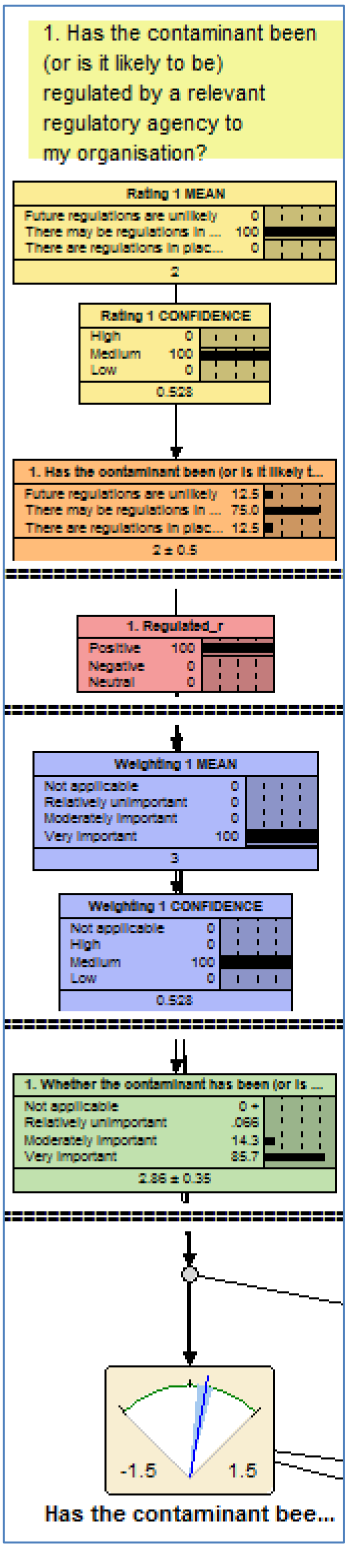
A Bayesian Network Model for Prioritization of Research on Constituents of Emerging Concern – Netica Version

A Bayesian Network model was developed using software package Netica (v6.0 Norsys Software Corp). This model can be used to prioritize (by numerical rank order) further research on trace chemical contaminants. The model incorporates current methodology for multi-criteria decision analysis (MCDA) with additional capabilities to incorporate uncertainty in criteria scores and weightings, and hence in final combined prioritization indicators. A fully functional freeware copy of Netica is available to run this Bayesian Network (www.norsys.com/download.html).

Problem Description

Many organizations, including water utilities, public health and environmental regulators, universities, and other research agencies, have an interest in developing new knowledge range emerging trace chemical contaminants. However, emerging contaminants are diverse in the issues that they present in terms of potential environmental, public health, or other risks. Furthermore, various organizations will have differing priorities, in terms of which knowledge gaps are most essential or urgent to address. Therefore, in order to develop a tool to assist prioritization, it is necessary to consider both the priorities of the organization, as well the attributes of each potential chemical to be investigated, as they relate to those priorities.



Prioritization Criteria

The importance of various prioritization criteria were assessed by means of a stakeholder survey. This survey was conducted online using SurveyMonkey and responses were received from 36 organizations. Respondents were asked to rate the importance of each of a list of 20 potential criteria and were provided an opportunity to identify additional criteria. From this approach, a final 11 criteria were selected for priority assessment of emerging contaminants. These are (in no specific order):

- Whether the contaminant has been (or is it likely to be) regulated by a relevant regulatory agency to the organization;
- Whether the contaminant is known or suspected to cause acute or chronic health risks;
- Whether there is a lack of information regarding health risks;
- Whether there is evidence for widespread or emerging public concern regarding this contaminant;
- Whether the contaminant is known to be present (occasionally or consistently) in water the organization supplies or discharges;
- Whether the contaminant is known or suspected to be produced during drinking water treatment;
- Whether the contaminant is known or suspected to be produced during wastewater treatment;
- Whether the contaminant is expected to present (somewhat) unique challenges to the organization;
- Whether there are existing analytical techniques available with suitable limits of detection;
- Whether the contaminant appears to be difficult to manage by existing/conventional treatment processes;
- Whether there is information available regarding removal by existing/conventional treatment processes.

Multi-Criteria Decision Analysis (MCDA)

Established techniques for multi-criteria decision analysis (MCDA) were coded into a Bayesian Network. This enables to user (or "prioritizing decision maker") to consider each chemical against each of the 11 criteria. These considerations are then combined with user-defined weightings for the importance of each criterion. These two types of information are then combined for each of the 11 criteria to provide a final combined prioritization indicator for each contaminant.

Prioritization Criteria Posed as Questions to the Prioritizing Decision Maker

To facilitate the assessment of each specific contaminant against each criterion, the 11 criteria are presented in the form of questions to the prioritizing decision maker. The 11 questions are then:

- Has the contaminant been (or is it likely to be) regulated by a relevant regulatory agency to my organization?
- Is the contaminant known or suspected to cause acute or chronic health risks?
- Is there a lack of information regarding health risks?
- Is there is evidence for widespread or emerging public concern regarding this contaminant?
- Is the contaminant known to be present (occasionally or consistently) in water my organization supplies or discharges?
- Is the contaminant known or suspected to be produced during drinking water treatment?
- Is the contaminant known or suspected to be produced during wastewater treatment?
- Is the contaminant expected to present (somewhat) unique challenges to this organization?
- Are there existing analytical techniques available with suitable limits of detection?
- Does the contaminant appear to be difficult to manage by existing/conventional treatment processes?
- Is there current information available regarding removal by existing/conventional treatment processes?

Chemical-Specific Criteria Scores (Selection and Uncertainty)

For each chemical, the prioritizing decision maker must consider the degree to which each criteria question is true or not. For example, for the criterion question "*Has the contaminant been (or is it likely to be) regulated by a relevant regulatory agency to my organization?*", the user may select whichever of the following responses most accurately represents the answer for the particular chemical:

- Future regulations are unlikely (score = 1)
- There may be regulations in the future (score = 2)
- There are regulations in place or under development (score = 3)

Uncertainty in this response is captured by the user then rating their degree of "confidence" that the most appropriate criteria score was selected. Degrees of confidence may be rated as "high", "medium", and "low". This choice applies a variable degree of "spillage" from the selected criteria score to adjacent criteria scores. If "high" confidence is selected, 100% of the probability density is allocated to the criteria score selected. If "medium" confidence is selected, 75% of the probability density is allocated to the selected criteria score and the remaining 25% is shared equally between adjacent criteria scores. If "low" confidence is selected, 50% of the probability density is allocated to the selected criteria score and the remaining 50% is shared equally between adjacent scores.

Type of Influence (Positive, Negative, Neutral)

For each criteria question, the user must consider what type of influence that criterion would have on their prioritization. For example, the fact that there are existing analytical techniques available for a chemical may make one organization more likely to want to undertake research on it (a "positive" influence). However, it may also make another organization less likely to want to undertake research on it (a "negative" influence). For a third organization, this may not be a significant consideration at all (a "neutral" influence). The prioritizing decision maker is required to define each criterion by designating its influence as positive, negative or neutral.

Criteria Weighting (Selection and Uncertainty)

Having defined the type of influence each criterion has for the prioritizing decision maker, it is then necessary to define a weighting for the relative degree of importance for each criterion. The following weightings are available:

- Not applicable (weighting = 0, automatically selected for criteria selected to have "neutral" influence)
- Relatively unimportant (weighting = 1)
- Moderately important (weighting = 2)
- Very important (weighting = 3)

Uncertainty is again captured by allowing the user to indicate their degree of confidence in this weighting selection as a rating of "high", "medium", or "low". This step has the same effect for the criteria weightings as the uncertainty descriptions had for chemical-specific scores, as described above.

Criterion-Specific Prioritization Indicator

The contribution of each criterion is then calculated according to the selections made by the prioritizing decision maker. Chemical-specific criteria sores are applied (1-3), with spillage to adjacent scores, as described by the application of uncertainty ratings. Similarly, criteria weightings are applied (0-3) with spillage to adjacent weightings according to selected uncertainty ratings. These data are then combined to determine the overall contribution of each criterion to the final prioritization indicator for each chemical.

Combined Prioritization Indicator

The calculated contributions of each criterion are combined to produce an overall combined prioritization indicator for each contaminant, according to the 11 criteria used in the assessment. Combined prioritization indicators may be negative or positive, depending on whether negative or positive types of influence were predominantly selected and the relative weightings that were applied to them. The combined prioritization indicator is presented with a mean and standard deviation, to indicate both the magnitude the degree of confidence in the combined prioritization indicator. The graphical representation of the result provides further insights to how much of combined prioritization density is determined to be contained within each one of six discretized bands (<-2, -2 to -1, -1 to 0, 0 to 1, 1 to 2, >2). Combined prioritization indicators and their associated uncertainty may then be compared among multiple chemicals to produce a prioritization ranking.

Batch Processing of Contaminants

Prioritization ranking is only meaningful when combined prioritization indicators of multiple contaminants are compared against one another. To facilitate this, batch processing of contaminants is possible. In this case, the chemical-specific criteria scores (and associated confidence rating) may be conveniently entered in an Excel spreadsheet, referred to as the "case file", as shown below.

	A	B	C	D	E
1	IDnum	Criterion 1 Score	Criterion 1 Confidence	Criterion 2 Score	Criterion 2Confidence
2	1	3	Low	1	Low
3	2	2	High	1	Low
4	3	2	Medium	2	High
5	4	1	High	3	Medium

Once complete, the prioritizing decision maker can batch process these contaminants in Netica by selecting [Cases] → [process cases] and then selecting a "control file", which governs the output (and is provided along with the Netica software file). Then the "case file" may be selected and finally a "results" file designated. This approach will quickly generate a prioritization score for all contaminants.

Bayesian Network Model for Prioritization of Constituents of Emerging Concern (CECs)