SHIPSL:
A New Health Index for Puget Sound Lowland Streams

Grace Chiu*, Peter Guttrop**

*NSERC Postdoctoral Fellow, Statistics, U of Washington

**Chair and Professor, Statistics, U of Washington

April 13, 2004
Outline:

1. photos of impacted PSL streams
Outline:

1. photos of impacted PSL streams
2. *biomonitoring* and the *IBI*
Outline:

1. photos of impacted PSL streams
2. *biomonitoring* and the *IBI*
3. statistician’s take:
   - *perhaps something more intuitive and practical?*
Outline:

1. photos of impacted PSL streams
2. *biomonitoring* and the *IBI*
3. statistician’s take:
   - perhaps something more intuitive and practical?
   - Revamping the IBI definition — **SHIPSL**
Outline:

1. photos of impacted PSL streams
2. *biomonitoting* and the *IBI*
3. statistician’s take:
   - *perhaps something more intuitive and practical?*
   ⇒ *Revamping the IBI definition — SHIPSL*
4. bootstrap analyses
Outline:

1. photos of impacted PSL streams
2. *biomonitoring* and the *IBI*
3. statistician’s take:
   - *perhaps something more intuitive and practical?*
   - ⇒ *Revamping the IBI definition — SHIPSL*
4. bootstrap analyses
   - bias
   - uncertainty
Outline:

1. photos of impacted PSL streams
2. *biomonitoring* and the *IBI*
3. statistician’s take:
   - perhaps something more intuitive and practical?
   - Revamping the IBI definition — *SHIPSL*
4. bootstrap analyses
   - bias
   - uncertainty
5. insight
   - *Guess which index we advocate!*
Two Puget Sound Lowland Streams
Two Puget Sound Lowland Streams
Two Puget Sound Lowland Streams

Juanita Creek

Rock Creek
Which Stream Is Healthier??

- Rock Creek
- Juanita Creek

Benthic invertebrate diversity and abundance = function of stream health (biotic integrity)

Benthic Index of Biotic Integrity (B-IBI)
Which Stream Is Healthier??

Check bugs...
Which Stream Is Healthier??

- Check bugs...

![Image of benthic invertebrates comparing Rock Creek and Juanita Creek](image_url)
Which Stream Is Healthier??

- Check bugs...

Rock Creek  |  Juanita Creek
Which Stream Is Healthier??

- Check bugs...

Rock Creek  
Juanita Creek

⇒ *benthic invertebrate diversity and abundance*  
  
  = *function of stream health (biotic integrity)*
Which Stream Is Healthier??

Check bugs...

Rock Creek
Juanita Creek

⇒ benthic invertebrate diversity and abundance
  = function of stream health (biotic integrity)
⇒ Benthic Index of Biotic Integrity (B-IBI)
Biomonitoring

track impact of human activities on ecological systems

Bi-IBI used as a “report card” measure

compare report cards over time and/or over space
Biomonitoring

- track impact of human activities on ecological systems
Biomonitoring

- track impact of human activities on ecological systems

⇒ B-IBI used as a “report card” measure
Biomonitoring

- track impact of human activities on ecological systems
- B-IBI used as a “report card” measure
- compare report cards over time and/or over space
History of IBI

Karr (1981): idea of multimetric index

Karr et al. (1986): fully developed original (Fish) IBI

Kerans & Karr (1994): Benthic-IBI

Karr (1998): B-IBI for PSL streams (10 metrics)

Others: US EPA, Env. Canada, France, Mexico, etc.

Localized versions of IBI

Sometimes non-water dwellers as indicators
History of IBI

Karr (1981): idea of multimetric index
History of IBI

- Karr (1981): idea of multimetric index
  - fish as indicators of stream health
  - Midwestern U.S. freshwater systems
History of IBI

- **Karr (1981):** idea of multimetric index
  - fish as indicators of stream health
  - Midwestern U.S. freshwater systems
- **Karr et al. (1986):** fully developed original (Fish-) IBI

Corvallis EPA Seminar, Page 6
History of IBI

- **Karr (1981):** idea of multimetric index
  - fish as indicators of stream health
  - Midwestern U.S. freshwater systems
- **Karr et al. (1986):** fully developed original (Fish-) IBI
- **Kerans & Karr (1994):** Benthic-IBI
History of IBI

- **Karr (1981):** idea of multimetric index
  - fish as indicators of stream health
  - Midwestern U.S. freshwater systems

- **Karr et al. (1986):** fully developed original (Fish-) IBI

- **Kerans & Karr (1994):** Benthic-IBI
  - 13 metrics for Tennessee Valley rivers
History of IBI

- **Karr (1981):** idea of multimetric index
  - fish as indicators of stream health
  - Midwestern U.S. freshwater systems
- **Karr et al. (1986):** fully developed original (Fish-) IBI
- **Kerans & Karr (1994):** Benthic-IBI
  - 13 metrics for Tennessee Valley rivers
- **Karr (1998):** B-IBI for PSL streams (10 metrics)

Others: US EPA, Env. Canada, France, Mexico, etc.

Localized versions of IBI sometimes non-water dwellers as indicators
History of IBI

- **Karr (1981):** idea of multimetric index
  - fish as indicators of stream health
  - Midwestern U.S. freshwater systems
- **Karr et al. (1986):** fully developed original (Fish-) IBI
- **Kerans & Karr (1994):** Benthic-IBI
  - 13 metrics for Tennessee Valley rivers
- **Karr (1998):** B-IBI for PSL streams (10 metrics)
- **Others:** US EPA, Env. Canada, France, Mexico, etc.
History of IBI

- **Karr (1981):** idea of multimetric index
  - fish as indicators of stream health
  - Midwestern U.S. freshwater systems

- **Karr et al. (1986):** fully developed original (Fish-) IBI

- **Kerans & Karr (1994):** Benthic-IBI
  - 13 metrics for Tennessee Valley rivers

- **Karr (1998):** B-IBI for PSL streams (10 metrics)

- **Others:** US EPA, Env. Canada, France, Mexico, etc.
  - localized versions of IBI
  - sometimes non-water dwellers as indicators
Main Idea of the \*\*-IBI (\* = B or FISH)

can identify attributes / metrics which best reflect diversity / abundance e.g. total taxa, % 3 most dominant taxa, etc.

metrics should be sensitive to impact of urbanization on environment

combine identified metrics

-IBI

Rating scheme for 10-metric PSL B-IBI:

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>18</td>
<td>26</td>
<td>28</td>
<td>36</td>
<td>38</td>
<td>44</td>
<td>46</td>
</tr>
</tbody>
</table>

10

|

VERY POOR POOR FAIR GOOD EXCELLENT
Main Idea of the *-IBI (* = B or FISH)

- identify attributes / metrics which best reflect diversity / abundance
Main Idea of the *-IBI (* = B or FISH)

- identify attributes / metrics which best reflect diversity / abundance

  e.g. total taxa, % 3 most dominant taxa, etc.
Main Idea of the $\ast$-IBI ($\ast$ = B or FISH)

- identify attributes / metrics which best reflect diversity / abundance
  
  e.g. \textit{total taxa}, \textit{\% 3 most dominant taxa}, etc.

- metrics should be sensitive to impact of urbanization on environment
Main Idea of the \( \ast \)-IBI (\( \ast \) = B or FISH)

- identify attributes / metrics which best reflect diversity / abundance
  - e.g. total taxa, % 3 most dominant taxa, etc.
  - metrics should be sensitive to impact of urbanization on environment
  - combine identified metrics \( \Rightarrow \) \( \ast \)-IBI
Main Idea of the **-IBI (** = B or FISH)

- identify attributes / metrics which best reflect diversity / abundance
  
  e.g. *total taxa, % 3 most dominant taxa*, etc.

- metrics should be sensitive to impact of urbanization on environment

- combine identified metrics ⇒ **-IBI

Rating scheme for 10-metric PSL B-IBI:

<table>
<thead>
<tr>
<th></th>
<th>VERY POOR</th>
<th>POOR</th>
<th>FAIR</th>
<th>GOOD</th>
<th>EXCELLENT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
<td>16</td>
<td>18</td>
<td>26</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>36</td>
<td>38</td>
<td>44</td>
<td>46</td>
<td>50</td>
</tr>
</tbody>
</table>
Details: CHOOSING METRICS

- Start with a large number of candidate metrics.
- Apply various statistical diagnostics to screen out irrelevant/redundant metrics as deemed appropriate.

Numerous issues on how to choose metrics will not be discussed here.
Details: CHOOSING METRICS

- start with large number of candidate metrics
- apply various statistical diagnostics* to screen out irrelevant / redundant metrics
Details: CHOOSING METRICS

- start with large number of candidate metrics
- apply various statistical diagnostics* to screen out irrelevant / redundant metrics

* as deemed appropriate
Details: CHOOSING METRICS

- start with large number of candidate metrics
- apply various statistical diagnostics* to screen out irrelevant / redundant metrics
  * as deemed appropriate
- numerous issues on how to choose metrics
  — *we shall not discuss them here*
Details: SCORING METRICS
Details: SCORING METRICS

— a statistical concept in disguise!
— a statistical concept in disguise!

i.e. “standardizing” metric values measured on different scales across different streams
Metric Scoring Roadmap

1. Select reference sites (least impacted)

2. For each metric:
   - Rank sites from worst to best health wrt metric

3. Trisect ranking => reference trisection

4. Metric of sampled site scored against reference trisection

5. Site’s metric score is 1, 3, or 5

6. Adjust ranking for relationship

7. Ranking associated with stream size?
Metric Scoring Roadmap

1. Select reference sites (least impacted)
2. For each metric:
   - Rank sites from worst to best health wrt metric
   - Trisect ranking ==> reference trisection
   - Metric of sampled site scored against reference trisection
   - Site's metric score is 1, 3, or 5
   - Adjust ranking for relationship
   - YES: Ranking associated with stream size?
     - YES: Adjust ranking for relationship
     - NO: Trisect ranking ==> reference trisection
e.g. Metric = Total # Taxa
e.g. Metric = Total # Taxa

(Reference Sites. From Karr et al. 1986.)
e.g. Metric = Total # Taxa

- graph includes repeated observations

(Reference Sites. From Karr et al. 1986.)
e.g. Metric = Total # Taxa

- graph includes repeated observations
- trisection often by eye

(Reference Sites. From Karr et al. 1986.)
e.g. Metric = Total # Taxa

- graph includes repeated observations
- trisection often by eye

!!! KUDOS to John Stoddard et al. for using regression !!

(Reference Sites. From Karr et al. 1986.)
e.g. Metric = Total # Taxa

- graph includes repeated observations
- trisection often by eye

!!! KUDOS to John Stoddard et al. for using regression !!

- score sampled site’s metric (AVERAGED OVER REPLICATES) against graph

(Reference Sites. From Karr et al. 1986.)
e.g. Metric = Total # Taxa

- graph includes repeated observations
- trisection often by eye

!!! KUDOS to John Stoddard et al. for using regression!!

- score sampled site’s metric (AVERAGED OVER REPLICATES) against graph
- metrics with small ranges don’t score 3!!

(Reference Sites. From Karr et al. 1986.)
e.g. Metric = Total # Taxa

- graph includes repeated observations
- trisection often by eye

!!! KUDOS to John Stoddard et al. for using regression !!

- score sampled site’s metric (AVERAGED OVER REPLICATES) against graph
- metrics with small ranges don’t score 3 !!

(Reference Sites. From Karr et al. 1986.)

\[
\text{site’s IBI} = \text{sum of its metric scores}
\]
Details: SCORING METRICS

- subjective definition of reference sites
- arbitrary discrete scale, cutpoints very subjective
  - continuous scale such as [0, 10] also involves cutpoint definitions
- requires recalibration in presence of spatial or temporal changes
  
  # perhaps favors “policy preferences” ?? (a la Robert Lackey)
Details: SCORING METRICS

- subjective definition of reference sites
- arbitrary discrete scale, cutpoints very subjective
  - continuous scale such as [0, 10] also involves cutpoint definitions
- requires recalibration\# in presence of spatial or temporal changes
  \# perhaps favors “policy preferences” ?? (a la Robert Lackey)

Scoring schemes for all versions of IBI have same basic structure.
Subjective definition of reference sites

Arbitrary discrete scale, cutpoints very subjective

Continuous scale such as [0, 10] also involves cutpoint definitions

Requires recalibration\# in presence of spatial or temporal changes

\# perhaps favors “policy preferences” ?? (a la Robert Lackey)

Scoring schemes for all versions of IBI have same basic structure.

Q: Really use IBI to monitor stream health over time??
Stream Health Index for the Puget Sound Lowland

SHIPS employs a statistically sound metric scoring mechanism not affected by spatial / temporal changes.

SIMPLE STANDARDIZATION!

score = \frac{\text{metric value}}{\text{metric mean} \pm \text{metric SD}}

adjusted for stream size (regression), if necessary

\text{SHIPSL} = \text{sum of all scores}

continuous scale centered at 0

reference sites unnecessary (as long as least- and most-impacted sites are present in the study)

non-subjective and requires no recalibration
Stream Health Index for the Puget Sound Lowland

SHIPSL employs a statistically sound metric scoring mechanism not affected by spatial / temporal changes
SHIPSL employs a statistically sound metric scoring mechanism not affected by spatial / temporal changes

*SIMPLE STANDARDIZATION!*
Stream Health Index for the Puget Sound Lowland

**SHIPSL** employs a statistically sound metric scoring mechanism *not* affected by spatial / temporal changes

**SIMPLE STANDARDIZATION!**

\[
\text{score} = \frac{\text{metric value}^\circ - \text{metric mean}^\circ}{\text{metric SD}^\circ}
\]

^\circ \text{ adjusted for stream size (regression), if necessary}
**Stream Health Index for the Puget Sound Lowland**

**SHIPSL** employs a statistically sound metric scoring mechanism not affected by spatial / temporal changes

**SIMPLE STANDARDIZATION!**

\[
\text{score} = \frac{\text{metric value} - \text{metric mean}}{\text{metric SD}}
\]

@ adjusted for stream size (regression), if necessary

⇒ **SHIPSL = sum of all scores**
**Stream Health Index for the Puget Sound Lowland**

**SHIPSL** employs a statistically sound metric scoring mechanism **not** affected by spatial / temporal changes

**SIMPLE STANDARDIZATION!**

\[
\text{score} = \frac{\text{metric value}_@ - \text{metric mean}_@}{\text{metric SD}_@}
\]

@ adjusted for stream size (regression), if necessary

⇒ SHIPSL = sum of all scores

- continuous scale centered at 0
SHIPSL employs a statistically sound metric scoring mechanism not affected by spatial / temporal changes

**SIMPLE STANDARDIZATION!**

\[
\text{score} = \frac{\text{metric value} - \text{metric mean}}{\text{metric SD}}
\]

@ adjusted for stream size (regression), if necessary

⇒ **SHIPSL = sum of all scores**

- continuous scale centered at 0
- reference sites unnecessary
  (as long as least- and most-impacted sites are present in the study)
**Stream Health Index for the Puget Sound Lowland**

**SHIPSL** employs a statistically sound metric scoring mechanism **not** affected by spatial / temporal changes

**SIMPLE STANDARDIZATION!**

\[
\text{score} = \frac{\text{metric value} - \text{metric mean}}{\text{metric SD}}
\]

[@] adjusted for stream size (regression), if necessary

⇒ **SHIPSL** = sum of all scores

- continuous scale centered at 0
- reference sites unnecessary
  (as long as least- and most-impacted sites are present in the study)
- non-subjective and requires no recalibration
Q: How to monitor single site over time without reference sites?

A1: Use sites from previous studies as "reference"

A2: If available, use historical data observed when the best sites were still "pristine"

A3: Define gold standard values for each metric mean and metric SD in the scoring formula, perhaps based on historical data — requires occasional recalibration.
Q: How to monitor single site over time without reference sites??

A1: use sites from previous studies as “reference”
Q: How to monitor single site over time without reference sites??

A1: use sites from previous studies as “reference”

A2: if available, use historical data observed when the best sites were still “pristine”
**Q:** How to monitor single site over time without reference sites??

**A1:** use sites from previous studies as “reference”

**A2:** if available, use historical data observed when the best sites were still “pristine”

**A3:** define **gold standard** values for each **metric mean** and **metric SD** in the scoring formula
Stream Health Index for the Puget Sound Lowland

Q: How to monitor single site over time without reference sites??

A1: use sites from previous studies as “reference”

A2: if available, use historical data observed when the best sites were still “pristine”

A3: define gold standard values for each metric mean and metric SD in the scoring formula

► perhaps based on historical data
— requires occasional recalibration
SHIPSL vs B-IBI: Case Study

SHIPSL vs B-IBI: Case Study

- Study region: Puget Sound Lowland (PSL)
  - 16 watersheds (> 16 sites), 3 replicates per site
SHIPSL vs B-IBI: Case Study

- study region: Puget Sound Lowland (PSL)
  - 16 watersheds (> 16 sites), 3 replicates per site
- scoring criteria calibrated for PSL:
SHIPSL vs B-IBI: Case Study

- study region: Puget Sound Lowland (PSL)
  - 16 watersheds (> 16 sites), 3 replicates per site
- scoring criteria calibrated for PSL:

<table>
<thead>
<tr>
<th>metric</th>
<th>1</th>
<th>3</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) total taxa</td>
<td>[0,14)</td>
<td>[14,28)</td>
<td>≥28</td>
</tr>
<tr>
<td>5) % dominance</td>
<td>&gt;75</td>
<td>(55,75]</td>
<td>[0,55]</td>
</tr>
<tr>
<td>10) % predators</td>
<td>[0,4.5)</td>
<td>[4.5,9)</td>
<td>≥9</td>
</tr>
</tbody>
</table>
### SHIPSL vs B-IBI: Case Study

- Study region: Puget Sound Lowland (PSL)
- 16 watersheds (> 16 sites), 3 replicates per site
- Scoring criteria calibrated for PSL:

<table>
<thead>
<tr>
<th>Metric</th>
<th>1</th>
<th>3</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Total taxa</td>
<td>[0,14)</td>
<td>[14,28)</td>
<td>≥ 28</td>
</tr>
<tr>
<td>5) % dominance</td>
<td>&gt;75</td>
<td>(55,75]</td>
<td>[0,55]</td>
</tr>
</tbody>
</table>

@ Most recent calibration due to change in taxonomic identification rules
SHIPSL vs B-IBI: Case Study

- Study region: Puget Sound Lowland (PSL)
- 16 watersheds (> 16 sites), 3 replicates per site
- Scoring criteria calibrated for PSL:

<table>
<thead>
<tr>
<th>metric</th>
<th>1</th>
<th>3</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>total taxa</td>
<td>[0,14)</td>
<td>[14,28)</td>
<td>≥28</td>
</tr>
<tr>
<td>% dominance</td>
<td>&gt;75</td>
<td>(55,75)</td>
<td>[0,55]</td>
</tr>
<tr>
<td>% predators</td>
<td>[0,4.5)</td>
<td>[4.5,9)</td>
<td>≥9</td>
</tr>
</tbody>
</table>

@ Most recent calibration due to change in taxonomic identification rules
SHIPSL vs B-IBI: Case Study

- metric value = average over replicates
SHIPSL vs B-IBI: Case Study

metric value = average over replicates

except \# long-lived taxa and \# intolerant taxa for B-IBI:
metric value = pooled over replicates
(i.e. value for supersample)
SHIPSL vs B-IBI: Case Study

- metric value = average over replicates

  - except # long-lived taxa and # intolerant taxa for B-IBI:
    metric value = pooled over replicates
    (i.e. value for supersample)

- measure of human influence:
  
  % urbanized area
SHIPSL vs B-IBI: Case Study

1997 B-IBI

Frequency

10 20 30 40 50

0 1 2 3 4 5 6
SHIPSL vs B-IBI: Case Study

1997 B-IBI

<table>
<thead>
<tr>
<th>Frequency</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1997 SHIPSL

<table>
<thead>
<tr>
<th>Frequency</th>
<th>-20</th>
<th>-10</th>
<th>0</th>
<th>10</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>0</td>
</tr>
</tbody>
</table>

Correlation: \( corr = 0.965 \)

SHIPSL retains general distribution shape highly correlated with B-IBI on a continuous scale with no tied sites.
SHIPSL vs B-IBI: Case Study

1997 B-IBI

1997 SHIPSL

\[ \text{corr} = 0.965 \]

SHIPSL retains general dist’n shape
highly correlated with B-IBI
continuous scale \(\Rightarrow\) no tied sites
SHIPSL vs B-IBI: Case Study

1998 B-IBI

1998 SHIPSL

corr = 0.979

SHIPSL gives clearer distinction of healthy sites??
SHIPSL vs B-IBI: Case Study

1998 B–IBI

1998 SHIPSL

corr = 0.979

SHIPSL gives clearer distinction of healthy sites...
SHIPSL vs B-IBI: Case Study

1998 B-IBI

\[\text{corr} = 0.979\]

1998 SHIPSL

gap \Rightarrow SHIPSL gives clearer distinction of healthy sites ??
SHIPSL vs B-IBI: Case Study

1997

+: BIBI

% urban

% urban
SHIPSL vs B-IBI: Case Study

1997

+: BIBI / o: SHIPSL (rescaled)

% urban

Corvallis EPA Seminar, Page 22
SHIPSL vs B-IBI: Case Study

1997

+: BIBI / o: SHIPSL (rescaled)

% urban

1998

+: BIBI

% urban
SHIPSL vs B-IBI: Case Study

1997
+: BIBI / o: SHIPSL (rescaled)

1998
+: BIBI / o: SHIPSL (rescaled)

SHIPSL indicates smaller disparity among mid-ranked sites
SHIPSL missed REAL disparity? or...
B-IBI too variable??

Corvallis EPA Seminar, Page 24
SHIPSL vs B-IBI: Case Study

SHIPSL indicates smaller disparity among mid-ranked sites
SHIPSL vs B-IBI: Case Study

SHIPSL indicates smaller disparity among mid-ranked sites

SHIPSL missed *REAL* disparity?
SHIPSL vs B-IBI: Case Study

SHIPSL indicates smaller disparity among mid-ranked sites

SHIPSL missed *REAL* disparity?

or...
SHIPSL vs B-IBI: Case Study

SHIPSL indicates smaller disparity among mid-ranked sites

SHIPSL missed *REAL* disparity?

or... B-IBI *too variable*??
A Bootstrap Study

GOAL:

Compare accuracy and precision between B-IBI and SHIPSL

METHOD:

resample organisms from (observed) field samples (a la Fore et al. (1994))

obtain 10,000 bootstrap samples

DATA:

Morley's 1997 data
A Bootstrap Study

GOAL: Compare accuracy and precision between B-IBI and SHIPSL
A Bootstrap Study

GOAL:
Compare accuracy and precision between B-IBI and SHIPSL

METHOD:
resample organisms from (observed) field samples (a la Fore et al. (1994))
A Bootstrap Study

**GOAL:**
Compare *accuracy* and *precision* between B-IBI and SHIPSL

**METHOD:**
resample organisms from (observed) field samples (a la Fore et al. (1994))

obtain 10,000 bootstrap samples
A Bootstrap Study

**GOAL:**
Compare **accuracy** and **precision** between B-IBI and SHIPSL

**METHOD:**
resample organisms from (observed) field samples
(a la Fore et al. (1994))

▶ obtain 10,000 bootstrap samples

**DATA:**
Morley’s 1997 data
### A Bootstrap Study

<table>
<thead>
<tr>
<th>TAXON ID</th>
<th>TAXON ATTRIBUTES</th>
<th>REP 1 COUNT</th>
<th>REP 2 COUNT</th>
<th>REP 3 COUNT</th>
<th>SITE TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>…</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>…</td>
<td>3</td>
<td>12</td>
<td>8</td>
<td>23</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
<tr>
<td>80</td>
<td>…</td>
<td>12</td>
<td>21</td>
<td>7</td>
<td>31</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td></td>
<td><strong>925</strong></td>
<td><strong>2321</strong></td>
<td><strong>542</strong></td>
<td><strong>3788</strong></td>
</tr>
</tbody>
</table>

**Limitations:**
- Observed 0's <就像一个结构上的0一样, 总是产生bootstrap 0's
- 等概率采样在生物意义上合理吗？

---

*Corvallis EPA Seminar, Page 26*
### A Bootstrap Study

<table>
<thead>
<tr>
<th>TAXON ID</th>
<th>TAXON ID</th>
<th>REP 1 COUNT</th>
<th>REP 2 COUNT</th>
<th>REP 3 COUNT</th>
<th>SITE TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>…</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>…</td>
<td>3</td>
<td>12</td>
<td>8</td>
<td>23</td>
</tr>
<tr>
<td>80</td>
<td>…</td>
<td>12</td>
<td>21</td>
<td>7</td>
<td>31</td>
</tr>
</tbody>
</table>

**TOTAL**

925 2321 542 3788

**Limitations**:
- Observed 0's will always produce bootstrap 0's when equal-probability sampling makes biological sense??
## A Bootstrap Study

### Table: Observed Taxon Counts

<table>
<thead>
<tr>
<th>TAXON ID</th>
<th>TAXON ID</th>
<th>REP 1 COUNT</th>
<th>REP 2 COUNT</th>
<th>REP 3 COUNT</th>
<th>SITE TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>...</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>...</td>
<td>3</td>
<td>12</td>
<td>8</td>
<td>23</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>80</td>
<td>...</td>
<td>12</td>
<td>21</td>
<td>7</td>
<td>31</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>925</td>
<td>2321</td>
<td>542</td>
<td>3788</td>
</tr>
</tbody>
</table>

### Table: Bootstrap Taxon Counts

<table>
<thead>
<tr>
<th>TAXON ID</th>
<th>BOOTSTRAP REP 2 COUNT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>17</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>80</td>
<td>14</td>
</tr>
<tr>
<td>TOTAL</td>
<td>2321</td>
</tr>
</tbody>
</table>
A Bootstrap Study

### Observed Table

<table>
<thead>
<tr>
<th>TAXON ID</th>
<th>TAXON ID</th>
<th>REP 1 COUNT</th>
<th>REP 2 COUNT</th>
<th>REP 3 COUNT</th>
<th>SITE TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>...</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>...</td>
<td>3</td>
<td>12</td>
<td>8</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>...</td>
<td>12</td>
<td>21</td>
<td>7</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>925</td>
<td>2321</td>
<td>542</td>
<td>3788</td>
</tr>
</tbody>
</table>

### Limitations:

- Observed 0's should be treated as structural 0's.
- Equal-probability sampling makes biological sense?

### Bootstrap Table

<table>
<thead>
<tr>
<th>TAXON ID</th>
<th>REP 2 COUNT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>14</td>
</tr>
<tr>
<td>TOTAL</td>
<td>2321</td>
</tr>
</tbody>
</table>
A Bootstrap Study

<table>
<thead>
<tr>
<th>TAXON ID</th>
<th>TAXON</th>
<th>REP 1 COUNT</th>
<th>REP 2 COUNT</th>
<th>REP 3 COUNT</th>
<th>SITE TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>...</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>...</td>
<td>3</td>
<td>12</td>
<td>8</td>
<td>23</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>80</td>
<td>...</td>
<td>12</td>
<td>21</td>
<td>7</td>
<td>31</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>925</td>
<td>2321</td>
<td>542</td>
<td>3788</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TAXON ID</th>
<th>BOOTSTRAP</th>
<th>REP 2 COUNT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>17</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td>...</td>
</tr>
<tr>
<td>80</td>
<td></td>
<td>14</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>2321</td>
</tr>
</tbody>
</table>

Limitations:
- observed 0’s always produce bootstrap 0’s

Corvallis EPA Seminar, Page 26
## A Bootstrap Study

<table>
<thead>
<tr>
<th>TAXON ID</th>
<th>TAXON Attributes</th>
<th>REP 1 Count</th>
<th>REP 2 Count</th>
<th>REP 3 Count</th>
<th>SITE Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>...</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>...</td>
<td>3</td>
<td>12</td>
<td>8</td>
<td>23</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>80</td>
<td>...</td>
<td>12</td>
<td>21</td>
<td>7</td>
<td>31</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td></td>
<td><strong>925</strong></td>
<td><strong>2321</strong></td>
<td><strong>542</strong></td>
<td><strong>3788</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TAXON ID</th>
<th>BOOTSTRAP Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0 &lt;========= like a structural 0</td>
</tr>
<tr>
<td>2</td>
<td>17</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>80</td>
<td>14</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>2321</strong></td>
</tr>
</tbody>
</table>

**Limitations:**
- Observed 0's always produce bootstrap 0's.
- Equal-probability sampling makes biological sense??
A Bootstrap Study: RESULTS

First, site-wise bias:
... each site has one observed B-IBI, one observed SHIPSL. 10,000 bootstrap field samples, 10,000 bootstrap B-IBI, SHIPSL values. Mean of bootstrap B-IBI, mean of bootstrap SHIPSL values. Compare observed value and bootstrap mean. Difference is (bootstrap) bias.
A Bootstrap Study: RESULTS

First, site-wise bias:
... each site has
A Bootstrap Study: RESULTS

First, site-wise bias:
... each site has
- one observed B-IBI, one observed SHIPSL
A Bootstrap Study: RESULTS

First, site-wise bias:

... each site has

- one observed B-IBI, one observed SHIPSL
- 10,000 bootstrap field samples
First, site-wise bias:
... each site has

- one observed B-IBI, one observed SHIPSL
- 10,000 bootstrap field samples

⇒ 10,000 bootstrap B-IBI, SHIPSL values
A Bootstrap Study: RESULTS

First, site-wise bias:
... each site has
  - one observed B-IBI, one observed SHIPSL
  - 10,000 bootstrap field samples
    ⇒ 10,000 bootstrap B-IBI, SHIPSL values
    ⇒ mean of bootstrap B-IBI, mean of bootstrap SHIPSL
A Bootstrap Study: RESULTS

First, site-wise bias:
... each site has

- one observed B-IBI, one observed SHIPSL
- 10,000 bootstrap field samples
  ⇒ 10,000 bootstrap B-IBI, SHIPSL values
  ⇒ mean of bootstrap B-IBI, mean of bootstrap SHIPSL

Compare observed value and bootstrap mean
A Bootstrap Study: RESULTS

First, site-wise bias:
... each site has

- one observed B-IBI, one observed SHIPSL
- 10,000 bootstrap field samples
  ⇒ 10,000 bootstrap B-IBI, SHIPSL values
  ⇒ mean of bootstrap B-IBI, mean of bootstrap SHIPSL

Compare observed value and bootstrap mean
  ⇒ difference is (bootstrap) bias
A Bootstrap Study: SITE-WISE BIAS

**B - I B I**

$r = -0.63$

<table>
<thead>
<tr>
<th>Site-wise bias</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>-4 -3 -2 -1 0 1</td>
<td>0 2 4 6 8</td>
</tr>
</tbody>
</table>

observed B-IBI

**S H I P S L**

$r = -0.22$

<table>
<thead>
<tr>
<th>Site-wise bias</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.0 -0.5 0.0 0.5 1.0</td>
<td>0 2 4 6 8</td>
</tr>
</tbody>
</table>

observed SHIPSL
Next, **sample SD:**
Next, sample SD:

... take $SD(\text{B-IBI})$ and $SD(\text{SHIPSL})$ over sample of 18 sites
A Bootstrap Study: RESULTS

Next, sample SD:

... take $SD(\ B-IBI\ )$ and $SD(\ SHIPSL\ )$ over sample of 18 sites

do same for each bootstrap sample
A Bootstrap Study: RESULTS

Next, sample SD:

... take $SD(\ B-IBI\ )$ and $SD(\ SHIPSL\ )$ over sample of 18 sites

- do same for each bootstrap sample

$\Rightarrow\ 10,000\ BSD(B-IBI);$

$10,000\ BSD(SHIPSL)$
A Bootstrap Study: RESULTS

Next, sample SD:

... take $SD(\ B-IBI\ )$ and $SD(\ SHIPSL\ )$ over sample of 18 sites

- do same for each bootstrap sample

⇒ 10,000 $BSD(\ B-IBI\ )$;
  10,000 $BSD(\ SHIPSL\ )$

- compare distribution of $BSD$’s to observed $SD(\ B-IBI\ )$ and $SD(\ SHIPSL\ )$
A Bootstrap Study: SAMPLE SD’s

**SHIPSL**

- Frequency distribution
- Standard Error (se) = 0.12

**B–IBI**

- Frequency distribution
- Standard Error (se) = 0.35

---

B-IBI SD has large bias!! need rescaling (index) and bias correction (SD) to compare variabilities.

Bootstrap index value

Central bootstrap SD

get SD’s for these (dist’n now centered at 1)

Correct new bootstrap SD dist’n for bias.
A Bootstrap Study: SAMPLE SD’s

**SHIPSL**
- Frequency
- se = 0.12

**B–IBI**
- Frequency
- se = 0.35

### Notes
- **B-IBI SD has large bias!!**
- Need rescaling (index) and bias correction (SD) to compare variabilities

*Corvallis EPA Seminar, Page 30*
A Bootstrap Study: SAMPLE SD’s

SHIPSL

B−IBI

B-IBI SD has large bias!!

need rescaling (index) and bias correction (SD) to compare variabilities
A Bootstrap Study: SAMPLE SD’s

SHIPSL

Frequency

0 500 1000 1500

7.6 7.8 8.0 8.2 8.4

se = 0.12

B–IBI

Frequency

0 1000 2000

7.0 7.5 8.0 8.5 9.0 9.5

se = 0.35

**B–IBI SD has large bias!!**

need rescaling (index) and bias correction (SD) to compare variabilities

\[
\text{bootstrap index value} \quad \text{central bootstrap SD} \quad \Rightarrow \text{get SD’s for these (dist’n now centered at 1)}
\]

\[
\text{correct new bootstrap SD dist’n for bias}
\]
A Bootstrap Study: SAMPLE SD’s
A Bootstrap Study

SO FAR:

Site-wise:
- SHIPSL reduces Bias (index value)

Sample-wise:
- SHIPSL eliminates Bias (sample mean)
- SHIPSL reduces SD (sample)
- SHIPSL reduces Bias (sample SD) and SE (sample SD)

Q: Is SHIPSL really measuring stream health?
A Bootstrap Study

SO FAR:

- Site-wise: SHIPSL reduces *Bias* (index value)
A Bootstrap Study

SO FAR:

- **Site-wise:** SHIPSL reduces *Bias(index value)*
- **Sample-wise:**

Corvallis EPA Seminar, Page 32
A Bootstrap Study

SO FAR:

- **Site-wise**: SHIPSL reduces $\text{Bias( index value )}$
- **Sample-wise**:
  - SHIPSL eliminates $\text{Bias( sample mean )}$
A Bootstrap Study

SO FAR:

- **Site-wise**: SHIPSL reduces *Bias(index value)*
- **Sample-wise**:
  - SHIPSL eliminates *Bias(sample mean)*
  - SHIPSL reduces *SD(sample)*
A Bootstrap Study

SO FAR:

- **Site-wise**: SHIPSL reduces $Bias(\text{index value})$
- **Sample-wise**:  
  - SHIPSL eliminates $Bias(\text{sample mean})$
  - SHIPSL reduces $SD(\text{sample})$
  - SHIPSL reduces $Bias(\text{sample SD})$ and $SE(\text{sample SD})$
A Bootstrap Study

SO FAR:

- Site-wise: SHIPSL reduces $Bias(\text{index value})$

- Sample-wise:
  - SHIPSL eliminates $Bias(\text{sample mean})$
  - SHIPSL reduces $SD(\text{sample})$
  - SHIPSL reduces $Bias(\text{sample SD})$ and $SE(\text{sample SD})$

Q: Is SHIPSL really measuring stream health?
A Bootstrap Study: \( \text{Corr}(\text{SHIPSL, B-IBI}) \)

The diagram shows a histogram of bootstrap correlation values. The x-axis represents the bootstrap correlation values ranging from 0.92 to 0.98, while the y-axis indicates frequency. The histogram peaks around a correlation value of 0.96, suggesting that the majority of bootstrap samples resulted in correlations close to this value. The distribution is slightly skewed towards higher correlation values.
A Bootstrap Study: \( \text{Corr}( \text{SHIPSL}, B-IBI ) \)

Both contain the same information about biotic integrity (without considering bias and scale difference).

Corvallis EPA Seminar, Page 33
Q: How much more efficient is SHIPSIL than B-IBI in practice?
A Bootstrap Study: Power Analysis

\[ \delta = \text{true difference in health} \at \text{between two sites} \]
A Bootstrap Study: Power Analysis

\[ \delta = \text{true difference in health@ between two sites} \]

@ mean B-IBI or SHIPSL value
A Bootstrap Study: Power Analysis

\[ \delta = \text{true difference in health}\@ \text{ between two sites} \]

\@ mean B-IBI or SHIPSL value

10,000 bootstrap \( \delta^*_k \)'s \[ \Rightarrow \] 100(1 - \( \alpha \)) % bootstrap C.I. for \( \delta \)
A Bootstrap Study: Power Analysis

\[ \delta = \text{true difference in health} \atop \text{mean B-IBI or SHIPSL value} \]

10,000 bootstrap \( \delta_k^* \)'s \( \Rightarrow \) 100(1 - \alpha) \% bootstrap C.I. for \( \delta \)

\[
\text{Power}_\alpha(\delta) = P \left\{ (1 - \alpha) \text{ C.I. for } \delta \text{ excludes } 0 \right\} \\
\approx \frac{\# \left\{ (1 - \alpha) \text{ bootstrap C.I. for } \delta \text{ excludes } 0 \right\}}{\# \text{ bootstrap samples}}
\]
A Bootstrap Study: Power Analysis

\( \delta = \text{true difference in health@ between two sites} \)

@ mean B-IBI or SHIPSL value

10,000 bootstrap \( \delta^*_k \)'s \( \Rightarrow \) 100(1 - \( \alpha \)) % bootstrap C.I. for \( \delta \)

\[
\text{Power}_\alpha(\delta) = P\left\{ (1 - \alpha) \ C.I. \ for \ \delta \text{ excludes } 0 \right\}
\approx \frac{\# \ \left\{ (1 - \alpha) \text{ bootstrap C.I. for } \delta \text{ excludes } 0 \right\}}{\# \text{ bootstrap samples}}
\]

But this involves bootstrapping the bootstrap!

(since we have only ONE bootstrap C.I. for every pair of sites)
No double bootstrap, *thank you.*
A Bootstrap Study: Power Analysis

No double bootstrap, thank you.

\[ \text{ad hoc Power} = \frac{\# \text{ C.I.'s excluding 0}}{\text{total \# pairs}} \]

<table>
<thead>
<tr>
<th></th>
<th>Excludes 0</th>
<th>Includes 0</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>C.I.'s</td>
<td>65.4%</td>
<td>7.8%</td>
<td>73.2%</td>
</tr>
<tr>
<td>B-I 1</td>
<td>16.3%</td>
<td>10.5%</td>
<td>26.8%</td>
</tr>
<tr>
<td>Total</td>
<td>81.7%</td>
<td>18.3%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>
No double bootstrap, thank you.

**ad hoc Power**

\[ \text{ad hoc Power} = \frac{\# \text{ C.I.'s excluding 0}}{\text{total # pairs}} \]

Take \( \alpha = 5\% \):

<table>
<thead>
<tr>
<th></th>
<th>SHIPSL C.I.'s</th>
<th>B-IBI C.I.'s</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>excludes 0</td>
<td>includes 0</td>
</tr>
<tr>
<td>excludes 0</td>
<td>65.4%</td>
<td>16.3%</td>
</tr>
<tr>
<td>includes 0</td>
<td>7.8%</td>
<td>10.5%</td>
</tr>
<tr>
<td>Total</td>
<td>73.2%</td>
<td>26.8%</td>
</tr>
</tbody>
</table>

SHIPSL is about 1.12 times more "powerful" than B-IBI.
Picking up what the other has missed: SHIPSL is 2X more efficient.
No double bootstrap, *thank you.*

**ad hoc Power**

\[
\text{ad hoc Power} = \frac{\# \text{ C.I.'s excluding 0}}{\text{total } \# \text{ pairs}}
\]

Take \( \alpha = 5\%: \)

<table>
<thead>
<tr>
<th></th>
<th>SHIPSL C.I.'s</th>
<th>B-IBI C.I.'s</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>excludes 0</td>
<td>includes 0</td>
</tr>
<tr>
<td><strong>excludes 0</strong></td>
<td>65.4%</td>
<td>16.3%</td>
</tr>
<tr>
<td><strong>includes 0</strong></td>
<td>7.8%</td>
<td>10.5%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>73.2%</td>
<td>26.8%</td>
</tr>
</tbody>
</table>

⇒ SHIPSL is about 1.12 times more “powerful”
A Bootstrap Study: Power Analysis

No double bootstrap, thank you.

\[ \text{ad hoc Power} = \frac{\# \text{ C.I.'s excluding 0}}{\text{total } \# \text{ pairs}} \]

Take \( \alpha = 5\% \):

<table>
<thead>
<tr>
<th></th>
<th>SHIPSL C.I.'s</th>
<th>B-IBI C.I.'s</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>excludes 0</td>
<td>includes 0</td>
<td>Total</td>
</tr>
<tr>
<td>excludes 0</td>
<td>65.4%</td>
<td>16.3%</td>
<td>81.7%</td>
</tr>
<tr>
<td>includes 0</td>
<td>7.8%</td>
<td>10.5%</td>
<td>18.3%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>73.2%</strong></td>
<td><strong>26.8%</strong></td>
<td><strong>100.0%</strong></td>
</tr>
</tbody>
</table>

⇒ SHIPSL is about 1.12 times more “powerful”
⇒ Picking up what the other has missed:

SHIPSL is 2X more efficient
A Bootstrap Study: Sensitivity to Noise

know B-IBI is more sensitive (less powerful) biologist’s possible concern: extra field noise throws off health rating based on B-IBI (v. poor, poor, etc.)?

know noise metric score IBI value rating we jitter the trisection cutoffs
A Bootstrap Study: Sensitivity to Noise

know B-IBI is more sensitive (less powerful)
A Bootstrap Study: Sensitivity to Noise

- know B-IBI is more sensitive (≡ less powerful)
- biologist’s possible concern:
A Bootstrap Study: Sensitivity to Noise

- Know B-IBI is more sensitive (less powerful)
- Biologist’s possible concern:

  ⇒ *extra field noise throws off health rating based on B-IBI (v. poor, poor, etc.)* ?
A Bootstrap Study: Sensitivity to Noise

- know B-IBI is more sensitive (≡ less powerful)
- biologist’s possible concern:
  ⇒ extra field noise throws off health rating based on B-IBI (v. poor, poor, etc.)?
- know: noise ⇒ metric score ⇒ IBI value ⇒ rating
A Bootstrap Study: Sensitivity to Noise

- know B-IBI is more sensitive (≡ less powerful)
- biologist’s possible concern:
  ⇒ extra field noise throws off health rating based on B-IBI (v. poor, poor, etc.)?
- know: noise ⇒ metric score ⇒ IBI value ⇒ rating
  - we jitter the trisection cutoffs
## A Bootstrap Study: Sensitivity to Noise

<table>
<thead>
<tr>
<th>metric</th>
<th>original</th>
<th>“jittered”</th>
</tr>
</thead>
<tbody>
<tr>
<td>total # taxa</td>
<td>0, 14, 28</td>
<td>0, 13.53, 27.50</td>
</tr>
<tr>
<td># Eph. taxa</td>
<td>0, 3.5, 7</td>
<td>0, 3.26, 7.22</td>
</tr>
<tr>
<td># Ple. taxa</td>
<td>0, 2.7, 5.3</td>
<td>0, 2.46, 5.07</td>
</tr>
<tr>
<td># Tri. taxa</td>
<td>0, 2.7, 5.3</td>
<td>0, 2.87, 4.99</td>
</tr>
<tr>
<td># LL taxa</td>
<td>0, 4, 8</td>
<td>0, 3, 7</td>
</tr>
<tr>
<td># intol. taxa</td>
<td>0, 2, 4</td>
<td>0, NA, 3</td>
</tr>
<tr>
<td>% tol. taxa</td>
<td>44, 27, 0</td>
<td>45.16, 25.86, 0</td>
</tr>
<tr>
<td>% pred. taxa</td>
<td>0, 4.5, 9</td>
<td>0, 4.61, 9.11</td>
</tr>
<tr>
<td># cl. taxa</td>
<td>0, 8, 16</td>
<td>0, 7.79, 15.70</td>
</tr>
<tr>
<td>% 3 dom. taxa</td>
<td>75, 55, 0</td>
<td>76.22, 54.16, 0</td>
</tr>
</tbody>
</table>
A Bootstrap Study: Sensitivity to Noise
A Bootstrap Study: Sensitivity to Noise

<table>
<thead>
<tr>
<th></th>
<th>TH1</th>
<th>LB4</th>
<th>BB1</th>
<th>MA1</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;1%</td>
<td>43%</td>
<td>&lt;1%</td>
<td>&lt;1%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>LB2</th>
<th>BB5</th>
<th>SW3</th>
<th>SW2</th>
</tr>
</thead>
<tbody>
<tr>
<td>2%</td>
<td>13%</td>
<td>7%</td>
<td>32%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>SW1</th>
<th>BS1</th>
<th>RO1</th>
</tr>
</thead>
<tbody>
<tr>
<td>39%</td>
<td>52%</td>
<td>38%</td>
<td></td>
</tr>
</tbody>
</table>
Insight:  **USE SHIPSL !!**

*because ...*
Insight: **USE SHIPSL !!**

*because ...*

- SHIPSL is non-subjective
Insight: **USE SHIPSL!!**

*because ...*

- SHIPSL is non-subjective
  - yields stronger scientific conclusions, more “policy-neutral” (a la Lackey)
Insight: **USE SHIPSL!!**

*because...*

- SHIPSL is non-subjective
  - yields stronger scientific conclusions, more “policy-neutral” (a la Lackey)
- SHIPSL scale gives intuitive interpretation of index values
Insight: **USE SHIPSL!!**

*because* ...

- SHIPSL is non-subjective
  - ⇒ yields stronger scientific conclusions, more “policy-neutral” (a la Lackey)
- SHIPSL scale gives intuitive interpretation of index values
- SHIPSL drastically reduces time (a.k.a. $) required to
Insight: **USE SHIPSL!!**

*because...*

- SHIPSL is non-subjective
  - \[\Rightarrow\] yields stronger scientific conclusions, more "policy-neutral" (a la Lackey)
- SHIPSL scale gives intuitive interpretation of index values
- SHIPSL drastically reduces time (a.k.a. \$) required to
  - choose reference sites
Insight: **USE SHIPSL !!**

*because* ...

- SHIPSL is non-subjective
  - ⇒ yields stronger scientific conclusions, more “policy-neutral” (a la Lackey)
- SHIPSL scale gives intuitive interpretation of index values
- SHIPSL drastically reduces time (a.k.a. $) required to
  - ▶️ choose reference sites
  - ▶️ (re)calibrate metric scoring mechanism
Insight: **USE SHIPSL !!**

because ...

- SHIPSL is non-subjective
  - yields stronger scientific conclusions, more “policy-neutral” (a la Lackey)
- SHIPSL scale gives intuitive interpretation of index values
- SHIPSL drastically reduces time (a.k.a. $) required to
  - choose reference sites
  - (re)calibrate metric scoring mechanism
  - can be easily localized to other regions
Insight: *USE SHIPSL!!*

*because ...*

- **SHIPSL is non-subjective**
  - $\Rightarrow$ yields stronger scientific conclusions, more "policy-neutral" (a la Lackey)
- **SHIPSL scale gives intuitive interpretation of index values**
- **SHIPSL drastically reduces time (a.k.a. \$) required to**
  - choose reference sites
  - (re)calibrate metric scoring mechanism
  - $\Rightarrow$ can be easily localized to other regions
- **SHIPSL measures biotic integrity more accurately and precisely**
Insight: **USE SHIPSL !!**

*because ...*

- SHIPSL is non-subjective
  - yields stronger scientific conclusions, more “policy-neutral” (a la Lackey)
- SHIPSL scale gives intuitive interpretation of index values
- SHIPSL drastically reduces time (a.k.a. $) required to
  - choose reference sites
  - (re)calibrate metric scoring mechanism
  - can be easily localized to other regions
- SHIPSL measures biotic integrity more accurately and precisely
  - year-to-year difference in SHIPSL for fixed site likely due to changing stream health
Insight: **USE SHIPSL!!**

*because ...*

- SHIPSL is non-subjective
  - yields stronger scientific conclusions, more “policy-neutral” (a la Lackey)
- SHIPSL scale gives intuitive interpretation of index values
- SHIPSL drastically reduces time (a.k.a. $) required to
  - choose reference sites
  - (re)calibrate metric scoring mechanism
  - can be easily localized to other regions
- SHIPSL measures biotic integrity more accurately and precisely
  - year-to-year difference in SHIPSL for fixed site likely due to changing stream health
  - not necessarily so for B-IBI
Insight: **USE SHIPSL !!**

*because ...*

- SHIPSL is non-subjective
  - ⇒ yields stronger scientific conclusions, more “policy-neutral” (a la Lackey)
- SHIPSL scale gives intuitive interpretation of index values
- SHIPSL drastically reduces time (a.k.a. $) required to
  - ▶ choose reference sites
  - ▶ (re)calibrate metric scoring mechanism
  - ⇒ can be easily localized to other regions
- SHIPSL measures biotic integrity more accurately and precisely
  - ▶ year-to-year difference in SHIPSL for fixed site likely due to changing stream health
  - ▶ not necessarily so for B-IBI

**SHIPSL achieves all of the above WITHOUT adding any technical requirement to conventional biomonitoring !!**
SHIPSL is made NOT for statisticians but for YOU!

Q: Would you prefer less subjective input in constructing a health index?— includes choosing reference sites, devising scoring scheme, defining gold standard
Biologists’ Opinions ??

SHIPSL is made NOT for statisticians but for YOU!
Biologists’ Opinions ??

SHIPSL is made NOT for statisticians but for YOU!

Q: Would you prefer less subjective input in constructing a health index?

— includes choosing reference sites, devising scoring scheme, defining gold standard
Biologists’ Opinions ??

SHIPSL is made NOT for statisticians but for YOU!

Q: If you prefer a gold-standard approach, how would you minimize “policy preference”? 

Corvallis EPA Seminar, Page 44
Biologists’ Opinions ??

SHIPSL is made NOT for statisticians but for YOU!

Q: Would you prefer a statistically standardized scale over one that involves some form of discretization?

— includes \{1,3,5\} and \[0,10\] schemes
Biologists’ Opinions ??

SHIPSL is made NOT for statisticians but for **YOU**!

**Q:** Would you prefer a **statistically standardized** scale over one that involves some form of discretization?

— includes \{1,3,5\} and \[0,10\] schemes

**NOTE:** (Co)incidentally, economists Zarnowitz & Boschan (1975) used similar standardization to construct the **composite index of leading indicators**

(published monthly by the Bureau of Economic Analysis (BEA))
SHIPSL is made NOT for statisticians but for YOU!

Q: Would you use equal or different weights for each standardized metric?
Biologists’ Opinions ??

SHIPSL is made NOT for statisticians but for YOU!

Q: Would you use equal or different weights for each standardized metric?

— Auerbach (1982): equal weighting tends to smooth out fluctuations of the relationships between the BEA index components and economic conditions (e.g. unemployment or another economic index).
Biologists’ Opinions ??

SHIPSL is made NOT for statisticians but for **YOU**!

Let’s work closely together to devise an index which
Biologists’ Opinions ??

SHIPSL is made NOT for statisticians but for **YOU**!

Let’s work closely together to devise an index which

- makes **statistical** AND **biological** sense
- will be used by environmental scientists for making policies
Current Work: Modeling Field Data
## Current Work: Modeling Field Data

<table>
<thead>
<tr>
<th>TAXON ID</th>
<th>TAXON ID</th>
<th>REP 1 COUNT</th>
<th>REP 2 COUNT</th>
<th>REP 3 COUNT</th>
<th>SITE TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>...</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>...</td>
<td>3</td>
<td>12</td>
<td>8</td>
<td>23</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>80</td>
<td>...</td>
<td>12</td>
<td>12</td>
<td>7</td>
<td>31</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td></td>
<td><strong>925</strong></td>
<td><strong>2321</strong></td>
<td><strong>542</strong></td>
<td><strong>3788</strong></td>
</tr>
</tbody>
</table>
### Current Work: Modeling Field Data

<table>
<thead>
<tr>
<th>TAXON ID</th>
<th>TAXON ATTRIBUTES</th>
<th>REP 1 COUNT</th>
<th>REP 2 COUNT</th>
<th>REP 3 COUNT</th>
<th>SITE TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>...</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>...</td>
<td>3</td>
<td>12</td>
<td>8</td>
<td>23</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>80</td>
<td>...</td>
<td>12</td>
<td>12</td>
<td>7</td>
<td>31</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>925</td>
<td>2321</td>
<td>542</td>
<td>3788</td>
</tr>
</tbody>
</table>
### Current Work: Modeling Field Data

#### Table 1: Site Data Summary

<table>
<thead>
<tr>
<th>TAXON ID</th>
<th>TAXON ATTRIBUTES</th>
<th>REP 1 COUNT</th>
<th>REP 2 COUNT</th>
<th>REP 3 COUNT</th>
<th>SITE TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>…</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>…</td>
<td>3</td>
<td>12</td>
<td>8</td>
<td>23</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
<tr>
<td>80</td>
<td>…</td>
<td>12</td>
<td>12</td>
<td>7</td>
<td>31</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td></td>
<td><strong>925</strong></td>
<td><strong>2321</strong></td>
<td><strong>542</strong></td>
<td><strong>3788</strong></td>
</tr>
</tbody>
</table>

#### Table 2: TAXON Data Summary

<table>
<thead>
<tr>
<th>TAXON ID</th>
<th>REP 1 COUNT</th>
<th>REP 2 COUNT</th>
<th>REP 3 COUNT</th>
<th>SITE TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$n_{11}$</td>
<td>$n_{12}$</td>
<td>$n_{13}$</td>
<td>$n_{1+}$</td>
</tr>
<tr>
<td>2</td>
<td>$n_{21}$</td>
<td>$n_{22}$</td>
<td>$n_{23}$</td>
<td>$n_{2+}$</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
<tr>
<td>80</td>
<td>$n_{80,1}$</td>
<td>$n_{80,2}$</td>
<td>$n_{80,3}$</td>
<td>$n_{80+}$</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>$n_{+1}$</td>
<td>$n_{+2}$</td>
<td>$n_{+3}$</td>
<td>$n_{++}$</td>
</tr>
</tbody>
</table>
Current Work: Modeling Field Data

Data: matrix $\mathbb{N}$

$\Rightarrow (n_1^+, \ldots, n_{80}^+), (n_1^+, n_2^+, n_3^+), \text{ and } n_{++}$
Current Work: Modeling Field Data

Data: matrix $\mathbb{N}$

$$\Rightarrow (n_{1+}, \ldots, n_{80+}), (n_{+1}, n_{+2}, n_{+3}), \text{ and } n_{++}$$

SHIPSL and B-IBI are statistics of $\mathbb{N}$
Current Work: Modeling Field Data

Models:

Karr & Chu (1999): columns of $N$ are independent multinomial $80(n + j; p)$, estimate $p$ by $(n_1 + \ldots + n_{80} + \ldots)/n$.

Bunea et al. (1999): (1) tends to underestimate B-IBI variability, add: $n_1, n_2, n_3$ iid $\sim$ neg. bin.($\theta; \phi$) estimate $\theta; \phi$ by mean and SD of observed $n_j$'s.

Chiu (current): neither (1) or (2) matches bootstrap results. Trying two approaches.
Current Work: Modeling Field Data

Models:

(1) Karr & Chu (1999):

- columns of $\mathbb{N}$ are independent multinomial$_{80}(n_{+j}, \underline{p})$,
- estimate $\underline{p}$ by $\left( n_{1+}, \ldots, n_{80+} \right)/n_{++}$
Models:

(1) Karr & Chu (1999):
columns of $\mathbb{N}$ are independent $\text{multinomial}_{80}(n_{+j}, p)$, estimate $p$ by $(n_{1+}, \ldots, n_{80+})/n_{++}$

(2) Bunea et al. (1999):
(1) tends to underestimate B-IBI variability
⇒ add: $n_{+1}, n_{+2}, n_{+3} \overset{\text{iid}}{\sim} \text{neg. bin.}(\mu, \sigma)$
estimate $\mu, \sigma$ by mean and SD of observed $n_{+j}$’s
Models:

(1) Karr & Chu (1999):

Columns of $\mathbb{N}$ are independent $\text{multinomial}_{80}(n_{+j}, p)$, estimate $p$ by $(n_{1+}, \ldots, n_{80+})/n_{++}$

(2) Bunea et al. (1999):

(1) tends to underestimate B-IBI variability

$\Rightarrow$ add: $n_{+1}, n_{+2}, n_{+3} \overset{\text{iid}}{\sim} \text{neg. bin.}(\mu, \sigma)$

Estimate $\mu, \sigma$ by mean and SD of observed $n_{+j}$’s

(3) Chiu (current):

Neither (1) or (2) matches bootstrap results

$\Rightarrow$ trying TWO approaches
Current Work: (i) Linguistical Idea

\[(n_{1j}, \ldots, n_{80,j}) \sim \text{multinomial}_{80}(n_{+j}, p)\] \quad \forall \; j = 1, 2, 3
Current Work: (i) Linguistical Idea

\[ (n_{1j}, \ldots, n_{80,j}) \overset{\text{ind}}{\sim} \text{multinomial}_{80}(n_{+j}, \underline{p}) \atop \forall j = 1, 2, 3 \]

\[ \underline{p} \sim \text{Dirichlet}_{80}(\varphi) \]
Current Work: (i) Linguistical Idea

\( (n_{1j}, \ldots, n_{80j}) \overset{\text{ind}}{\sim} \text{multinomial}_{80}(n_{+j}, \underline{p}) \)

\( \forall j = 1, 2, 3 \)

\( p \sim \text{Dirichlet}_{80}(\varphi) \)

@ \textit{sparse multinomial} (i.e. involving structural 0’s)

\( \Rightarrow \) linguistics literature:
Current Work: (i) Linguistical Idea

- \((n_{1j}, \ldots, n_{80,j}) \overset{\text{ind}}{\sim} \text{multinomial}_{80}(n+j, \underline{p})\) \(\forall j = 1, 2, 3\)
- \(\underline{p} \sim \text{Dirichlet}_{80}(\varphi)\)

@ **sparse multinomial** (i.e. involving structural 0’s) ??

⇒ linguistics literature:
  - generalization of Laplace’s **Law of Succesion**
  - explicit formula for \([ \underline{p} | \mathbb{N} ]\)
Current Work: (ii) Generalized Linear Mixed-Effects Model

\[ Y_{ijk} = \mu + \bar{\alpha}_j + \bar{\beta}_j x_j + \bar{\gamma}_i + U_{j(i)} + \varepsilon_{ijk} \]

This is a linear mixed-effects model, combining to give "health" for site \( i \).
write metric value $Y$ as function of
Current Work: (ii) Generalized Linear Mixed-Effects Model

- write metric value $Y$ as function of
  - urbanization $x$ (covariate)
  - metric effect (main factor)
  - site effect (random blocking factor)
  - metric-within-site effect (random interaction)
Current Work: (ii) Generalized Linear Mixed-Effects Model

- write metric value $Y$ as function of
  - urbanization $x$ (covariate)
  - metric effect (main factor)
  - site effect (random blocking factor)
  - metric-within-site effect (random interaction)

$$Y_{ijk} = \mu + \beta_0j + \beta_1j x_j + \alpha_j + H_i + U_{j(i)} + \varepsilon_{ijk}$$

- $(\mu)$ (overall mean)
- $(\beta_0j + \beta_1j x_j)$ (each metric has own regression line)
- $(\alpha_j)$ ($j$-th metric’s deviation from $\mu$)
- $(H_i)$ ($i$-th site’s deviation from $\mu$)
- $(U_{j(i)})$ (deviation due to interaction)
- $(\varepsilon_{ijk})$ (unexplained error)
write metric value $Y$ as function of

- urbanization $x$ (covariate)
- metric effect (main factor)
- site effect (random blocking factor)
- metric-within-site effect (random interaction)

$Y_{ijk} = \mu + \beta_{0j} + \beta_{1j} x_j + \alpha_j + H_i + U_{j(i)} + \varepsilon_{ijk}$

⇒ this is a LINEAR mixed-effects model
Current Work: (ii) Generalized Linear Mixed-Effects Model

- write metric value $Y$ as function of
  - urbanization $x$ (covariate)
  - metric effect (main factor)
  - site effect (random blocking factor)
  - metric-within-site effect (random interaction)

\[
Y_{ijk} = \mu + \beta_{0j} + \beta_{1j} x_j + \alpha_j + H_i + U_{j(i)} + \varepsilon_{ijk}
\]

- (overall mean)
- (each metric has own regression line)
- ($j$-th metric's deviation from $\mu$)
- ($i$-th site's deviation from $\mu$)
- (deviation due to interaction)
- (unexplained error)

⇒ this is a LINEAR mixed-effects model
Current Work: (ii) Generalized Linear Mixed-Effects Model

Need to generalize our model to

- Poisson regression for metrics that are counts, e.g., # taxa, # clinger taxa
- Logistic / Probit regression for metrics that are percentages, e.g., % 3 most dominant taxa

Other formulations of the regression model also considered.
Current Work: (ii) Generalized Linear Mixed-Effects Model

Need to generalize our model to

- Poisson regression for metrics that are counts
  e.g. # taxa, # clinger taxa
Need to generalize our model to

- **Poisson regression** for metrics that are **counts**
  - e.g. # taxa, # clinger taxa

- **Logistic / Probit regression** for metrics that are **percentages**
  - e.g. % 3 most dominant taxa
Current Work: (ii) Generalized Linear Mixed-Effects Model

Need to generalize our model to

- Poisson regression for metrics that are counts
e.g. # taxa, # clinger taxa

- Logistic / Probit regression for metrics that are percentages
e.g. % 3 most dominant taxa

Other formulations of the regression model also considered.
Farewell

This presentation is downloadable from
www.stat.washington.edu/grace/corvallis.pdf

The original manuscript is available from
www.nrcse.washington.edu/pdf/trs78.pdf

THANK YOU!