PICES-2018 Annual Meeting Yokohama, Japan Oct. 25-Nov. 4, 2018
Toward Integrated Understanding of Ecosystem Variability in the North Pacific

Effects of sample size and distribution characteristics of survey data on estimation of abundance index of fish population using delta distribution model

Jian Liu, Jing Wang, Binduo Xu*, Chongliang Zhang, Ying Xue, Yiping Ren

Lab of Fisheries Ecosystem Monitoring and Assessment
College of Fisheries
Ocean University of China

Yokohama, Japan
Nov. 2, 2018

Today's content

- Background
- Effect of sample size
- Effect of distribution characteristics of survey data
- Summary

Background

- Fish abundance index is important for stock assessment and fishery management, which is usually derived from fishery-independent bottom trawl survey.
- Due to the heterogeneous distribution of fish populations, the survey data are always variable and contain many zeroes and extreme large values.

Background

The reliability of conventional methods in estimating abundance indices decreases and may cause problems if the extreme values are ignored.

Problem?

Ways?

reduce the impact of extreme values by increasing the sample size covered or conducting replicate surveys?

hard to conduct this survey practically due to the limitation of survey time and cost

Background

- Feasible way to reduce the influence of extreme values by treating zero and nonzero values separately and conducting a log-transformation for nonzero values.
- Delta distribution model is a practicable choice, widely used in estimating abundance index for stock assessment.
- Survey data can be described by delta distribution model when data contain a proportion of zero value and the nonzero values are lognormally distributed.

Delta distribution model

The unbiased estimates of the population mean ρ¯ and variance are given in the following formula（Aichison, 1957）:

$$\bar{\rho}_k=\begin{cases}\frac{m}{n}\exp(\bar{y})g_m(\frac{s^2}{2}), & m>1\\ \frac{x}{n}, & m=1\\ 0, & m=0\end{cases}$$

$$\mathrm{var}(\rho)_{\mathrm{est}}=\begin{cases}\frac{m}{n}\exp(2\bar{y})\left\{\frac{m}{n}g^2{}_m\left(\frac{s^2}{2}\right)-(\frac{m-1}{n-1})g_m(\frac{m-2}{m-1}\cdot s^2)\right\}, & m>1\\ (\frac{x}{n})^2, & m=1\\ 0, & m=0\end{cases}$$

$$g_m(t)=1+\frac{m-1}{m}\cdot t+\sum_{j=2}^{\infty}\frac{(m-1)^{2j-1}}{m^j(m+1)(m+3)K(m+2j-3)}\cdot\frac{t^j}{j\,!}$$

Where, *n* is the number of tows, *m* is the number of nonzero values, $\bar{y}$ and s^2 are the sample mean and variance of the nonzero $\log_e$ values

Question 1

How does the sampling effort affect estimation of fish abundance index using delta distribution model?

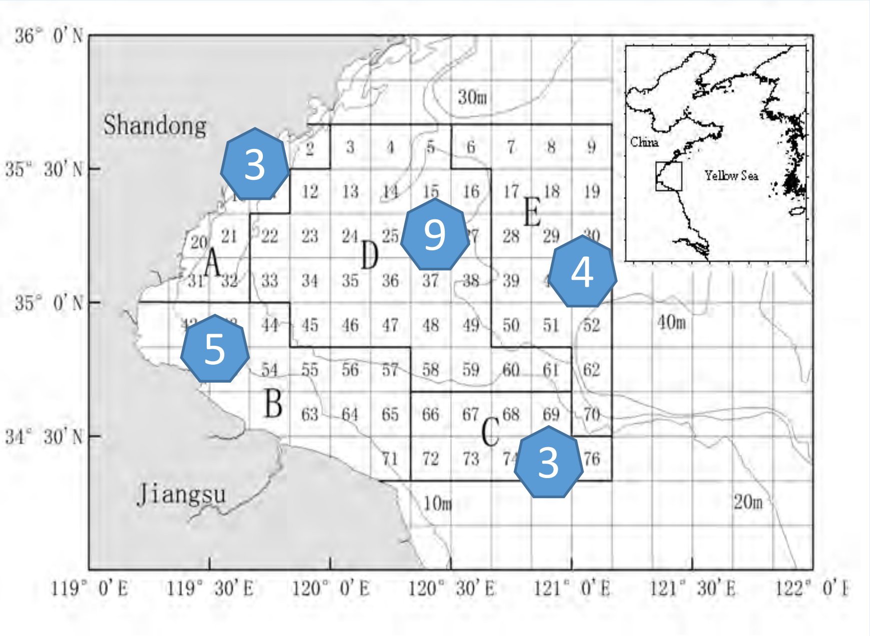
Delta distribution model is often used for the survey data when sample size is large ($>$50 sampling stations) (e.g., Pennington, 1983; Smith, 1990; Li, 2008). What is the performance of this model at relatively small sample size?

(Liu et al. in prep)

Aim of this study

To examine the performance of delta distribution model in estimating abundance indices of fish populations with different sample size.

Data source



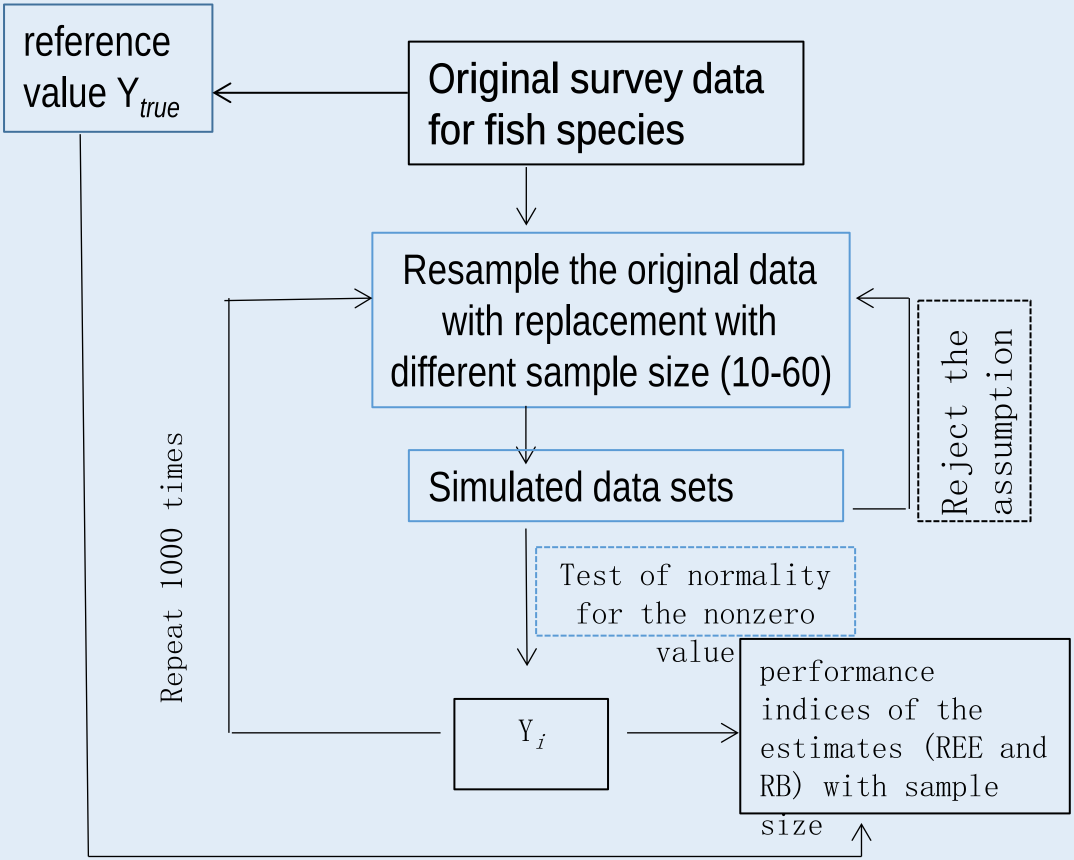
Bottom trawl survey in the Haizhou Bay

- May, December 2011
- 24 sampling stations
- 10 survey data sets of fish species

Original survey data

Survey data	Skewness	Number of zero values
Callionymus sagittal (May)	1.32	6
Chaeturichthys hexanema (May)	1.33	6
Lophius litulon (Dec.)	2.08	8
Hexagrammos otakii (May)	2.36	9
Callionymus sagittal (Dec.)	2.37	6
Conger myriaster (Dec.)	2.55	3
Conger myriaster (May)	2.77	10
Konosirus punctatus (Dec.)	4.12	9
Thryssa chefuensis (Dec.)	4.34	4
Syngnathus acus (May)	4.88	4

Simulation study flowchart



Measures for evaluating performances

REE（relative estimation error）:

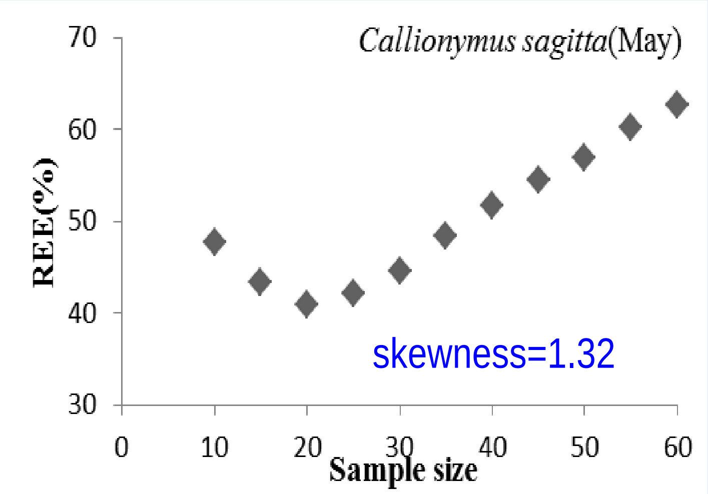
$$\mathrm{REE}=\frac{\sqrt{\sum_{i=1}^{R}(Y_i-Y_{\mathrm{true}})^2/R}}{Y_{true}}\times 100\%$$

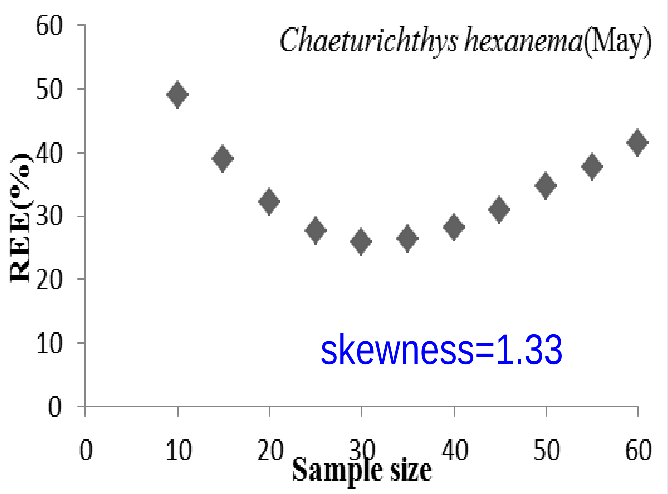
RB（relative bias）:

$$\mathrm{RB}=\frac{\sum_{i=1}^{R} Y_i/R-Ytrue}{Y_{true}}\times 100\%$$

Relative estimation errors (REEs)

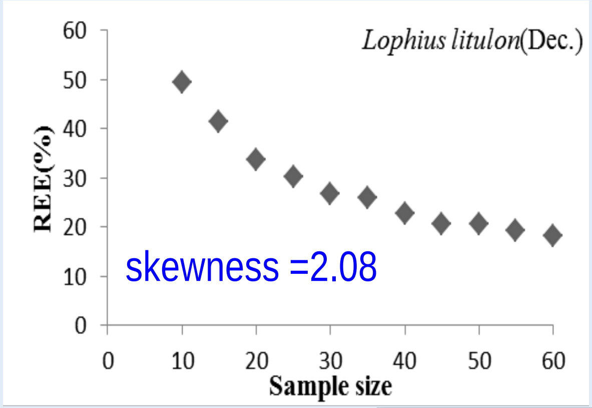
① skewness<1.5

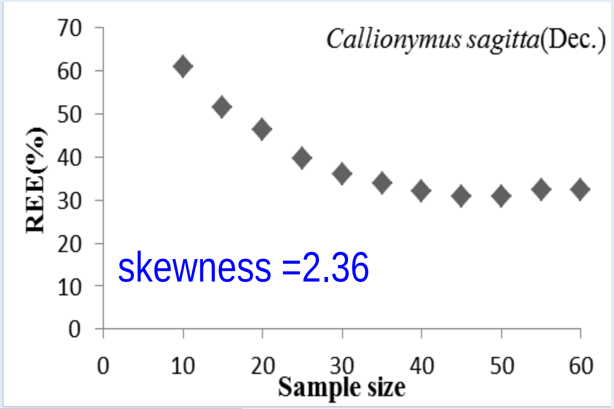


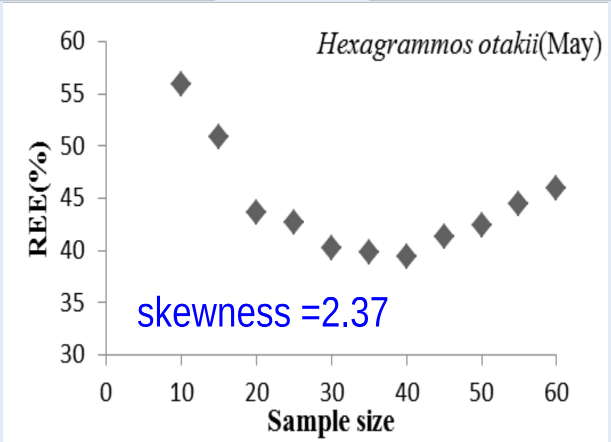


Relative estimation errors (REEs)

② 1.5≤Skewness<2.5

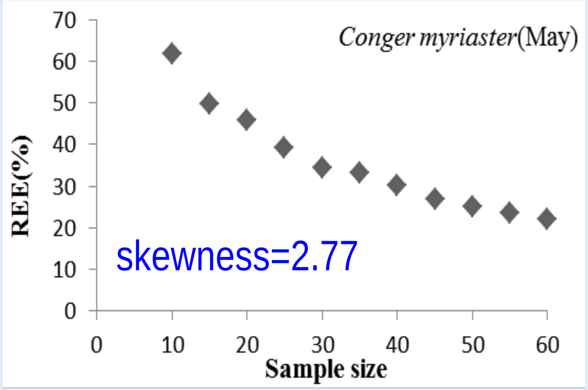


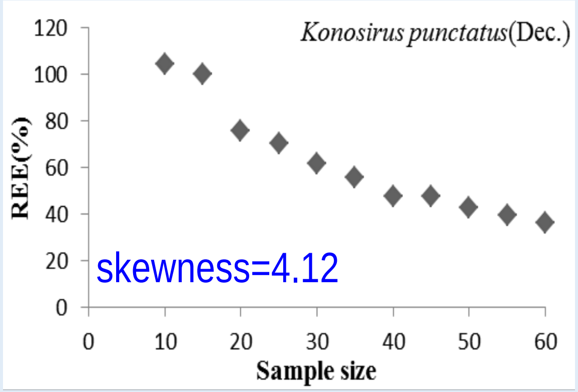


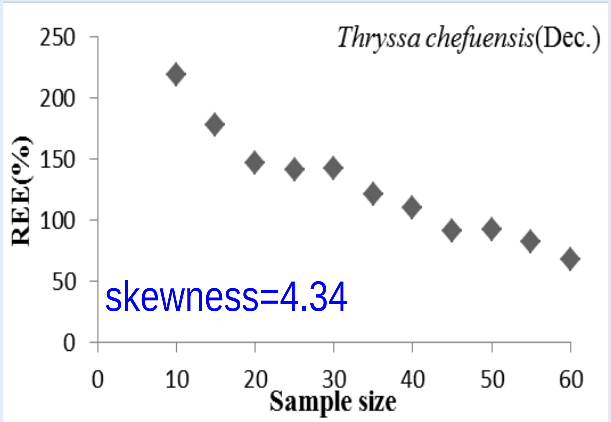


Relative estimation errors (REEs)

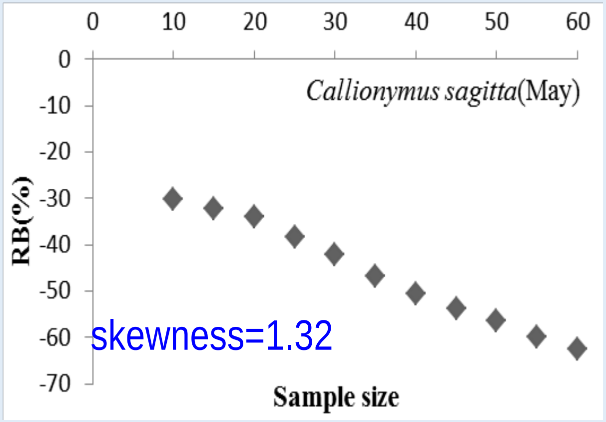
③ skewness>2.5

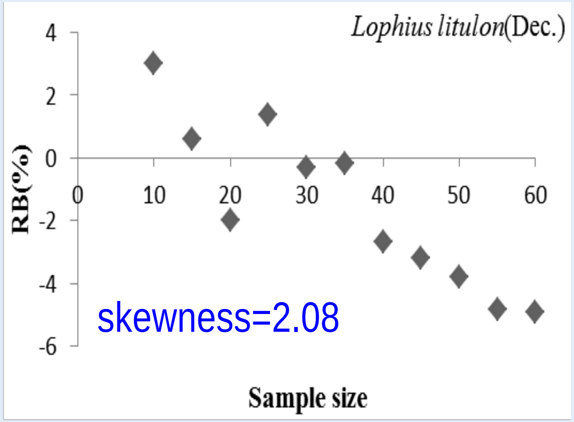


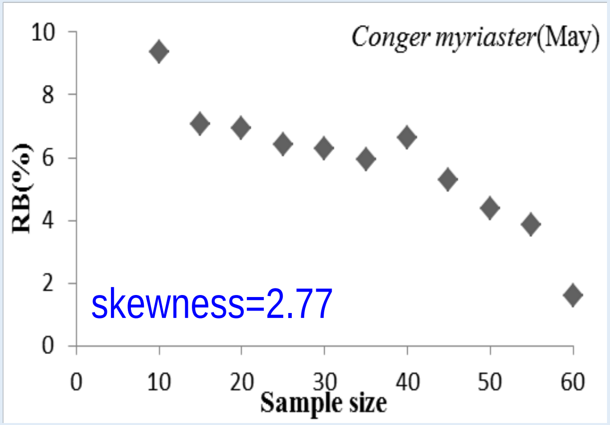


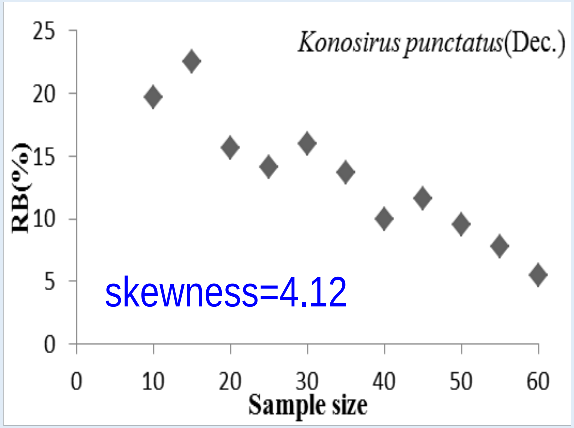


Relative bias (RB)









Key points

- Sample size did affect the estimation of abundance index using delta distribution model， varying with different skewness of data
- Get relatively precise abundance index estimates at relatively small sample size when the original survey data was relatively evenly distributed and at large sample size when the original data were skewed.
- Abundance indices derived from the simulated data based on the delta-distribution model decreased with sample size increasing

Question 2

What is the effect of data distribution characteristics on estimation of fish abundance index using delta distribution model?

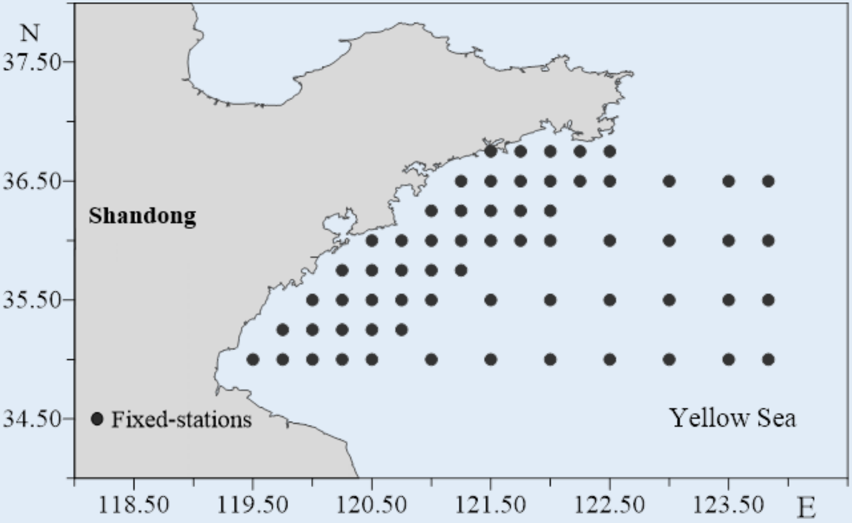
The spatial distribution range and level of aggregation are different for fish species.
For survey data, proportion of zeros and skewness were used to measure the spatial distribution features in this study.

(Xu et al. in prep)

Aim of this study

To examine the effect of distribution characteristics of survey data on estimation of abundance indices of fish populations using delta distribution model.

Data source



Sampling station in the southern waters off Shandong in the Yellow Sea

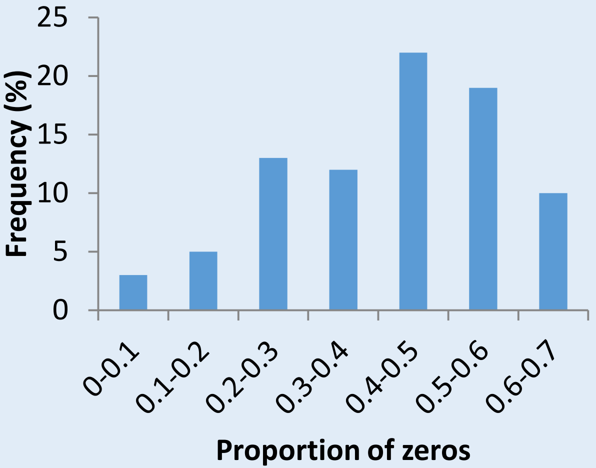
- Oct. 2016, Jan., May, August, 2017
- 63 sampling stations

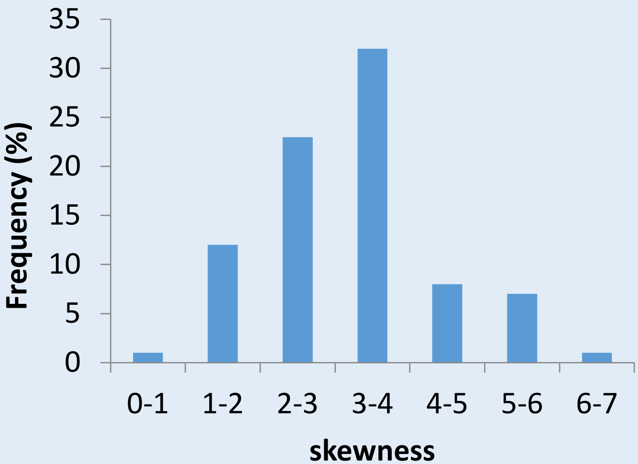
Distribution characteristics of data

84 data sets

Proportion of zero values: 3.17%-63.48%

Skewness: 0.88-6.3

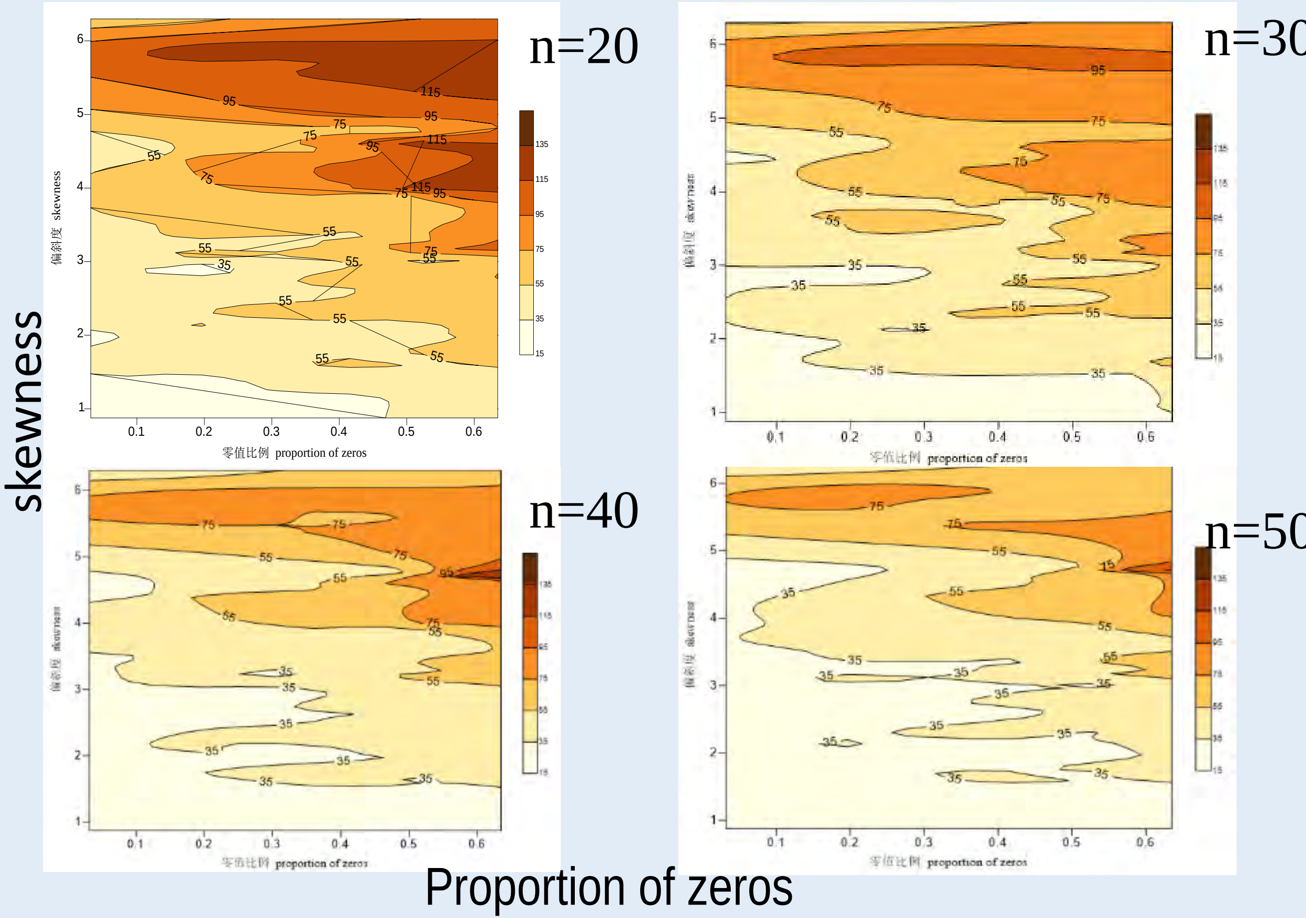




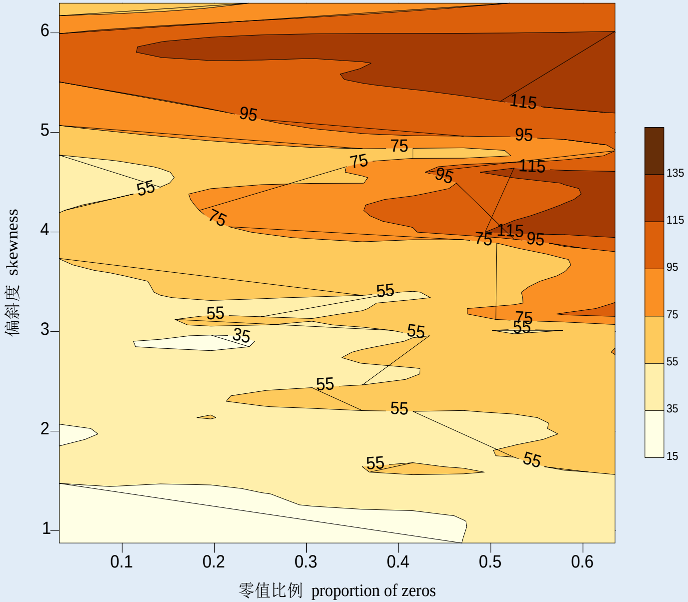
Data analysis and simulation study

- Survey data as original data, calculate the reference value Y_{true}
- Sampling with replacement from the original survey data for each species to generate the simulated data at sample size of 20, 30,40, 50. Abundance index Yi was calculated if the simulated data were lognormally distributed. The resampling was repeated 1000 times.
- Relative estimation error and relative bias were used to evaluate the accuracy and precision of the estimates of abundance indices
- Mapping relationship between REE, RB and proportion of zeros and skewness

Relative estimation error (REE)



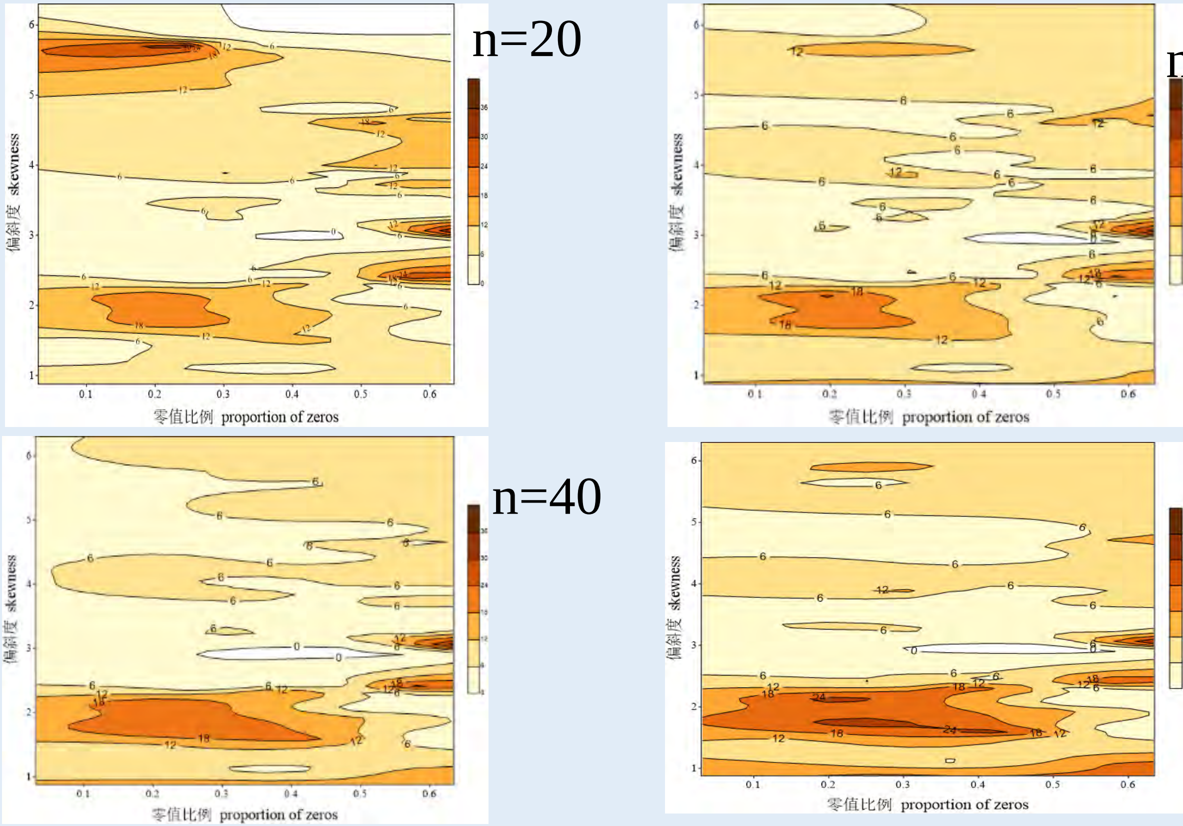
Relative estimation error (REE)



n=20

- REE varied greatly, nonlinear relationship with proportion of zeros and skewness
- When skewness increased, REE fluctuated, increased first and then decreased; when proportion of zeros increased, REE always increased
- REE varied greatly with proportion of zeros, indicating proportion of zeros having higher effect than skewness

Relative bias



skewness

Proportion of zeros

Summary

- Delta-distribution model could get relatively precise estimates at relatively small sample size when the original survey data was relatively evenly distributed and at large sample size when the data were rightly-skewed.

- Distribution characteristics of survey data obviously influenced the performances of the delta-distribution model for estimating abundance index.

Summary

- Increase of zero values proportion, skewness of non-zero values would cause decline in stability and accuracy for estimating abundance indices using delta-distribution model.

- The effects of any of these factors on the estimation of abundance index could not be described by simple linear function.

Acknowledgements

We would like to give thanks to all the members who help collecting the data and the funding projects.

Thank you for your attention!

Binduo Xu
E-mail: bdxu@ouc.edu.cn