Scientific Standards and Data Quality

Dr. Long Fu, Director
Standards, Quality and Innovation
Environmental Monitoring and Science Division

CPANS Conference
May 9th, 2017
Life is made up of a series of judgments on insufficient data - US Judge Learned Hand (1872 – 1961)

- Science (accuracy and precision)
- Policy (proportional and appropriate)
- Society (Peace, order and good government)
The Era of Technology Black Boxes
A Matter of Trust
People are entitled to their own opinion, but not their own facts.

- Data Quality Act (USA)
- Canadian Access to Information Act (1985)
- Freedom of Information and Protection of Privacy Act (Alberta)
- Security Exchange laws and regulations.
- Environmental Impact Assessment.
- Regulatory requirements on monitoring and reporting.
- EMSD’s Standards and Quality Program
JOSM Scientific Integrity Review

• Rigor
  – Relevant questions are addressed through the scientific method using appropriate and consistent methodology
• Transparency
  – Data and metadata are openly available
• Standards and Protocols
  – Adoption and implementation of internationally recognized standards and protocols, where applicable.
Canada-Alberta Joint Oil Sands Monitoring External Expert Peer Review Report

Philip K. Hopke
Panel Chair
Presented on February 16th, 2016
Panel review: May 2015 – February 2016
Develop and document a uniform QA approach that is implemented and tracked across all monitoring activities

- A more rigorous approach to the QA process, including full independent auditing, could likely increase stakeholder confidence in the reported data.

- A suite of QA documents could include an overall QA program plan, detailed standard operating procedures and a QA annual report documenting the implementation of the program plan and the completion of any QA audit.
Functions and Responsibilities

• Standard Operating Procedures (SOPs)
  – Maintain and update SOPs Inventory (Air, Water, Land and Biodiversity).
  – Identify Gaps, Revise or Develop SOPs.

• Data Quality Objectives and Data Quality Systems.
  – Implement Data Quality Requirements and Initiatives.

• Emerging Science and Technology
  – Incorporate emerging technologies and methods into EMSD’s Standards and Quality Programs as appropriate.
## Data Quality System

<table>
<thead>
<tr>
<th>Objectives</th>
<th>Approaches</th>
</tr>
</thead>
</table>
| Ensure implementation of credible standards                               | • Contracting process  
                            • Accreditation                                             |
| Provide oversight and guidance on quality management                      | • External Peer Review and Audit  
                            • Site Visits  
                            • QAPP Review  
                            • DQO Assessment                                             |
| Maintain consistent and comparable methodology among Partners             | • Agreements on Standards and Protocols  
                            • MOUs                                                       |
| Empower Communities and Citizens                                         | • Tools and support  
                            • Educational Materials                                      |
Eureka Moment and Aufgehoben

Immediate Knowledge

Mediated Knowledge
Scientific Paradigm

- Pre-paradigm
- Normal Science (general agreement on methods and protocols)
- Crisis period due to anomalies
- Paradigm shift (Copernican Revolution)
- Post revolution
Science is the reduction of uncertainties

<table>
<thead>
<tr>
<th></th>
<th>Field Instrument</th>
<th>Lab Analysis</th>
<th>DNA Technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type I, there is a true value</td>
<td>FRM for PM2.5</td>
<td>Instrument Detection Limit</td>
<td>Linear dose-response relationship</td>
</tr>
<tr>
<td>to be determined</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type II, variability +</td>
<td>Continuous PM2.5 Monitor</td>
<td>Method Detection Limit</td>
<td>DNA expression</td>
</tr>
<tr>
<td>multiple measurements</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type III, complex causal</td>
<td>Epidemiological Studies</td>
<td>Secondary use of multi-source water</td>
<td>System to system effects.</td>
</tr>
<tr>
<td>relationships + random factors</td>
<td></td>
<td>quality data</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Federal Reference Method and TEOM
Sample Data from Toronto – Etona Region

\[
Ra = 1.5654 - 0.0086 \times T \\
R^2 = 0.3599
\]
Social Benefits of Having Quality Assured Data

TEOM-FDMS (+50 to 100%)

SHARP

TEOM (−25%)

CM' = (\(\alpha\hat{^\prime}+\beta \hat{T}\))×CM

FRM/CM = \(\alpha + \beta T + \varepsilon\)

“A two-step approach for relating tapered element oscillating microbalance and dichotomous air sampler PM2.5 measurements”

J of A&WMA September 2014
A “time series” epidemiologic study compares

– the day-to-day variation in the counts of health events in an unenumerated population (open cohort)

– with the day-to-day variation in a pollutant of interest measured at a central site

– while adjusting for
  • time (trend, season, day of week)
  • weather (temperature, humidity, precipitation)
Challenges with secondary use of multi-source water-quality data in the United States

- Lori A. Sprague a, *, Gretchen P. Oelsner b, Denise M. Argue c

• 25 million nutrient records by 488 organizations in US since 1899.
• 14.5 million had missing or ambiguous information
  – sample fraction, chemical form, parameter name, units of measurement, precise numerical value, and remark codes
• Value of ambiguous data is ca. $US12 billion.
• Value of unambiguous data is ca. $US8.2 billion.

• Water Research, December 2016
A test will produce 99% true positive results for drug users and 1% false positive results for non-drug users.

Suppose that 0.5% of people are users of the drug. If a randomly selected individual tests positive, what is the probability that he is a user?

\[
P(\text{user}|+) = \frac{P(+|\text{user}) \times P(\text{user})}{(P + |\text{user}) \times P(\text{user}) \times P(+|\text{non} - \text{user}) \times P(\text{non} - \text{user})}
\]

\[
= \frac{0.99 \times 0.005}{((0.99 \times 0.005) + (0.01 \times 0.995))} = 33.2%
\]
Strength and Weakness

Consider this:

When 1.0% of people are users of the drug 50.0%
When 2.0% of people are users of the drug 66.9%
When 3.0% of people are users of the drug 75.4%
Data of Sufficient Quality and Quantity
Random Raindrops
Data of Sufficient Quality and Quantity

By McGeddon - Own work, CC BY-SA 4.0,
https://commons.wikimedia.org/w/index.php?curid=53081927

By http://www-history.mcs.st-andrews.ac.uk/PictDisplay/Wald.html, PD-US,
Number of Citations for “Bayes’ Theorem”
Expert Re-defined
The Era of Post Truth
Science offers little in the way of cheap thrills. The standards of evidence are strict. But when followed they allow us to see far, illuminating even a great darkness.

Questions, suggestions, comments? 😊

• Thank you for your attention!