Optimization of sampling design for a fishery-independent survey with multiple objectives

Binduo Xu\(^1\), Yiping Ren\(^1\), Yong Chen\(^2\), Chongliang Zhang\(^1,2\), Ying Xue\(^1\) and Rong Wan\(^1\)

\(^1\)College of Fisheries, Ocean University of China, Qingdao, China.
\(^2\)School of Marine Sciences, University of Maine, Orono, ME, USA
Outline

• Background

• Optimization of sampling efforts

• Optimization of stratification scheme
A basic requirement for a regular fisheries stock assessment and management is an estimate of fish abundance index of target fish populations at a defined tempo-spatial scale (Hilborn and Walters 1992).

The data are usually obtained through well-designed fishery-independent survey programs (Cochran 1977).

Stratified random sampling design can improve the precision of estimates when fish distributions are heterogeneous (Cochran 1977; Manly et al. 2002; Miller et al. 2007).
Introduction

• In general, fishery-independent survey programs tend to be more costly and time-consuming than commercial fishery-dependent programs (Scheirer et al. 2004).

• For fish populations with low abundance and aggregated distribution in a coastal ecosystem, high intensity bottom trawl surveys may result in extra mortality and disturbance of benthic community, imposing large negative impacts on the populations and ecosystem.
Introduction

• Therefore, optimization of sampling design is necessary to acquire cost-effective sampling efforts for a fishery-independent survey.

• Computer simulation studies are often used for evaluating sampling strategies in determining an optimal sampling design to achieve the goals of a survey program (Simmonds and Fryer 1996; Liu et al. 2009; Yu et al. 2012).

• Most studies tend to be focused on the optimization of a survey design with a single goal such as yielding high quality of the abundance index for one or a few important fish species, which may differ from the design optimization when multiple goals need to be considered.
Is it possible to reduce the sampling efforts while.....?

Optimization of sampling efforts
Objectives

• Develop a framework for evaluating and optimizing design for a fishery-independent survey for which the main objective is to estimate abundance and species composition in a shallow and dynamic coastal ecosystem with low fish abundances and variable spatial distributions;

• Compare the performance of sampling design with different sample sizes in quantifying the spatial and temporal variability in fish population abundance and species diversity;

• Minimize the impacts of the sampling survey on depleted populations while still achieving reasonable levels of precision for survey estimates.
Stratified random sampling in Haizhou Bay

The map of the study area, stratified random sample stations and bathymetric contours in the Haizhou Bay. The geographic location of Haizhou Bay in the Yellow Sea indicated by the inserted map.
Flow chart of optimization of sampling efforts

1. Survey data
2. Select three types of index (multiple objectives)
3. Calculate the mean/CV of index with original data ($Y_{true}$)
4. Allocation of sample size among strata following the current design
5. Define number of sites in each stratum with reduced sample size
6. Repeat 1000 times (bootstrap)
7. Resample original data at the defined sample size with replacement
8. Calculate mean of each index with resampled data ($Y_{estimated}$)
9. Calculate performance indices of the estimates
10. Performance indices with sample size/CV
11. Compare the performance of different sampling designs for different indices over different seasons

Flow chart of optimization of sampling efforts
The distribution of the total sample size among strata (from A-E) defined in this study. The italic and bold numbers indicate the first new sample sizes after the sampling efforts were reduced. $N_h$ is the number of possible sample units in stratum $h$. $W_h$ is weighting factor of stratum $h$.

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Strata description</th>
<th>$W_h$</th>
<th>$N_h$</th>
<th>The total sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24</td>
</tr>
<tr>
<td>A</td>
<td>&lt;20m, northern, coastal currents</td>
<td>0.13</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>B</td>
<td>&lt;20m, central, coastal currents</td>
<td>0.21</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>C</td>
<td>&lt;20m, southern, coastal currents</td>
<td>0.13</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>D</td>
<td>20-30m, cold water mass</td>
<td>0.38</td>
<td>29</td>
<td>9</td>
</tr>
<tr>
<td>E</td>
<td>&gt;30m, cold water mass</td>
<td>0.17</td>
<td>18</td>
<td>4</td>
</tr>
<tr>
<td>Types of index</td>
<td>Specific index</td>
<td>Species/groups codes</td>
<td>CV (%)</td>
<td>Mean index value (g/h for species and fish groups)</td>
</tr>
<tr>
<td>----------------</td>
<td>----------------</td>
<td>----------------------</td>
<td>-------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Abundance index of individual species</td>
<td>Hairtail (<em>Trichiurus lepturus</em>)</td>
<td>TL¹</td>
<td>29.9 (29.1-30.7)</td>
<td>49.2 (0-170.7)</td>
</tr>
<tr>
<td></td>
<td>Small yellow croaker (<em>Larimichthys polyactis</em>)</td>
<td>LP¹</td>
<td>15.5 (10.8-24.1)</td>
<td>718.5 (0-1426.1)</td>
</tr>
<tr>
<td></td>
<td>Fat greenling (<em>Hexagrammos otakii</em>)</td>
<td>HO¹</td>
<td>17.7 (8.0-23.6)</td>
<td>378.6 (66.9-1145.1)</td>
</tr>
<tr>
<td></td>
<td>Whitespotted conger (<em>Conger myriaster</em>)</td>
<td>CM¹</td>
<td>18.4 (15.5-23.9)</td>
<td>329.7 (38.5-651.5)</td>
</tr>
<tr>
<td></td>
<td>Blenny (<em>Pholis fangi</em>)</td>
<td>PF¹</td>
<td>14.1 (7.8-24.8)</td>
<td>1032.6 (107.4-1935.2)</td>
</tr>
<tr>
<td></td>
<td>Pinkgray goby (<em>Amblychaeturiaichthys hexanema</em>)</td>
<td>AH¹</td>
<td>16.1 (11.5-21.8)</td>
<td>244.0 (61.7-785.1)</td>
</tr>
<tr>
<td></td>
<td>White-hair rough shrimp (<em>Trachypenaeus curviostris</em>)</td>
<td>TC²</td>
<td>17.6 (13.5-23.0)</td>
<td>497.2 (0.2-1187.9)</td>
</tr>
<tr>
<td></td>
<td><em>Metapenaeopsis dalei</em></td>
<td>MD²</td>
<td>22.6 (14.2-27.8)</td>
<td>1221.6 (69.3-2111.3)</td>
</tr>
<tr>
<td></td>
<td><em>Palaemon graviere</em></td>
<td>PG²</td>
<td>22.2 (20.5-23.0)</td>
<td>144.9 (43.9-268.0)</td>
</tr>
<tr>
<td></td>
<td><em>Charybdis bimaculata</em></td>
<td>CB³</td>
<td>13.3 (8.5-15.9)</td>
<td>291.0 (58.1-784.5)</td>
</tr>
<tr>
<td></td>
<td>Squid (<em>Loligo japonica</em>)</td>
<td>LJ⁴</td>
<td>13.5 (9.8-17.1)</td>
<td>1163.7 (101.0-2468.4)</td>
</tr>
<tr>
<td>Abundance index of fish groups</td>
<td>Finfish group</td>
<td>FI</td>
<td>6.4 (4.34-8.50)</td>
<td>8456.2 (2602.5-12220.4)</td>
</tr>
<tr>
<td></td>
<td>Cephalopod group</td>
<td>CE</td>
<td>10.8 (9.3-13.5)</td>
<td>3036.9 (944.2-4284.9)</td>
</tr>
<tr>
<td></td>
<td>Shrimp group</td>
<td>SH</td>
<td>12.2 (8.0-15.5)</td>
<td>3290.4 (1336.8-4235.8)</td>
</tr>
<tr>
<td></td>
<td>Crab group</td>
<td>CR</td>
<td>14.5 (7.2-26.8)</td>
<td>1834.1 (539.7-3412.5)</td>
</tr>
<tr>
<td>Species diversity index</td>
<td>Margalef’s richness index</td>
<td>d</td>
<td>2.1 (1.7-2.6)</td>
<td>3.2 (3.1-3.3)</td>
</tr>
<tr>
<td></td>
<td>Pielou’s evenness index</td>
<td>J’</td>
<td>1.5 (0.9-2.4)</td>
<td>0.6 (0.6-0.7)</td>
</tr>
<tr>
<td></td>
<td>Shannon’s diversity index</td>
<td>H’</td>
<td>1.4 (1.3-2.4)</td>
<td>2.2 (2.0-2.4)</td>
</tr>
</tbody>
</table>
Measures for evaluating performance

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

\[
RB = \frac{\sum_{i=1}^{R} Y_i^{estimated}}{R - Y_{true}} / R \times 100\%
\]

\(Y_{true}\) is the “true” value of abundance index of the individual species, fish groups or species diversity indices calculated from the original survey data, \(Y_i^{estimated}\) is the estimated value from the resampled data in the i\(^{th}\) simulation run, and \(R\) is the number of simulation runs (i.e., 1000 in this study)
Comparison of REE

• REE for all the indices increased when the sample size decreased from 24 to 15 in the four surveys.

• REE for abundance index of the individual species was the largest, followed by that of the fish groups, and REE of the species diversity indices was lowest.

• For most indices, the REE was stable or slightly increased when the sample size decreased from 24 to 22, and had a relatively distinct increase at a sample size of about 21, and then the REE was relatively constant until it showed a large increase at sample size of 15 or 17.
For all the three types of indices combined, a higher CV of indices tended to lead to a higher REE of the indices.
The changes in the REE (%) and the trawl survey catch (in weight) (%) when sample size decreased from 24 to 18 for species/species groups for the stratified random survey.

<table>
<thead>
<tr>
<th>Species/species group</th>
<th>March Catch</th>
<th>March REE</th>
<th>May Catch</th>
<th>May REE</th>
<th>September Catch</th>
<th>September REE</th>
<th>December Catch</th>
<th>December REE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hairtail (<em>Trichiurus lepturus</em>)</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>-24.5</td>
<td>7.8</td>
<td>-27.2</td>
<td>13.6</td>
</tr>
<tr>
<td>Small yellow croaker(<em>Larimichthys polyactis</em>)</td>
<td>/</td>
<td>/</td>
<td>-26.5</td>
<td>4.2</td>
<td>-28.1</td>
<td>6.6</td>
<td>-24.5</td>
<td>4.5</td>
</tr>
<tr>
<td>Fat greenling (<em>Hexagrammos otakii</em>)</td>
<td>-23.7</td>
<td>4.0</td>
<td>-24.1</td>
<td>3.1</td>
<td>-25.4</td>
<td>8.1</td>
<td>-22.4</td>
<td>3.4</td>
</tr>
<tr>
<td>Whitespotted conger (<em>Conger myriaster</em>)</td>
<td>-28.8</td>
<td>8.9</td>
<td>-25.4</td>
<td>7.1</td>
<td>-25.6</td>
<td>3.6</td>
<td>-24.8</td>
<td>5.2</td>
</tr>
<tr>
<td>Blenny (<em>Pholis fangi</em>)</td>
<td>-25.6</td>
<td>4.3</td>
<td>-23.2</td>
<td>3.1</td>
<td>-24.7</td>
<td>6.1</td>
<td>-24.5</td>
<td>2.3</td>
</tr>
<tr>
<td>Pinkgray goby (<em>Amblychaeturchthys hexanema</em>)</td>
<td>-24.4</td>
<td>3.8</td>
<td>-29.1</td>
<td>4.8</td>
<td>-30.7</td>
<td>6.8</td>
<td>-27.7</td>
<td>5.9</td>
</tr>
<tr>
<td>White-hair rough shrimp (<em>Trachypenaeus curviostris</em>)</td>
<td>/</td>
<td>/</td>
<td>-31.2</td>
<td>9.2</td>
<td>-27.5</td>
<td>5.4</td>
<td>-23.8</td>
<td>5.4</td>
</tr>
<tr>
<td><em>Metapenaeopsis dalei</em></td>
<td>-23.3</td>
<td>6.4</td>
<td>-24.7</td>
<td>10.1</td>
<td>-29.4</td>
<td>7.5</td>
<td>-22.9</td>
<td>4.0</td>
</tr>
<tr>
<td><em>Palaemon gravieri</em></td>
<td>-25.8</td>
<td>7.0</td>
<td>-30.5</td>
<td>7.9</td>
<td>-24.9</td>
<td>4.6</td>
<td>-22.9</td>
<td>7.3</td>
</tr>
<tr>
<td><em>Charybdis bimaculata</em></td>
<td>-24.1</td>
<td>2.2</td>
<td>-31.1</td>
<td>4.9</td>
<td>-27.6</td>
<td>3.7</td>
<td>-23.9</td>
<td>3.3</td>
</tr>
<tr>
<td>Squid (<em>Loligo japonica</em>)</td>
<td>-24.0</td>
<td>7.1</td>
<td>-22.9</td>
<td>3.3</td>
<td>-26.1</td>
<td>5.0</td>
<td>-23.4</td>
<td>3.2</td>
</tr>
<tr>
<td>Fish group</td>
<td>-25.4</td>
<td>2.9</td>
<td>-24.7</td>
<td>1.0</td>
<td>-26.1</td>
<td>1.5</td>
<td>-24.6</td>
<td>1.6</td>
</tr>
<tr>
<td>Cephalopod group</td>
<td>-25.6</td>
<td>2.5</td>
<td>-23.6</td>
<td>2.7</td>
<td>-26.4</td>
<td>4.6</td>
<td>-24.0</td>
<td>2.3</td>
</tr>
<tr>
<td>Shrimp group</td>
<td>-24.3</td>
<td>7.0</td>
<td>-25.3</td>
<td>3.8</td>
<td>-26.0</td>
<td>4.1</td>
<td>-23.9</td>
<td>1.7</td>
</tr>
<tr>
<td>Crab group</td>
<td>-24.6</td>
<td>2.3</td>
<td>-25.1</td>
<td>9.3</td>
<td>-28.1</td>
<td>2.6</td>
<td>-23.4</td>
<td>2.9</td>
</tr>
</tbody>
</table>
Is the stratification scheme currently used optimal for the survey?

Optimization of stratification scheme
Objectives

- Develop a framework for evaluating and optimizing stratification schemes in a fishery-independent survey with the main target estimating abundance of individual species and species;
- Compare the performances of different stratification schemes in quantifying the spatial and temporal variability in fish population abundance and species diversity;
- Compare the performance of different stratification schemes when the target indices differ in their spatial distributions;
- Evaluate the consistency of performances for different stratification schemes over time.
Simulation procedure

1. Survey data
2. Follow the sampling effort allocation scheme currently used
3. Define different scenarios of stratification schemes
4. Resample original data for different stratification schemes with replacement
5. Calculate mean of each index with resampled data
6. Calculate performance indices of the estimates
7. Performance indices with stratification schemes
8. Compare the performance of different stratification schemes for different indices over different survey seasons

Repeat at different levels of sampling efforts

2015/11/13 PICES-2015, Oct. 14-25, Qingdao, China
The distribution of the total sample size among strata (from A to E) defined in this study. $N_h$ is the total number of possible sample unit in stratum $h$. $W_h$ is weighting factor of stratum $h$.

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Strata description</th>
<th>$W_h$</th>
<th>$N_h$</th>
<th>24</th>
<th>21</th>
<th>18</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>&lt;20m, northern, coastal currents</td>
<td>0.13</td>
<td>8</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>&lt;20m, central, coastal currents</td>
<td>0.21</td>
<td>12</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>C</td>
<td>&lt;20m, southern, coastal currents</td>
<td>0.13</td>
<td>9</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>D</td>
<td>20-30m, cold water mass</td>
<td>0.38</td>
<td>29</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>E</td>
<td>&gt;30m, cold water mass</td>
<td>0.17</td>
<td>18</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>
Different designs of stratification schemes in the sampling design for the fishery-independent survey.

<table>
<thead>
<tr>
<th>Design</th>
<th>Stratification schemes</th>
<th>Strata description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1(1)</td>
<td>ABCDE</td>
<td>Simple random sampling</td>
</tr>
<tr>
<td>2(2)</td>
<td>ABC/DE</td>
<td>&lt;20m, &gt;20m</td>
</tr>
<tr>
<td>3(2)</td>
<td>ABCD/E</td>
<td>&lt;30m, &gt;30m</td>
</tr>
<tr>
<td>4(3)</td>
<td>AB/C/DE</td>
<td>&lt;20m northern-central, &lt;20m southern, &gt;20m</td>
</tr>
<tr>
<td>5(3)</td>
<td>A/BC/DE</td>
<td>&lt;20m northern, &lt;20m central-southern, &gt;20m</td>
</tr>
<tr>
<td>6(3)</td>
<td>ABC/D/E</td>
<td>&lt;20m, 20-30m, &gt;30m</td>
</tr>
<tr>
<td>7(4)</td>
<td>A/B/C/DE</td>
<td>&lt;20m northern, &lt;20m central, &lt;20m southern, &gt;20m</td>
</tr>
<tr>
<td>8(4)</td>
<td>AB/C/D/E</td>
<td>&lt;20m northern-central, &lt;20m southern, 20-30m, &gt;30m</td>
</tr>
<tr>
<td>9(4)</td>
<td>A/BC/D/E</td>
<td>&lt;20m northern, &lt;20m central-southern, 20-30m, &gt;30m</td>
</tr>
<tr>
<td>10(5)</td>
<td>A/B/C/D/E</td>
<td>&lt;20m northern, &lt;20m central, &lt;20m southern, 20-30m, &gt;30m</td>
</tr>
</tbody>
</table>
Comparison of CVs

4 levels of sampling efforts, 10 stratification scheme designs
The CVs of selected indices were obviously reduced by stratification in comparison with simple random sampling design (i.e., Design 1. The CVs of most indices generally kept relatively stable or showed slight decrease from stratification scheme Designs 2 to 10 in the four survey months.
Comparison of REEs

4 levels of sampling efforts, 10 stratification scheme designs
The REE values of all indices were reduced by different stratification schemes compared with simple random survey design (Design 1). The REE values of the selected indices were relatively constant or exhibited decrease to a certain extent from stratification scheme Designs 2 to 10 in the four survey months.
Comparison of RBs

4 levels of sampling efforts, 10 stratification scheme designs
For four levels of sampling efforts, the RB values for all the indices with different Designs were low, ranging between -8% and 10% without exhibiting consistent positive or negative trends with Designs in all the four sampling months. This result indicated that the estimation of all the indices was unbiased.
Combination of different strata designs and sampling efforts

Contour plots with different stratification schemes and sampling efforts for CV of abundance index of fat greenling (*Hexagrammos otakii*)
Summary

• A simulation approach was developed to evaluate and optimize stratification schemes and sampling efforts for a stratified random survey with multiple goals including estimation of abundance indices of individual species and fish groups and species diversity indices.

• Gains in precision of survey estimates from the stratification schemes were acquired compared to simple random sampling design for most indices.

• The loss of precision of survey estimates due to the reduction of sampling efforts could be compensated by improved stratification schemes.
Summary

• Sampling efforts in a stratified random survey could be reduced while still achieving relatively high precision and accuracy for most indices measuring abundance and biodiversity, which would reduce the cost and negative impacts of survey trawling on those species with low abundance and aggregated distribution in the coastal ecosystem.

• This study also showed that optimization of sampling design for a fishery-independent survey might vary with different survey objectives.

• A post-survey analysis, such as this study, could improve survey designs to achieve the most important survey goals.
Acknowledgements

Ocean University of China
Chen Lab, University of Maine, Orono
China Scholarship Council
Thanks for your patience

Any question?