Experiences with the NUSAP system for multidimensional uncertainty assessment

J. van der Sluijs*, M. Craye, S. Funtowicz**, P. Kloprogge*, J. Ravetz*** and J. Risbey******

*Copernicus Institute for Sustainable Development and Innovation, Utrecht University, Heidelberglaan 2, 3584 CS Utrecht, The Netherlands (E-mail: *j.p.vandersluijs@chem.uu.nl; p.kloprogge@chem.uu.nl*)

** Knowledge Assessment Methodologies (KAM), Institute for the Protection and Security of the Citizen (IPSC), European Commission – Joint Research Centre (EC–JRC), TP 361, 21020, Ispra (VA), Italy (E-mail: *matthieu.craye@jrc.it; silvio.funtowicz@jrc.it*)

*** Research Method Consultancy (RMC), 106 Defoe House, Barbican, London EC2Y 8ND, UK (E-mail: *jerry_ravetz@lineone.net*)

**** School of Mathematical Sciences, PO Box 28M, Monash University, Clayton, Victoria 3800, Australia (E-mail: *james.risbey@sci.monash.edu.au*)

Water Science & Technology **Vol 52 No 6 pp 133–144** © IWA Publishing 2005

Abstract This paper discusses recent experiences with the NUSAP system for multidimensional uncertainty assessment, based on case studies that vary in complexity. We show that the NUSAP method, and especially the pedigree analysis part of it, is useful to assess not only parameter uncertainty but also to systematically reflect upon (model) assumptions and problem frames. A diagnostic diagram can be used to synthesize results of quantitative and qualitative analysis. It provides an analytic tool to prioritize uncertainties according to quantitative and qualitative insights in the limitations of available knowledge. We further show that extension of the pedigree scheme to include societal dimensions of uncertainty, such as problem framing and value-laden assumptions, further promotes reflexivity and collective learning. When used in a deliberative setting, NUSAP pedigree assessment has the potential to foster a deeper social debate and a negotiated management of complex environmental problems.

Keywords Controversy; NUSAP; pedigree; problem frames; uncertainty; value-laden assumptions

Introduction

According to the mainstream conception of policy-related research, uncertainty hampers the assessment and foresight of complex environmental problems. However, it is a fact of life that the available knowledgebase comprises inputs differing in status, covering the entire spectrum from well established scientific knowledge to practical knowledge, including judgments, educated guesses, anecdotal evidence, and tentative assumptions. This relevant knowledge is co-produced and used in a societal context where decisions need to be made before conclusive supporting evidence is available, while at the same time the potential error costs of wrong decisions can be huge. Usually, controversies surround these issues in which three interrelated factors play a key role: uncertainty in the knowledge base, differences in framing of the problem, and the inadequacy of the institutional arrangement at the science policy interface (Craye *et al.*, 2001). This societal context implies an urgent need for a deliberative, reflexive, and multidimensional approach in which uncertainty is not treated as a deficit but as an opportunity (Funtowicz and Ravetz, 1993; Van der Sluijs, 1997; Craye *et al.*, 2001; Van der Sluijs, 2002). Its discussion should be extended to include non-scientific actors and should take place within a process, taking into account the different perspectives on the issue. Problem framing is seen as a crucial element in uncertainty assessment.

Mainstream uncertainty methods such as Monte Carlo Analysis alone are not suitable for such assessments because a main characteristic of the issues addressed is that unquantifiable uncertainties tend to dominate the quantifiable ones. Unquantifiable uncertainties include those associated with problem framings, model structures, assumptions, system boundaries, indeterminacies, and value-ladenness. Although quantitative techniques (see Saltelli *et al.*, 2000, 2004 for a state of the art overview) are essential in any uncertainty analysis, they can only account for what can be quantified in a meaningful way and thus provide only a partial insight in what usually is a very complex mass of uncertainties. Key dimensions of uncertainty are technical (inexactness), methodological (unreliability), epistemological (ignorance), and societal (social robustness). Each of these dimensions can be manifest at different locations in environmental assessments, for instance, in context, problem frames, indicator choice, model assumptions, model structure, model parameters, and data (Van der Sluijs, 1997). Regarding the technical dimensions, inexactness refers to the numerical precision in for instance data that can be hampered by limited resolving power in measuring equipment or is caused by spread in repeated measurements, but can also arise from natural variability. A next dimension of uncertainty is produced by the limitations of the methods we use. If a number is obtained by the method of educated guessing, it is less reliable than if it was based on a large sample of direct measurements. Further, risk assessment studies are often conditioned on a large number of assumptions made in the analysis, whereas the validity of these assumptions is usually poorly known. The third dimension, epistemological uncertainty, stems from limits to our capacity to know and understand: ignorance. It can take an active or passive form. Active ignorance is for instance at stake when a risk assessor is aware of the aspects in which his knowledge is limited, but has no insight in the relative importance of those aspects so that he, therefore, has no good way to include these in the analysis. Passive ignorance is at stake when one is not aware of what one does not know. As a fourth dimension we can distinguish the societal dimension. The concept here is social robustness (Nowotny, 1999). If scientists fail to negotiate credibility and support for their risk assessment within the public, the assessment can be perceived to be biased or perceived to be so uncertain that it has no guiding value. In risk controversies one often sees that scientific assessments fall apart when a key assumption is revealed in societal debate to be vulnerable to the suspicion of partiality and bias.

Present day quantitative uncertainty methods focus on the inexactness dimension only and, in practice, their application is usually limited to uncertainties in data and parameters (Van der Sluijs, 1997). In many complex environmental problems the other dimensions (unreliability, ignorance, social robustness) and the other locations (problem framing, model structure, and model assumptions) dominate the overall uncertainty, making these methods of limited value. These methods can however be complemented with new approaches addressing these qualitative dimensions of uncertainty that are hard to quantify and were therefore largely under-addressed in the past. In a number of cases we have implemented, demonstrated and tested the Numeral Unit Spread Assessment Pedigree (NUSAP) system that complements quantitative uncertainty analysis with systematic qualitative uncertainty assessment. This paper presents and discusses some of our experiences with the application of the NUSAP method, building on case studies that vary in complexity. In the case studies that are published separately elsewhere we addressed all qualifiers of the NUSAP system. In this paper, the main focuses are on the most innovative part of NUSAP: the pedigree analysis.

Methods

NUSAP is a notational system proposed by Funtowicz and Ravetz (1990), which aims to provide an analysis and diagnosis of uncertainty in the knowledge base of complex (environmental) policy problems. It captures both quantitative and qualitative dimensions of uncertainty and enables one to communicate these in a standardized and self-explanatory way. The basic idea is to qualify quantitative information by means of the five qualifiers of the NUSAP acronym: Numeral, Unit, Spread, Assessment, and Pedigree. We will discuss these five qualifiers. The first is Numeral; this will usually be an ordinary number; but when appropriate it can be a more general quantity, such as the expression "a million" (which is not the same as the number lying between 999,999 and 1,000,001). The second qualifier is Unit, which may be of the conventional sort, but which may also contain extra information, as the date at which the unit is evaluated (most commonly with money). The middle category is Spread, which generalizes from the "random error" of experiments or the "variance" of statistics. Although Spread is usually conveyed by a number (either $\pm$, % or "factor of") it is not an ordinary quantity, for its own inexactness is not of the same sort as that of measurements. Methods to address Spread can be statistical data analysis, sensitivity analysis or Monte Carlo analysis possibly in combination with expert elicitation.

The remaining two qualifiers constitute the more qualitative side of the NUSAP expression. Assessment expresses qualitative judgments about the information. In the case of statistical tests, this might be the significance level. In the case of numerical estimates for policy purposes, it might be the qualifier "optimistic" or "pessimistic." It can also reflect on possible systematic error of which the magnitude can only be estimated in retrospect. Finally there is P for Pedigree, which conveys an evaluative account of the production process of information, and indicates different aspects of the underpinning of the numbers and scientific status of the knowledge used. Pedigree is expressed by means of a set of pedigree criteria to assess these different aspects. Assessment of pedigree involves qualitative expert judgment. To minimize arbitrariness and subjectivity in measuring strength, a pedigree matrix is used to code qualitative expert judgments for each criterion into a discrete numeral scale from 0 (weak) to 4 (strong) with linguistic descriptions (modes) of each level on the scale. Each special sort of information has its own aspects that are key to its pedigree, so different pedigree matrices using different pedigree criteria can be used to qualify different sorts of information, as we will see in the case studies below.

In earlier work, we used NUSAP to complement quantitative analysis with expert judgment of reliability (Assessment) and systematic multi-criteria evaluation of the different phases of production of a given knowledge base (Pedigree) (Van der Sluijs *et al.*, 2002, 2005; Van Gijlswijk *et al.*, 2004). Table 1 gives an example of a pedigree matrix. Van der Sluijs *et al.* (2005) provide a systematic procedure for expert elicitation of

Table 1 Pedigree matrix for emission monitoring data (adapted from Ellis *et al*., 2000 by Risbey *et al*., 2001; Van der Sluijs *et al*., 2005)

Score	Proxy	Empirical	Method	Validation
4	Exact measure	Large sample direct measurements	Best available practice	Compared with indep. measurements of same variable
3	Good fit or measure	Small sample direct measurements	Reliable method commonly accepted	Compared with indep. measurements of closely related variable
2	Well correlated	Modeled/derived data	Acceptable method limited consensus on reliability	Compared with measurements not independent
1	Weak correlation	Educated guesses/ rule of thumb estimate	Preliminary methods unknown reliability	Weak/ indirect validation
0	Not clearly related	Crude speculation	No discernible rigor	No validation

pedigree scores. In the two recent case studies (assumptions in quantitative environmental foresight and controversies on environmental health risks), the approach is further extended to cover societal dimensions such as controversy, problem framing, institutional dimensions, and stakeholder views in a deliberative and reflexive way.

NUSAP provides insights in two independent properties related to uncertainty in numbers, namely spread and strength. *Spread* expresses inexactness whereas *strength* expresses the methodological and epistemological limitations of the underlying knowledge base. The two metrics can be combined in a Diagnostic Diagram mapping *strength* of for instance model parameters and sensitivity of model outcome to *spread* in these model parameters. The Diagnostic Diagram is based on the notion that neither spread alone nor strength alone is a sufficient measure for quality of quantitative information. Robustness of model output to parameter strength could be good even if parameter strength is low, provided that the model outcome is not critically influenced by the spread in that parameter. In this situation, our ignorance of the true value of the parameter has no immediate consequences because it has a negligible effect on model outputs. Alternatively, model outputs can be robust against parameter spread even if its relative contribution to the total spread in the model is high provided that parameter strength is also high. In the latter case, the uncertainty in the model outcome adequately reflects the inherent irreducible uncertainty in the system represented by the model. Uncertainty then is a property of the modeled system and does not stem from imperfect knowledge on that system. Mapping components of the knowledge base in a diagnostic diagram thus reveals the weakest spots and helps in the setting of priorities for improvement.

Figure 1 presents an example of a diagnostic diagram for emissions of acidifying substances in The Netherlands (Van Gijlswijk *et al.*, 2004). The rank correlations squared that resulted from a Monte Carlo assessment of the monitoring calculations express the sensitivity of total emission to inexactness in input data whereas strength – measured by averaged pedigree scores, in our case assuming equal weights for each criterion– expresses the methodological and epistemological limitations of the underlying

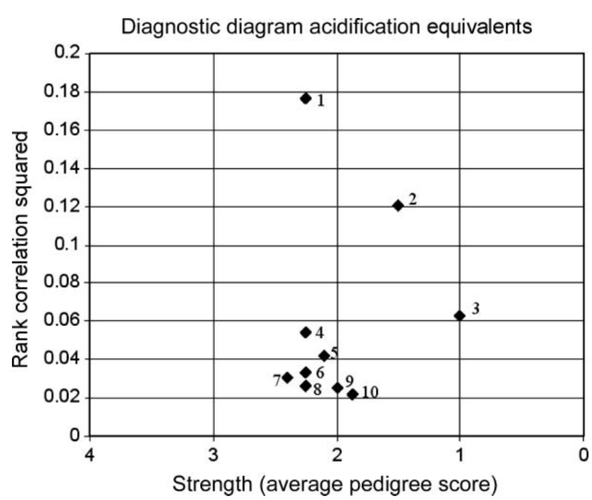

Figure 1 Example diagnostic diagram for the 10 most sensitive monitoring input data (source-activity combinations in this case) for total emission of acidification equivalents. Labels of source-activity combinations plotted: 1. NH_3 dairy cows, application of manure, 2. NO_x mobile sources agriculture, 3. NO_x agricultural soils, 4. NH_3 meat pigs, application of manure, 5. NO_x highway: gasoline personal cars, 6. NH_3 dairy cows, animal housings and storage, 7. NO_x highway: truck trailers, 8. NH_3 breeding stock pigs, application of manure, 9. NH_3 calves, yearlings, application of manure, 10. NH_3 application of synthetic fertilizer

knowledge base. One can, of course, adopt other weighting schemes to reflect relative relevance of the pedigree criteria used.

The diagram clearly identifies three source-activity combinations as the most problematic (high contribution to overall uncertainty combined with a weak knowledge base, the upper right corner of the diagram).

One reviewer of this paper raised the question whether NUSAP is more about "quantifying qualities" than about "qualifying quantities." NUSAP addresses uncertainty in outcomes of interest of scientific assessment. These outcomes of interest are usually expressed as a number and a unit, for instance "*the emission of* NH_3 *in The Netherlands in the year 2000 was 152 kton.*" NUSAP combines analysis of quantitative dimensions of uncertainty (inexactness) in such outcomes of interest with analysis of qualitative dimensions of uncertainty (unreliability, ignorance, social robustness). Pedigree analysis uses the semi-quantitative ordinal scales as defined in the pedigree matrices to elicit, score, and code qualitative dimensions of uncertainty. In the elicitation procedure, the reasoning and arguments by which experts arrive at the pedigree scores they give are documented and special attention is given to disagreement between experts on pedigree scores. The information that one gets is thus much richer than the pedigree scores alone. The use of ordinal scales makes ranking easier than would be possible if one only had the qualitative expert reasoning.

Results and discussion

The first case focused on uncertainty in quantitative environmental indicators presented in the Netherlands 5th Environmental Outlook (EO5). These indicators are based on scenario calculations with a whole chain of soft-linked computer models – varying in complexity. Many assumptions have been made in combining research results in these calculation chains, especially since the output of one computer model often does not fit the requirements of input for the next model (scales, aggregation levels).

As a test and demonstration, we applied NUSAP to two EO5 indicators: "change in length of the growth season" and "deaths and emergency hospital admissions due to tropospheric ozone." We identified implicit and explicit assumptions in the calculation chain by systematic mapping and deconstruction of the calculation chain, based on document analysis, interviews and critical review. This "assumption hunting" resulted in a list of key assumptions that was reviewed and completed in an expert workshop. Table 2 presents the pedigree matrix used in this study. In the workshop, the experts indicated on scoring cards (one card for each assumption) how they judge the assumption on the pedigree criteria and how much influence they think the assumption has on results. A moderated group-discussion takes place in which arguments for high or low scores per criterion are exchanged and discussed. In this way experts in the group remedy each other's blind spots, which enriches the quality of the individual expert judgments. We deliberately did not ask a consensus judgment of the group, because we consider expert disagreement a relevant dimension of uncertainty.

Assumptions that have at the same time a high influence on the outcomes of interest and a low pedigree can be qualified as "weak links" in the chain of which the user of the assessment results needs to be particularly aware.

Analysis of the calculation chain of the indicator "deaths and hospital admittances due to exposure to ozone" yielded a list of 24 assumptions. Fourteen key-assumptions were selected by the workshop participants as the most important ones, and prioritized. Combining the results of pedigree analysis and estimated influence in a diagnostic diagram (Figure 2) identified the following assumptions as the weakest links of the calculation chain: assumption that uncertainty in the indicator is only determined by the uncertainty

Table 2 Pedigree matrix for reviewing the knowledge base of assumptions

Score →	2	1	0
Plausibility	plausible	acceptable	fictive or speculative
Inter-subjectivity peers	many would make same assumption	several would make same assumption	few would make same assumption
Inter-subjectivity stakeholders	many would make same assumption	several would make same assumption	few would make same assumption
Choice space	hardly any alternative assumptions available	limited choice from alternative assumptions	ample choice from alternative assumptions
Influence situational limitations (time, money, etc.)	choice assumption hardly influenced	choice assumption moderately influenced	totally different assumption when no limitations
Sensitivity to view and interests of the analyst	choice assumption hardly sensitive	choice assumption moderately sensitive	choice assumption sensitive
Influence on outcomes of interest	only local influence	greatly determines the results of link in chain	greatly determines the results of the indicator

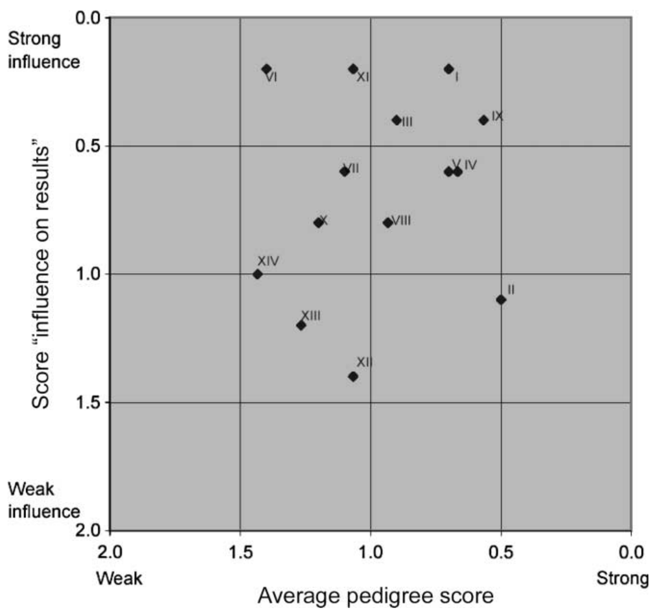

Figure 2 Diagnostic diagram for the 14 assumptions selected as important by the workshop participants (Kloprogge *et al.*, 2005). Labels of Assumptions plotted: I Uncertainty in Relative Risk (RR) dominates, II emissions abroad will not differ in EC & GC scenarios, III O_3 is homogeneously distributed over model grid cells, IV worst case metrological circumstances will not change over time, V developments in emission factors and volume growth are harmonized in European context, VI changes in the composition of the air pollution mix will not lead to changes in RR, VII linear dose-effect relationship, VIII O_3 concentration is representative for exposure, IX global background concentration of ozone is constant, X 2010 emissions abroad meet national emission ceilings, XI Dutch epidemiological data are adequate for the whole of the Netherlands, XII 2010-'30 Sectoral emissions abroad follow NL pattern, XIII direct causal relationship between ozone and death, XIV CPB scenarios were suitable for exploring NL societal-demographic developments

in the Relative Risk (RR is the probability of developing a disease in an exposed group relative to those of a non-exposed group as a function of ozone exposure) and the assumption that the global background concentration of ozone is constant over the 30 year time horizon. The full EO5 case and method for the review of assumptions is documented in Kloprogge *et al.* (2005).

The second case focused on uncertainty in environmental health risk science and policy. Near the city of Antwerp, an intense controversy has developed on the potential health effects of a waste incinerator. In a neighbourhood near the incinerator, an unusual high number of children had congenital defects. Local population and health workers pointed to the incinerator's (dioxin) emissions as the cause. The incinerator's management, supported by local authorities, deemed these accusations as "irrational, meaning purely hypothetical and not scientifically proven".

Following years of heated debate, involving citizens' committees, policymakers (both local and regional) and scientific experts, the conflict evolved to a phase in which all parties realised that a business as usual style will not work any longer. This led to the establishment of the Flemish Centre of Expertise on Environment and Health (CEEH) and initiatives to renew interactions between science, policy and society (Keune *et al.*, 2003).

Against this background, a workshop was held to explore how NUSAP Pedigree schemes can support and structure deliberations on uncertainties in environmental risk assessment (Craye *et al.*, 2004). The workshop involved experts and actors directly involved and external experts from the Netherlands Environmental Assessment Agency

(RIVM) and representatives of stakeholders in environmental health issues. The workshop session discussed here focused on an epidemiological study that had been used in the sociopolitical discussions on the incinerator's impact on the environment and local health. The study investigated whether there were increased health risks among children whose parents lived or had lived in the particular neighbourhood (Aelvoet *et al.*, 1998). The tailored pedigree matrix that was designed to structure the deliberative uncertainty assessment of this study is presented in Table 3.

Through analysis of the study reports and interviews with their authors, the main phases in knowledge production to be covered in the pedigree assessment were devised. The choice of these phases reflects the complementarity between more cognitive and more social aspects. A phase related to problem framing was explicitly added. In this way, a discussion was triggered on the "status" of the used problem definition in relation to other disciplinary framings and socio-political perspectives and on the "process" through which the expert framing and other socio-political framings had (not) been matched.

The other phases had already been used in earlier applications of the pedigree scheme (Funtowicz and Ravetz, 1990). "Data-definitions" relates to all those decisions, logically prior to actual enumeration, concerning the establishment of the relevant conceptual objects and the set of operating procedures. It deals with the question of how is defined what has to be enumerated and how the way to do the counting is determined. "Data-collection" deals mostly with the more technical aspect of quality of databases in function of what has to be enumerated. Quality has also to be seen as being influenced by the culture of the institute which is charged with the data collection.

"Analysis" highlights the reliability of the chosen method of analysis in function of how the problem was approached. It focuses on the scientific discussion about the choice of methods and on possible "extra-scientific" reasons for preference for certain methods. "Review" is related to the process of quality control of the research with particular attention to its interactions with the socio-political context in which the research took place.

The modes for each pedigree criterion referred to how the process to deal with that phase was run. A higher ranked mode had to be interpreted as a better chance to increase the quality of the study along three dimensions that were deemed essential to achieve "results fit for function" ("a serviceable truth"):

- scientific robustness, in a perspective of constructive knowledge development;
- practical functioning of the science–policy interface to create a social basis for the policy measures and their knowledge base;
- normative considerations on a legitimate knowledge support, i.e., taking account of plural perspectives and rendering agency to all actors involved.

The hierarchical ranking of the modes was sometimes a subject of discussion, which was not a surprise as "quality" is a generic concept (on which all agree), whose dimensions and concrete meaning (on which there is no a priori consensus) as a standard in a particular context can only be filled in through deliberation (Craye *et al.*, 2004).

Table 3 Pedigree matrix for the epidemiological study

Score	Problem framing	Data – definitions	Data – collection	Analysis	Review
4	Negotiation	Negotiation	Task Force	Established	Extended
3	Scientific	Science	Direct	Discussion	External
2	Compromise	Pragmatic	Bureaucratic	Competition	Independent
1	Inertia	Symbolic	Indirect	Embryonic	Internal
0	Controversy	Unknown	Fiat	No info	None

The discussions were shaped as a reasoned, structured debate focusing on underlying assumptions and frame-dependent choices in the different studies. In each session, a discussion leader had to prevent only technical features of uncertainty being covered. The protocol followed in the three sessions is elaborated in more detail by Craye *et al.* (2004).

The aggregation level chosen for the assessment by the panel was the "study as a whole." Given that high aggregation level, it was not feasible to display on a disaggregated level all the types of uncertainty involved. Two illustrative critical aspects were presented for each of the phases. These related to choices that had been made in the framing and the design of the study, and subsequently had been criticized by other experts and relevant actors (e.g., under the phase "data-definitions": "how to define the exposed population?"). Also included were other aspects that had not been openly debated in the past but could have led to a more reflexive knowledge development had they been approached with openness (e.g., under the phase "data-definition": "who is competent to define a congenital defect? a family doctor, a parent, a professor in epidemiology, an operator of a database… ?").

Two experts – the author of the relevant study and an "opponent" or "critical judge" – introduced each topic. Then, the discussion extended to include the views and reactions of the stakeholders, citizens, and policymakers in the panel. The session leader and another social scientist had to guarantee an informed and fair debate took place. He also had a list of possible questions in order to (re)focus the discussion if necessary. These model questions were based on insights on the structure of argumentations (Toulmin, 1958; van de Graaf, 2000), the content of actors' frames of meaning (Grin *et al*., 1997) and the different types of scientific debate and controversy when uncertainty is salient (Von Schomberg, 1997). In these ways, the protocol enhanced the reflexivity of the process, both in terms of content, i.e. opening up the problem definition and the scope of argumentation, and in terms of process, i.e. placing the participants in new roles and rules of interaction. This setting challenged the traditional division between the scientist as a provider of facts versus policymakers and the public as defenders of values. The questions promoted discussion on the validity of assumptions, which could reveal particular biases in the framing of the risk. They were intended to deliver insight in the deeper debate on plausible hypotheses, distinguishing it from the more factual discussions on the empirical basis and the methodological work. Included were cycles of typical why?-questions, e.g. "What is the right (research) approach to this problem? …..Why is this the adequate approach (asks for the definition of the (research) problem)? ….Why do you define the (research) problem in this way (asks for underlying and supporting 'theories")? ….Why do you use these theories in this case (asks for the fundamental features of framing, the preferences and convictions)?"

The discussion of each pedigree phase was concluded by giving a score. The scoring was a collective exercise of deliberation, which enabled the main points of discussion to be summarized, to explain disagreement and to clarify any ambiguity in the pedigree scheme. The resulting pedigree score for epidemiological study was (1-2, 1-2, 2, 2-3, 0). The notation n-m signifies disagreement in the group.

The low score given to the epidemiological study was consistent with its failure to deliver robust insights and to play a relevant role in the policy debate at the time its results were communicated. In short, it was consistent with its quality for use in the socio-political context.

Through discussing the critical aspects for each phase in the pedigree scheme and through providing and sharpening argumentation for specific pedigree scores, insight was gained in the reasons why the study was not fit for purpose and in the dynamics of controversy to which the study contributed. Themes discussed through the pedigree assessment included the incongruity of the epidemiological and socio-political framings, the plurality of disciplinary framings, the relation between the scientific problem and the socio-political issue, the value loadings of the research methods, and issues of agency in knowledge production and review.

In this sense, the session promoted reflexivity and collective learning. It showed the potential of pedigree assessment to foster a deeper debate on problem framing and institutional aspects of knowledge development, possibly influencing the positions and perspectives held by different actors, their mutual expectations and the future design of policy relevant research.

Whereas the problem definition used in the epidemiological study and the choice of data sources and methods of data collection had been intensively discussed between the research team and the Ministry of Public Health, the resulting framing failed to address the concerns of the local population and was quite meaningless from the perspective of the incinerator's management.

The reactions on this framing ranged from "an inadequate use of epidemiology" to "a complete irrelevance of the epidemiological approach." The problem definition used, implicitly called into question the existence of the cluster of congenital diseases in the neighbourhood by statistically testing the significance of these diseases' incidence in the area, compared to the whole Flemish region. Opponents of the study argued a more correct and relevant use of epidemiology would have been to test the relation between these diseases and possible causing factors.

It turned out that during the discussions on the other phases the participants often referred to the frame-dependency of certain choices, thus confirming the crucial importance of problem frames. This suggested that in the still emerging environmental health science, ignorance and indeterminacy are the predominant forms of uncertainty, largely outweighing in importance methodological and technical aspects.

Overall, the session confirmed the centrality of the issue of framing in this kind of environmental health risk assessment. Participants took more time to discuss the framing than any other phase. The non-scientists also felt that their contribution was most relevant with respect to framing and felt less need to intervene in the more "technical" phases dealing with the choices of data sources and of methods. However, they remained very interested and followed with attention the expert discussions on these issues.

The session raised awareness about the complexity of the issues to be studied and the resulting inherent uncertainty and ignorance. As participants learned that choices and assumptions could not be based exclusively on objective science, questions were raised about who is competent and "entitled" to make the necessary choices.

Many participants suggested the approach could be applied in a constructive way, i.e., when policy supporting research is being developed, thus contributing to a negotiated management of environmental health risks. However, others argued that the method still was too science-centered, thereby devaluing the contributions by citizens and other lay knowledge providers. The lessons learnt during the workshop are being used to develop a set of pedigree schemes that can be deployed in distinct processes dealing with framing, research design, and extended review. The experience also points to the need to reflect on the integration of these processes and their results in an overall inclusive approach.

Conclusions

We have presented experiences and results with the NUSAP method for multidimensional uncertainty assessment in cases that vary in complexity. The cases have shown that the NUSAP method, and especially the pedigree analysis part of it, is useful not to only assess parameter uncertainty, but also to systematically reflect upon (model) assumptions and problem frames. A diagnostic diagram synthesizes results of quantitative and qualitative analysis and helps to prioritize uncertainties. Extension of the pedigree scheme to include societal dimensions of uncertainty, such as problem framing and value loadings, further promotes reflexivity and collective learning. The task of quality control in the knowledge base of complex and controversial (environmental) policy issues is a complicated one and the NUSAP method disciplines and supports this process by facilitating and structuring a creative reflexive process and in-depth review of the limitations of a given knowledge base. NUSAP promotes explicitation of and reflection on the various dimensions of uncertainty. It provides a diagnostic tool for assessing the robustness of a given knowledge base for policymaking and promotes criticism by users of all sorts, expert and lay, and will thereby support extended peer review processes. Similar to a patient information leaflet alerting the patient to risks and unsuitable uses of a medicine, NUSAP enables the delivery of policy-relevant quantitative information together with the essential warnings on its limitations and pitfalls. It thereby promotes the responsible and effective use of the information in policy processes.

References

Aelvoet, W., Nelen, V., Schoeters, G., Vanoverloop, J., Vlietinck, R. and Wallijn, E. (1998). *Risico op gezondheidsschade bij kinderen van de Neerlandwijk te Wilrijk*, Vito-onderzoeksrapport 1998/TOX/R/030, Mol, Belgium.

Craye, M., Goorden, L., Vandenabeele, J. and Van Gelder, S. (2001). *Milieu en gezondheid in Vlaanderen: naar een adequate dialoog tussen overheid, bevolking en wetenschap*, UFSIA-STEM onderzoeksrapport, Antwerpen, Belgium.

Craye, M., Funtowicz, S. and Van der Sluijs, J. (2004). A reflexive approach to dealing with uncertainties in environmental health risk science and policy. *Int. J. Risk Assessment and Management*, **5**(2).

Ellis, E., Gang Li, R., Zhang Yang, L. and Cheng, X. (2000). Long-term change in village-scale ecosystems in China using landscape and statistical methods. *Ecological Applications*, **10**(4), 1057–1089.

Funtowicz, S. and Ravetz, J. (1990). *Uncertainty and Quality in Science for Policy*, Kluwer, Dordrecht.

Funtowicz, S. and Ravetz, J. (1993). Science for the Post-Normal age. *Futures*, **25**(7), 735–755.

Grin, J., Graaf van der, H. and Hoppe, R. (1997). *Technology Assessment through Interaction – a guide*, Rathenau Instituut, Den Haag, The Netherlands.

Keune, H., Goorden, L., Mertens, R. and Loots, I. (2003). *Communicatie, Interactie en Reflectie over Biomonitoring* Nota Opzet Communicatiestrategie, Lokale contacten en Reflectie. Biomonitoringcampagne, UA-Steunpunt Milieu & Gezondheid, Luik Sociaal en Gezondheidseconomisch Onderzoek, Luik, Belgium.

Kloprogge, P., Van der Sluijs, J.P. and Petersen, A. (2005). *A method for the analysis of assumptions in assessments applied to two indicators in the fifth Dutch Environmental Outlook*, Research report, Department of Science Technology and Society, Utrecht University, Utrecht.

Nowotny, H. (1999). The place of people in our knowledge. *European Review*, **7**(2), 247–262.

Risbey, J., Van der Sluijs, J. and Ravetz, J. (2001). *Protocol for Assessment of Uncertainty and Strength of Emission Data*, Report E-2001-10 Department of Science Technology and Society, Utrecht University, Utrecht, The Netherlands. Available at www.nusap.net.

Saltelli, A., Chan, K. and Scott, M. (2000). *Sensitivity Analysis*, John Wiley & Sons publishers, Probability and Statistics series.

Saltelli, A., Tarantola, S., Campolongo, F. and Ratto, M. (2004). *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*, John Wiley & Sons publishers.

Toulmin, S.E. (1958). *The uses of argument*, Cambridge University Press, Cambridge, UK.

Van der Sluijs, J. (1997). *Anchoring amid uncertainty; On the management of uncertainties in risk assessment of anthropogenic climate change*. PhD thesis, Utrecht University, Utrecht, The Netherlands.

Van der Sluijs, J. (2002). A way out of the credibility crisis of models used in integrated environmental assessment. *Futures*, **34**, 133–146.

Van der Sluijs, J., Potting, J., Risbey, J., Van Vuuren, D., De Vries, B., Beusen, A., Heuberger, P., Corral Quintana, S., Funtowicz, S., Kloprogge, P., Nuijten, D., Petersen, A. and Ravetz, J. (2002). Uncertainty assessment of the IMAGE/TIMER B1 CO_2 emissions scenario, using the NUSAP method, NRP Report 410 200 104, Bilthoven, The Netherlands. Available at www.nusap.net.

Van der Sluijs, J., Risbey, J. and Ravetz, J. (2005). Uncertainty Assessment of VOC emissions from Paint in The Netherlands, *Environmental Monitoring and Assessment*, **105**, 229–259.

Van Gijlswijk, R., Coenen, P., Pulles, T. and Van der Sluijs, J. (2004). *Uncertainty assessment of NOx, SO2 and NH3 emissions in the Netherlands*, TNO and Copernicus Institute Research Report R2004/100, Apeldoorn, The Netherlands (www.nusap.net).

Von Schomberg, R. (1997). *Argumentatie in de context van een wetenschappelijke controverse*, Universiteit Twente WMW-publicatie 27, Eburon, Delft.