Science, trans-science and policy: How to build a dredged material decision framework

6th International SedNet Conference, 6-8 October, 2009, Hamburg, Germany

Sabine E. Apitz

SEA Environmental Decisions, Ltd.

*1 South Cottages, The Ford; Little Hadham, Herts, SG11 2AT, UK; +44 (0)1279 771890;
drsea@cvrl.org

SedNet
European Sediment Research Network

- Whilst most DM decision frameworks are risk-based and built upon our scientific understanding of ecological risks of various processes, they are tools for implementing policy
- Many aspects of these frameworks, such as how lines of evidence (LOEs) will be combined, and what decisions they lead to, are quite clearly policy decisions
- What is less clear is that even more seemingly scientific aspects, such as the development of toxic risk standards and the selection of bioassays is permeated with policy choices
- "…contemporary science can provide only partial answers to pressing environmental problems, (but that) this limitation is esoteric and often escapes the lay observer"* (and many scientists)

*Wagner WE. 1995. The Science Charade in Toxic Risk Regulation. Columbia Law Review 95(7):1613-1723

- The development of standards and tools may fall victim to a “science charade”* in which
- “the capabilities of science susceptible to …overstatement”*, and
- the role of science, trans-science (questions which can be asked of science and yet which cannot be answered by science, and are thus addressed by policy) and policy can be unclear
- When these lines are blurred, we lose our ability to be adaptive, and this poses risks as tools are applied to different management decisions, regulatory frameworks and policy priorities.

*Wagner WE. 1995. The Science Charade in Toxic Risk Regulation. Columbia Law Review 95(7):1613-1723

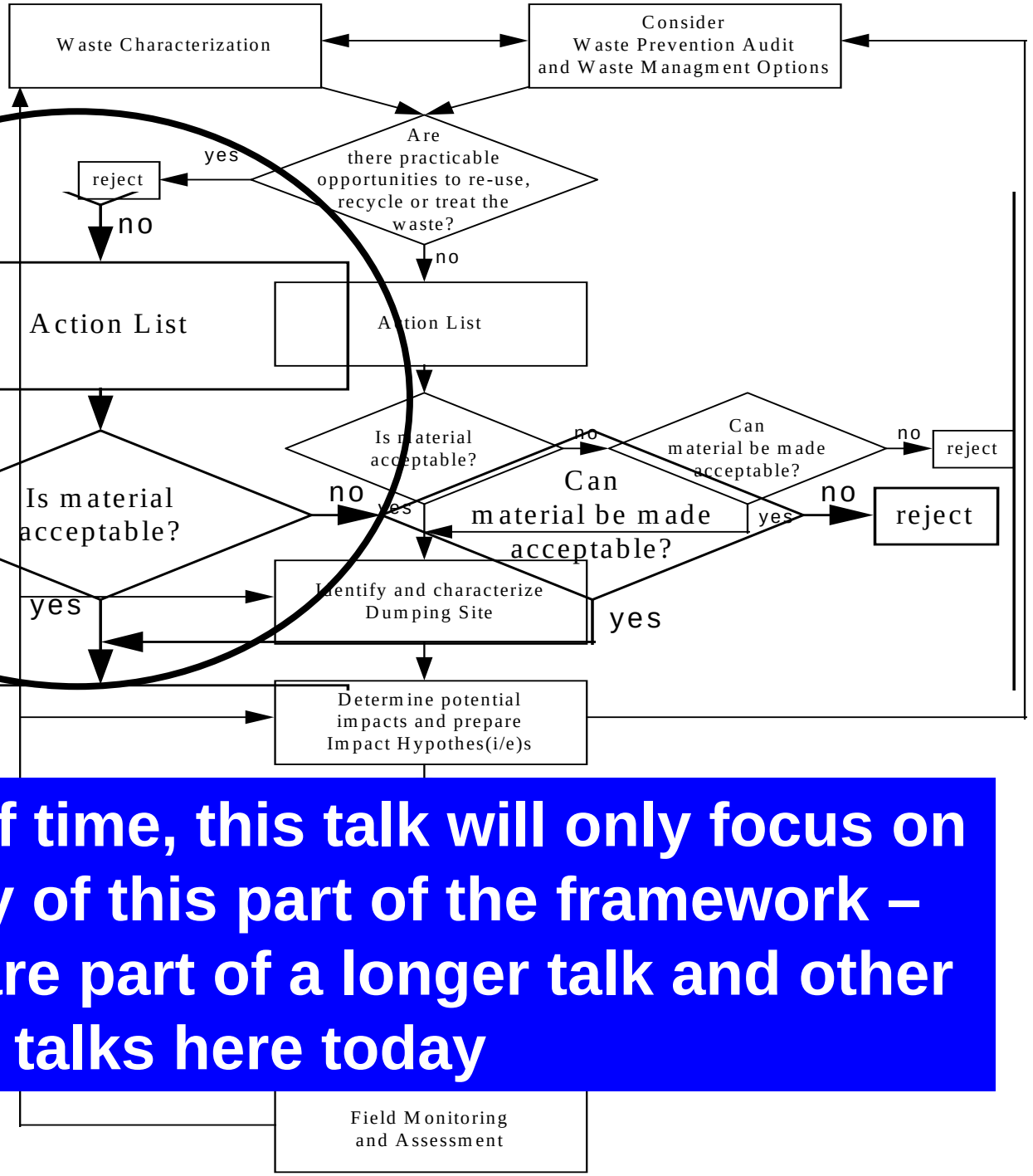
International review of Dredged Material assessment/ management frameworks and approaches

- Reviewed frameworks for dredged material management
- Examined the technical and policy drivers
- Seemingly subtle differences can result in significantly different decisions
  - Frameworks are ***not*** interchangeable without careful analysis of decision drivers and program needs
  - This presentation will provide a sampling of a few of the countless issues that can affect DM disposal decisions

London Convention 1972 and the 1996 Protocol to the London Convention

- Objectives:
  - "…to protect and preserve the marine environment from all sources of pollution and
  - take effective measures...to prevent, reduce, and where practicable eliminate pollution caused by dumping...at sea..."

Assessment Framework for the London Convention and Protocol



In the interest of time, this talk will only focus on the complexity of this part of the framework – other aspects are part of a longer talk and other talks here today

Definition of Terms (London Convention)

- A **characteristic** is an attribute of the dredged material (e.g., copper, mercury, silt, petroleum compounds, pathogens) or a biological response to the dredged material (e.g., mortality, growth, bioaccumulation).
- A **metric** is a measurement that can be made on the characteristic (e.g., concentration, percent survival).
- A **benchmark** is a point on the range of the metric (e.g. 4 mg/kg copper, 20% amphipod mortality) that is used to identify where environmental concern may be low or high for that characteristic. These can be referred to as the lower benchmark and upper benchmark.
- An **Action List** comprises of a number of characteristics to be considered for measurement in the dredged material.
- An **Action Level** is a decision rule based on the findings of one or more characteristics in comparison to the respective benchmarks.

Action Levels establish thresholds that provide decision points that determine whether sediments can be disposed of at sea

- Action Levels specify an Upper Level (UAL) and may also specify a Lower Level (LAL)
  - Upper Level should be set so as to avoid acute or chronic effects on human health or on sensitive marine organisms representative of the marine ecosystem
  - Below the Lower Level, there should be little concern for disposal at sea
  - Between the two, more detailed assessment is required

From Action Levels to Frameworks

- Different countries and programmes define action levels differently, using various lines of evidence
- Many programmes have developed tiered systems to prevent unnecessary analyses
- These are designed so that more detailed analyses are only needed to reduce uncertainty
- Which metrics are used and combined to assess parameters in the action list defines the action levels and the DM framework

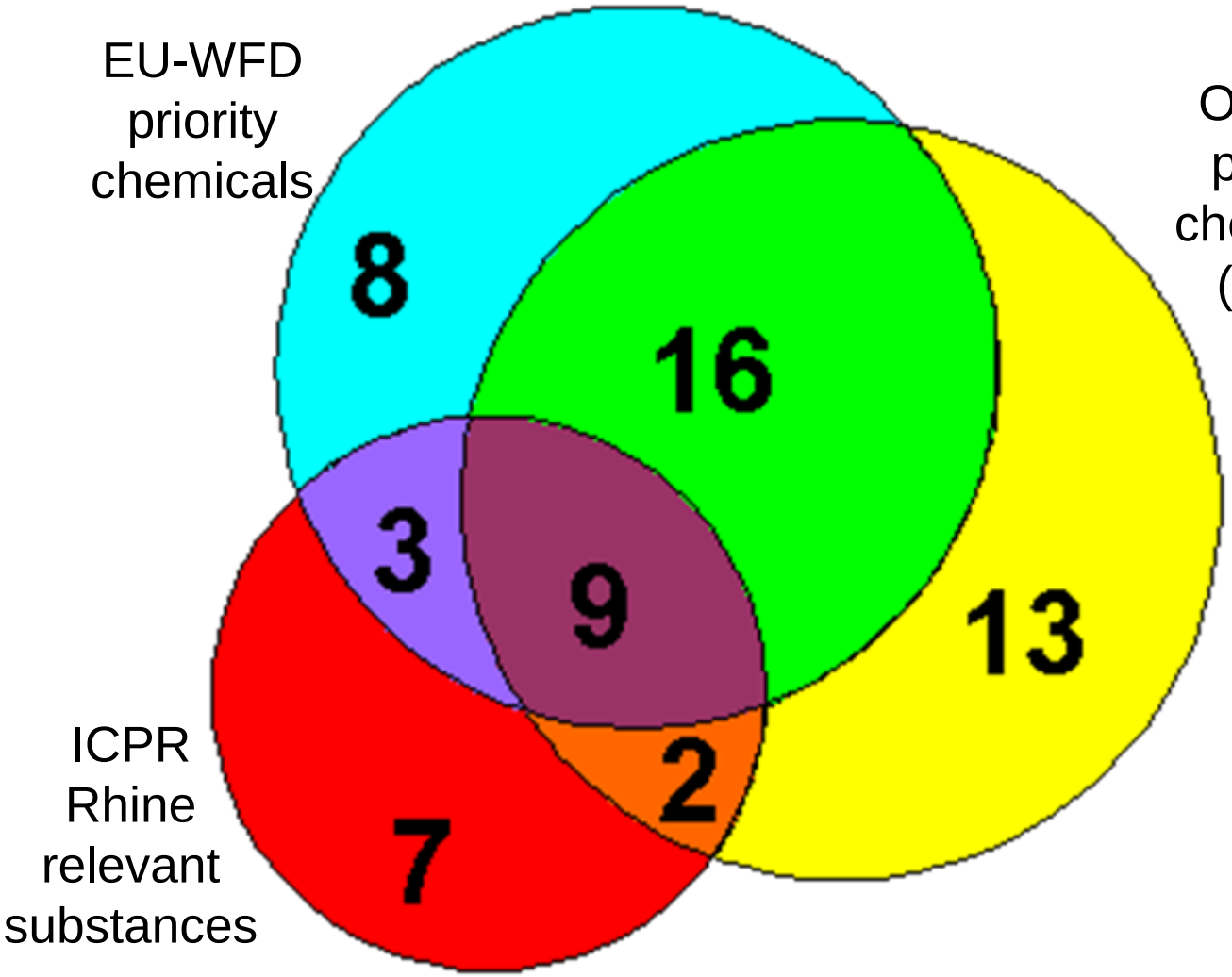
Country	General approaches used			Notes
		No. categories in action level approach	*Methods used in development of action levels*	
Belgium	Action level	3	1) Sediment chemistry 2) Bioassays	Action levels based on mean contaminant concentration in marine navigation channels + scientific study
Denmark	Action level	3	Sediment chemistry	Not yet implemented in law
Finland	Case-by-case	---	---	Legislation for action levels in preparation
France	Action level	3	Sediment chemistry	
Germany	Action level	3	1) Sediment chemistry 2) Bioassays	
Portugal	Action level and case-by-case	5	sediment chemistry	
Netherlands	Action level	1 limit level	Sediment chemistry	Biological effects methods being used.
Norway	Action level + case-by-case	5	---	
R. of Ireland	Case-by-case	---	---	3 categories under development
Spain	Action level	3	Sediment chemistry	Sediment chemistry not yet implemented in law. Sediment bioassays under development
Sweden	Case-by-case?	---	---	
UK	England and Wales (E+W) and Scotland: Case-by-case approach.	--- ---	---	E+W: Under review. 3-category action level approach in preparation. Scotland: Data assessed against OSPAR BRCs

Action list – what do we measure?

- The priority pollutants lists used in programmes are not necessarily the most risky or important contaminants
- There is a “skewed prioritization system”*
  - Many believe that “…agencies assign priority to the worst risks first,”
  - But agencies “…appear to default to an ad hoc system in which the substances with more scientifically established health effects are selected over less-studied substances, many of which are believed to present greater risks at lower concentrations,”
  - Regulation is often withheld until a standard is supported by the "weight of evidence“,
  - but slow pace of studies can potentially cause delays of several decades for each toxic substance*

*Wagner WE. 1995. The Science Charade in Toxic Risk Regulation. Columbia Law Review 95(7):1613-1723

Comparison of lists of priority chemicals in different programmes affecting Rhine Sediments



(PORII, Port of Rotterdam)
Slide Courtesy of Jos Brils, TNO

Many of the toxicants in European Rivers are not Priority Pollutants

Compound	Priority substance [a]	Confirmed [b]	Site/basin	Reference
Mutagenicity/genotoxicity				
benzo[a]pyrene, benzo[k]fluoranthene, indeno[1,2,3-cd]pyrene, benzo[ghi]perylene	Yes	Yes	Many sites worldwide	e.g. (Brack et al. 2005b)
perylene, benz[a]fluoranthene	No	Yes	Neckar basin (Germany)	Brack et al. 2005b
11H-indeno[2,1,7-cde]pyrene, methyl benz[a]anthracenes and perylenes	No	No	Neckar basin (Germany)	Brack et al. 2005b
polar polycyclic compounds including benz[a]anthracenquinone, pyrenequinone, nitropyrenequinone, nitroanthraquinone, nitrobenzanthracenedione, 6-nitrochrysene, nitrobenzo[a]-pyrenes, nitroindeno[1,2,3-cd]pyrene	No	No	Mediterranean Sea, coastal zone at Barcelona (Spain)	(Fernandez et al. 1992)
Ah-receptor-mediated effects				
PCDD/Fs, PCBs	No	Yes	Western Scheldt (The Netherlands), Spittelwasser (Elbe basin, Germany)	(Stronkhorst et al. 2002, Klamer et al. 2005, Brack et al. 2002)
benzo[a]pyrene, benzo[k]fluoranthene, indeno[1,2,3-cd]pyrene, benzo[ghi]perylene	Yes	Yes	Morava river (Danube basin, Czech Republic)	(Vondracek et al. 2004, Hilscherova et al. 2001, Machala et al. 2001c)
dinaphthofurans, 2-(2-naphthalenyl)benzothiophene, 9-methylbenz[a]anthracene, 1-methylchrysene	No	Yes	Spittelwasser (Elbe basin, Germany)	(Brack and Schirmer 2003)
Estrogenicity				
nonylphenol	Yes	Yes	Llobregat (Spain), river Neckar	(Céspedes et al. 2005, Hollert et al. 2005)
benzophenone, phthalates, dehydroabietic acid, sitosterol, 3-(4-methylbenzylidine)camphor, 6-acetyl-1,1,2,4,4,7-hexa-methyltetralin	No	No	Rivers Neckar, Rhine (Germany), Thames (United Kingdom)	(Fastall et al. 2006)
tributyltin	Yes	Yes	Elbe (Germany)	(Brack et al. 1999, Schulte-Oehlmann et al. 2001)
17β-estradiol, estrone, estriol	No	Yes	United Kingdom estuaries, different rivers in the Netherlands, Swiss wastewater treatment plant effluents	(Thomas et al. 2001, Houtman et al. 2004, Aerni et al. 2004)
Androgenicity				
dehydrotestosterone, androstenedione, androstanedione, 5β-androstane-3α,11β-diol-17-one, androsterone, epi-androsterone	No	Yes	United Kingdom estuaries	(Thomas et al. 2002b)
Green algae				
N-phenyl-2-naphthylamine, prometryn	No	Yes	Spittelwasser (Elbe basin, Germany)	(Brack et al. 1999)
priority PAHs, tributyltin	Yes	Yes	Spittelwasser (Elbe basin, Germany)	(Brack et al. 1999)
Invertebrates				
methyl parathion	No	Yes	Spittelwasser (Elbe basin, Germany)	(Brack et al. 1999)
pentachlorophenol, atrazine	Yes	No	United Kingdom estuaries	(Thomas et al. 1999)
tri-, tetra-chlorophenol, 4-chloro-3,5-dimethylphenol, nonylphenol, 4-chloro-3,5-xylenol, dieldrin, carbophenothion methylsulfoxide	No	No	United Kingdom estuaries	(Thomas et al. 1999)

[a] according to EU-WFD
[b] confirmed as a cause of the measured effect

Brack W, Klamer HJC, Alda ML, Barceló D. 2007b. Effect-Directed Analysis of Key Toxicants in European River Basins A Review. Env Sci Pollut Res 14(1):30-38.

What we can analyze is only a fraction of what is there

“not everything that can be measured is worth measuring, and not everything worth measuring is measurable.” *

So, how can we address these issues in a DM framework?

*Daughton CG. 2002. Pharmaceuticals and Personal Care Products (PPCPs) as Environmental Pollutants - Pollution from Personal Action. Las Vegas, NV: Environmental Sciences Division, National Exposure Research Laboratory, Office of Research and Development, USEPA

Option 1: limited list, chemistry alone in first tier

- If the action list characteristics
  - are the cause of toxicity, or
  - co-vary with unmeasured contaminants that are the primary cause of toxicity
    - then a chemical “fail” will lead to further biological assessment and the approach will catch the contaminant risk
    - (if the mode of toxicity is one for which the bioassay battery is sensitive)
- On the other hand, if the chemicals analyzed
  - are below action levels,
  - but undetected compounds are at levels that cause toxicity,
- then this will be missed
  - the sediment may pass and
  - never be subject to further bioassessment, and
  - potentially toxic sediments may be disposed of at sea.

Short list, chemistry first

- Advantages:
  - well-established
  - provides a level of consistency
- On the other hand:
  - Applicants may be penalised for site knowledge, if screening lists are only expanded based on knowledge of other current or historical sources
  - If knowledge of such issues is good, then this assumption is a sensible, efficient one.
  - But monitoring data which inform the screening list may only look for what is known, and important contaminants might be missed
    - This can result in self-perpetuating programs in which regulation, monitoring and decisions focus on what is deemed of importance only because it is what has been examined, perpetuating in the “skewed prioritisation”

Option 2: Longer Action List, Chemistry Alone in first Tier

- How effective this is depends upon
  - The chemical selected (and the basis for choice)
  - Whether the added chemicals are drivers for impacts that the current list does not catch
  - Whether selected bioassays are sensitive to the modes of toxicity important in chemicals selected
    - “standard” bioassays have co-evolved with priority pollutant list and are often confirmatory, not complementary to chemical measurements

Option 3: Screening biotest in first tier

- Depends upon the biotest selected
- Some have suggested that a sediment extract or elutriate in the first tier of a tiered framework
  - Ideally the bioassays would be aimed at various toxicity mechanisms, with bioassays in later tiers looking at bioavailability issues.
  - A commonly proposed candidate for this biotest is acute bioluminescence inhibition of the bacteria *Vibrio fischeri*
  - However, short-term tests detect non-specific effects and long-term effects are not detectable
  - A bioassay that could be deemed “universally sensitive” is unlikely

Option 4: Add bioeffects in action levels

- To address the problem of emerging or ignored contaminants with different biological effects, specific bioeffects can be used as pass/fail criteria in action level, even if chemistry is not identified
  - Examples of effect-types are cytotoxicity, neurotoxicity, mutagenicity, estrogenic or photosynthesis inhibiting potencies
- If designed to catch contaminants and modes of toxicity not currently addressed, these tests ***must*** be included in a screening tier
  - Otherwise, IF the analyzed *chemicals* are below LAL, but there are (unmeasured) effects due to other contaminants, the sediment will pass without other contaminants being detected.
- If effects-bases tools are used in the first tier, then other classes of chemicals will be flagged, leading to further assessment
- Critical questions then are:
  - Can DM be passed or failed only on bioeffects?
  - Do chemicals need to be identified?
  - How do we address the effects from natural chemicals?
  - How do we define references?
  - How do we standardise?
- Each of these questions has detailed technical and policy implications

Benchmarks and chemical data

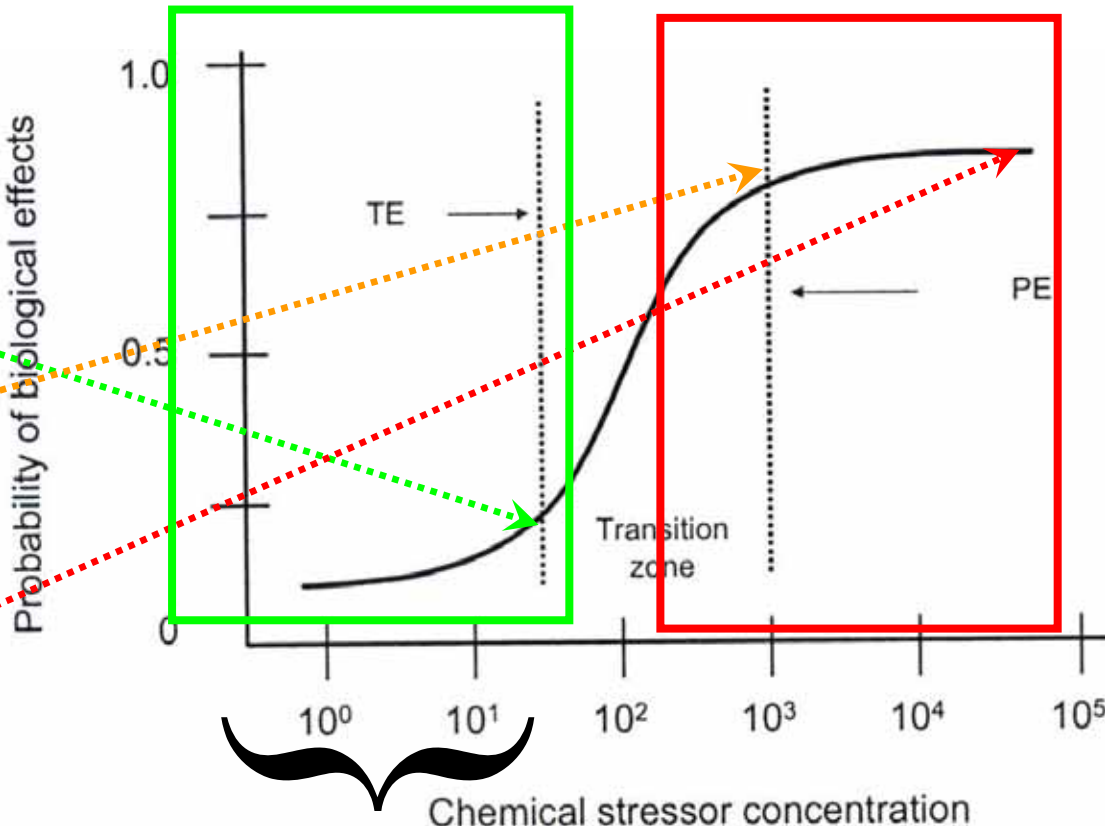
- How does one derive benchmarks if none are available?
- How can chemical analysis results (which may include levels for thirty or more substances) be interpreted in the screening and first tier of a decision framework?

Chemical benchmarks can be broken into four general categories

- Background (BK)
  - Not based upon toxicity, al
  - Below these levels, effects are unlikely (e.g., TEL)
- Probable effects (PE)
  - Above these levels, effects are probable (e.g.
- I
  - probable effects; these can be used to trigger immediate action

Often the basis of UALs

Often the basis of LALs



Note: this figure illustrates the relative position of benchmark classes on the probability curve; their actual generation may be based on much more

Figure from Wenning et al. (Eds.), 2005. Use of Sediment Quality Guidelines & Related Tools for the Assessment of Contaminated Sediments (SQG). SETAC Press, Penasacola, FL.

Chemical benchmark derivation

- Empirically Derived (e.g., ERM, ERL, PEL, TEL, AET)
- Equilibrium Partitioning (EqP) Approaches
- Consensus Approaches
- Background sediment chemistry approaches
- Tissue residues or bioaccumulation potential approaches

- Details of these approaches outside scope today, but derivation details (sampling, tests, statistics, safety factors, etc) grounded in both science and policy

Selection of LAL Benchmarks

- Background-based benchmarks are one option
- This approach would support the *pollution prevention* aspect of London Convention objectives
  - Below regionally appropriate LALs, the probability is that sediments pose no net risk (no greater risk than all sediments in the region), and no further analysis is required
  - If contaminant levels are higher than background levels, then higher-tier assessments evaluate whether these levels present a risk

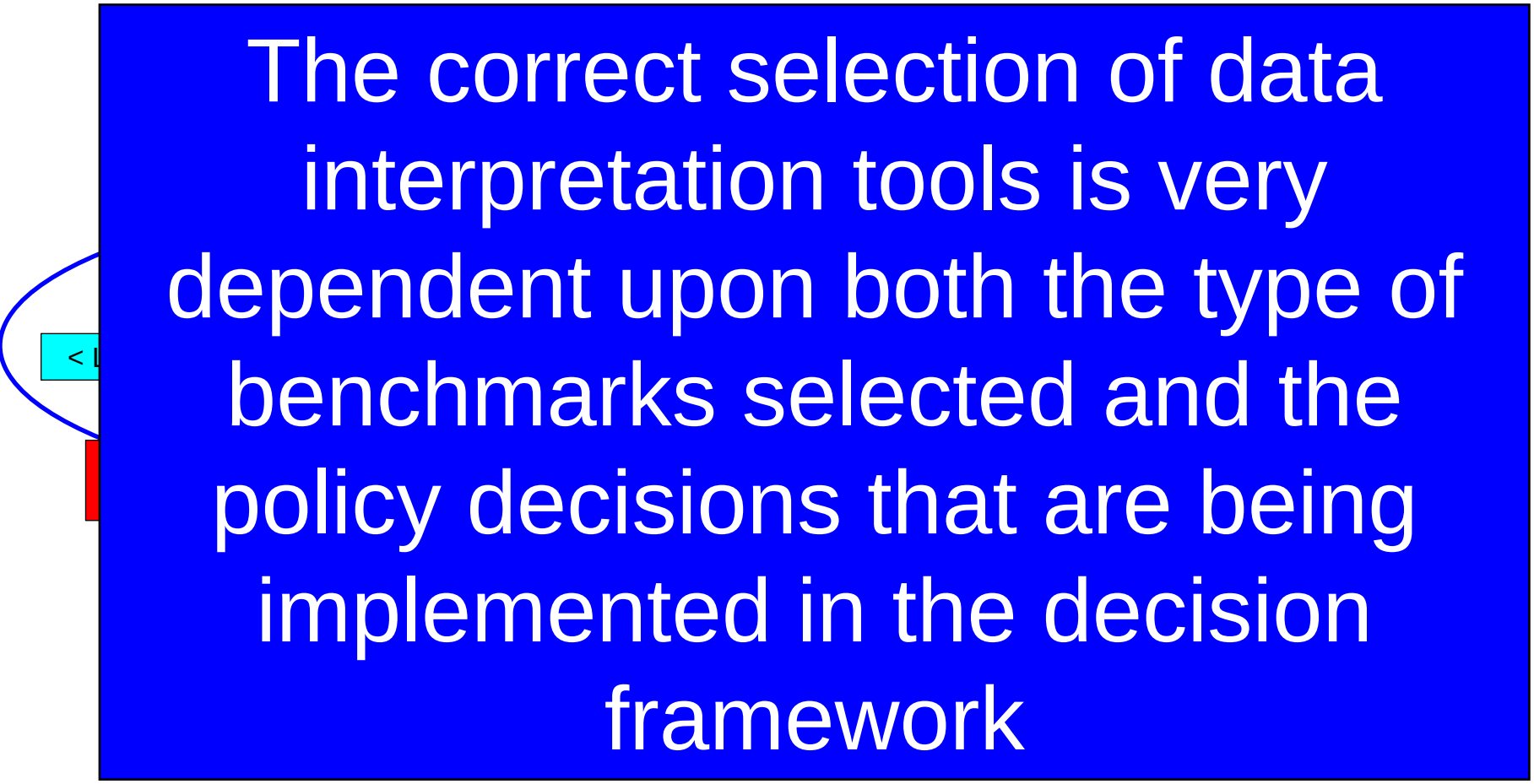
Background vs risk-based LALs: hypothetical outcomes

- If LALs are risk-based (rather than background-based), three hypothetical scenarios:
  - Risk-based LALs the same as background levels;
    - no difference in outcome between the approaches
  - Risk-based LALs below background levels,
    - may trigger failures in sediments with metals at background level
    - screened out in higher tier evaluations of background levels in many frameworks
    - Unlikely use
  - Risk-based LALs higher than background levels.
    - Such “passes” would not be examined in higher tiers
    - Even if appropriate, may raise concerns in terms of the objective of pollution (not risk) prevention

Selection of UAL Chemical Benchmarks

- Many frameworks use a chemical benchmark as a UAL “rejection level”
  - Should be set to minimise false positives
  - If chemical, should be set at the point where biological testing is also likely to fail
  - If too conservative, may cause excessive cost or push unwanted outcomes
- If chemical benchmarks are used as sole pass/fail criteria, without other LOEs, flexibility is lost

How can chemical analysis results (which may include levels for thirty or more substances) be interpreted in the screening and first tier of a decision framework?

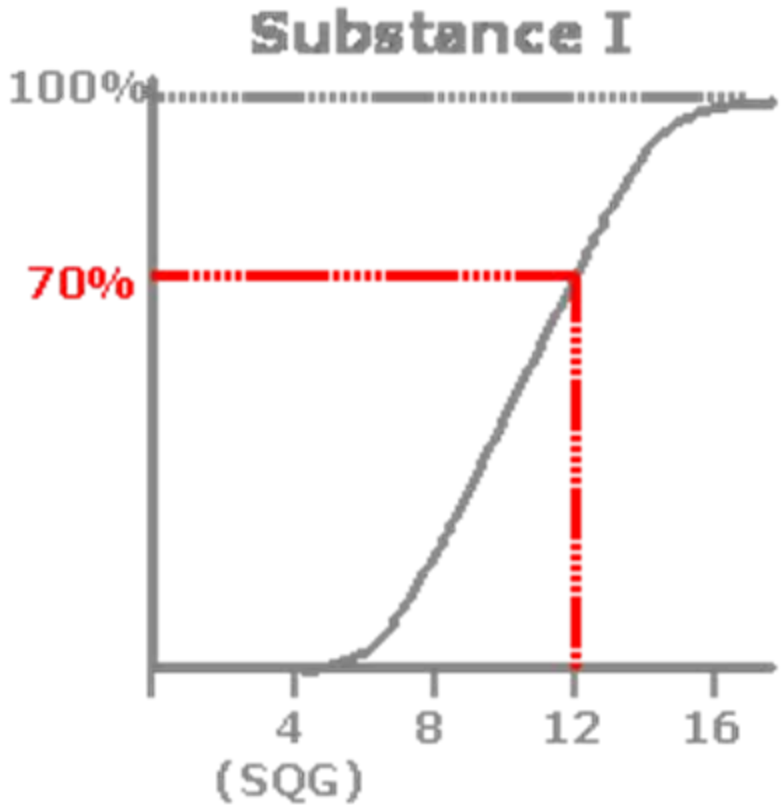


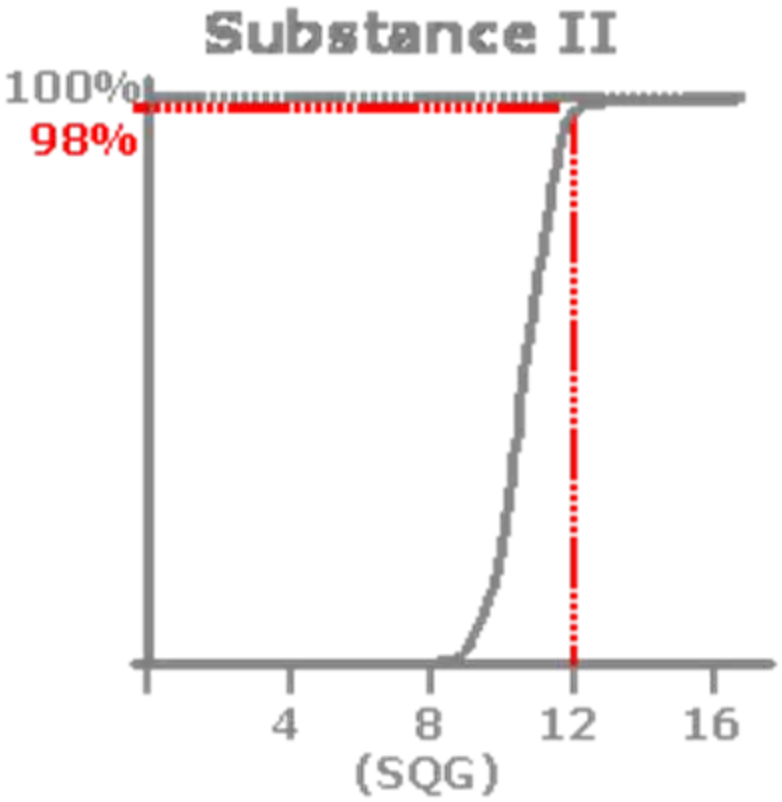
The correct selection of data interpretation tools is very dependent upon both the type of benchmarks selected and the policy decisions that are being implemented in the decision framework

Different ways of using sediment chemistry to evaluate sediment quality. Formulae are for sediment j with contaminant i at concentration $[C]_{ij}$.

Criteria	Advantages	Disadvantages
Pass/fail If $[C]_{ij} > SQG_i$, even by a small amount, the sediment "fails"	•Easy to use •Unambiguous	•Single compound evaluation •Does not account for variability/uncertainty (unless statistically designed to do so)
Set exceedance $[C]_{ij}$-SQG_i Some set exceedance can be allowed	•Easy to use •Unambiguous	•Same as above •Effectively a slightly less conservative SQG
SQG Quotient $SQGQ_{ij} = [C]_{ij} / SQG_i$ If SQG quotient >1 the SQG has been exceeded	•Easy to use •Magnitude of quotient = degree of exceedance •Quotients useful for examining variability/uncertainty •Allows site ranking and comparisons for one contaminant	•Single compound evaluation •If quotient is set at 1, same as pass/fail in decisions •If many contaminants are being considered, comparisons can get confusing
SQG sum $\Sigma_{i=1-n}(SQGQ_{ij})$	•Evaluation of additive risk (if SQGs used are appropriate) •Good tool for graphically showing risk "drivers" •Allows site ranking and comparisons for multiple contaminants	•Depends on the number of contaminants analysed •Less clear as a decision tool •Not all risks are additive
Mean SQG quotient $mSQGQ_j = (\Sigma_{i=1-n}(SQGQ_{ij}))/n$	•Evaluation of combined risk •Can be used to define multiple compound decision criteria •Allows site ranking and comparisons for multiple contaminants	•Depends on the number of contaminants analysed •Can obscure high exceedances of single compounds •If only this number is reported, it obscures the cause(s) of risk

Absolute risk of a given contaminant is a function of its dose-response curve. The fact that both benchmark quotients are well above 1 indicates risk from both substances, but it does not quantify that risk





$SQGQ=[C]_I/SQG_I = 12/4 = 3$

RISK = 70% affected

$SQGQ=[C]_{II}/SQG_{II} = 12/8 = 1.5$

RISK = 98% affected

(adapted from Ragas and Leuven 2007),

Criteria	Advantages	Disadvantages
Rules-based (combinations of above for decisions)	•Easy to use •Unambiguous •Provides decision rules when dealing with multiple contaminants and multiple samples •Allows for variability/uncertainty •Allows for flexible implementation to achieve objectives	•Can be seen as more complex •Can be more difficult to show graphically
Indices (combine different measures into single metric)	•Provide a single metric for comparing sediments (for trends analysis or risk ranking) •Allow for the combination of various criteria (e.g., number of fails, degree of exceedance, etc)	•Indices obscure the reason for failure •Less clear-cut for a pass/fail decision •Can have artefacts
Statistical methods (compare data on sediment being considered with reference or target sediment chemistry using statistical tools)	•Can help compare complex datasets •Useful if decision criteria are reference-based •Can help spot unusual sediments that pass criteria but are still contaminated	•Not risk-based •Difficult to defend •Difficult to communicate

How can chemical analysis results be interpreted in the screening and first tier of a decision framework?

- Is the sediment below LAL levels?
  - LALs based upon (appropriately designed) background values
    - Decisions based upon a contaminant-by-contaminant pass/fail criterion
  - LALs based upon regionally-appropriate empirically-derived TE benchmarks,
    - Rules-based approach, based upon the predictive power of benchmarks
    - Selection of rules should consider the scientific basis and narrative intent of the benchmarks selected and the policy objectives of this decision level

How can chemical analysis results be interpreted in the screening and first tier of a decision framework?

- Is the sediment above UAL levels?
  - UALs based upon regionally-appropriate empirically-derived PE benchmarks
    - A rules-based approach, based upon the predictive power of the benchmarks is a good approach
    - Rules should be based upon the scientific basis and narrative intent of the selected benchmarks and the policy objectives of this decision level
  - However, this decision should be designed to avoid false positives

Chemical benchmarks based upon tissue residues or bioaccumulation potential

- Reflect sediment concentrations associated with appropriate tissue concentrations
- Two common approaches (others exist):
  - direct guidelines based on tissue residue effects data
  - guidelines that incorporate indirect effects through the action of trophic transfer
- Apart from the Netherlands, no other country has actively considered health-based national guidelines,
  - State of Washington has developed human health criteria for bioaccumulative compounds in Puget Sound sediments
- Many of the factors affecting bioaccumulation and biomagnification can be highly site-specific
- Thus, benchmarks developed for this purpose must either use
  - very generic (and necessarily conservative) assumptions, or
  - extensive site-specific information, obviating their use as an early Tier tool.
  - However, a tiered decision structure can provide balance

Biomagnification – some issues

- Most action levels developed do NOT address potential biomagnifiers (e.g., organic mercury; PCBs; DDT; and, 2,3,7,8-TCDD)

Approaches to addressing biomagnification in early tiers:

- **Precautionary approach:** LAL based upon background levels (essentially zero in some cases), and UAL = LAL, essentially banning ocean disposal for sediments with known biomagnifying substances above background levels
  - This approach is conservative and does not benefit from the tiered assessments
- **Intermediate approach:** LAL based upon background levels, UAL = LAL, with an allowance for site-specific, peer-reviewed modelling to demonstrate that potential biomagnifiers in the sediments do NOT pose a biomagnification risk.
  - Detailed field work and modelling may be too costly, so some sediments may unnecessarily be sent to upland disposal sites, which, could ***increase*** biomagnification risk, albeit to terrestrial food chains.
- **Detailed modelling approach:** Conservative screening, background comparisons, and then preliminary quantitative and detailed quantitative assessments for biomagnification.
  - TGD (EC 2003) provides guidance for the assessment of food chain bioaccumulation and vertebrate secondary poisoning
  - Because the purpose is primarily for prognostic assessments calculations are very conservative, and extensively use assessment factors to account for uncertainty.
  - This approach is not ideal for the site-specific assessment of DM, other tools needed
- In all these, biomagnifiers above background trigger a failure or higher tiered assessment of biomagnification potential.

Implications for a DM programme

- A review of the history of the disposal at sea permit applications, and the range of chemical results encountered would allow for an evaluation of the potential risks, costs and benefits of various approaches
  - For instance, if it is determined that a given decision rule would result in a certain percentage of applications going on to comparative risk assessment of treatment or disposal options (or permit refusal), one can decide if this potential outcome is consistent with its objectives.
- Where both chemical and biological data are available, the predictive power and policy consequences of various decision rules can be tested.
- An important part of the design of a rules-based decision framework will be the relationship between chemistry, any screening bioassay and bioaccumulation potential test results
- The exact tests used, and the relationship between chemical, biaccumulation and biological LOEs in the decision framework all have important implications
- When we are explicit about these choices and their implications, we can be more adaptive

Acknowledgements

- This paper has results from a number of reviews and projects
- Partial funding for various aspects is acknowledged from:
  - Department of Food and Rural Affairs
  - Environment Agency of England and Wales
  - Environment Canada
  - Environ Milan
  - Office of Naval Research
  - Venice Water Authority