

## UNCERTAINTY ASSESSMENT OF VOC EMISSIONS FROM PAINT IN THE NETHERLANDS USING THE NUSAP SYSTEM

JEROEN P. VAN DER SLUIJS<sup>1,\*</sup>, JAMES S. RISBEY<sup>2</sup> and JERRY RAVETZ<sup>3</sup>

<sup>1</sup>*Copernicus Institute for Sustainable Development and Innovation, Utrecht University,  
Heidelberglaan 2, Utrecht, The Netherlands;* <sup>2</sup>*School of Mathematical Sciences,*

*Monash University, Clayton, Australia;* <sup>3</sup>*Research Methods Consultancy, London, UK*

(\*author for correspondence, e-mail: j.p.vandersluijs@chem.uu.nl)

(Received 13 November 2003; accepted 16 July 2004)

**Abstract.** We demonstrate an innovative approach to uncertainty assessment known as the NUSAP system, to assess qualitative and quantitative uncertainty for the case of emissions of volatile organic compounds (VOC) from paint in The Netherlands. Using expert elicitation, we identified key sources of error, critical assumptions, and bias in the monitoring process. We assessed pedigree and probabilistic uncertainty of all input quantities. We used four pedigree criteria to assess the strength of the knowledge base: proxy representation, empirical basis, methodological rigour and degree of validation. Using Monte Carlo analysis, we assessed sensitivity and propagation of uncertainty. Results for sensitivity and pedigree were combined in a ‘NUSAP Diagnostic Diagram’, which effectively highlighted the assumption for VOC percentage of imported paint as the weakest spot in the monitoring of VOC emissions. We conclude that NUSAP facilitates systematic scrutinization of method and underlying assumptions and structures creative thinking on sources of error and bias. It provides a means to prioritise uncertainties and focus research efforts on the potentially most problematic parameters and assumptions, at the same time identifying specific weaknesses in the knowledge base. With NUSAP, nuances of meaning about quantities can be conveyed concisely and clearly, to a degree that is not possible with statistic methods only.

**Keywords:** data quality, emission monitoring, NUSAP, pedigree, quality control, quality evaluation, uncertainty, VOC

### 1. Introduction

Recent national and international reviews and audits of The Netherlands Environmental Assessment Agency (RIVM/MNP) have recommended the intensification of efforts to deal with uncertainty in emission monitoring in The Netherlands (RIVM, 2001; Van der Sluijs, 2002). Internationally, the Intergovernmental Panel on Climate Change (IPCC) developed “Good Practice Guidance and Uncertainty Management in National Greenhouse Gas Inventories” (IPCC, 2000). In the same vein, the workshop “Monitoring of Greenhouse Gases in the Netherlands: Uncertainty and Priorities for Improvement” made a number of recommendations for a more systematic treatment of data uncertainty and quality in emission monitoring practices (Van Amstel *et al.*, 2000). In response to these developments, the RIVM and Utrecht University initiated a project on the reliability of emission figures,

aiming at the selection, testing and implementation of a methodology to assess and document uncertainty and quality of emission figures as produced for the annual Netherlands 'Milieubalans' that documents the state of the environment.

Uncertainty is a multi-dimensional concept involving quantitative and qualitative dimensions. The quantitative dimension can be characterized as technical and can be expressed in terms of inexactness. The qualitative dimensions can be subdivided in methodological, epistemological and societal dimensions. The methodological dimension can be expressed in terms of unreliability, the epistemological dimension in terms of ignorance and the societal dimension in terms of social robustness. (Funtowicz and Ravetz, 1990; Van der Sluijs, 1997; and for social robustness: Nowotny, 2003). Uncertainty in each of these dimensions can manifest itself at different locations in a given knowledge base of an environmental assessment, such as system boundaries, indicator choice, model structure, model parameters and input data (Van der Sluijs, 1997; Walker *et al.*, 2003; Van der Sluijs *et al.*, 2003).

In recent years, an increasing body of work in the field of uncertainty management has been accomplished to assess different types of uncertainties at different locations in the knowledge base. A number of useful formal and informal methods, tools and approaches are now available. Most of these tools address quantitative dimensions of uncertainty in model parameters and (input) data using sophisticated algorithms for sensitivity analysis and Monte Carlo techniques (e.g., Saltelli, 2002, 2004). Recently, a more integrated framework for uncertainty analysis in model-based environmental risk assessment has been developed by an international team of uncertainty experts (Walker *et al.*, 2003). Further available tools include quality assistance checklists (Risbey *et al.*, in press), uncertainty typologies (see Van der Sluijs, 1997 for an overview), uncertainty management checklists (Van der Sluijs *et al.*, 2003), and good practice guidances (IPCC, 2000; Van der Sluijs *et al.*, 2003).

Although the various quantitative techniques are essential to analyse the relative importance of various uncertainty sources, they tend to ignore unquantifiable uncertainty and thus provide only a partial insight into what is a very complex mass of uncertainties. This paper demonstrates a novel approach known as NUSAP, which extends the statistical techniques with qualitative analysis to partly overcome this limitation. NUSAP is a notational system proposed by Funtowicz and Ravetz (1990), which aims to provide an analysis and diagnosis of uncertainty in science for policy. It captures both quantitative and qualitative dimensions of uncertainty and enables one to display these in a standardized and self-explanatory way. The basic idea is to qualify quantities using the five qualifiers of the NUSAP acronym: numeral, unit, spread, assessment, and pedigree. By adding expert judgment of reliability (assessment) and systematic multi-criteria evaluation of the production process of numbers (pedigree), NUSAP extends the statistical approach to uncertainty (inexactness) with the methodological (unreliability) and epistemological (ignorance) dimensions.

NUSAP acts as a heuristic for good practice in science for policy by promoting reflection on the various dimensions of uncertainty and making these explicit. It provides a diagnostic tool for assessing the robustness of a given knowledge base for policymaking and promotes criticism by clients and users of all sorts, expert and lay and will thereby support extended peer review processes.

NUSAP yields insight on two independent properties related to uncertainty in numbers, namely spread and strength. Spread expresses inexactness whereas strength expresses the quality of the underlying knowledge base, in view of its methodological and epistemological limitations. The two metrics can be combined in a diagnostic diagram mapping strength and sensitivity of model parameters (see Section 13). Mapping model parameters in a diagnostic diagram reveals the weakest critical links in the knowledge base of the model with respect to the model outcome assessed and helps in the setting of priorities for model improvement.

## 2. Method

For demonstration and testing of the NUSAP system for uncertainty assessment, the case of emissions of volatile organic compounds (VOC) from paint in The Netherlands was selected. The study focussed on the VOC emission in the year 1998 as reported in RIVM's State of the Environment report ('Milieubalans') for that year. The method developed for and applied in this case study is documented step-by-step in a separate application protocol (Risbey *et al.*, 2001). The present paper documents this VOC emissions case study.

The objective of this case study is to demonstrate the application of NUSAP in the field of emission monitoring. The implementation of NUSAP in this study involved the following steps in the analysis for which the following set of uncertainty tools were jointly used:

- identification of sources of error and expert elicitation of unquantifiable uncertainty (NUSAP-based elicitation protocol);
- data qualification (pedigree analysis; Funtowicz and Ravetz, 1990);
- expert elicitation of quantifiable uncertainty (Stanford Protocol, Spetzler and Steal von Holstein, 1975);
- analysis of propagation of uncertainty (Monte Carlo);
- analysis of propagation of pedigree (pedigree calculator, Ellis, 2000a, 2000b).

We developed a NUSAP-based expert elicitation protocol for the systematic identification of sources of error and bias. For the qualification of input data and parameters, we used the NUSAP method and tailored a pedigree matrix specifically applicable to emission monitoring to map the strength and scientific status of the numbers and assumptions used in the emission monitoring calculations.

Pedigree conveys an evaluative account of the production process of a quantity and indicates different aspects of the underpinning of the numbers and scientific status of the knowledge where it originates from. Pedigree is expressed by means of a set of pedigree criteria to assess these different aspects. Criteria used in this study are proxy, empirical basis, methodological rigour and degree of validation (see Section 9). These criteria are used as indicators for data- and parameter strength. We developed a questionnaire for expert elicitation to draft pedigree scores (Van der Sluijs *et al.*, 2002). To assess the pedigree of the total national VOC emission number, we used the pedigree calculator developed by Ellis *et al.* (2000a, 2000b).

For the quantitative assessment, we assessed probability density functions (PDFs) by expert elicitation following the Stanford Protocol (Spetzler and Steal von Holstein, 1975) for all input data and parameters in the calculation. To analyse the propagation of data- and parameter uncertainty in the calculation of the total VOC emission from paint in The Netherlands, we used Monte Carlo simulation (using the Crystal Ball software package).

We would like to stress that pedigree and spread (c.q. PDFs) are seen as in principle independent properties. All combinations of high and low pedigree and spread are conceivable. There may seem to be some correlation between the two if uncertainty is mainly caused by lack of knowledge. However, uncertainty can also stem from variability and randomness of the phenomenon studied. A high spread can thus coincide with a high pedigree, if the spread is mainly caused by such intrinsic uncertainty. Also, it is not necessarily the case that a low pedigree implies a high spread.

Overall, the uncertainty assessment method we used involves the following steps (see Risbey *et al.*, 2001):

- Exploration of problem context.
- Identification and classification of variables, parameters and data.
- Disassembly and aggregation of key input.
- Assessment of assumptions.
- Identification of expertise.
- Identification of sources of error.
- Qualitative assessment of data- and parameter strength (using expert elicitation).
- Quantitative assessment of data- and parameter spread (using statistics or expert elicitation).
- Calculation of Monte Carlo and pedigree results.
- Synthesis.

The methodological aspects of these steps and the elicitation protocol that we used are described in detail in Risbey *et al.* (2001). For the purpose of our analysis, we focus on the emission figures of VOC in paint for the year 1998. In the following sections, we will discuss our findings for each step.

### 3. Exploration of Socio-political Problem Context

It is widely acknowledged that quality of knowledge can not be assessed independently of the function of that knowledge. Quality relates primarily to fitness for function (Funtowicz and Ravetz, 1990; Corral Quintana, 2000; Risbey *et al.*, 2001). The socio-political context influences what sorts of uncertainty are salient in a given problem. To assess the uncertainty and quality of emission figures of VOC, we therefore first need to explore the socio-political context: Why are VOC emission being monitored and what are VOC emission figures used for within policymaking processes?

VOCs play a role in a number of environmental atmospheric and air quality problems, including the formation of tropospheric ozone and photochemical smog. VOCs from paint consist of a number of different organic substances (solvents), some of which may impose direct health risks. Over the past decades, policies for VOC emission reduction have been formulated and implemented in The Netherlands. The problem entered the policy agenda in the 1980s. A major effort has been the so-called KWS2000 policy project (VROM, 1989), aiming at a substantial reduction of VOC emissions in The Netherlands. The KWS2000 reduction plan came into effect in the year 1988. It aimed at a reduction of VOC emissions by 50% in the year 2000, relative to the VOC emissions in the year 1981. For the paint branch, specific reduction targets and strategies have been formulated and implemented. After the year 2000, the reduction policies have continued. Agreements are now incorporated in law by specifying them in the Dutch Emission Guideline for air. The National Environmental Policy Plan 3 (NMP-3, 1998) has set a new target, namely to achieve a reduction of VOC emissions by 80% in the year 2010 relative to the year 1980.

In general, the KWS2000 target to accomplish a 50% reduction for the KWS2000 sectors appears to have been achieved. When one includes the non-KWS2000 sectors (mainly agriculture and traffic), in the year 2000, the emission of VOC in The Netherlands was 50% higher than the absolute NMP-3 target of 193 million kg, despite a reduction by 40% in the period 1990–2000.

KWS2000 VOC policies have gone through a policy life cycle of problem recognition (1980s), gaining control (late 1980s), solving (up from 1988) and maintaining control (1990s). With the formulation of new reduction targets for 2010 and the exploration of further reduction potentials in various sectors, the VOC problem is undergoing a new iteration of the policy life cycle. One tendency here is to more strongly link VOC policies to occupational health issues and indoor environment issues.

The major roles of the VOC monitoring efforts in the policy process have thus been twofold. On the one hand, the monitoring was part of the KWS2000 strategy and served as an instrument to monitor progress and compliance. On the other hand, the monitoring has been used in the context of formulating new policies beyond 2000.

Fitness for function hence relates to questions of whether the VOC monitoring framework is sufficiently sensitive to detect whether the required progress has been achieved; whether the monitoring has provided sufficient insights for the purpose of formulating new reduction targets and reduction policies; whether the monitoring instruments meet requirements to serve as an adequate source of knowledge in the maintaining control phase for the new policies; i.e. whether the nature and magnitude of uncertainties do not hamper conclusions about whether required emission reductions have been achieved and policy goals have been met.

#### 4. Identification and Classification of Input and Parameters

VOC emissions from paint in The Netherlands are being assessed on an annual basis by the RIVM. The VOC emissions are calculated from national paint sales (defined as paint that is both produced *and* sold in The Netherlands) statistics, provided by The Netherlands Association of Paint Producers VVVF (VVVF, 1999) and paint import statistics provided by the Statistics Netherlands (CBS). The VVVF (through its members) directly monitors VOC in paint whereas CBS monitors paint. These CBS paint statistics are combined with an overall estimate of the thinner (VOC) percentage in imported paint. On top of that, estimates have been made for emissions from paint-related thinner use.

One implicit assumption of the way the VOC emissions from paint are determined is that all the VOC in paint that is produced and sold in The Netherlands and that is imported in a given year is totally emitted in that same year. In other words, stocks and the possibility of delayed or partial emissions are not accounted for or implicitly assumed to balance one another from year to year.

The sectoral aggregation level at which VOC from paint is monitored takes the application area of the paint as starting point. These application areas are:

- ship construction (SHI),
- decoration (buildings and steel) (B&S),
- do it yourself (DIY),
- car refinishing (CAR),
- industry (IND),

The 1998 VOC emission figures from paint as compiled by the RIVM are given in Table I. This table formed the starting point for our uncertainty analysis.

The numbers in the first column of Table I, national sales (NS), were obtained by RIVM from the annual paint statistics produced by the VVVF and directly gives VOC content. The numbers for import in the second column are calculated from a lumped figure for total paint import provided by CBS, using a series of assumptions: Based on CBS import statistics, it was inferred that 80 ktonne paint was imported

TABLE I  
RIVM's monitoring spreadsheet for VOC emissions from paint

Source	National sales	Import	AddTh%	Thinner emission	Total
SHI	1.8	<i>0.6</i>	0%	<i>0.0</i>	<i>2.4</i>
B&S	11.9	<i>4.0</i>	11%	<i>1.7</i>	<i>17.6</i>
DIY	6.6	<i>2.2</i>	28%	<i>2.5</i>	<i>11.3</i>
CAR	2.5	<i>0.8</i>	19%	<i>0.6</i>	<i>4.0</i>
IND	9.4	<i>3.1</i>	30%	<i>3.8</i>	<i>16.3</i>
Total	32.2	10.8		8.6	51.6

Year 1998 (all figures in ktonne VOC). Numbers printed in italics are calculated, the other numbers are input.

in 1998. The composition of the paint has not been monitored, hence RIVM made an educated guess, assuming that the average VOC percentage in the imported paint is similar to the average VOC percentage in paints produced and sold in The Netherlands. RIVM assumed imported paint to consist of 13.5% VOC, leading to 10.8 ktonne VOC in total from imported paint (the figure at the bottom of the second column in Table I). No detailed knowledge was available on how the imported paint is distributed over the application categories distinguished in the monitoring. As a first approximation, it was assumed by RIVM that the 10.8 ktonne VOC from imported paint can be attributed to the five application categories in the same ratio as the VOC from NS is distributed over these categories. This is how the VOC figures from imported paint for each of the categories in Table I were calculated.

The third column of the table, AddTh%, lists estimates for additional use of thinner, related to the use of paint in each application category. This is expressed as a percentage that adds to the VOC in the paint. The exact figures for additional thinner use are not well known. The figures used by RIVM are educated guesses, jointly established by several parties involved in the KWS2000 project several years ago and have not been updated since.

The fourth column list the VOC emissions from additional thinner use per category. These figures were calculated by adding the VOC emissions from NS and import for each category and then applying the additional thinner (%) assumed for that category. In the final column, the sum of NS, import and additional thinner use is listed. Adding these figures over all the categories yields the total emission of VOC from paint in The Netherlands in 1998 as monitored with the RIVM method to be 51.6 ktonne VOC.

## 5. Disassembly and Aggregation

In Section 4, we identified the following quantities and assumptions as main inputs to the monitoring:

- paint sales statistics per application category, drafted by VVVF;
- a lump figure for imported paint, drafted by CBS;
- an assumption for the VOC percentage in imported paint, drafted by RIVM;
- an assumption for the attribution of imported paint to the categories, drafted by RIVM;
- estimates for additional thinner use per application category, jointly drafted by parties involved in KWS2000.

We can disassemble each of these key inputs to identify its underlying components.

### 5.1. VVVF PAINT STATISTICS

The VVVF uses annual questionnaires from their members to produce paint statistics. These questionnaires monitor a number of things, including annual sales figures. Three metrics are used for the annual sales figures: amount of paint (unit: tonnes); amount of VOC in that amount of paint (unit: tonnes); and volume of trade (unit: 1000 guilders). The questionnaire distinguishes 13 application categories. In each category, different paint types are distinguished such as high solvent, low solvent, and the like. In total, the questionnaire is differentiated into 43 paint types for each of which the three metrics have to be filled in by each VVVF member. Note that the way the questionnaire is set means that the paint industry itself directly produces the monitoring figures for the VOC amount in the paint. The idea is that they know very well how much VOC goes in the paint, yielding more reliable figures than would be obtained if the paint volumes were monitored and VOC was calculated by applying a VOC content percentage to the tonnes of paint.

The VVVF questionnaire is only circulated amongst the members of the VVVF. The VVVF estimates that its members cover 95% of the total paint sales volume in The Netherlands. For the 1998 emission monitoring, 55 VVVF members filled in VOC figures in the questionnaire, whereas 30 members did not provide VOC figures. However, all the big firms were within the group of responders, so the VVVF estimates the respondents to cover 87% of the total sales volume of all VVVF members (VVVF, 1999).

All members provide at least total sales figures (in guilders) because this is coupled to the membership fee and the required audit certificate. This means that this information is available also for the 30 members that did not provide VOC figures. This total monetary figure is however not differentiated into the 43 paint types in the 13 application categories. In the case of non-responders for the VOC figures, VVVF uses the files on that firm and available data from previous years to split out and attribute the total sales figures to categories and paint types. The

VOC content is then estimated using the monetary sales figures in combination with available knowledge on the category and paint types.

## 5.2. CBS IMPORT STATISTICS

The Business Surveys Department at CBS develops statistics on international trade. They collect data on paint and varnish along with 10,000 other trade items *via* an electronic survey (the Goederennaamlijst) sent to every firm in the NL with annual imports or exports in excess of Dfl. 500 K. They estimate that only a small fraction of imported paint is accounted for by firms with imports smaller than Dfl. 500 K. The response rate to the survey is approximately 60% for imports from EU member countries and 75% for non-EU countries after 6 weeks. One year after the survey has been sent out, the response rates have risen typically to approximately 90 and 95%, respectively. For Dutch firms, responses are required by law. For non-respondents, the missing data is estimated by CBS using standard state-of-the-art techniques for imputation and macro-editing (Van de Pol and Diederer 1996; Slootbeek, 1997). In some cases, it is possible to generate the appropriate estimates by cross-referencing with data provided from the Tax Department tapes. When this is not possible, imputation and weighting methods are used to infer the response from historical reports from the firms (Slootbeek, 1997).

All items in the Goederennaamlijst survey are in a standard record layout and are quality controlled electronically. A macro-editing algorithm searches for anomalies in the data that deviate from the history of responses for each firm. Typical errors such as using the wrong units are corrected internally. Other errors are corrected by phoning the reporting firms to query the anomalous data items. The data in the Goederennaamlijst are collected at various levels of aggregation (eight, six, and four digit codes in order of increasing aggregation). The data at the highest levels of aggregation (four and six digit codes) are subject to more rigorous checking procedures than the eight digit code data. Survey responses from the largest firms and from those with problematic reporting are checked periodically by performing visiting audits with the firms.

The survey data is cross-checked with statistics from other agencies, including the monthly tapes with tax and turnover figures from the Tax Bureau and Ministry of Finance, and with statistical agencies in other countries (principally Germany, Belgium, and England). Cross-checks with other Dutch agencies are typically performed on firm-level aggregate data and so would not single out individual items such as paint. These cross-checks are performed routinely, while cross-checks on aggregate imports and exports are performed with foreign statistical agencies on an occasional basis. Such cross-checks may also focus on particular products where discrepancies are suspected (e.g. crude oil). Within CBS, another check that is performed compares estimates for particular goods to ensure that production plus imports equals consumption plus exports. While this has been problematic for some

commodities (e.g. flowers), paints have not generated notable imbalances at this level.

### 5.3. ASSUMPTIONS

The RIVM assumption for the VOC% in imported paint applied to the total amount of imported paint and was not differentiated into the individual paint categories in the Goederennaamlijst used by CBS. The assumption for the attribution of imported paint over the categories and the estimates for additional thinner use per application category can also not be further disassembled, as there are no underlying aggregation levels to these assumptions.

### 5.4. AGGREGATION LEVELS FOR OUR ANALYSIS

The diagram in Figure 1 shows the various aggregation levels that we identified in the figures used for the monitoring of VOC emissions from paint.

For the assessment of assumptions and identification of errors, we searched for sources of error on all aggregation levels identified in the previous sections. For the elicitation of pedigree scores and PDFs, we have chosen to work at the same aggregation level as the one used in the original RIVM spreadsheet table (Table II), but we set up our elicitation procedure in such a way that we stimulated the respondents to consider consequences of errors on lower aggregation levels in drafting PDFs and pedigree scores.

## 6. Assessment of Assumptions

In Table II, we have listed the main assumptions that we identified in the process by which annual figures for VOC emissions from paint are compiled. The second column in the table discusses briefly the validity and tenability of each assumption. In the third column, we have indicated what possible consequences for the final emissions result and the usefulness of these results in the policy process might be in cases where the assumption is (hypothetically) found to be invalid.

## 7. Identification of Expertise

The major input data for the monitoring calculations are produced by two institutions: The Netherlands Association of Paint Producers (VVFV), providing statistics on national paint production and NS statistics of paint produced in The Netherlands, and Statistics Netherlands (CBS), producing statistics on paint import. An expert elicitation was conducted with selected experts at CBS on 3 November 2000. Experts consulted were G. Slootbeek (Project Manager for Trade, Transport

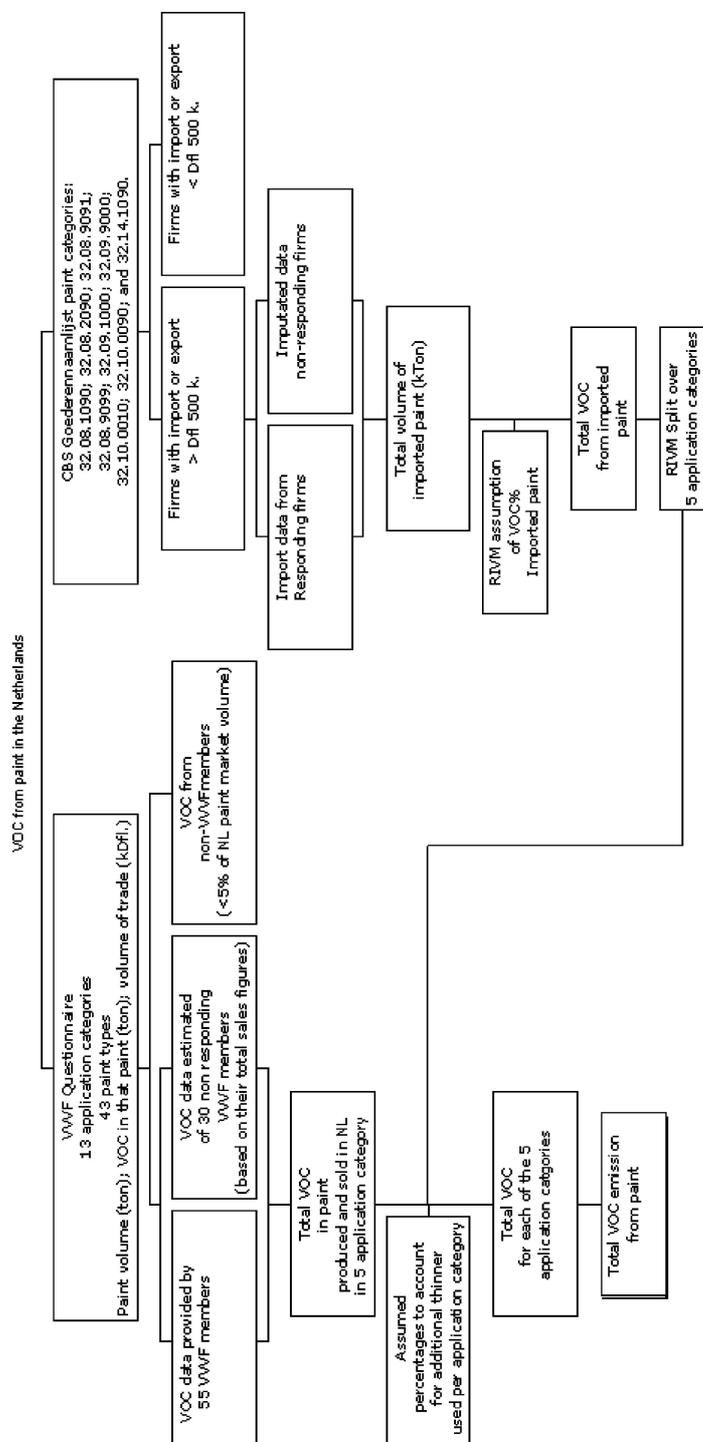


Figure 1. Aggregation levels in the monitoring of VOC from paint in The Netherlands.

TABLE II  
Assessment of assumptions identified in the monitoring calculations

Assumption	Validity and tenability	Consequences
Tonnes of VOC emission is a good indicator for air quality and health risks policies	VOCs enclose a very large number of different chemical compounds. Each of these may have different toxic properties, atmospheric life times, atmospheric chemical reaction pathways and the like. Hence, it need not be true that environmental damage is proportional to the amount of tonne total VOC emitted However, chemical composition of bulk solvents in paint may not have varied too much over time, making tonnes a reasonable indicator for environmental impact	If environmental damage per tonne of VOC has changed substantially over time due to changes in chemical composition, the indicator may fail the policy objective of protecting the environment: The possibility cannot be excluded that damage from VOC emissions goes up while tonnes of emissions go down and <i>vice versa</i>
Paint sales are a good proxy for VOC emissions from paint	Assumes all VOC released from paint and the VOC in unused amount of paint each year is the same. Reasonable assumption unless a big shift in use to different paint types changes stock levels	Paint sales and VOC could be out of sync for some period if big change in buying patterns, but likely corrected over time
Overall VOC content in imported paint same as for the NL.	Valid so long as imports from countries with similar production standards and similar VOC emission reduction policies	If imports favour countries producing paint with higher VOC content then undercounting VOC amount
Imported VOC split across categories as per NL	Valid so long as imports not favour specific domestic categories	If imports going preferentially into high VOC application categories (e.g. car refinishing) then undercounting VOC. This seems to be the case
Non-respondents and non-members have similar sales patterns as similar size responding member firms	Valid so long as non-members engaged in similar operations as members	Could lead to under or over estimates depending on non-member activities. Worst case would be an unknown big producer among non-members, which seems unlikely

and Services), R. Bakkes (Economist for Business Services), and J. Heylands, all from the Division of Business Statistics, Business Surveys Department of CBS, Heerlen. The three CBS experts were interviewed as a group. Another expert elicitation was held on 8 December 2000 and continued till 18 December 2000 with A. Winkelaar from The Netherlands Association of Paint Producers (VVVF).

The major basis for the expertise of our respondent at VVVF is the following: VVVF has many years of experience in monitoring NS and figures are warranted by audit certificate. VVVF has inside knowledge of the paint market and has sufficient opportunities to look into the kitchen of the paint sector to judge sales statistics. The respondent considered his expertise on VOC% in paint medium because knowledge on imported paint is at a highly aggregated level. VVVF has only limited knowledge on how imported paint is distributed over sectors. However, VVVF has sufficient knowledge of the paint market to be able to make educated expert judgements.

The goal of the elicitations was to identify sources of uncertainty in the emission monitoring, to assess the pedigree of the quantities generated for the monitoring and to quantify the uncertainty in these quantities. The details of the elicitation procedure are documented in Risbey *et al.* (2001).

## 8. Sources of Error

The major sources of error identified in the elicitations at CBS and VVVF are summarized in Table III (in no particular order), along with an assessment of the likely direction of bias introduced by each error source.

We also have searched for conceivable sources of motivational bias. Motivational bias may occur when people have an incentive to reach a certain conclusion or see things a certain way. Generic reasons for occurrence of motivational bias include: (a) a person may want to influence a decision to go a certain way; (b) the person may perceive that he will be evaluated based on the outcome and might tend to be conservative in his estimates; (c) the person may want to suppress uncertainty that he actually believes is present in order to appear knowledgeable or authoritative; and (d) the expert has taken a strong stand in the past and does not want to appear to contradict himself by producing a distribution that lends credence to alternative views. Table IV summarises the specific conceivable sources of motivational bias that we identified in our case.

Summarising the past two tables, our assessment is that overall the sources of error we identified tend to produce a underestimation of the total VOC emissions from paint. We did not attempt to quantify the size of this underestimation. However, we found no direct indications to imply a large underestimation.

Several of the potential errors identified influence the split of paint among application categories without influencing the total amount of paint. However, the total VOC emission calculated is sensitive to the split in the application sectors because

TABLE III  
Sources of error in monitoring VOC emissions from paint

Name of error	Qualitative description	Random or systematic?
Definitional inconsistency	CBS uses paint definitions outlined in the Goederennaamlijst. This list was drafted in the nineteen fifties and hence has a number of obsolete paint types whereas it misses a number of new paint types. VVVF uses their own (different) definitions of paint, following state of the art in the paint industry	Systematic, leading to some over-counting and some undercounting if not corrected
Interpretation of definitions	VVVF experience with VOC monitoring shows that interpretation of definitions has varied across the people that fill in the questionnaires at the paint industries	Typically may lead to underestimates of paint sales due to omission
Fuzzy boundaries between raw materials, products, assortment	It is hard to draw unambiguous boundaries between raw materials, products and assortment. There are many boundary products. However, the major products in each category will always be captured by the definitions	Unknown but likely to be small
Miscategorisation	Firms may report some paint products in non-paint categories and vice versa. There is a number of VOC containing products which do not fall in the definition of paint (e.g. wax-pens for furniture repair)  In cases where the number of firms in a sub-category is so small that the statistics can be traced back to individual firms, the numbers are not reported under that sub-category but in a category "general"	Can be both random and systematic. Can work in both directions
Misreporting <i>via</i> unit confusion	There may be unit confusion between liters and kilograms. The sales figures are in tonnes whereas the paint is actually sold per liter  Over-reporting by a factor 1000 occurs when kg are used instead of tonnes. This is likely to be caught by the macro-editing process. Under-reporting typically results in only small deviations in the aggregate data and is less likely to be caught by the macro-editing process. This results in some net bias towards underestimates	Overestimate by factor 1000, but usually caught and corrected
Miscoding	Errors in the data coded by firms are checked by the macro-editing routine. Checks on the history of firm reports would tend to catch the most egregious errors of this kind	Random and typically caught by cross-checking with past responses

(Continued on next page.)

TABLE III  
(Continued)

Name of error	Qualitative description	Random or systematic?
Non-response	At CBS, for EU firms the non-response rate is about 10%. Non-respondents are far more likely to be smaller firms than larger firms, so the net contribution to paint imports of non-respondents is likely much less than this. Methods are used to estimate data for non-respondents, further minimizing this source of error. At VVVF the response rate is 65%. Non-respondents are mainly smaller firms. The VVVF estimates that the respondents cover 87% of the total sales volume of all VVVF members together	Random However, if motivational bias exists, the error could be systematic
Not counting small firms	No attempt is made to estimate the contribution to imported paint of small firms with less than 500K in annual imports. This is a source of underestimation, though its magnitude is likely to be small because the paint market is assumed to be dominated by large firms	Systematic, leading to underestimate of imports, but likely small because market dominated by larger firms
Not counting of non-VVVF members	VVVF estimates that its members together cover 95% of the total NS volume. This estimate has not been validated	Systematic, leading to underestimate of NS, but likely small because market dominated by VVVF members
Firm structure	For some firm holdings, individual daughter groups report to the main firm holding, who then reports to the CBS. This data is harder to quality control as the CBS interacts only with the parent firm. Daughter groups may report paint sales in inconsistent units, which will likely only be corrected if the parent corrects it. Otherwise, it would be a source of systematic error in the data, assuming the same use of inconsistent units by daughter groups from year to year	Systematic, leading to some undercounting if not corrected

(Continued on next page.)

TABLE III  
(Continued)

Name of error	Qualitative description	Random or systematic?
Firm dynamics	Firms are bought, sold, merge, and split. This presents some confound, since the data are reported by firms. The CBS tracks firm changes, updating records to reflect changes. There are some lags in receiving this information at the CBS. Further, some firm changes involve financial reconciliations that span a period of many months. Thus there is a lag of some months in capturing and reconciling firm changes. This source of error may result in some double-counting and some missed-counting at any given instant in time. Barring major structural changes in the paint industry, the magnitude of this error is likely to be small. Changes in firm size could also result in some firms moving over or under the reporting threshold through time. Barring a trend toward smaller or larger firms in general in the industry, this source of error is likely to be small and random.	Random
Paint market dynamics	The paint market is very dynamic. Firms are changing all the time, new markets appear (e.g. specialized paints for high tech applications) and new products appear on the market annually whereas other paint types may become obsolete over time	Systematic, may lead to underreporting if new paints are not captured by old definitions
Unofficial imports and parallel markets	There may be cross-border paint purchases not captured by the monitoring method There may also be a parallel market in paint: paint produced in The Netherlands and exported to abroad may be bought abroad by paint users because it is cheaper (due to sharper competition in Germany). The size of this market is unknown	Systematic, leading to some underestimate
Disagreement on attribution	We observed disagreement on the extent to which VOC from extra use of thinner during paint use by the end user should be attributed to the paint sector. For instance, the choice of an end-user to use thinner to clean and degrease a surface before it painting it is something beyond the sphere of influence of the paint producer	Systematic, leading to some overestimation for the VOC emissions from the paint sector
Computer code errors	The computer code used to manage the CBS data may have errors in it that affect the imported paint calculation. While the code is implicitly checked when CBS data is cross-checked with data from other agencies, these checks are performed at a more aggregate level than that of paint imports. Thus errors in the paint calculation may persist	Could lead to systematic errors but the direction of such errors is unknown

TABLE IV  
Conceivable sources of motivational bias in monitoring VOC emissions from paint

Conceivable sources of motivational bias	Qualitative description	Bias
Deliberate misreporting	Firms may conceal the true level of their imports when incentives exist to avoid taxation or inflate expenses. Such incentives may exist in this case if the firms believe that governmental departments share or cross-check databases (e.g. cross-check between CBS data by tax authorities)	May lead to systematic underreporting
Membership due to VVVF depends on sales figures	The membership due to the VVVF depends on the sales figures and ranges from 4000 Dutch Guilders up to about 100.000 Dutch Guilders. Firms consequently may have an incentive to underreport their sales figures to apply for a lower membership due	May lead to systematic underreporting
Paint sales statistics are market-sensitive information	There is a lot of competition in the paint market. In this context, paint sales statistics are market sensitive information. This may introduce incentives for strategic reporting to mislead competitors. The raw data from the VVVF questionnaire is confidential. In the published version of the statistics firms are anonymous. For some categories, there are very few players, meaning that the figures can be easily traced back to them. This may give these firms an incentive to report more in unspecified rest categories to restore anonymity	Unknown
Reputation of paint industry	The paint industry is a main player in the KWS2000 project. As polluters they have been subject to criticism from a number of societal groups. It is in their interest to negotiate and maintain a good reputation. This may provide an incentive to utilise the interpretative space in the data to work towards a monitoring outcome that shows an emission reduction trend over time	May lead to systematic underreporting

in the monitoring calculation, the assumed percentages for additional thinner use differ across categories.

### 9. Qualitative Assessment: Elicitation of Pedigree Scores

We used the input numbers and assumptions identified in Sections 4 and 5 as the starting point for the elicitation of pedigree scores and subjective PDFs. We used four pedigree criteria: proxy, empirical basis, methodological rigour and degree of validation (see Table VII; we refer to Risbey *et al.*, 2001 for methodological details). The quantities and their aggregation used in the elicitation are:

- NS figures for each of the five categories:
  - SHI (ship construction),
  - B&S (decoration/buildings),
  - DIY (do it yourself),
  - CAR (car refinishing),
  - IND (industry),
- percentages assumed for additional thinner use for each of the five categories;
- the lump figure for the amount of paint imported;
- the assumed VOC percentage in the imported paint.

During the elicitation, we decided to add three quantities for inclusion in our Monte Carlo simulation because the respondent considered these sources of error quantifiable:

- import below reporting threshold;
- gap between NS VVVF members and total NS;
- overlap between CBS import figure/VVVF NS figure.

For the categories DIY and B&S, the expert we interviewed at VVVF preferred to split his estimates of “additional thinner use” into *attributable to paint* and *attributable to other sectors* (further referred to as ‘rest’), because he disagreed with RIVM attributing thinner used for cleaning a surface to the *paint* sector.

We also asked for differences in state of knowledge across the five categories. The respondent preferred to treat SHI and B&S as one category in the pedigree elicitation. NS in the SHI and B&S are equally well known with a limited number of well-defined paints and well-defined applications. DIY is the most diffuse category. A very broad assortment of paints and a wide spectrum of applications makes this category less well known than the other categories. CAR is a homogenous market characterized by a limited number of paints and a limited number of well-known steps in the application. It is a stable industry sector with regard to paint types and their uses. IND is a quite diffuse and dynamic category. Applications are not well

known because statistics are specified according to client, which are anonymous and heterogeneous, Also, the category is quite dynamic. For instance, if there is a recession in steel one year then the paint is applied to for instance glass. Also, a number of new high-tech applications complicates the industry category.

We elicited pedigree scores for each of the quantities. For that purpose, we used a comprehensive elicitation template (Van der Sluijs *et al.*, 2002). The pedigree matrix used is documented and discussed in Risbey *et al.* (2001) and is displayed in Table V.

The pedigree scores we elicited for each of the quantities assessed are presented in Table VI. The last column of the table combines the scores for the four pedigree criteria by calculating the sum and dividing by the maximum attainable score (16 in this case). We will briefly discuss the backgrounds of the pedigree scores.

The proxy pedigree criterion refers to how good or close the quantity that we measure is to the actual quantity about which we seek information. What is measured for the first five entries in the tables is tonnes of VOC in national paint sales in The Netherlands of paint produced in The Netherlands as reported in the questionnaires by the members of VVVF who responded.

TABLE V

Pedigree matrix for emission monitoring data (adapted from Ellis *et al.*, 2000a, 2000b; ORNL and RFF, 1994; see Risbey *et al.*, 2001)

Score	Proxy	Empirical	Method	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 established 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 approximation rule of thumb estimate	Preliminary methods unknown reliability	Weak and very indirect validation
0	Not correlated and not clearly related	Crude speculation	No discernible rigour	No validation performed

TABLE VI  
Pedigree scores elicited

	Proxy (0–4)	Empirical (0–4)	Method (0–4)	Validation (0–4)	Strength (normalised to a scale 0–1)
NS-SHI	3	3.5	4	0	0.7
NS-B&S	3	3.5	4	0	0.7
NS-DIY	2.5	3.5	4	3	0.8
NS-CAR	3	3.5	4	3	0.8
NS-IND	3	3.5	4	0.5	0.7
AddTh%-SHI	2	1	2	0	0.3
AddTh%-B&S	2	1	2	0	0.3
AddTh%-DIY	1	1	2	0	0.25
AddTh%-CAR	2	1	2	0	0.3
AddTh%-IND	2	1	2	0	0.3
Imported-paint	3	4	4	2	0.8
VOC%-Import	1	2	1.5	0	0.3
Attribution-import	1	1	2	0	0.25
Imp-Below-Threshold	2	1	1	0	0.25
Gap-VVVFNS-RNS	2	1	1	0	0.25
Overlap-VVVF-CBS	2	1	1	0	0.25

For the VOC% in imported paint, the proxy score is low because what is measured here is thinner percentage in paint produced in The Netherlands, which our respondent considered a weak proxy for thinner (%) in imported paint.

In elaborating the proxy scores for VOC% in imported paint and the attribution of imported paint over the categories, the respondent argued that paint is imported for certain reasons. According to the respondent, hardly any paint is imported for the DIY category, whereas cheap plasters are imported from Germany for building. For car refinishing, about half of the paint in the category is imported. In the industry, some paint is imported.

Empirical quality typically refers to the degree to which direct observations are used to estimate the variable. The VVVF has 85 paint producing members. For the 1998 figures, 55 members returned the questionnaire (65%). Because the non-respondents are small firms, the VVVF estimated that the 55 respondents covered 87% of the NS. The questionnaire is normally filled in by account managers or by a person from the purchasing department of the firm.

The NS figures are based on a medium to large sample of direct measurements and hence score between 3 and 4 on the empirical scale. Estimates for additional thinner use and attribution of imported paint to the categories are educated guesses and hence score a 1. The figure for VOC% in imported paint is derived from measurements of VOC% in Dutch paint and hence scores a 2 for empirical pedigree.

With regard to methodological rigour, the VVVF has developed its own practice for estimating the missing figures to handle non-response. Further, the administrator at VVVF, who enters the survey results in the computer, checks the data for anomalies. Unit confusion (kg in stead of tonne of guilders in stead of million guilders) is likely to be flagged. There also is a check with historic data for each firm and knowledge on recent developments in the paint market.

VVVF knows her members quite well. Major changes in the market, such as burn down of a firm or merger of firms are known and are reflected in the sales statistics. If unexpected growth or shrink is observed, the data are checked with the firm that provided them.

For NS figures, VVVF uses questionnaire techniques and employ imputation techniques following best available practice in the branch. Therefore, our respondent gave a score of 4 for the methodological rigour. For other quantities, the methodological rigour is much lower. No systematic reliable method was used for estimating additional thinner use and attribution of paint over the categories, and the method used to estimate VOC% in imported paint was judged by our respondent to have unknown reliability.

With regard to validation, there is some checking for internal anomalies in the data. Validation is complicated. The VVVF has a small secretariat and has about 100 members with complex administration of hundreds of different paint products.

Meta-information on shapes of distributions obtained from earlier years is used together with knowledge of the category to perform consistency checks. The paint market is however very dynamic. Firms are changing all the time, new markets appear (e.g. specialized paints for high-tech applications such as CDs) and new products appear on the market, making validation by comparison with data from earlier years unreliable.

For most categories, there are no consistency checks with independent data because such data are not available. The figures by CBS are not usable because the VOC content of the paint categories of the Goederennaamlijst are not well defined. This list was drafted in the 1950s. Since then the chemical formula of paint have changed substantially. For the DIY and CAR categories, the branch produces its own figures. For the IND category, there are some partial data (namely on timber furniture) available for validation. For all the other categories and numbers, no validation against independent data takes place.

There is a limited amount of peer review: RIVM reviews the VVVF efforts and the figures produced are discussed with the other parties involved in the KWS2000 monitoring. Overall, the validation scores are low, with the exception of paint sales figures for DIY and CAR.

The CBS paint import figure was assessed to be of high pedigree quality for most of the pedigree metrics. The CBS has a highly quality control oriented culture and uses best available practice state-of-the-art methods for imputation and macro-editing. Response rates are fairly high, leading to a high score on empirical pedigree. The weakest dimension of the data's quality appears to be the validation

metric. However, that weakness is largely intrinsic to the problem of collecting this data and is not a function of the methods or effort applied by the CBS.

Overall, the parameters with the lowest strength are the assumed VOC percentage in imported paint and the assumed attribution of imported paint over the categories. The NS figures have the highest parameter strength. There is also difference in strength in the knowledge base across the categories with CAR and DIY having the strongest knowledge base.

### 10. Quantitative Assessment: Elicitation of PDFs

We elicited PDFs for all inputs using the Stanford Protocol (Spetzler and Steal von Holstein, 1975). We started by asking the expert to state the extreme minimum and maximum plausible values for the variable and then asked them to try to envision ways or situations in which the extremes might be broader than stated. We then let the expert revise the extreme values accordingly. Then we asked them whether the expert considered the distribution to be symmetric or not. Next we asked to estimate the 5th, 50th and 95th percentiles of the distribution. The results of the elicitation are given in Table VII.

TABLE VII  
Probability density functions elicited

Quantity	5th percentile	50th percentile	95th percentile	Initial point value	Unit	Coded in Crystal Ball
NS-SHI	1.6	1.85	1.9	1.8	ktonne VOC	Custom
NS-B&S	11.3	12.3	12.5	11.9	ktonne VOC	Custom
NS-DIY	6.1	6.3	6.7	6.6	ktonne VOC	Custom
NS-CAR	2.3	2.6	2.7	2.5	ktonne VOC	Custom
NS-IND	9	9.6	9.8	9.4	ktonne VOC	Custom
AddTh%-SHI	1	2	3	0	%	Custom
AddTh%-B&S	2	3	4	11	%	Custom
AddTh%-B&S-rest	7	8	9		%	Custom
AddTh%-DIY	1	2	3	28	%	Custom
AddTh%-DIY-Rest	25	28	30		%	Normal
AddTh%-CAR	15	19	25	19	%	Custom
AddTh%-IND	20	25	35	30	%	Custom
Imported-paint	78	80	85	79.7	ktonne paint	Extreme Value
VOC%-Imp	10	15	18	13.5	%	Custom
Imp-Below-Threshold	3	5	8	0	ktonne paint	Custom
Gap-VVVFNS-RNS	3	4.5	6	0	ktonne paint	Custom
Overlap-VVVF-CBS	10	15	20	0	ktonne paint	Custom

The initial point values are the values listed in Table I.

### 11. Calculation of Monte Carlo and Pedigree Results

We used the Crystal Ball software package to perform Monte Carlo analysis to trace out how the uncertainties in the inputs propagate in the calculation and to find out what the contribution of each parameter is to the variance in the calculated total VOC emission from paint.

Table VIII presents the Monte Carlo results for the total VOC emission from paint. In summary, the mean was found to be 52 ktonne VOC and a range  $\pm 15\%$  around that number captures 95% of the distribution whereas a range of  $\pm 40\%$  captures the entire calculated distribution.

Figure 2 presents a sensitivity chart, which summarises the results in terms of relative contribution to variance in the calculated total VOC emission from paint.

We used the pedigree calculator developed by Ellis *et al.* (2000a, 2000b) to calculate how the individual pedigree scores of the inputs and parameters propagate in the calculation chain to obtain a pedigree score for the calculated total VOC emission from paint. The pedigree calculator uses the following rules to determine the pedigree of a number calculated from other numbers (see Box 1). When variables are added, the pedigree for their sum is the mean weighted average (MWA) of their individual pedigrees. When variables are multiplied or divided, a “weak-link” principle is applied and the minimum pedigree of all variables in the calculation is used as the pedigree of the result. Pedigrees for subtraction depend on whether the spread ( $S$ ) of the variables is similar to the difference between them ( $D$ ). When the spread of variables is similar in magnitude to their difference, as tested by the inequality  $D/S < 5$ , then subtraction of the variables yields a result that is more uncertain than the original variables themselves. In this case, the pedigree for the

TABLE VIII  
Results of Monte Carlo calculations of total  
VOC emission from paint

Statistics	Value
Trials	200,000
Mean	51.6 ktonne
Median	51.8 ktonne
Standard deviation	4.1 ktonne
Variance	16.9 ktonne <sup>2</sup>
Skewness	-0.15
Kurtosis	4.64
Coefficient of variability	0.08
Range minimum	32.2 ktonne
Range maximum	74.1 ktonne
Range width	41.9 ktonne
Mean S.E.	0.01

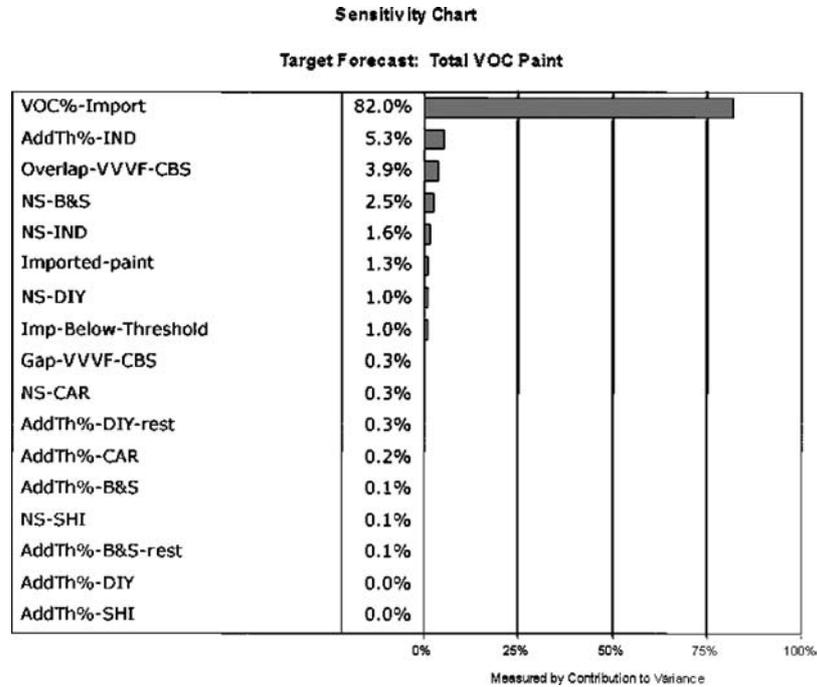


Figure 2. Sensitivity chart ranking the input quantities according to their relative contribution to variance in the calculated distribution (Table VIII) for the total VOC emission from paint.

results of subtraction should be decreased below the MWA, using the rules in Box 1. For a more detailed discussion of the pedigree calculation rules, we refer to Ellis *et al.* (2000a, 2000b).

In all calculations, each of the scores within each pedigree is calculated independently.

**Box 1: Rules for calculating the propagation of pedigree in model calculations**

- I. Multiplication and division
    - A. Use the minimum of all variables.
  - II. Addition (simple, no comparisons, averaging, etc.)
    - A. Use the MWA pedigree of the variables.
  - III. Subtraction (simple). Decrease the pedigrees when the difference between variables is similar to their spread, as judged from  $D/S$ :
    - A. If  $D/S > 5$ , use the MWA.
    - B. If  $D/S < 2$ , reduce the MWA by 50%.
    - C. If  $5 > D/S > 2$ , reduce the MWA by:  $[(5 - D/S) \times 50/3]$ .
- From: Ellis *et al.* (2000b, <http://www.esapubs.org/archive/appl/A010/006/default.htm>).

The resulting pedigree of the total VOC emission from paint scores 2 for proxy, 3 for empirical basis, 3 for methodological rigour and 1 for validation (for scales see Table V).

## 12. Synthesis: The Diagnostic Diagram

For each input quantity, we now have a quantitative and a qualitative metric for uncertainty, namely the relative contribution to the variance in the calculated total VOC emission (Figure 2) and a strength figure obtained by aggregating the pedigree scores assuming equal weight (last column Table VI). These two metrics can be combined in a NUSAP Diagnostic Diagram. The diagnostic diagram is based on the notion that neither spread alone nor strength alone is a sufficient measure for quality. Robustness of calculated 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 uncertainty on the true value of the parameter has no immediate consequences because it has a negligible effect on calculated model outputs. Alternatively, calculated model outputs can be considered robust against parameter spread even if its relative contribution to the total spread in model is high, provided that parameter strength is high. In the latter case, it is likely that the uncertainty in inputs and parameters was not misspecified so that the model outcome adequately reflects the inherent irreducible uncertainty in the model inputs and parameters. In other words, the uncertainty then is a property of the modelled system and does not stem from imperfect knowledge on 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 the setting of priorities for model improvement.

Figure 3 shows the NUSAP Diagnostic Diagram for VOC emissions from paint. For the vertical axis, a logarithmic scale was chosen, to make the diagram maximum discriminative between the inputs and parameters plotted.

The diagnostic diagram clearly shows that the assumption for VOC percentage of imported paint is the weakest spot in the monitoring of VOC emissions. The results are highly sensitive to this assumption, but its pedigree is poor. Further inspection of the diagram draws the attention to a cluster of three of the input quantities whose uncertainty is important: the assumed percentage of additional thinner use for paint applied in industry (AddThin% IND), the overlap between the CBS import number and the VVVF NS numbers, and the import below the reporting threshold for firms to report their imports to the CBS. Finally, there is a cluster of parameters whose uncertainty contributes little to the variance in output but whose pedigree is low, indicating that we may well be ignorant of the true uncertainty ranges in these numbers. Therefore, it is not warranted to conclude that the uncertainties in these quantities are not important, as we would conclude from the sensitivity chart alone.

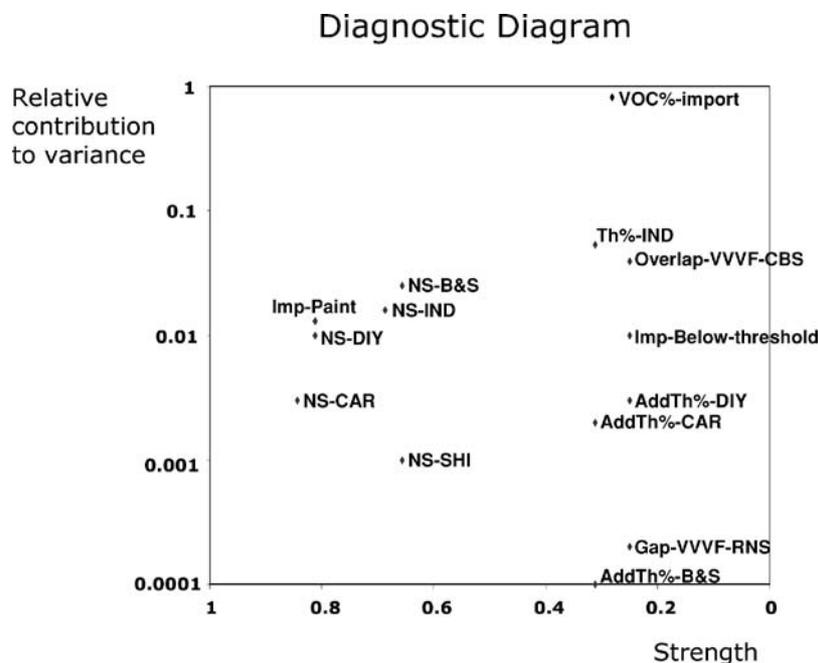


Figure 3. Synthesises our findings of the qualitative and quantitative assessment of uncertainties in VOC emissions from paint.

The quantities are: percentage of additional thinner use for paint applied in DIY and in car refinishing, and the NS by firms that are not a member of the VVVF (Gap VVVF NS statistics and Real NS).

### 13. Discussion

Our Monte Carlo analysis did not take into account uncertainty in the assumption that imported paint can be allocated to the application categories in the same ratio as the national paint sales is distributed over these categories. A change in this assumption affects the VOC percentage in imported paint, as VOC percentage of paint varies across application categories. This allocation assumption was made as a first approximation by RIVM because information on the distribution of imported paint over the categories was absent. We did not include it in the Monte Carlo analysis because we could not think of a meaningful way to represent this uncertainty by a PDF. Instead, we assessed this uncertainty by a sensitivity analysis. As a first indication of the importance of this uncertainty, we tested the extreme cases: allocating all imported paint (the 10.9 ktonne at the bottom of the third column in Table II) to one category. Testing this for all categories yielded a range in total VOC from paint from 49.0 ktonne (if all import is in the ship category) to 52.3 ktonne (if all

import is in the industry category). In other words, the error caused by potential misallocation of imported paint lies between an underestimation of total VOC by less than 1.3% or an overestimation by less than 5.3%.

From the elicitation however, we attained some information on the real allocation. We learned that for the car refinishing category, half of the paint used is imported whereas the other half is produced in The Netherlands. This means that 2.5 ktonne of the 10.9 ktonne VOC in imported paint should be allocated to the car refinishing sector. If we repeat the extreme case sensitivity assessment using this information by allocating 2.5 ktonne VOC imported in the car refinishing category and then putting the remaining 8.4 ktonne VOC in the remaining categories, respectively, the extremes range for total VOC from paint narrows down to 49.5–52.0 ktonne. This corresponds to a relative error caused by potential misallocation of imported paint that lies between an underestimation of total VOC by less than 0.8% or an overestimation by less than 4%. In this calculation, we did not consider that the assumed VOC percentage in imported paint should be higher if such a large share exists of paint for car refinishing.

Our respondent at VVVF further considered it plausible that a substantial part of the imported paint consists of cheap wall-plasters from Germany. He was not able to quantify this amount, but it is important to notice that wall plasters have a low VOC content, meaning that this might imply also that the assumed VOC% in imported paint is somewhat overestimated.

A weak point in our study is the small number of experts from whom we elicited pedigree scores and PDFs for the quantities. Most of the quantities, namely all that are produced by the VVVF, were elicited from one expert only. This is mainly due to the fact that the national paint statistics is an in-house activity of VVVF, meaning that they are the only ones who have detailed expertise on national paint statistics.

The approach we used and especially the expert elicitation step is time intensive. Our elicitation sessions took 3–6 h per session. When applied to cases with many more input quantities than the VOC case this may become a problem. Therefore, in more complex cases, it may be recommendable to take a two-step approach: first a quick-scan to make a selection of parameters deemed relevant and a full analysis for the parameters selected. We implemented such a two-step procedure in another project where we assessed uncertainties in the IMAGE/TIMER Energy model (Van der Sluijs *et al.*, 2001). In that study, we used the Morris algorithm for global sensitivity analysis in combination with expert elicitation to select the 30 most critical parameters out of 300 parameters. By grouping these 30 parameters we brought down the set to 18 items to focus the elicitation of pedigree scores and uncertainty ranges on.

In view of the broader conceptual and methodological uncertainty literature as briefly reviewed in the introduction, this study has covered uncertainty located in data and parameters, and partly in system boundary. It has however not systematically addressed uncertainty in model structure, nor in aggregation, nor in problem

framing and indicator choice. In view of uncertainty dimensions, we addressed inexactness (by PDFs) and unreliability (for as far as it is captured by the four pedigree criteria that we used). We did not systematically address ignorance (we did it partly by drafting a comprehensive inventory of conceivable sources of error) and we addressed only limitedly the societal dimensions of uncertainty, in terms of conceivable motivational biases.

## 14. Conclusions

In this study, we demonstrated application of the NUSAP method in combination with expert elicitation to assess qualitative and quantitative aspects of data- and parameter uncertainty in monitoring VOC emissions from paint.

By means of expert elicitation, we identified about five basic assumptions with limited validity and tenability in the monitoring calculation, about fifteen sources of error and about four conceivable sources of motivational bias. We elicited pedigree and PDFs of all input quantities at the aggregation level that RIVM uses to monitor the VOC emissions from paint. The four pedigree criteria we used were: proxy, empirical basis, methodological rigour and validation. The scores for validation were poor for most input quantities. The NS statistics and import statistics scored reasonable-to-good for proxy, empirical basis and methodological rigour. Weak on all pedigree criteria were the estimates for additional thinner used when paint is applied and the assumption for the VOC percentage in imported paint.

By means of Monte Carlo analysis, we traced out how uncertainty in inputs propagated in the calculation of total VOC emission from paint. We found that a range of  $\pm 15\%$  around the average for total VOC emission from paint (52 ktonne) captures roughly 95% of the calculated distribution, whereas a range of  $\pm 40\%$  captures the entire calculated distribution (after 200,000 Monte Carlo realisations). Note that these calculated figures only capture the uncertainties we were able to quantify as PDFs. We were not able to quantify all error sources identified and we cannot rule out that we overlooked sources of error, so the range is indicative only. Furthermore, one should realise that for a number of quantities the pedigree scores were relatively low, indicating that we may well be ignorant of the actual uncertainty ranges on these quantities.

We also assessed the propagation of the pedigree scores in the calculation of VOC emissions. The resulting pedigree of the total VOC emission from paint scores 2 for proxy, 3 for empirical basis, 3 for methodological rigour and 1 for validation, all on a scale from 0 (poor) to 4 (excellent).

We further analysed the uncertainty by combining the results for sensitivity (as obtained from Monte Carlo analysis, using relative contribution to variance in outcome as metric) and strength (as obtained by aggregating the scores for the pedigree criteria elicited) in a NUSAP Diagnostic Diagram.

The diagnostic diagram identified the assumption for VOC percentage of imported paint as the weakest spot in the monitoring of VOC emissions. The results are highly sensitive to this assumption whereas its pedigree strength is low. Other input quantities in the VOC monitoring calculations whose uncertainty was diagnosed to be 'important' are: the assumed percentage of additional thinner use for paint applied in industry, the overlap between the CBS paint import number and the VVVF national paint sales numbers, and the import below the reporting threshold for firms to report their imports to the CBS.

Although NUSAP broadens the scope of uncertainty assessment compared to the use of statistical approaches only, it does not address all relevant dimensions of uncertainty and does not solve the problem of unquantifiable uncertainty. Various uncertainty methods each address and emphasise other dimensions of uncertainty in a given knowledge base. Although different methods can complement each other, we should not forget that each method provides only a partial insight in what usually is a very complex mass of uncertainties involving technical, methodological, epistemological and societal dimensions.

Overall, we conclude that NUSAP provides a strong diagnostic tool that yields a richer insight in sources and nature of uncertainty than Monte Carlo analysis alone. The NUSAP elicitation procedure stimulates scrutinization of method and underlying assumptions and effectively facilitates structured creative thinking on sources of error. The combination of quantitative and qualitative aspects of uncertainty in data and parameters is helpful in setting priority for uncertainty management and quality improvement.

### Acknowledgements

This study was funded by and carried out in commission of RIVM as a part of the project "Emissie registratie," sub-project "Betrouwbaarheid emissiegegevens" M/888885/04/AA. We thank Mark van Oorschot, Peter Janssen, Kees Peek, Adri Winkelaar, Vera Sol, Raoul Bakkes, G. Slootbeek, and Jo Heylands for contributions to this study.

### References

- 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.
- Ellis, E. C., Li, R. G., Yang, L. Z. and Cheng, X.: 2000a, 'Long-term change in village-scale ecosystems in china using landscape and statistical methods', *Ecol. Appl.* **10**, 1057–1073.
- Ellis, E. C., Li, R. G., Yang, L. Z. and Cheng, X.: 2000b, 'Long-term change in village-scale ecosystems in china using landscape and statistical methods', *Ecol. Appl.* **10**, 1057–1073; 'Supplement 1: Data quality pedigree calculator', *Ecol. Arch.* A010-006-S1. (<http://www.esapubs.org/archive/appl/A010/006/default.htm>).

- Funtowicz, S. O. and Ravetz, J. R.: 1990, *Uncertainty and Quality in Science for Policy*, Kluwer Academic Publishers, Dordrecht.
- InfoMil: 2000, KWS 2000 Annual Report 98–99, Informatiecentrum Milieuvergunningen (InfoMil), The Hague.
- IPCC: 2000, *Good Practice Guidance and Uncertainty Management in National Greenhouse Gas Inventories*, Intergovernmental Panel on Climate Change, Geneva.
- Nowotny, H.: 2003, 'Democratising expertise and socially robust knowledge', *Sci. Public Policy* **30**(3) 151–156.
- ORNL and RFF: 1994, Estimating Fuel Cycle Externalities: Analytical Methods and Issues, *Report 2*, prepared by Oak Ridge National Laboratory and Resources for the Future for the U.S. Department of Energy.
- Risbey, J. S., Van der Sluijs, J. P. and Ravetz, J.: 2001, Protocol for the Assessment of Uncertainty and Strength in Emission Monitoring, Department of Science, Technology and Society, Utrecht University, *STS report E-2001-10*, 18 pp. Available from [www.nusap.net](http://www.nusap.net)
- Risbey, J., van der Sluijs, J., Kloprogge, P., Ravetz, J., Funtowicz, S. and Corral Quintana, S.: 2004, 'Application of a checklist for quality assistance in environmental modelling to an energy model', *Env. Mod. Assess.* (in press).
- RIVM: 1999, Milieubalans 1998. Het Nederlandse milieu verklaard. RIVM, Bilthoven.
- RIVM: 2001, Scientific Audit of Integrated Environmental Policy Assessment (IEPA), RIVM, Bilthoven, 28 pp.
- Saltelli, A., Chan, K. and Scott, E. M. (eds): 2000, *Sensitivity Analysis, Probability and Statistics Series*, Wiley, New York, 504 p.
- Saltelli, A., Tarantola, S., Campolongo, F. and Ratto, M.: 2004, *Sensitivity Analysis in Practice, A Guide to Assessing Scientific Models*, Wiley, New York, 200 pp.
- Slootbeek, G.: 1997, *The Weighting Process of the Redesign of the Netherlands Foreign Trade Survey*, CBS Statistics Netherlands, Division of Research and Development, Heerlen.
- Spetzler, C. S. and Steal von Holstein, C. A. S.: 1975, 'Probability encoding in decision analysis' *Manage. Sci.* **22**(3) 340–358.
- Van Amstel, A. R., Olivier, J. and Ruysenaars, P. G. (eds): 2000, *Monitoring of Greenhouse Gases in the Netherlands: Uncertainty and Priorities for Improvement*, WIMEK/RIVM report 773201003, Bilthoven.
- Van de Pol, F. and Diederer, B.: 1996, *A Priority Index for Macro-Editing The Netherlands Foreign Trade Survey*, CBS Statistics Netherlands, Division of Research and Development, Heerlen.
- Van der Sluijs, J. P.: 1997, 'Anchoring Amid Uncertainty, on the Management of Uncertainty in Risk Assessment of Anthropogenic Climate Change', *Ph.D. Thesis*, 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. S., Van Vuuren, D., De Vries, B., Beusen, A., Heuberger, P., Corral Quintana, S., Funtowicz, S., Kloprogge, P., Nuijten, D., Petersen, A. and Ravetz, J.: 2001, Uncertainty Assessment of the IMAGE/TIMER B1 CO<sub>2</sub> Emissions Scenario, Using the NUSAP Method, Dutch National Research Program on Climate Change, *Report no. 410 200 104*, 227 pp. Available from [www.nusap.net](http://www.nusap.net)
- Van der Sluijs, J. P., Risbey, J. S. and Ravetz, J.: 2002, Uncertainty Assessment of VOC Emissions from Paint in the Netherlands, Department of Science, Technology and Society, Utrecht University, *STS report E-2002-13*, 27 pp. Available from [www.nusap.net](http://www.nusap.net)
- Van der Sluijs, J. P., Risbey, J. S., Kloprogge, P., Ravetz, J. R., Funtowicz, S. O., Corral Quintana, S., Guimaraes Pereira, A., De Marchi, B., Petersen, A. C., Janssen, P. H. M., Hoppe, R. and Huijs, S. W. F.: 2003, *RIVM/MNP Guidance for Uncertainty Assessment and Communication: Detailed Guidance*, Copernicus Institute for Sustainable Development and Innovation, Utrecht University and RIVM-MNP, Utrecht, ISBN 90-393-3536-2, 71 pp. Available from [www.nusap.net](http://www.nusap.net)

- VROM: 1989, *Project KWS 2000, Bestrijdingsstrategie voor de emissies van vluchtige organische stoffen*, Projectgroep Koolwaterstoffen 2000, Ministry of Housing, Physical Planning and Environment, The Hague.
- VROM: 1998, *National Environmental Policy Plan 3*, Ministry of Housing, Physical Planning and Environment, The Hague.
- VVVF: 1999, *Statistieken 1998 VVVF*, Vereniging van Verf- en Drukinktfabrikanten, Leiden.
- Walker, W. E., Harremoës, P., Rotmans, J., Van der Sluijs, J. P., van Asselt, M. B. A., Janssen, P. and Kreyer von Krauss, M. P.: 2003, 'Defining uncertainty a conceptual basis for uncertainty management in model-based decision support', *Integrated Assess.* **4**(1), 5–17.