

29

THE NUSAP APPROACH TO UNCERTAINTY APPRAISAL AND COMMUNICATION

Jeroen P. van der Sluijs

Introduction

The knowledge base available for decision-making on contemporary environmental and sustainability issues is often characterised by an imperfect understanding of the complex systems involved. Decisions will need to be made before conclusive scientific evidence is available, while at the same time the potential error costs of wrong decisions can be huge. This societal context of knowledge production and use for decision-making and risk management implies an urgent need for explicit appraisal and consideration of all dimensions of scientific uncertainty (Funtwicz and Ravetz, 1990, 1993; Van der Sluijs, 2002; Van der Sluijs et al., 2008, Saltelli et al., 2013); [see also Chapters 26–28].

The transdisciplinary nature of science for sustainability poses additional requirements with regard to the systematic analysis, documentation and communication of uncertainty. When quantitative information is produced in one disciplinary context and used in another, we often see that important caveats tend to be ignored, uncertainties compressed and numbers used at face value (Wynne, 1992). Knowledge utilisation for sustainability issues requires a full and public awareness of the various sorts of uncertainty and underlying assumptions. Knowledge needs to be robust both technically and socially (Nowotny, 1999).

The past record of science for policy has shown that omitting uncertainty assessment and communication can undermine public trust in the science (e.g. Keepin and Wynne, 1984). An example concerns the Netherlands National Institute for Public Health and the Environment (RIVM). In early 1999, De Kwaadsteniet, a senior statistician, accused the RIVM of “lies and deceit” in their State of the Environment Reports and Environmental Outlooks. In a quality newspaper (*Trouw*) he criticised RIVM for basing their studies on the ‘virtual reality’ of poorly validated computer models, while RIVM presented these results as point values with unwarranted significant digits and without elaborating the uncertainties. A vehement public debate was triggered on the credibility and reliability of environmental numbers and models. The case got front page and prime time coverage in the mass media and led to a debate in the Netherlands parliament (Van der Sluijs, 2002). The RIVM went through a learning process that led to the development of guidance for uncertainty management in the institute (Van der Sluijs et al., 2008; Petersen et al., 2011).

In this chapter I will explain how such incidents have led to a revision of the approach to uncertainty. The next section outlines how the understanding of uncertainty has changed and the

implications this has for policy. This background explains why new transdisciplinary methods for evaluating uncertainty have been developed. The chapter then focuses on the Numeral, Unit, Spread, Assessment and Pedigree (NUSAP) framework and explains its key features. Pedigree is given particular attention as the innovative aspect of this post-normal science system [see Chapter 28].

Understanding uncertainty

In the early phase of its development, the field of uncertainty analysis mainly evolved around mathematical statistical methods such as sensitivity analysis and Monte Carlo techniques for the assessment of error propagation in model calculations. These tools address quantitative dimensions of uncertainty using sophisticated algorithms (Saltelli et al., 2000, 2008). Although these quantitative techniques are essential in any uncertainty analysis, they provide only a partial insight into what usually is a very complex mass of uncertainties involving technical, methodological, epistemological and societal dimensions. Quantitative methods can however be complemented with new qualitative approaches addressing aspects of uncertainty that are hard to quantify and were therefore largely under addressed in the past. In their combination, the quantitative and qualitative methods provide a richer diagnosis of uncertainty than each of these methods alone.

Over the past decades, an increasing body of conceptual and theoretical work in the field of uncertainty management has been compiled. Key insights from the field include:

- Uncertainty is partly socially constructed and its assessment always involves subjective judgement;
- More research does not necessarily reduce uncertainty, it often reveals unforeseen complexities;
- Some uncertainty is irreducible (intrinsic or practically);
- High quality scientific knowledge for policy making does not require low uncertainty;
- Uncertainty is a multi-dimensional concept involving quantitative (technical inexactness) and qualitative dimensions (i.e., methodological unreliability, epistemological ignorance and societally limited robustness), and it can manifest itself at different locations (e.g., context, indicator choice, model structure, parameters and data).
- In problems that are characterised by high systems uncertainties, ignorance, and high decision stakes the qualitative dimensions of uncertainty may well dominate the quantitative dimensions.

Most of present day methodologies and practices for addressing uncertainty focus exclusively on quantitative uncertainty in model parameters and input data. Methods to address qualitative and societal dimensions of uncertainty are absent or in their early stage of development. Uncertainties relating to model structure, model assumptions and model context are largely ignored.

Scientists, policymakers and stakeholders now widely hold that uncertainty management in environmental assessment is essential. However, in the practice of uncertainty management there is little appreciation for the fact that uncertainty is more than a number. There are many different dimensions of uncertainty and there is a lack of understanding about their different characteristics, relevance and relative importance. Even within the different fields of decision support (such as integrated assessment, environmental risk assessment, environmental impact assessment, policy analysis, engineering risk analysis and cost-benefit analysis), there is neither a commonly shared terminology nor agreement on a generic typology of uncertainties (Walker et al., 2003).

A better understanding of the various dimensions of uncertainty is needed in order to provide an improved theoretical foundation for uncertainty assessment. Improved

conceptualisation of uncertainty is desirable for a number of reasons. First, it will aid better communication amongst the many disciplines involved. In the current situation, different analysts use different terms for the same kinds of uncertainty, and some use the same term to refer to different kinds. This makes it difficult for those who have not participated in the actual work to understand what has been done. Improved conceptualisation of uncertainty will further provide for better communication among scientists, policymakers and stakeholders. A common belief is that policymakers expect scientists to provide certainties and hence dislike uncertainty in the scientific knowledge base. However, uncertainty is a fact of life and a better understanding of its key dimensions and their implications for policy choices would be likely to lead to more trust in the scientists providing decision support, and ultimately to better policies. Finally, a better understanding of the different dimensions of uncertainty and their potential impact on the relevant policy issues at hand would help in identifying and prioritising effective and efficient research and development activities for improving the knowledge base.

Van der Sluijs (1997) concludes that in the practice of uncertainty management in integrated modelling of climate change, major gaps exist in the systematic analysis of unreliability of the knowledge about input data, model parameters and model assumptions, and also in the analysis of uncertainty about model structure. A major obstacle is that tools for assessing these types of uncertainty and how these might affect the outcomes of assessments, are either not available or in their early stage of development. Only recently have new tools been developed that focus on the qualitative (methodological and epistemological) dimensions of uncertainty using methods of expert elicitation, quality assistance checklists (Risbey et al., 2005), Pedigree analysis (using multiple criteria [Chapter 30]) assessing the strength of various underpinning components of the knowledge base by self-review, peer review or extended peer review (Funtowicz and Ravetz, 1990; Van der Sluijs et al., 2005a, 2005b, 2005c), and methods for the systematic identification and characterisation of critical assumptions in models (Kloprogge et al., 2011).

Addressing the multiple dimensions of uncertainty

Whereas quantitative methods are well developed, standardised and supported by handbooks (Morgan and Henrion, 1990, Saltelli et al., 2000, 2008) and software (@Risk, Crystal ball, Simlab, Analytica), qualitative and multi-dimensional methods have been demonstrated and tested but have not yet been widely disseminated and adopted. Multi-dimensional methods are those that address qualitative and quantitative aspects in a coherent way. They do so by assessing the technical (inexactness), methodological (unreliability), societal (social robustness), and epistemological (border with ignorance) dimensions of uncertainty, as shown in Table 29.1.

Numeral, Unit, Spread, Assessment, Pedigree (NUSAP)

NUSAP is a notational system, proposed in the context of post-normal science by Funtowicz and Ravetz (1990), which aims to provide an analysis and diagnosis of uncertainty in science for policy. The NUSAP system structures the systematic appraisal and communication of the various dimensions of uncertainty. It provides an heuristic for good practice addressing uncertainty in quantitative information. NUSAP has extended the statistical approach to uncertainty with methodological and epistemological dimensions by adding expert judgement of reliability (Assessment) and systematic multi-criteria evaluation of the underpinning of numbers (Pedigree). Examples of Pedigree criteria are empirical basis, methodological rigour, theoretical understanding, degree of validation and peer acceptance.

Table 29.1 Dimensions of uncertainty

<i>Dimension</i>	<i>Type</i>	<i>Can stem from or can be produced by</i>
Technical	Inexactness	<i>Intrinsic uncertainty:</i> Variability; stochasticity; heterogeneity <i>Technical limitations:</i> Error bars, ranges, variance; Resolution error (spatial, temporal); Aggregation error; Linguistic imprecision, unclear definitions
Methodological	Unreliability	<i>Limited internal strength of the knowledge base in:</i> Use of proxies; Empirical basis; Theoretical understanding; Methodological rigour (including management of anomalies); Validation
Epistemological	Ignorance	<i>Limited theoretical understanding</i> <i>System indeterminacy:</i> Open-endedness of system under study; Chaotic behaviour <i>Intrinsic unknowability with active ignorance:</i> Model fixes for reasons understood; Limited domain of validity of assumptions; Limited domains of applicability of functional relations; Numerical error; Surprises type A (some awareness of possibility exists) <i>Intrinsic unknowability with passive ignorance:</i> Bugs (software error, hardware error, typos); Model fixes for reasons not understood; Surprises type B (no awareness of possibility)
Societal	Limited social robustness	<i>Limited external strength of the knowledgebase in:</i> Completeness of set of relevant aspects; Exploration of rival problem framings; Management of dissent; Extended peer acceptance/stakeholder involvement; Transparency; Accessibility <i>Bias/Value ladenness:</i> Value laden assumptions; Motivational bias (interests, incentives); Disciplinary bias; Cultural bias; Choice of (modelling) approach (e.g. bottom up, top down); Subjective judgement

The NUSAP framework provides a means for synthesis and integration of findings on each of these dimensions, combining formal Monte Carlo and mathematical sensitivity analysis techniques with systematic qualitative uncertainty assessment. NUSAP enables providers and users of knowledge to be clear and transparent about its various uncertainties. This promotes critical reflection on the strengths and weaknesses of the underlying knowledge base by users of all sorts (e.g., experts, lay public) and thereby supports an extended peer review process. It aims to provide those who produce, use and are affected by policy-relevant knowledge with a set of diagnostic tools for a critical self-awareness of their engagement with that knowledge.

NUSAP extends the statistical approach to uncertainty (inexactness) by incorporating the methodological (unreliability) and epistemological (ignorance) dimensions using expert judgement of reliability (Assessment) and systematic multi-criteria evaluation of the process by which numbers are produced (Pedigree). Numbers are provided with a separate qualification for each dimension of uncertainty, allowing nuances of meaning about quantities to be conveyed concisely and clearly, to a degree that is quite impossible with reliance on statistical methods alone. NUSAP captures both quantitative and qualitative dimensions of uncertainty and enables one to display these in a standardised and self-explanatory way. The basic idea is to qualify quantities using the five aspects of the NUSAP system: (i) Numeral, (ii) Unit, (iii) Spread, (iv) Assessment and (v) Pedigree. Each of these dimension, or numeric qualifiers, is discussed in turn.

First is the Numeral, which is normally an ordinary number, but, when appropriate, 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). Second is the Unit, which may be of the conventional sort, but which may also contain extra information, such as the date on which the Unit is evaluated (e.g. a common qualification for monetary values subject to inflation). The third category is Spread, which is a generalisation from the 'random error' of experiments or the variance of statistics. Although Spread is usually conveyed by a number (either +, % or 'factor of') it is not an ordinary quantity, because its own inexactness is of a different sort from that of measurements. Methods to address Spread can be statistical data analysis, sensitivity analysis or Monte Carlo analysis, possibly in combination with expert consultation.

The remaining two qualifiers constitute the more qualitative side of the NUSAP framework. 'Assessment' expresses qualitative judgements 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 qualities of optimism or pessimism. In some experimental fields, information is supplied qualified by two + terms, of which the first is the Spread, or random error, and the second is the systematic error which must be estimated on the basis of the history of the measurement, and which corresponds to the use of Assessment in NUSAP. A frequently observed pitfall is to wrongly think that systematic error must always be less than any experimental error, or else a stated error bar would be meaningless or misleading. However, in many real life cases systematic error can be well estimated only in retrospect, and then it can produce surprising results that are far outside the error bar of the previously published number(s).

The fifth and final aspect of NUSAP is the Pedigree. This 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 as a set of criteria and assessed using qualitative expert judgement. Risbey et al. (2001) document a method to draft Pedigree scores by means of expert elicitation, on which Knol et al. (2010) provide guidance as to good practice. Arbitrariness and subjectivity in measuring strength are minimised by using a Pedigree matrix to code qualitative expert judgements for each criterion into a discrete Numeral scale from 0 (weak) to 4 (strong) accompanied by linguistic descriptors or modes.

Pedigree and its assessment

Each special sort of information has its own aspects that are key to its Pedigree, so different Pedigree matrices using different criteria can be used to qualify different sorts of information. An overview of the literature on Pedigree matrices and examples of questionnaires for eliciting Pedigree scores is available online at <http://www.nusap.net>. Ellis et al. (2000) have developed a Pedigree calculator to assess propagation of Pedigree in a calculation in order to establish Pedigree scores for quantities calculated from other quantities. Table 29.2 gives an example of a Pedigree matrix for emission monitoring data. Next I will briefly elaborate the four criteria employed.

Proxy

Sometimes measuring the thing we are interested in directly, or representing it by a parameter, is impossible, so some form of proxy measure is used. Proxy refers to how good or close a measure of the quantity that we measure or model is to the actual quantity we seek or represent. Examples are first order approximations, over simplifications, idealisations, gaps in aggregation levels, differences in definitions, non-representativeness and incompleteness issues.

Table 29.2 Example pedigree matrix for emission monitoring data

Score	Proxy	Empirical basis	Methodological rigour	Validation
4	An exact measure of the desired quantity	Controlled experiments and large sample direct measurements	Best available practice in well-established discipline	Compared with independent measurements of the same variable over long domain
3	Good fit or measure	Historical/field data uncontrolled experiments small sample direct measurements	Reliable method common within est. discipline Best available practice in immature discipline	Compared with independent measurements of closely related variable over shorter period
2	Well correlated but not measuring the same thing	Modelled/derived data Indirect measurements	Acceptable method but limited consensus on reliability	Measurements not independent proxy variable limited domain
1	Weak correlation but commonalities in measure	Educated guesses indirect approx. rule of thumb est.	Preliminary methods unknown reliability	Weak and very indirect validation
0	Not correlated and not clearly related	Crude speculation	No discernible rigour	No validation performed

Source: Risbey et al. (2001) adapted from Ellis et al. (2000a, 2000b).

Empirical basis

This typically refers to the degree to which direct observations, measurements and statistics are used to estimate a parameter. Sometimes directly observed data are unavailable, and the parameter or variable is estimated based on partial measurements or calculated from other quantities. Parameters or variables determined by such indirect methods have a weaker empirical basis and will generally score lower than those based on direct observations.

Methodological rigour

Parameter or variable estimates employ a method to collect, check, and revise the data. Methodological quality refers to the norms for methodological rigour in this process, as applied by peers in relevant disciplines. Well-established and respected methods for measuring and processing data would score high on this metric, while untested or unreliable methods would tend to score low.

Validation

This metric refers to the degree to which the analyst has been able to cross-check the data and assumptions used to produce the Numeral of the parameter against independent sources. In many cases, independent data for the same parameter over the same time period are unavailable and other data sets must be used for validation. This may require a compromise in the length or overlap of the data sets, or may require use of a related, but different, proxy variable for indirect

validation, or perhaps use of data that has been aggregated on different scales. The more indirect or incomplete the validation, the lower it will score on this metric.

Visualising Pedigree analysis

In general, Pedigree scores will be established using expert judgements from more than one expert. Two ways of visualising results of a Pedigree analysis are discussed here: radar diagrams and kite diagrams (Risbey et al., 2001; Van der Sluijs et al., 2002), as exemplified in Figure 29.1. Both representations use polygons with one axis for each criterion, having 0 in the centre and 4 on each corner point. In the radar diagrams a coloured line connecting the scores represents the scoring of each expert, whereas a black line represents the average scores.

The kite diagrams follow a traffic light analogy. The minimum scores by a group of experts for each Pedigree criterion span the green kite; the maximum scores span the amber kite. The remaining area is red. The width of the amber band represents expert disagreement on the Pedigree scores. In some cases the size of the green area can be strongly influenced by a single deviating low score given by one of the experts. In those cases the light green kite shows what the green kite would look like if that outlier had been omitted. Note that the algorithm for calculating the light green kite is such that outliers are evaluated per Pedigree criterion, so that outliers defining the light green area need not be from the same expert. A web-tool to produce kite diagrams is available from <http://www.nusap.net>.

The kite diagrams can be interpreted as follows: the green coloured area reflects the (apparent minimal consensus) strength of the underpinning of each parameter. The more green, the stronger is the underpinning. The orange coloured zone shows the range of expert disagreement on that underpinning. The remaining area is red. The more red, the weaker is the underpinning (all according to the assessment by a group of experts). A kite diagram captures the information from all experts in the group without the need to average expert opinion. Averaging expert opinion is a controversial issue in elicitation methodologies (Knol et al., 2010). Another advantage is that it provides a fast and intuitive overview of parameter strength, preserving the underlying information. However, kite diagrams can be misleading because the amount of red and green surface area can be sensitive to the order of the criteria in the diagram. As an alternative, bar charts can be used with error-bars to reflect the range of expert opinion; see Klopogge et al. (2007) and Wardekker et al., (2008) for further guidance.

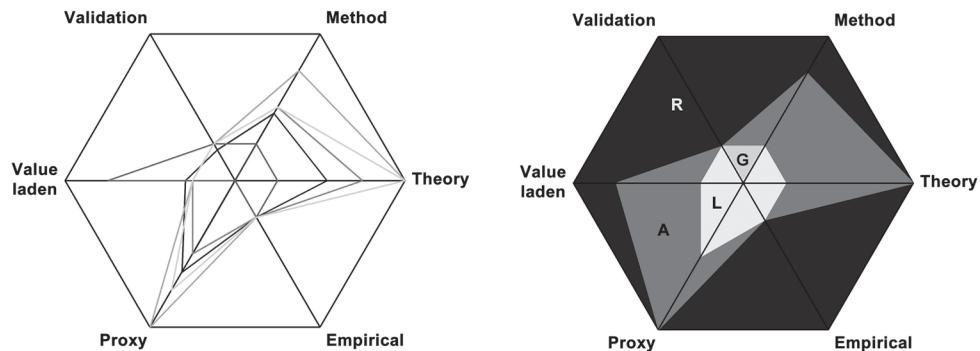


Figure 29.1 Graphic representation of pedigree scoring.
 Notes: Example of representations of same pedigree results as scored by six different experts by radar diagram (left) and kite diagram (right) G=green, L=light green, A=amber, R=red.
 Source: Van der Sluijs et al (2002).

Diagnostic Diagrams

There are two independent metrics that can be used for diagnostic purposes. First is the method chosen to address the Spread qualifier (typically sensitivity analysis or Monte Carlo analysis) which provides, for each input quantity, a quantitative metric of uncertainty contribution, or sensitivity (e.g., the relative contribution to the variance in a given model output). Second are the Pedigree scores that can be aggregated (by dividing the sum of the scores of the Pedigree criteria by the sum of the maximum attainable scores) to produce a metric for parameter strength. These two independent metrics can be combined in a NUSAP Diagnostic Diagram (see Figure 29.2).

The Diagnostic Diagram is based on the notion that neither Spread nor strength can alone provide a sufficient measure of quality. 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 a parameter has no immediate consequences because it has a negligible effect on calculated model outputs. Alternatively, model outputs can be robust against parameter Spread—even if its relative contribution to the total Spread in a model is high—provided that parameter strength is high. In the latter case, the uncertainty in the model outcome adequately reflects the inherent irreducible uncertainty in the system represented by the model. In other words, the uncertainty then is a property of the modelled system and does not stem from imperfect knowledge about that system. Mapping model parameters in the Assessment diagram thus reveals the weakest critical links in the knowledge base of the model with respect to the model outcome assessed, and helps in setting the priorities for model improvement.

Most of the Pedigree assessments in the literature have addressed uncertainties located in inputs and parameters, thereby focussing on the internal strength of the knowledge base. Kloprogge et al. (2011) extended Pedigree analysis to assess assumptions in models. Examples of putting the approach into practice include Laes et al. (2011) for evaluating the external costs of nuclear energy, De Jong et al. (2012) for quantified health risks of overhead power lines, and Boone et al. (2009) and Bouwknegt et al. (2014) for quantitative microbial risk assessment. Van der Sluijs et al. (2015) have further extended the application of NUSAP to assess modelling assumptions in a chain of integrated models in the context of decisions concerning local adaptation to climate change impacts.

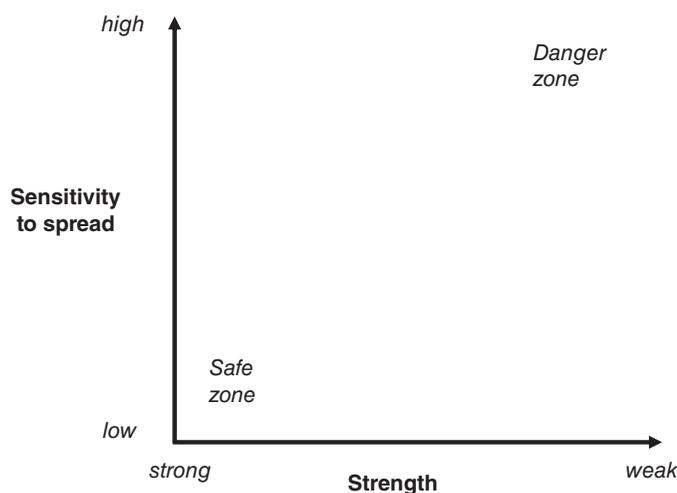


Figure 29.2 NUSAP diagnostic diagram.

Corral (2000), in his Ph.D. thesis, extended the Pedigree scheme to address uncertainties located in the socio-political context, focussing on the external strength of the knowledge base (i.e., its relationship to the world outside of science). The criteria that Corral used to assess the Pedigree of the processes of knowledge utilisation and institutional context of the analysts were inter alia: accessibility, terminology, completeness, source of information, verification, colleague consensus, extended peer acceptance, legitimation, experience, and flexibility.

Future directions

The NUSAP approach has a great potential to systematise the appraisal and consideration of uncertainty at the science-governance interface. Further tailoring and standardisation of Pedigree matrices and procedures for the elicitation of Pedigree scores is desirable but the main challenge is in dissemination. Successful pilots with inclusion of NUSAP in M.Sc. and Ph.D. teaching curricula at the universities of Utrecht, Bergen and Versailles Saint-Quentin-en-Yvelines can be scaled up. An open access course in knowledge quality assessment is also now available online (<https://proxy.eplanete.net/galleries/broceliande7/KQA>).

Concluding remarks

Overall, NUSAP has a strong foundation in the theory of knowledge and the philosophy of science connecting to post-normal science [see Chapter 28]. It provides a framework to systematically and coherently address and communicate three of the dimensions of uncertainty, namely: technical (inexactness), methodological (unreliability) and epistemological (bordering with ignorance). It provides a framework for synthesising qualitative and quantitative assessments of uncertainty and can act as a bridge between the quantitative mathematical disciplines/traditions and the qualitative discursive and participatory disciplines/traditions in the field of uncertainty management.

Key further readings cited

- Funtowicz, S.O. and Ravetz, J.R. (1990). *Uncertainty and Quality in Science for Policy*. Dordrecht: Kluwer.
- van der Sluijs, J.P., Craye, M., Funtowicz, S.O., Kloprogge, P., Ravetz, J.R. and Risbey, J. (2005a). Combining quantitative and qualitative measures of uncertainty in model based environmental assessment: the NUSAP system. *Risk Analysis* 25(2), 481–492.
- van der Sluijs, J.P., Risbey, J. and Ravetz, J.R. (2005b). Uncertainty assessment of VOC emissions from paint in the Netherlands. *Environmental Monitoring and Assessment* 105, 229–259.
- van der Sluijs, J.P., Petersen, A.C., Janssen, P.H.M., Risbey, J.S. and Ravetz, J.R. (2008). Exploring the quality of evidence for complex and contested policy decisions. *Environmental Research Letters*, 3, 024008 (9 pp.).
- van der Sluijs J.P. and Wardekker, J.A. (2015). Critical appraisal of assumptions in chains of model calculations used to project local climate impacts for adaptation decision support: The case of Baakse Beek. *Environmental Research Letters* 10, 045005.

Other literature cited

- Boone, I., Van der Stede, Y., Bollaerts, K., Vose, D., Maes, D., Dewulf, J., Messens, W., Daube, G. Aerts, M. and Mintiens, K. (2009). NUSAP method for evaluating the data quality in a quantitative microbial risk assessment model for Salmonella in the pork production chain. *Risk analysis* 29(4), 502–517.
- Bouwknegt, M., Knol, A.B., van der Sluijs, J.P., Evers, E.G. (2014). Uncertainty of population risk estimates for pathogens based on QMRA or epidemiology: a case study of *Campylobacter* in the Netherlands. *Risk analysis*, 34(5), 847–864.

- Corral Quintana, S.A., (2000). Una Metodología integrada de exploración y comprensión de los procesos de elaboración de políticas públicas. Ph.D. thesis, University of La Laguna.
- De Jong, A., Wardekker, J.A., van der Sluijs, J.P. (2012). Assumptions in quantitative analyses of health risks of overhead power lines. *Environmental science & policy* 16, 114–121.
- Ellis, E.C., Li, R. G., Yang, L. Z. and Cheng, X. (2000). Long-term change in village-scale ecosystems in China using landscape and statistical methods. *Ecological Applications* 10, 1057–1073.
- Funtowicz, S.O. and Ravetz, J.R. (1993). Science for the post-normal age. *Futures* 25(7), 735–755.
- Keepin B. and Wynne, B. (1984). Technical analysis of IIASA energy scenarios. *Nature* 312, 691–695.
- Kloprogge, P., van der Sluijs J.P., and Wardekker, A. (2007). Uncertainty communication: issues and good practice, report NWS-E-2007-199, Department of Science Technology and Society, Copernicus Institute, Utrecht University. 60 pp.
- Kloprogge, P., van der Sluijs J.P., and Petersen A.C. (2011). A method for the analysis of assumptions in model-based environmental assessments. *Environmental Modelling & Software* 26(3), 289–301.
- Knol, A.B., Slotje, P., van der Sluijs, J.P., and Lebre, E. (2010). The use of expert elicitation in environmental health impact assessment: a seven step procedure. *Environmental Health* 9, 1–16.
- Laes, E., Meskens G., and van der Sluijs, J.P. (2011). On the contribution of external cost calculations to energy system governance: The case of a potential large-scale nuclear accident. *Energy Policy* 39(9), 5664–5673.
- Morgan, M.G., and Henrion, M. (1990). *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*, Cambridge: Cambridge University Press.
- Nowotny, H. (1999). The place of people in our knowledge. *European Review* 7(2), 247–262.
- Petersen, A.C., Cath, A., Hage, M., Kunseler, E., and van der Sluijs, J.P. (2011). Post-normal science in practice at the Netherlands Environmental Assessment Agency. *Science Technology & Human Values* 36(3), 362–388.
- Risby, J. S., van der Sluijs, J.P., Ravetz, J. (2001). *Protocol for Assessment of Uncertainty and Strength of Emission Data*, Department of Science Technology and Society, Utrecht University, report nr. E-2001-10, 22 pp.
- Risbey, J., van der Sluijs, J. P., Kloprogge, P., Ravetz, J., Funtowicz, S. and Corral Quintana S. (2005). Application of a checklist for quality assistance in environmental modelling to an energy model. *Environmental Modelling & Assessment* 10(1), 63–79.
- Saltelli, A., Chan, K., Scott, E.M. (eds.). (2000). *Sensitivity Analysis*. Chichester: John Wiley & Sons.
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. and Tarantola, S. (2008). *Global Sensitivity Analysis: The Primer*. Chichester: Wiley.
- Saltelli, A., Guimarães Pereira, Â., van der Sluijs, J.P. and Funtowicz, S. (2013). What do I make of your Latinorum? Sensitivity auditing of mathematical modelling. *International Journal of Foresight and Innovation Policy* 9(2/3/4), 213–234.
- Van der Sluijs, J.P. (1997). *Anchoring Amid Uncertainty: On the Management of Uncertainties in Risk Assessment of Anthropogenic Climate Change*, Ph.D. dissertation, Utrecht: Utrecht University.
- Van der Sluijs, J.P. (2002). A way out of the credibility crisis around model-use in Integrated Environmental Assessment. *Futures* 34, 133–146.
- van der Sluijs, J.P., 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., Ravetz, J. (2002). *Uncertainty assessment of the IMAGE/TIMER B1 CO2 emissions scenario, using the NUSAP method*, Dutch National Research Program on Climate Change, Report no: 410 200 104, Bilthoven (available from www.nusap.net).
- van der Sluijs, J. P., Craye, M., Funtowicz, S., Kloprogge, P., Ravetz, J. and Risbey, J. (2005c). Experiences with the NUSAP system for multidimensional uncertainty assessment in Model based Foresight Studies. *Water Science and Technology* 52(6), 133–144.
- Walker, W.E., Harremoës, P., Rotmans, J., van der Sluijs, J.P., van Asselt, M.B.A., Janssen, P., Krayen von Krauss, M.P. (2003) Defining uncertainty, a conceptual basis for uncertainty management in model-based decision support. *Integrated Assessment* 4(1), 5–17.
- Wardekker, J.A., van der Sluijs, J.P., Janssen, P.H.M., Kloprogge, P., Petersen, A.C. (2008). Uncertainty communication in environmental assessments: Views from the Dutch science-policy interface. *Environmental Science and policy* 11, 627–641.
- Wynne, B. (1992). Uncertainty and environmental learning: Reconceiving science and policy in the preventive paradigm. *Global Environmental Change* 6, 111–127.