Perspectives on the CCL Classification Process and Prototype Modeling

Report for the NDWAC CCL Work Group
September 17, 2003

1

The Classification Process

- Review perspectives
  - on the CCL Classification Process
  - on the role of the models
- Address some of the issues related to the process and use of the models

2

Perspective

CCL Classification is a Judgment Process -

based on an evaluation whether or not a contaminant is likely to occur in drinking water - and be harmful if it does

- Models to aid in moving from PCCL to CCL make process more consistent and transparent
- Model is pattern recognition algorithm – to replicate past decisions (those made with the training data set)
- Automaton is probably a misnomer
- Attributes are ranked scales 1=good 10=bad
- Need to be careful not to imply unjustified precision

3

Perspective

- CCL Classification is a Judgment Process

"Quantification does not eliminate subjectivity, but it discourages vagueness"

From page 1, chapter 1 of "Environmental Systems Optimization" b y Doug Haith

4

Perspective

- On Attribute Scoring
  - If we had same data available for all contaminants over time this wouldn't be necessary
  - Imprecision embedded in the range of data elements to be used; uncertainty inherent
  - Attributes are ordered categories (ranks)
  - Some approaches to handling uncertainty/confidence imply false precision in the assessment

5

Calibration-Scales-Scoring for Attributes

- On Scales for Scoring Attributes
  - What should the scale be?
    - 1-3 – good, bad, ugly
    - 1-10
    - 1-100
    - The whole real number line…
    - 1-10 consistent with the level of precision inherent in this process

6

Calibration-Scales-Scoring for Attributes

- Defined scales/calibration and scoring processes for attributes improves transparency, consistency
- Establishing scales documents decisions, and increases transparency ("discourages vagueness")
- Defined Scales - Experts make 50 decisions up front versus trying to make individual decisions on 1,000 contaminants

7

Calibration-Scales-Scoring for Attributes

- Same scales must be used for the training data and all subsequent contaminants evaluated
- Calibration scale can be related to "1-10" in many ways - Data do not assume normal distributions across the scale
- Ensure end members (1 and 10) fit the important end members of the properties of the data element/attribute
- Scales must cover the range of conditions expected and shouldn't cluster around one or two scores

8

Independence

- On Magnitude, Potency and Independence of the Attributes
  - Concern about adjusting Magnitude related to Potency - they will no longer be independent variables
    - Model usefulness is not contingent on assumption of independence among attributes
    - Final model is likely to use a subset of the 5 attributes

9

Missing Data

- On Persistence-Mobility and Missing Data
  - NRC – "For PCCL contaminants that have demonstrated occurrence in water, the occurrence attributes of prevalence and magnitude should be scored and take precedence over their persistence-mobility scores. However, *in the absence* of data on occurrence persistence and mobility should be used to assess the potential for significant occurrence … … …"
  - Not all models being considered can deal with "missing" data; no models deal well with "either-or" situations

10

Microbes

- On Different Scales/Definitions for Microbiological and Chemical Attributes
  - There are many issues related to the differences between chemical and microbiological contaminants
    - What if it is not feasible to use/score all attributes for microbiological contaminants at this time?
    - What if the attributes require different definitions and scales for microbes than chemicals?

11

Microbes

- Several options:
  - Use separate models for chemicals, microbes
  - Include a “dummy” (indicator) attribute
- Final model will not likely use all the attributes
  - Test models and see what is necessary

12

Conclusions

- Don’t worry, be happy…