Validation of the SOP of the Environmental Specimen Bank for the representativeness and quality of the samples

2015. 7. 1

National Environmental Specimen Bank
National Institute of Environmental Research
Korea

Jong-chun Lee
Contents

- Some facts about the NESB, Korea
- In the making of SOPs
- How representative is the sample?
- Evaluation of Measurement Uncertainty
Some facts about the NESB, Korea

- National Environmental Specimen Bank (NESB):
  - Building established 2009 at NIER
  - Two story building with basement: 2,340 m²
Some facts about the NESB, Korea

- **Cryogenic Storage**
  - Currently 32 LN2 Freezer tanks (1,200L and 1,400L) for vials and field samples
  - Capacity up to 71 tanks
  - 7 electric deep freezers
Some facts about the NESB, Korea

- **Cryo-Homogenization Facility**
  - ISO Class 5 clean room for sample treatment and two ISO Class 7 clean rooms for cryo-milling and analysis

Vibrating Mill (Palla)

ZiO₂ Ceramic Ball Mill
Some facts about the NESB, Korea

- **2D Barcoding and Sampling tracking system**
  - Each vial is given ID and its stored location can be tracked
Some facts about the NESB, Korea

- **Types of Stored Specimens**
  - 7 ecological samples: pine needles, leaves, seagull eggs...
    - Mussel will be added in '15
  - Human samples (from some cohort studies on industrial area...)
  - Samples from some environmental accident sites

- **Analytical Labs**
  - Inorganic Lab: ICP-MS, Mercury Analyzer
  - Organic Lab: UPLC-QToF-MS, GC-MS
Standard Operation Procedures (SOPs)

- Based on German SOPs
  - Sharing similar objectives and scheme
  - Comparable species to represent different ecological types
  - Set prior to the full-scale operation
Out in the field

- **Arising questions**
  - Similar SOP for the similar species?
    - Big picture seemed OK but...
  - Justification of the sample size for a representativeness of a colony
  - Once set, stick to the rule....forever?
  - Ideal vs. Reality

Sun needle shoot  Shade needle shoot
Tweaks to the SOPs based on experience

**Before:**
*inconsistent, prone to human error*

Pick around 10 trees randomly from the colony
From each tree, 200~300 g of one year old shoot
is sampled respectively

**After:**
*Specific, consistent sampling*

Pick four main branches(four directions) around the tree trunk from down to top between the 2nd and the 4th branches. Cut 8 to 12 secondary branches from which 1-year-old branches are taken considering the morphology of the shoot(sun branch, shade branch)
To know what we are doing

- Verification of the SOP(...before it is too late)
  - to have confidence for ourselves
  - to eliminate doubts on the sample quality(or representativeness)
  - composite sample vs. individual samples

- Which is the weakest link, sampling or analysis?
  - Understand the sample and site characteristics:
    - especially prior to composite sampling
  - Limited resource: budget & man-power
  - Sampling vs. Analytical precision
  - Uncertainty on the measured values(for a site mean)
Duplicate sampling

- Study sites: two pine sites
  - Taehwa Mt.: close to Metropolitan area
  - Worak Mt.: National Park

- Duplicate sampling for five respective individual samples
Test Methods

- **Analysis of heavy metals: for Pb, Zn, Ni, Cu, Cr, Cd, As**
  - Duplicate analysis for each discrete sample using ICP-MS
    - randomized order to eliminate systematic bias

- **Statistic analysis**
  - site characterization: effect of two different settings (urban vs. natural)
  - Factor Analysis: behavioral characteristics between the two sites

- **Robust ANOVA**
  - balanced design with duplicate analyses on the duplicate sampling
Test Methods

- **Evaluation of uncertainty level on the estimated concentration**
  - **Definition of Measurement Uncertainty**
    "interval around the result of a measurement that contains the true value with high probability (95%)"
    \[ U_{meas} = 2s_{meas} = 2\sqrt{s^2_{samp} + s^2_{anal}} \]
  - **Separation of total variance into three variables**
    \[ s^2_{total} = s^2_{geochem} + s^2_{samp} + s^2_{anal} \]
    - \( s^2_{geochem} \): from the across variations throughout the site
    - \( s^2_{samp} \): from sampling method (or biological variation)
    - \( s^2_{anal} \): from precision of an analytical method
Findings - behavior of heavy metals

- For Red Pine (Taehwa Mt. Site, close to Seoul)
  - Relatively higher heavy metal concentrations
  - Strong correlation among Zn, Pb, and Cd: sharing same source of pollution (aerial deposition)

<table>
<thead>
<tr>
<th></th>
<th>Cd</th>
<th>Pb</th>
<th>Cu</th>
<th>Cr</th>
<th>Ni</th>
<th>Zn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pb</td>
<td>0.706</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cu</td>
<td>0.490</td>
<td>0.806</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cr</td>
<td>-0.245</td>
<td>-0.075</td>
<td>0.188</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ni</td>
<td>0.081</td>
<td>0.056</td>
<td>0.452</td>
<td>0.583</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zn</td>
<td>0.802</td>
<td>0.783</td>
<td>0.654</td>
<td>-0.086</td>
<td>0.125</td>
<td></td>
</tr>
<tr>
<td>As</td>
<td>0.100</td>
<td>0.128</td>
<td>-0.036</td>
<td>0.315</td>
<td>-0.189</td>
<td>0.160</td>
</tr>
</tbody>
</table>

Bold numbers indicate p ≤ 0.01 (two tailed)
Findings - behavior of heavy metals

- For Korean Pine (Worak Mt. Site, National Park)
  - Factor analysis discriminate Zn by the First factor
  - Washing off experiment: Zn distinguish itself from other heavy metals (accumulated mainly by uptake)

<table>
<thead>
<tr>
<th></th>
<th>Cd</th>
<th>Pb</th>
<th>Cu</th>
<th>Cr</th>
<th>Ni</th>
<th>Zn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pb</td>
<td>0.998</td>
<td>0.984</td>
<td>0.850</td>
<td>0.999</td>
<td>-0.317</td>
<td>1.000</td>
</tr>
<tr>
<td>Cu</td>
<td>0.998</td>
<td>0.990</td>
<td>0.875</td>
<td>0.990</td>
<td>0.998</td>
<td>0.998</td>
</tr>
<tr>
<td>Cr</td>
<td>0.923</td>
<td>0.990</td>
<td>0.923</td>
<td>0.990</td>
<td>0.998</td>
<td>0.986</td>
</tr>
<tr>
<td>Ni</td>
<td>0.869</td>
<td>0.850</td>
<td>0.856</td>
<td>0.999</td>
<td>0.999</td>
<td>0.984</td>
</tr>
<tr>
<td>Zn</td>
<td>-0.282</td>
<td>-0.264</td>
<td>-0.149</td>
<td>0.206</td>
<td>0.998</td>
<td>1.000</td>
</tr>
<tr>
<td>As</td>
<td>-0.305</td>
<td>0.000</td>
<td>0.986</td>
<td>0.856</td>
<td>0.998</td>
<td>0.984</td>
</tr>
</tbody>
</table>

Bold numbers indicate p ≤ 0.01 (two tailed)
RANOVA Results

<Taehwa Mt.>

- **Analytical Precision:** 
  \( (s^2_{\text{anal}}) \) less than 1\% of total variance except As

- **Weakest link: sampling(biological) variance**
  \( (s^2_{\text{samp}} \) explains most of the total variance

- **measurement uncertainty: for \( \text{Pb} \)**
  \( 1.472 \pm 0.815 \text{ ppm (at 95\% confidence level)} \)

<table>
<thead>
<tr>
<th></th>
<th>Cd</th>
<th>Pb</th>
<th>Cu</th>
<th>Cr</th>
<th>Ni</th>
<th>Zn</th>
<th>As</th>
</tr>
</thead>
<tbody>
<tr>
<td>( s_{\text{total}} )</td>
<td>0.019</td>
<td>0.407</td>
<td>0.588</td>
<td>0.404</td>
<td>0.593</td>
<td>3.968</td>
<td>0.044</td>
</tr>
<tr>
<td>( s_{\text{geoch}} )</td>
<td>0.017</td>
<td>0.000</td>
<td>0.000</td>
<td>0.277</td>
<td>0.329</td>
<td>2.111</td>
<td>0.027</td>
</tr>
<tr>
<td>((% \text{ of total variance}))</td>
<td>82.6</td>
<td>0.0</td>
<td>0.0</td>
<td>47.1</td>
<td>30.8</td>
<td>28.3</td>
<td>36.8</td>
</tr>
<tr>
<td>( s_{\text{samp}} )</td>
<td>0.008</td>
<td>0.407</td>
<td>0.587</td>
<td>0.282</td>
<td>0.493</td>
<td>3.339</td>
<td>0.024</td>
</tr>
<tr>
<td>((% \text{ of total variance}))</td>
<td>16.7</td>
<td>99.6</td>
<td>99.6</td>
<td>48.7</td>
<td>69.0</td>
<td>70.8</td>
<td>30.3</td>
</tr>
<tr>
<td>((% \text{ of measurement variance}))</td>
<td>96.2</td>
<td>99.6</td>
<td>99.6</td>
<td>92.0</td>
<td>99.7</td>
<td>98.7</td>
<td>47.9</td>
</tr>
<tr>
<td>( s_{\text{anal}} )</td>
<td>0.002</td>
<td>0.025</td>
<td>0.037</td>
<td>0.083</td>
<td>0.026</td>
<td>0.377</td>
<td>0.026</td>
</tr>
<tr>
<td>((% \text{ of total variance}))</td>
<td>0.7</td>
<td>0.4</td>
<td>0.4</td>
<td>4.3</td>
<td>0.2</td>
<td>0.9</td>
<td>32.9</td>
</tr>
<tr>
<td>((% \text{ of measurement variance}))</td>
<td>3.8</td>
<td>0.4</td>
<td>0.4</td>
<td>8.0</td>
<td>0.3</td>
<td>1.3</td>
<td>52.1</td>
</tr>
<tr>
<td>( s_{\text{mean}} = \sqrt{(s^2_{\text{samp}} + s^2_{\text{anal}})} )</td>
<td>0.008</td>
<td>0.407</td>
<td>0.588</td>
<td>0.294</td>
<td>0.493</td>
<td>3.360</td>
<td>0.035</td>
</tr>
<tr>
<td>( s_{\text{mean}} (\text{% of total variance}) )</td>
<td>17.4</td>
<td>100.0</td>
<td>100.0</td>
<td>52.9</td>
<td>69.2</td>
<td>71.7</td>
<td>63.2</td>
</tr>
<tr>
<td>mean</td>
<td>0.065</td>
<td>1.472</td>
<td>2.209</td>
<td>1.097</td>
<td>1.610</td>
<td>20.712</td>
<td>0.157</td>
</tr>
<tr>
<td>( U(\text{at the mean}) = 2s_{\text{mean}} )</td>
<td>0.016</td>
<td>0.815</td>
<td>1.177</td>
<td>0.587</td>
<td>0.986</td>
<td>6.720</td>
<td>0.071</td>
</tr>
<tr>
<td>( U% )</td>
<td>24.0</td>
<td>55.4</td>
<td>53.3</td>
<td>53.5</td>
<td>61.3</td>
<td>32.4</td>
<td>45.1</td>
</tr>
<tr>
<td>RSD</td>
<td>28.7</td>
<td>27.7</td>
<td>26.6</td>
<td>36.8</td>
<td>36.8</td>
<td>19.2</td>
<td>28.4</td>
</tr>
</tbody>
</table>
Findings - Separation of the sampling and the analysis

RANOVA Results

<Worak Mt.>

- **Analytical Precision:**
  \( s_{\text{anal}}^2 \) less than 5% of total variance except As and Cd

- **Weakest link: sampling(biological) variance**
  \( s_{\text{samp}}^2 \) explains most of the total variance

- **measurement uncertainty:** for Pb
  \( 1.334 \pm 0.623 \text{ ppm(at 95% confidence level)} \)

<table>
<thead>
<tr>
<th></th>
<th>Cd</th>
<th>Pb</th>
<th>Cu</th>
<th>Cr</th>
<th>Ni</th>
<th>Zn</th>
<th>As</th>
</tr>
</thead>
<tbody>
<tr>
<td>( s_{\text{total}} )</td>
<td>0.009</td>
<td>0.387</td>
<td>0.277</td>
<td>0.243</td>
<td>0.069</td>
<td>1.359</td>
<td>0.048</td>
</tr>
<tr>
<td>( s_{\text{geo}} )</td>
<td>0.007</td>
<td>0.229</td>
<td>0.257</td>
<td>0.072</td>
<td>0.060</td>
<td>0.593</td>
<td>0.000</td>
</tr>
<tr>
<td>(% of total variance)</td>
<td>64.1</td>
<td>35.1</td>
<td>85.6</td>
<td>8.8</td>
<td>74.0</td>
<td>19.0</td>
<td>0.0</td>
</tr>
<tr>
<td>( s_{\text{samp}} )</td>
<td>0.005</td>
<td>0.308</td>
<td>0.095</td>
<td>0.074</td>
<td>0.032</td>
<td>1.197</td>
<td>0.047</td>
</tr>
<tr>
<td>(% of total variance)</td>
<td>31.9</td>
<td>63.4</td>
<td>11.7</td>
<td>9.2</td>
<td>20.7</td>
<td>77.6</td>
<td>92.5</td>
</tr>
<tr>
<td>(% of measurement variance)</td>
<td>92.5</td>
<td>97.8</td>
<td>81.2</td>
<td>10.1</td>
<td>79.6</td>
<td>95.8</td>
<td>92.5</td>
</tr>
<tr>
<td>( s_{\text{anal}} )</td>
<td>0.001</td>
<td>0.046</td>
<td>0.046</td>
<td>0.220</td>
<td>0.016</td>
<td>0.251</td>
<td>0.013</td>
</tr>
<tr>
<td>(% of total variance)</td>
<td>2.6</td>
<td>1.4</td>
<td>2.7</td>
<td>82.0</td>
<td>5.3</td>
<td>3.4</td>
<td>7.5</td>
</tr>
<tr>
<td>(% of measurement variance)</td>
<td>7.5</td>
<td>2.2</td>
<td>18.8</td>
<td>89.9</td>
<td>20.4</td>
<td>4.2</td>
<td>7.5</td>
</tr>
<tr>
<td>( s_{\text{meas}} = \sqrt{s_{\text{samp}}^2 + s_{\text{anal}}^2} )</td>
<td>0.005</td>
<td>0.311</td>
<td>0.105</td>
<td>0.232</td>
<td>0.035</td>
<td>1.223</td>
<td>0.048</td>
</tr>
<tr>
<td>( s_{\text{meas}}\text{(% of total variance)} )</td>
<td>34.4</td>
<td>64.9</td>
<td>14.4</td>
<td>91.2</td>
<td>26.0</td>
<td>81.0</td>
<td>100.0</td>
</tr>
<tr>
<td>mean</td>
<td>0.042</td>
<td>1.334</td>
<td>1.912</td>
<td>2.411</td>
<td>0.481</td>
<td>11.987</td>
<td>0.187</td>
</tr>
<tr>
<td>U(at the mean)=2s_{\text{meas}}</td>
<td>0.011</td>
<td>0.623</td>
<td>0.210</td>
<td>0.464</td>
<td>0.071</td>
<td>2.446</td>
<td>0.097</td>
</tr>
<tr>
<td>U%</td>
<td>26.0</td>
<td>46.7</td>
<td>11.0</td>
<td>19.2</td>
<td>14.7</td>
<td>20.4</td>
<td>51.7</td>
</tr>
<tr>
<td>RSD</td>
<td>22.1</td>
<td>29.0</td>
<td>14.5</td>
<td>10.1</td>
<td>14.4</td>
<td>11.3</td>
<td>25.9</td>
</tr>
</tbody>
</table>
Validation of a measured concentration for the representativeness
• the main source of total variance rises from the sampling (or biological) variation (small-scale variation)

\[ S^2_{\text{geochem}} < 50\% \text{ of total variance (and } s^2_{\text{geochem}} < s^2_{\text{samp}}) \]
- in other words, variation across the site is rather homogeneous

Findings – representativeness of a measured value for a site

<Tahwa Mt. L: Pb R: Zn>

<Worak Mt. L: Pb R: Zn>
Pb was used to test the representativeness of the sample taken by the SOP

- Most of Pb originated from aerial deposition: from washing off experiment
- Result from the RANOVA for the two sites confirms,

\[ s^2_{geochem} < s^2_{samp} \]

- Taehwa Mt. Pb: \[ s^2_{geochem} : s^2_{samp} : s^2_{anal} = 0\% : 99.6\% : 0.4\% \]
- Worak Mt. Pb: \[ s^2_{geochem} : s^2_{samp} : s^2_{anal} = 35.1\% : 63.4\% : 1.4\% \]

➤ need to increase the amount of a discrete(individual) sample

\[ s'_{samp} = s_{samp}/\sqrt{n} \]

- If \( s^2_{geochem} > s^2_{samp} \), then add discrete sampling points(n)
Juggling with a limited resource

To decrease the measurement uncertainty with a limited resource:

to sampling vs. analysis

- find the weakest link in a SOP
- then can put more resource to improve the weaker part
Merci

FIN