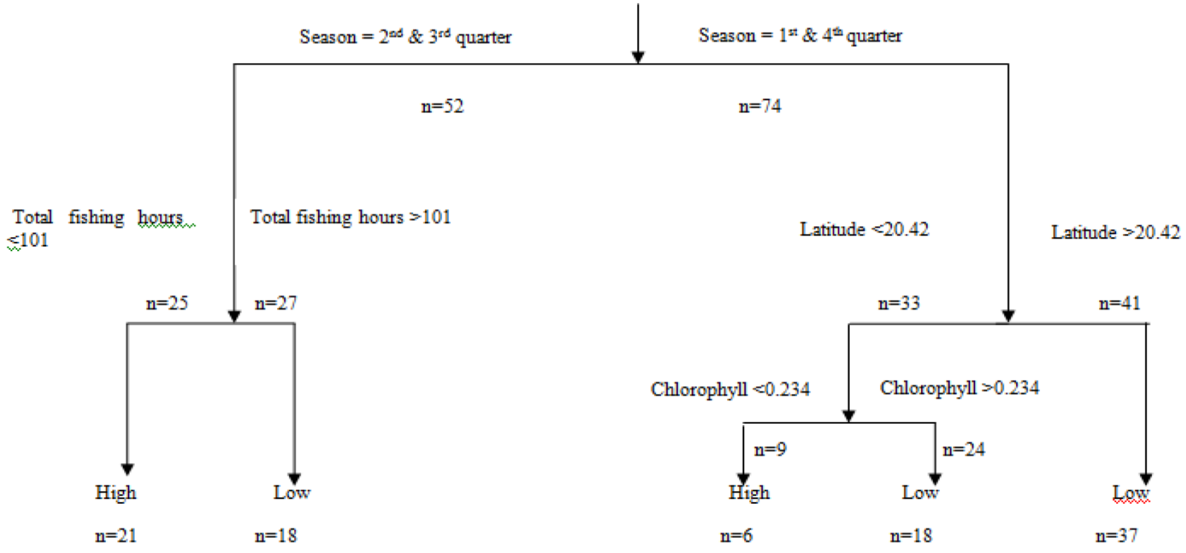
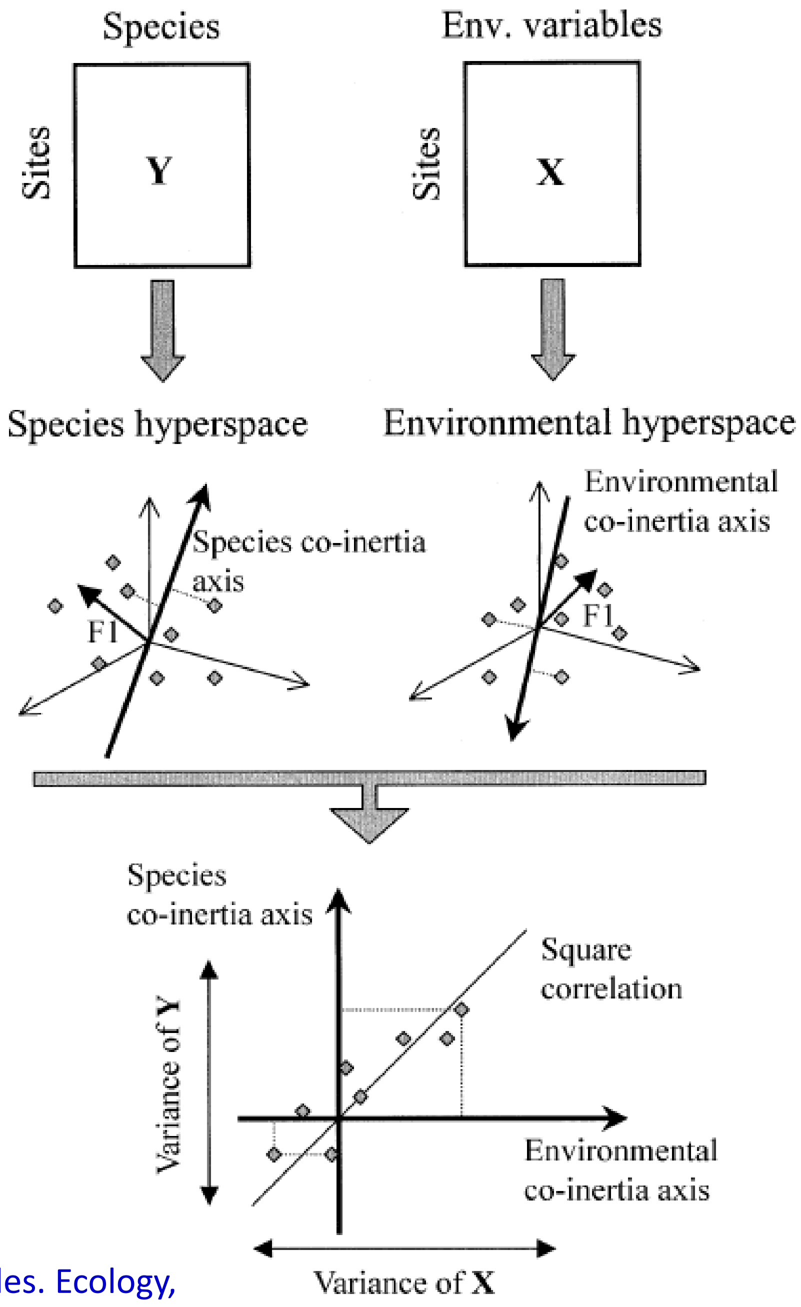


Fig. 3 Classification Tree and Associated Results for Predicting CPUE abundance

04.1 Ordination (Co-inertia analysis)

CART (Classification And Regression Tree)

- Also known as “Double-PCA”, PCA: Principal Component Analysis, since it ordinates and studies the relationship between the two set of variables, i.e. oceanic environments and fishes in our course.
- Analyses species-environment tables, and attempts to find new axes in both so that the covariance between the new sets of scores is maximised.
- The maximized covariance means a maximal correlation and maximal standard deviations of the new environmental and species scores.
- Interpretation: High when two structures vary simultaneously (and also when they vary inversely) AND Low when they vary independently or do not vary.



Dray, S., Chessel, D. and Thioulouse, J., 2003. Co-inertia analysis and the linking of ecological data tables. Ecology, 84(11), pp.3078-3089.

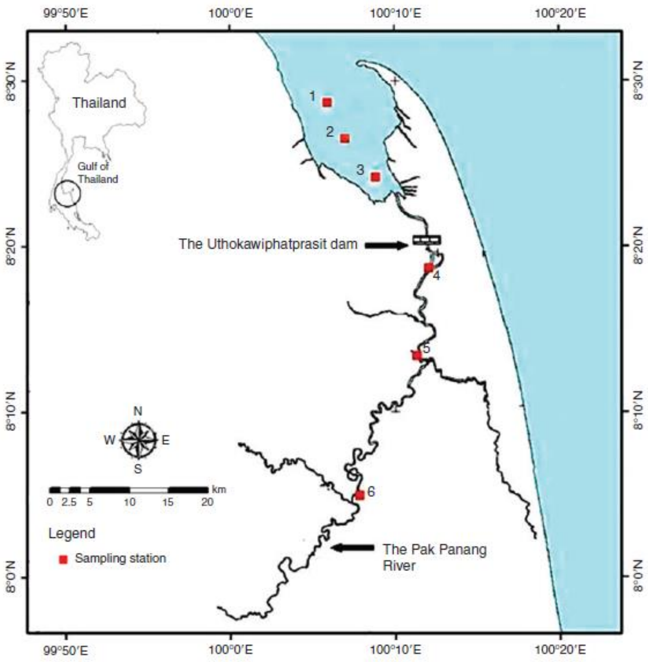
04.1 Ordination (Co-inertia analysis)

CSIRO PUBLISHING

Marine and Freshwater Research, 2010, **61**, 288–301 www.publish.csiro.au/journals/mfr

Effects of an anti-salt intrusion dam on tropical fish assemblages

Tuantong Jutagate[A,F], Amonsak Sawusdee[B], Thanitha Thapanand-Chaidee[C], Sovan Lek[D], Gaël Grenouillet[D], Sutheera Thongkhoa[B] and Piyapong Chotipuntu[E]



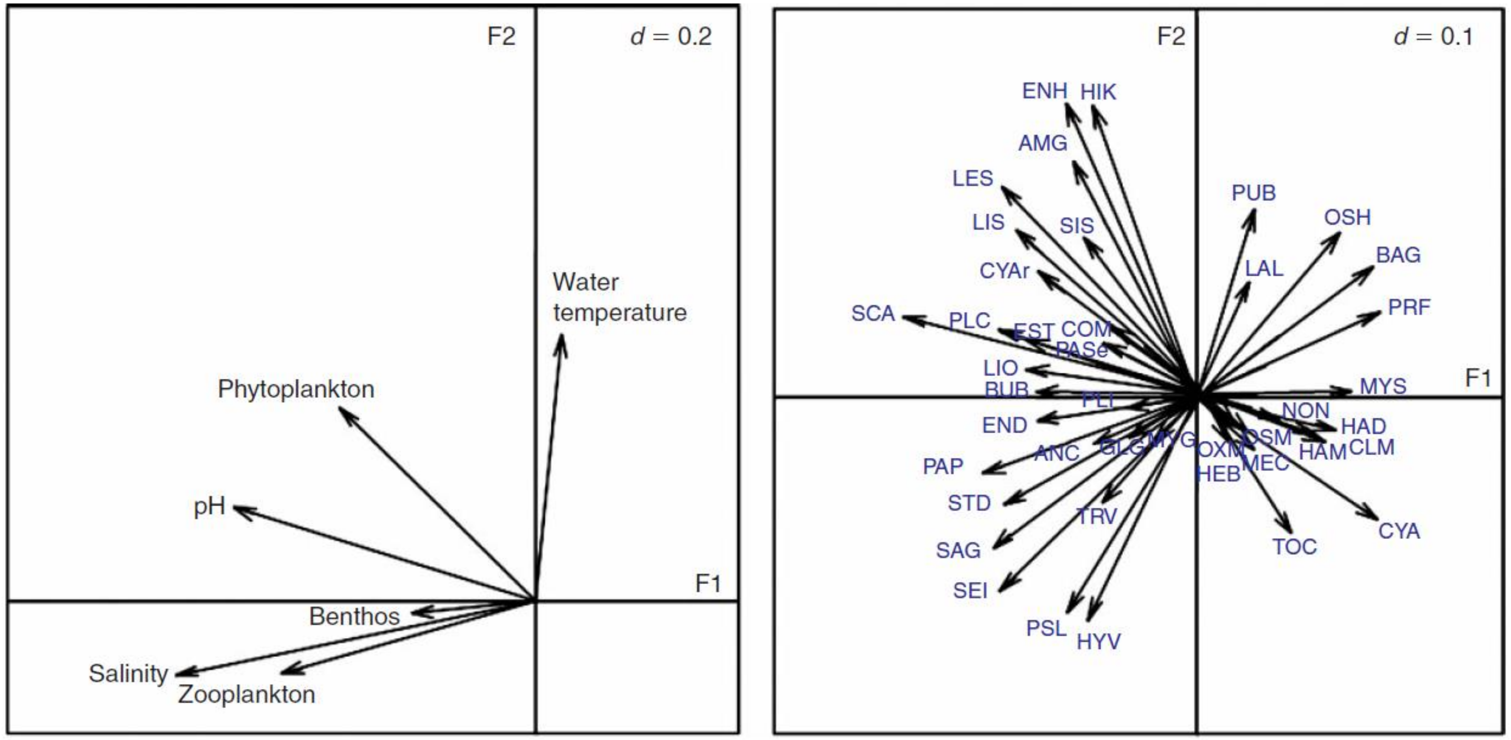


Fig. 6. Results of the co-inertia analysis of environmental variables to fish species found in the study (showing only the ecologically dominant species)

JTF

TRUST FUND - FISHERIES AGENCY OF JAPAN

SEAFDEC

มหาวิทยาลัยอุบลราชธานี