

1 **EUFRAM**

2 **Concerted action to develop a European Framework for probabilistic risk**
3 **assessment of the environmental impacts of pesticides¹**

4 **EUFRAM REPORT, VOLUME 1**

5 **INTRODUCING PROBABILISTIC METHODS INTO THE**
6 **ECOLOGICAL RISK ASSESSMENT OF PESTICIDES²**

7 **Version 6, 11 May 2005**

8
9 **THIS IS THE FIRST PUBLIC VERSION OF THIS DOCUMENT**

10 **PLEASE HELP US TO IMPROVE IT**

11 **BY PROVIDING FEEDBACK AND COMMENTS VIA THE EUFRAM WEBSITE**

12 **www.eufram.com**

13
14 **Feedback can be provided in 3 ways:**

- 15 • **Comments on EUFRAM's Provisional Conclusions. These are highlighted in**
16 **boxes in the text and numbered PC1-PC9**
- 17 • **Responses to Questions highlighted in the text, numbered Q1-Q3**
- 18 • **Comments on any other part of the document, by line number**

19 **Please submit comments using the feedback database at www.eufram.com**

20

¹ EUFRAM is supported by European Commission's 5th Framework Programme (www.cordis.lu), contract number QLK5 - CT 2002 01346. Further information and news about EUFRAM and downloads of this and other project documents are available at www.eufram.com.

² This document is the responsibility of its publishers and does not represent the views of the European Commission or its services.

20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45

TABLE OF CONTENTS

1	Summary	3
2	Introduction.....	7
2.1	Origin, purpose and status of this document	7
2.2	What is probabilistic risk assessment?.....	7
2.3	Deterministic methods.....	9
2.4	Proposed role of probabilistic methods under 91/414/EEC	10
2.5	Validation of probabilistic methods	17
3	Framework for probabilistic risk assessment.....	18
3.1	Interactions between risk assessors and decision-makers.....	18
3.2	Problem definition.....	19
3.3	Planning the assessment model	23
3.4	Key methods for probabilistic assessment	25
3.5	Interpreting the results of probabilistic assessments.....	35
3.6	Reporting the results of probabilistic assessments.....	38
3.7	Communicating results of probabilistic assessments	39
3.8	Potential for harmonised approaches.....	43
3.9	Summary checklist	43
4	Case studies of probabilistic assessment.....	44
5	List of Contents for Volume 2	44
6	List of Contents for Volume 3	44
7	Glossary and abbreviations	44
8	References	47
	Annex 1. Members of the EUFRAM project consortium.	50

45 1 SUMMARY

46 Directive 91/414/EEC requires that if a plant protection product³ fails to pass
47 preliminary, first-tier assessment criteria for environmental risk, then it may not be
48 authorised for use unless an “appropriate risk assessment” shows that it will cause
49 no unacceptable impact. Various options for these refined, higher tier risk
50 assessments are identified in existing EU Guidance Documents, including
51 probabilistic approaches. Until now, however, probabilistic approaches have gained
52 only limited acceptance, partly due to a lack of guidance on how to implement and
53 evaluate them.

54 Therefore, **this draft document aims to provide a framework of basic concepts,**
55 **principles and methods that will help users to conduct, report, evaluate and**
56 **communicate probabilistic assessments in appropriate ways. It is aimed**
57 **primarily at risk assessors in government, industry and consultancy**
58 **companies.** Volumes 2 and 3 provide more detail and worked examples (case
59 studies). The three volumes are being developed by the EUFRAM project, an EU-
60 funded concerted action involving 29 organisations including regulatory authorities,
61 government research institutes, agro-chemical companies, consultancy companies
62 and universities.

63 **EUFRAM does not have a formal status in relation to Directive 91/414/EEC, so**
64 **this document should not be regarded as formal guidance.** It will be refined and
65 revised to take account of feedback from all interested parties, including two
66 workshops scheduled for October 2005 and July 2006⁴.

67 **The defining feature of probabilistic risk assessments is that they quantify**
68 **variability and/or uncertainty.** Potential benefits of quantifying variability include:

- 69 • Increased realism, representing real-world variation in factors that influence risk
- 70 • The opportunity to replace or refine worst-case assumptions
- 71 • Provides an alternative to conducting higher tier laboratory or field studies
- 72 • Makes more use of the available data.

73 Potential benefits of quantifying uncertainty include:

- 74 • Provides an objective basis for discussions about reducing uncertainty factors
75 (e.g. the TER thresholds of 10 and 100) when additional data is provided
- 76 • Indicates the influence of quantified uncertainties on the assessment outcome
- 77 • May help increase the cost-effectiveness of higher tier studies by targeting them
78 on major sources of uncertainty.

79 Almost all current approaches for first-tier environmental risk assessment under
80 Directive 91/414/EEC already incorporate some probabilistic elements, e.g. the use
81 of a 90th percentile estimate for spray drift. After considering their strengths and
82 weaknesses, **EUFRAM concludes that there is scope for deploying probabilistic**
83 **approaches to a greater extent, as one of several alternatives for higher-tier**
84 **assessment.** They can be applied either to exposure assessment, or effects
85 assessment, or both.

³ Plant protection product is the formal term for pesticides, safeners and plant growth regulators that are within the scope of Directive 91/414/EEC. For simplicity, the word “pesticide” is used in this document to cover all types of plant protection product.

⁴ See www.eufram.com for details of the EUFRAM project and workshops.

86 The main steps of a probabilistic assessment are as follows:

- 87 1. **Define the assessment objectives.** The objectives should reflect as closely as
88 possible the information needs and protection goals of decision-makers. They
89 should define the pesticide use, non-target organisms and types of effects to be
90 considered, what types of variation are of interest, and whether confidence
91 intervals are required with the outputs.
- 92 2. **Define the assessment endpoint,** that is, the primary output of the probabilistic
93 assessment. In particular, define what measure to use for the magnitude of
94 hazard, exposure or risk, for what ensemble or group of entities (e.g. species or
95 community) it should be estimated. It is essential to ensure that these choices
96 provide an output that will be meaningful and relevant to decision-makers.
- 97 3. **Identify the key factors and mechanisms** that influence the effect that is to be
98 assessed, and how they interact, and develop an assessment model to represent
99 them.
- 100 4. Consider each part of the model in turn, and **decide which factors and**
101 **mechanisms might contribute significantly to variability** in the assessment
102 endpoint. These will be represented by distributions in the assessment model.
- 103 5. **For each input variable, identify the appropriate ensemble.** This depends on
104 the ensemble of the assessment endpoint and the model structure. For example,
105 if the output is frequency of effects in an ensemble of water bodies, then the
106 appropriate ensemble for exposure is concentrations for different water bodies.
- 107 6. **Identify what data are available** that can help in quantifying each factor and
108 mechanism, including distributions for those that will be treated as variables.
- 109 7. **Decide whether any extrapolations or adjustments are needed** to model the
110 key factors and mechanisms from the available data, e.g. to account for lab-to-
111 field extrapolation or non-random sampling. If adjustment or extrapolation factors
112 are required, then they should be identified as part of the assessment model.
- 113 8. Consider each part of the model in turn to **identify possible sources of**
114 **uncertainty.** Decide which uncertainties might contribute significantly to
115 uncertainty in the assessment endpoint, and which of these will be quantified
116 using distributions (if any). Decide what to do about important uncertainties that
117 cannot or will not be quantified (e.g. use conservative values).
- 118 9. Consider carefully for each input distribution whether it should be treated as
119 contributing uncertainty or variability to the assessment output.
- 120 10. **Identify potential dependencies** affecting the assessment. Dependencies occur
121 where the value of one variable depends upon the value of another variable (e.g.
122 food intake may be positively related to body weight) and can have a major
123 impact on the assessment outcome.
- 124 11. Express the entire assessment model as a set of mathematical equations, so that
125 they can be used for calculations.
- 126 12. **Select appropriate computational or graphical methods** for combining the
127 input distributions to obtain the assessment output. This choice will depend on
128 various factors including the number of input distributions, whether a confidence
129 interval is required for the output, the approaches that will be used to handle

130 uncertainties and dependencies, and ease of use. More specific
131 recommendations will be developed during the remainder of the project.

132 **13. Specify distributions for the input variables.** This requires expertise in
133 statistics as well as expert knowledge of each variable and how it relates to the
134 assessment output.

135 **14. Consider conducting sensitivity analyses.** These may help in various ways,
136 e.g. in deciding which sources of variation and uncertainty should be quantified.

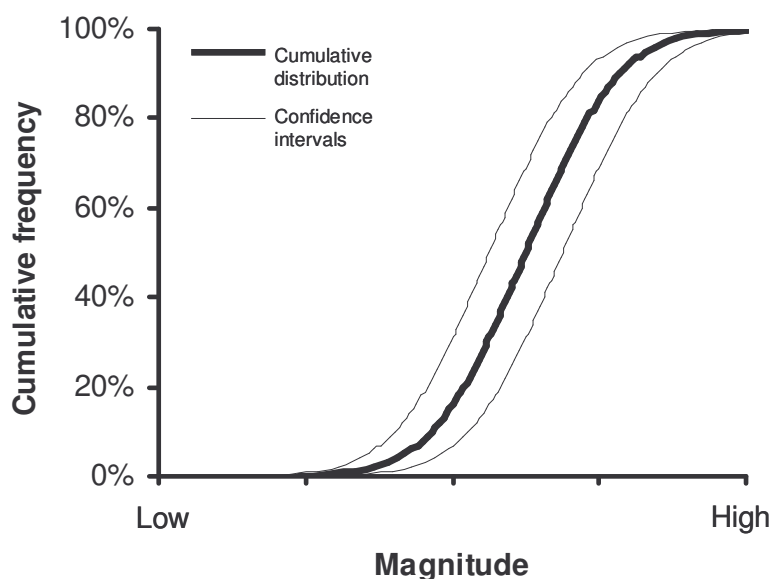
137 **15. Select appropriate software** to carry out the computations and generate
138 outputs. EUFRAM has defined desirable characteristics for probabilistic software
139 and databases. Existing tools meet these criteria to a limited extent and can be
140 used for some types of pesticide assessment, but further development is highly
141 desirable.

142 **16. Check the outputs of the probabilistic assessment and consider their**
143 **implications for decision-making.** This will have to be done case-by-case, as
144 for other refined assessments, unless standard criteria for decision-making are
145 established.

146 Probabilistic methods can produce many different types of output. Consultations with
147 EUFRAM end-users indicate a desire to facilitate communication by adopting a single
148 format as standard. **EUFRAM provisionally recommends cumulative**
149 **distributions (with confidence intervals if required) as a preferred choice for**
150 **the primary output of probabilistic assessments.** Cumulative distributions can be
151 used to represent three basic dimensions of risk: magnitude, frequency and
152 uncertainty (e.g. Figure 1). The x-axis shows the magnitude of hazard, exposure or
153 risk (how bad), the y-axis shows the frequency (how often), and the confidence
154 intervals show uncertainty (how sure). It can be used to read off the frequency of
155 effects below (or above) any given magnitude, together with confidence intervals if
156 required. Different measures of magnitude and frequency can be used, according to
157 the needs of the assessment. For example, one possible output for an aquatic
158 assessment might be the percentage of water bodies (frequency) with more than 5%
159 of species affected (magnitude). One possible output for an avian assessment might
160 be the percentage of individual skylarks (frequency) that have toxicity-exposure ratios
161 less than 10 (magnitude). In addition, quantitative results should always be
162 accompanied with a written statement, expressing and interpreting them in language
163 that can be understood by non-specialists.

164 It is never possible to quantify all sources of uncertainty and variability in an
165 assessment, but quantifying a few may be sufficient to reach a regulatory decision.
166 However, the quantitative output of an assessment should always be accompanied
167 by a list of **unquantified sources of variability, uncertainty and dependency**, and
168 a qualitative assessment of their potential influence on the assessment endpoint.

169 **The results of probabilistic assessments should be considered together with**
170 **conventional deterministic results and other lines of evidence** (e.g. field studies
171 or monitoring), to arrive at overall conclusions. This may include consideration of the
172 **wider ecological consequences** of predicted impacts (e.g. extrapolation from
173 effects on individual organisms to consequences for the wider population).



174
 175 *Figure 1. Generalised example of a cumulative distribution. This is provisionally*
 176 *recommended by EUFRAM as the preferred format for graphical output of probabilistic*
 177 *assessments.*

178 **The formal report of a probabilistic assessment** should document and justify the
 179 methods, results and conclusions clearly and concisely, but in sufficient detail to
 180 enable critical evaluation of all stages by other specialists (e.g. peer reviewers). It
 181 should include an executive summary, communicating the main points required for
 182 decision-making. Sufficient background information for other specialists to duplicate
 183 the assessment should be provided in appendices.

184 **Effective communication is essential** if probabilistic approaches are to be
 185 accepted. Different approaches are required when communicating to different
 186 audiences, e.g. technical specialists, decision-makers and the public. The format of
 187 graphical, tabular and textual outputs needs careful and detailed consideration, to
 188 maximise their effectiveness. Provisional recommendations in this document will be
 189 refined in the light of further experience and feedback as EUFRAM progresses.

190 **Validation** is possible only to a limited extent for both deterministic and probabilistic
 191 approaches. Consequently, thorough **peer review** by relevant technical experts will
 192 be a key requirement for the acceptance of new approaches, and for the acceptance
 193 of individual probabilistic assessments.

194 Probabilistic approaches are still evolving and it would be premature to attempt
 195 harmonisation at the present state of the art. Nevertheless, if a probabilistic
 196 assessment has been conducted for one pesticide and accepted by the relevant
 197 authorities, then much of the approach may be directly transferable to other, similar
 198 pesticides with similar use patterns. This transferability of approaches opens the
 199 possibility of establishing **generic, peer-reviewed probabilistic approaches and**
 200 **tools for scenarios that frequently require refined assessment under**
 201 **91/414/EEC**, analogous to the FOCUS models and scenarios for exposure
 202 assessment. This would increase efficiency both for people conducting probabilistic
 203 assessments, and also for people evaluating them. However, **flexibility is important**
 204 and it should remain open for assessors and decision-makers to select other
 205 approaches where appropriate.

206 2 INTRODUCTION

207 2.1 Origin, purpose and status of this document

208 This draft document has been developed by an EU-funded “concerted action” project
209 called EUFRAM. The project consortium is listed in Annex 1 and comprises 29
210 organisations including regulatory authorities, government research institutes, agro-
211 chemical companies, consultancy companies and universities.

212 The overall aim of EUFRAM is to assist the implementation of probabilistic methods
213 for assessing the environmental risks of plant protection products in Europe. The
214 main outputs are being developed as a report in 3 volumes:

- 215 • **Volume 1 (this document)** – a framework of basic principles and methods for
216 probabilistic assessment, together with summaries of selected examples,
- 217 • **Volume 2** – eight chapters providing more detail on selected aspects of
218 probabilistic approaches,
- 219 • **Volume 3** – detailed case studies developed or evaluated during the project.

220 These documents are aimed primarily at potential users and evaluators of
221 probabilistic approaches in government, industry and consultancy companies. The
222 ultimate objective is to provide **a framework of basic concepts, principles and**
223 **methods** that will help users to **conduct, report, evaluate and communicate**
224 probabilistic assessments in appropriate ways.

225 The framework will be refined and revised as the EUFRAM project progresses. Early
226 drafts aim to help potential users explore the range of probabilistic approaches and
227 evaluate their usefulness for pesticide assessments. Users are encouraged to
228 provide feedback via the EUFRAM website and via 2 workshops in October 2005 and
229 July 2006 (see www.eufram.com for details). The feedback will help EUFRAM
230 partners to refine the documents and identify which approaches are most promising.
231 It is hoped that the final version of the framework at the end of the EUFRAM project
232 (December 2006) will assist in the implementation of probabilistic approaches within
233 the context of European pesticide authorisation.

234 **EUFRAM does not have a formal status in relation to Directive 91/414/EEC, so**
235 **this document should not be regarded as formal guidance. It is currently in an**
236 **early draft form so later versions may change substantially.**

237 2.2 What is probabilistic risk assessment?

238 EUFRAM defines **probabilistic risk assessment** as:

239 ***Risk assessments that use probabilities or probability distributions to***
240 ***quantify one or more sources of variability and/or uncertainty in***
241 ***exposure and/or effects and the resulting risk.***

242 This definition itself contains several terms that require definition.

243 **Variability** is defined as ***Real variation in factors that influence risk***. For example,
244 toxicity varies between species, and exposure varies in time and space. Variability

245 matters because risk assessment usually needs to address a range of relevant
246 species and exposures, not just one particular species and one exposure.

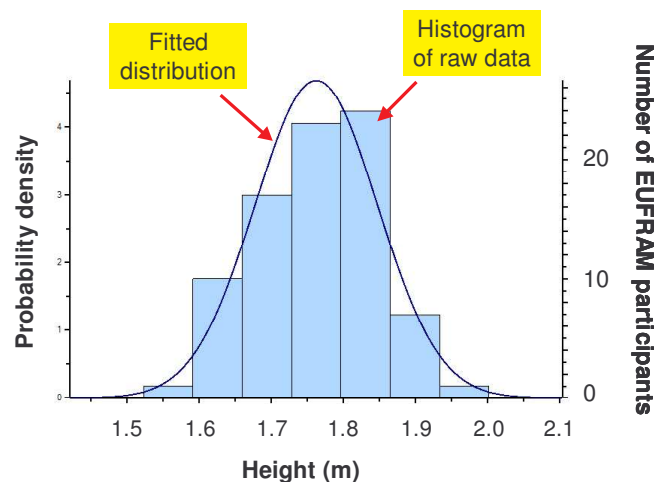
247 **Uncertainty** is defined as **Limitations in knowledge about factors that influence**
248 **risk**. For example, there is uncertainty when we extrapolate toxicity from a small
249 number of tested species to other, untested species, and uncertainty when we
250 extrapolate from mathematical models of exposure to the real world. Uncertainty
251 matters because decision-makers and stakeholders need to know how different the
252 real impacts might be from the scientists' best estimates.

253 An important practical difference between uncertainty and variability is that
254 uncertainty can often be reduced by obtaining further data, whereas variability can be
255 better quantified but not reduced by further data.

256 **Probabilities** can be used to quantify variability and/or uncertainty. The probability of
257 something can be defined as its frequency in repeated independent trials, or as the
258 degree of belief that it will occur, expressed as a proportion (i.e. a number between 0
259 and 1). For example, when tossing an unbiased coin, it is expected to land on
260 "heads" on 50% of occasions, so the probability of obtaining "heads" is 0.5. Similarly,
261 if you know or believe that half the people in a population are taller than 1.8m, then
262 you can express the probability of heights above this value as 0.5.

263 The most familiar form for a **probability distribution** is a graph showing the relative
264 probabilities of different values of a variable such as height. For example, Figure 2
265 shows the variation in height of participants at the first EUFRAM workshop in two
266 ways: as a histogram showing the raw frequencies (the number of participants in
267 particular height ranges), and as a curve showing a "Normal" distribution⁵ fitted to the
268 raw data. This curve is a "probability density function" (PDF) and shows the relative
269 frequency of occurrence for each point on the x-axis.

270



271

272 *Figure 2. An example of a probability distribution: variability in the height of*
273 *participants at the first EUFRAM workshop, in March 2005.*

274 So, **probabilistic risk assessment** uses probabilities and probability distributions to
275 take account of variability and uncertainty in risk assessment. It can be applied to any

⁵ The Normal distribution is a symmetrical, bell-shaped, theoretical distribution, which fits well to many biological variables such as height and weight and can be defined mathematically by its mean and variance (or standard deviation).

276 type of risk calculation, including those already used for plant protection products
277 under Directive 91/414/EEC. The basic steps of a probabilistic assessment are:

- 278 • Define the objective of the assessment and decide what form of probability or
279 distribution is required for the assessment output;
- 280 • Identify one or more inputs to the risk assessment, for which variability and/or
281 uncertainty is to be considered, and quantify them using appropriate probabilities
282 or distributions;
- 283 • Use appropriate methods to combine the different input distributions and produce
284 the distribution for the assessment output, showing the variability and uncertainty
285 of the predicted impacts.
- 286 • Interpret and communicate the results.

287 In principle, distributions can be used to quantify variability and uncertainty for any
288 number of inputs to a risk assessment, and for any output of the assessment that
289 may be of interest to the decision-maker.

290 **2.3 Deterministic methods**

291 **Deterministic methods** are defined as:

292 ***Methods that use point estimates⁶ to represent one or more factors in a***
293 ***risk assessment and treat them as if they were fixed and precisely known.***

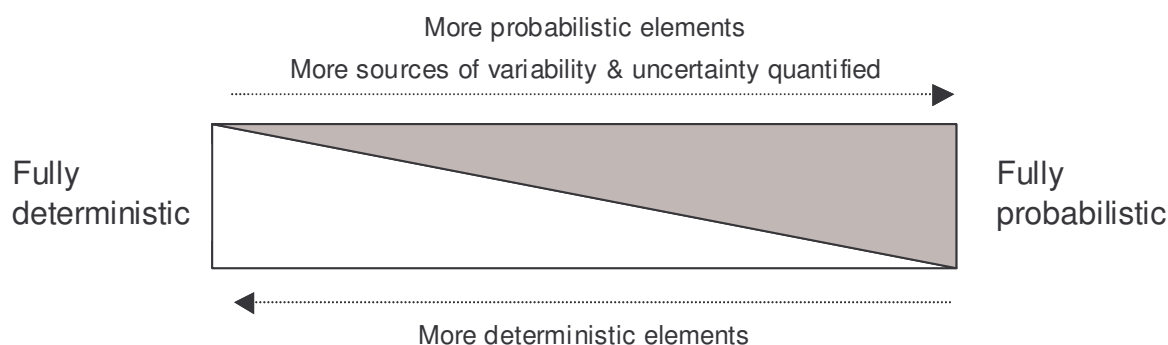
294 Most risk assessments include at least some deterministic elements, but few are
295 wholly deterministic. This is because many of the point estimates that are used for
296 exposure and/or toxicity have in fact been derived from probabilistic calculations.

297 For example, the first tier aquatic assessment for spray drift under Directive
298 91/414/EEC estimates the TER for a scenario using the 90th percentile of a
299 distribution for spray drift; and the first tier avian assessment incorporates estimates
300 of residues on food items that are intended to be approximate 90th percentiles.
301 Assessments for most taxonomic groups use LC50, LD50 or EC50 measures of
302 toxicity, which are derived from probit curves representing variation in sensitivity
303 between individuals. Similarly, fixed uncertainty or assessment factors used in risk
304 assessment may be derived from quantitative analysis of uncertainties⁷. Of the first
305 tier assessments for ecological risk under Directive 91/414/EEC, perhaps only the
306 approach for soil micro-organisms is entirely deterministic.

307 Thus almost all assessments under 91/414/EEC already incorporate at least some
308 probabilistic elements. On the other hand, as will be seen later, it is practically
309 impossible to quantify absolutely every source of variability and uncertainty affecting
310 an assessment. In fact, most assessments are neither fully deterministic nor fully
311 probabilistic, but somewhere in between (Figure 3).

⁶ Point estimates represent a measured or estimated quantity by a single number (e.g. the minimum, mean or maximum value), rather than a distribution.

⁷ For example, assessment factors for TERs are specified in Annex VI of Directive 91/414/EEC, although it is not specified which uncertainties they address nor how they were derived.



312
313 *Figure 3. Continuum between wholly deterministic and wholly probabilistic risk*
314 *assessments.*

315 So the question is not whether to start doing probabilistic assessments, but whether it
316 may be helpful to include more probabilistic elements than we already do and – if so
317 – when and how to do it.

318 **2.4 Proposed role of probabilistic methods under 91/414/EEC⁸**

319 **2.4.1 Compatibility with existing legislation and guidance**

320 Probabilistic risk assessment is not mentioned in Directive 91/414/EEC or its
321 Annexes, but the “unless” clauses in Annex VI can be interpreted as providing the
322 possibility for probabilistic methods to be used. For example, in relation to aquatic
323 organisms Annex VI states that a plant protection product that fails the preliminary,
324 first-tier assessment shall not be authorised “unless it is clearly established through
325 an appropriate risk assessment that under field conditions no unacceptable impact
326 on the viability of exposed species occurs”. This opens the way for probabilistic
327 assessments at higher tiers of the assessment process if they are considered
328 “appropriate” by the responsible authorities. Furthermore, using distributions to
329 represent variation in exposure and toxicity can be regarded as one way to take
330 account of “field conditions” and “exposed species”.

331 Current EU Guidance Documents for both aquatic and terrestrial ecotoxicology state
332 that traditional deterministic assessment methods have limitations that could be
333 overcome by probabilistic approaches. The Aquatic Guidance Document (European
334 Commission, 2002a) states that probabilistic risk assessment is usually a tool for
335 higher-tier assessments and hence its suitability needs to be considered case-by-
336 case. The Terrestrial Guidance Document (European Commission, 2002b) states
337 that probabilistic methods are promising tools and already now there may be
338 situations where their use could be envisaged.

339 **2.4.2 Potential benefits of probabilistic methods**

340 Probabilistic approaches enable the risk assessor to quantify and account for some
341 major sources of variability and uncertainty affecting risk assessment.

⁸ Detailed discussion of the role and outputs of probabilistic methods can be found in Appendix 2, Work Package 3. Additional material comes from feedback received at the EUFRAM workshop in Brussels, March 2005.

342 Potential benefits of quantifying variability include:

- 343 • Increased realism through representing more fully variation in the real world and
- 344 its influence on the risk
- 345 • The opportunity to replace worst-case assumptions with more realistic ones
- 346 • May be a more cost-effective option for refining the assessment than conducting
- 347 higher tier laboratory or field studies
- 348 • Makes more use of the available data (i.e. variances or individual data points).

349 Potential benefits of quantifying uncertainty include:

- 350 • Provides an objective basis for discussions about reducing uncertainty factors
- 351 (e.g. the TER thresholds of 10 and 100) when additional data is provided
- 352 • Provides an indication of the combined influence of the quantified uncertainties on
- 353 the assessment outcome
- 354 • By identifying major sources of uncertainty, it may help the targeting of higher tier
- 355 studies so as to maximise their cost-effectiveness in reducing uncertainty.

356 **2.4.3 Potential weaknesses of probabilistic methods**

357 The Aquatic and Terrestrial Guidance Documents (European Commission 2002a,b)

358 list some potential disadvantages of probabilistic approaches:

- 359 • A lack of reliable information for specifying distributions of many input parameters,
- 360 • Concerns about the validity of assumptions (e.g. representativeness of tested
- 361 species),
- 362 • The lack of common standard methods for the statistical calculations.

363 The EUPRA workshop gave a detailed analysis of potential disadvantages of

364 probabilistic methods (Table 1 in Hart, 2001), under the following headings:

- 365 • Probabilistic methods are more complex,
- 366 • Some probabilistic methods require more data,
- 367 • Probabilistic approaches and outputs are difficult to communicate,
- 368 • There is a risk of misleading results,
- 369 • There is no established guidance on what outputs are required,
- 370 • Validation of probabilistic methods is difficult.

371 EUPRA also provided a detailed list of recommended actions that could be taken to

372 address these potential disadvantages (Table 1 in Hart, 2001). These can be

373 summarised as:

- 374 • Develop a framework for the appropriate use of probabilistic methods in the
- 375 regulatory process, including key principles for how to conduct, report and
- 376 communicate probabilistic assessments,
- 377 • Develop a set of case studies to illustrate the use of the framework,
- 378 • Improve access to existing data,
- 379 • Evaluate probabilistic approaches for limited datasets,
- 380 • Validate probabilistic approaches, to the extent that is possible,
- 381 • Provide training and expert advice,
- 382 • Adopt standard computer software and databases.

383 EUFRAM includes activities aimed at progressing or facilitating all of these

384 recommended actions. The extent to which these actions are succeeding in

385 addressing the potential weaknesses of probabilistic methods, and what remains to

386 be done, will be reviewed at the remaining EUFRAM workshops in October 2005 and
387 July 2006.

388 Both EUPRA (Hart, 2001) and the Aquatic Guidance Document (European
389 Commission, 2002a) identify an additional potential difficulty: the lack of established
390 criteria for using probabilistic results in decision-making (e.g. what percentage of
391 species may be affected?). This is a very important issue but requires balancing
392 ecological risk against other relevant factors that are outside the scope of EUFRAM,
393 such as the benefits of pesticide use⁹ and social and aesthetic considerations.
394 EUFRAM is trying to contribute by encouraging the use of probabilistic outputs that
395 are as relevant as possible to the concerns of decision-makers, but ultimately
396 decision criteria would need to be set by the responsible authorities in consultation
397 with other interested parties as appropriate. While such criteria are lacking, decisions
398 will have to be considered case-by-case, as is already done for most other types of
399 refined assessment.

400 **2.4.4 Provisional EUFRAM recommendations on the role of probabilistic** 401 **methods**

402 Having considered the potential benefits and weaknesses of probabilistic
403 approaches, the EUFRAM project proposes the following provisional conclusions:

404 **PC1. EUFRAM provisionally concludes that there is scope for more use of**
405 **probabilistic approaches, as one of several alternatives for higher-tier refined**
406 **assessment.** They can be used to investigate potential risks identified by
407 conventional lower-tier assessments, provided that the methods and assumptions
408 are fully documented and justified. They help to increase the realism of the
409 assessment and can be used in conjunction with other higher-tier approaches such
410 as additional ecotoxicity testing and refined exposure assessment.

411 **PC2. It is not necessary or possible to quantify all sources of uncertainty and**
412 **variability.** In some cases, a probabilistic assessment that quantifies only one
413 source of uncertainty or variability in either exposure or effects may be sufficient to
414 enable a regulatory decision to be reached, provided that other sources of
415 uncertainty and variability are adequately addressed in other ways (e.g. by the use of
416 appropriate assumptions).

417 **NOTE:** These conclusions are provisional and will be reviewed at future EUFRAM
418 workshops. Comments are welcome at any stage via the project website
419 www.eufram.com.

420 **2.4.5 What types of output are required?**

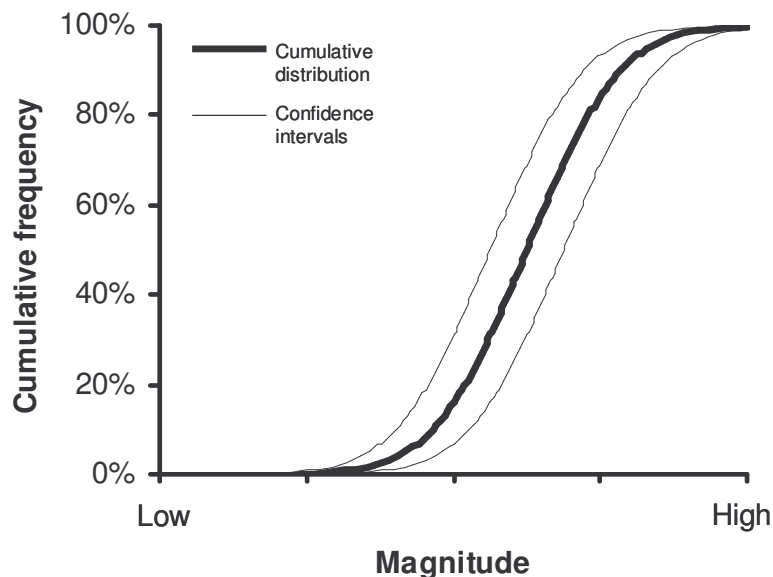
421 The type of outputs required can be considered at several levels. In this section,
422 EUFRAM proposes a general preference for the format of the immediate quantitative
423 outputs of probabilistic assessments. The particular form of this output required for
424 individual assessments should be specified case by case, as part of “problem
425 definition” (Section 3.2). The format and content of documents reporting probabilistic
426 assessments is discussed in Section 3.5, and strategies for communicating the
427 results to different audiences are discussed in Section 3.6.

⁹ Note, however, that recital no. 9 in the opening part of Directive 91/414/EEC states that risk should take priority over (i.e. be given more weight than) the objective of improving plant production.

428 Probabilistic risk assessments in which the calculations are done with point values
429 derived from probability distributions will produce point values as the immediate
430 output. When presenting this type of output it is important to ensure that the user is
431 aware which point values were taken from each underlying distribution (e.g. 90th
432 percentile of spray drift distribution, EC50 from a toxicity test) as this affects the
433 conservatism of the assessment and should therefore be taken into account in
434 interpretation and decision-making.

435 Probabilistic risk assessments in which the calculations are done directly with
436 distributions will generally produce distributions for the outputs. Output distributions
437 could potentially be presented in several alternative formats. One possible strategy
438 would be to use different formats for different purposes, according to what is most
439 effective in each situation. However, feedback at the first EUFRAM workshop
440 suggests that using multiple graphical formats would be confusing and that it may be
441 better to standardise on a single graphical format for all purposes¹⁰.

442 Feedback at the first EUFRAM workshop suggested that the most popular choice for
443 a single graphical format would be the cumulative distribution function (CDF) showing
444 the magnitude and frequency of hazard, exposure or risk, together with confidence
445 intervals. A generalised example is shown in Figure 4. Outputs of this type can be
446 generated with existing software, although improvements are required. The proposal
447 to recommend this format is provisional and will be reviewed at the next EUFRAM
448 workshop.



449

450 *Figure 4. Generalised example of a cumulative distribution function (CDF), together*
451 *with confidence intervals. The y-axis shows the frequency with which the magnitude*
452 *of hazard, exposure or risk is less than any value on the x-axis. The measure of*
453 *hazard, exposure or risk will vary from assessment to assessment, as appropriate.*

454 The cumulative distribution function illustrated in Figure 4 can be understood in
455 several ways, ranging from informal to formal:

¹⁰ It will generally be helpful to present the preferred graphical output together with non-graphical formats including narrative text and tables (see later in this Section and also Section 3.6).

- 456 • Informally, the graph can be viewed as expressing three basic dimensions of risk:
457 magnitude, frequency and uncertainty. The x-axis shows magnitude (how bad),
458 the y-axis shows frequency (how often), and the confidence intervals show
459 uncertainty (how sure). For example, an aquatic assessment might estimate the
460 percentage of water bodies (how often) in which different proportions of species
461 are affected (how bad).
- 462 • The graph can be viewed as the result of stacking up all the estimated
463 magnitudes in ascending order, from lowest to highest, marking the y axis as
464 proportion of the ensemble of entities, and then adding a confidence interval at
465 each point on the curve.
- 466 • The y-axis shows the proportion of values less than or equal to any point on the x-
467 axis.
- 468 • The confidence intervals show a minimum estimate of the degree of uncertainty
469 about the true position of the CDF. The estimate is a minimum because it only
470 represents those uncertainties that have been quantified.
- 471 • Formally, the CDF is a function $F(x)$ expressing the probability that a variable X
472 assumes a value less than or equal to some value x . It can be obtained from the
473 probability density function (PDF, e.g. Figure 2) by integrating the probability
474 density from the minimum value of X up to any value x , and is equal to the area
475 under the PDF curve from the minimum value of X up to any value x .

476 The CDF is a less familiar graphical format for a distribution than the histogram and
477 probability density function (PDF) shown in Figure 2, but it has two big advantages.
478 First, it is easy to read off the proportion of the population that is below any value x
479 (how bad), and second, it shows confidence intervals clearly, for each point on curve
480 (how sure).

481 **EUFRAM envisages that probabilistic assessments for pesticides will generate**
482 **four main types of output using cumulative distributions**, depending on the type
483 of question posed by the assessment objectives:

- 484 • A CDF for hazard (effects), of which the most familiar is a species sensitivity
485 distribution (SSD, Posthuma et al. 2002), showing variation in sensitivity between
486 species¹¹. This can be compared to a point estimate for exposure (e.g. Figure 5a).
- 487 • A CDF for exposure, showing variation of exposure, e.g. between water bodies.
488 This can be compared to a point estimate for effects (Figure 5b).
- 489 • Two CDFs, one for exposure and one for effects, so that they can be compared
490 with one another (Figure 5c).
- 491 • A single CDF for “risk” or impacts, which takes account of both exposure and
492 toxicity (Figure 5d). This can be used by decision-makers to make judgements
493 about the acceptability of the risk, or compared to standard decision criteria for
494 risk, if these have been established.

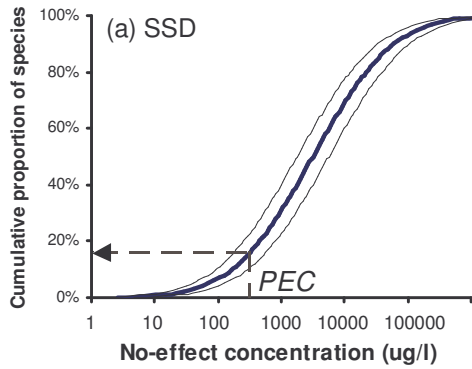
495 The x and y axes in Figure 5 show possible examples for measures of “magnitude”
496 and “frequency”. In practice, the measures of magnitude and frequency should be

¹¹ For SSDs, the measure of magnitude is sensitivity (in this case the no-effect concentration), and the measure of frequency is the percentage of species with no-effect concentrations equal to or less than the x-axis value. Note that for sensitivity, and also some other measures of magnitude (e.g. TER), low values indicate a greater hazard or risk than high values.

497 chosen to serve the objectives of the assessment as part of “problem definition” (see
498 Section 3.2). Presentational aspects (e.g. axis labelling and the use of tables and
499 other graphical formats to facilitate communication) are considered in section 3.6
500 (communicating results of probabilistic assessments).

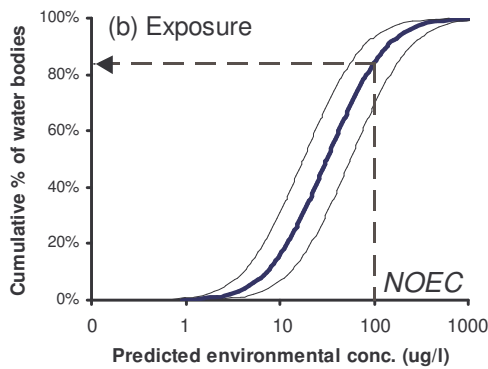
501 An important limitation of probabilistic outputs is that they cannot quantify all sources
502 of variability and uncertainty. This is because (a) there are simply too many sources
503 of variability and uncertainty for it to be practical to quantify them all, and (b) many
504 uncertainties can only be assessed subjectively (e.g. by expert judgement) and are
505 difficult to quantify. Consequently there will always be some sources of variability and
506 uncertainty that are not accounted for in the quantitative output of a probabilistic
507 assessment. Therefore, **the outputs of probabilistic assessments should include**
508 **a list of unquantified sources of variability and uncertainty, and a qualitative**
509 **evaluation of how they might affect the assessment outcome.** Ideally, this should
510 also be done for deterministic assessments, to show which sources of uncertainty
511 and variation they cover (e.g. via worst-case assumptions or uncertainty factors) and
512 which they omit.

513 Even though conventional regulatory assessment already includes some probabilistic
514 inputs, introducing probabilistic outputs will be challenging. The EUFRAM project is
515 considering ways to facilitate this. One recommendation is that **when a probabilistic**
516 **risk assessment is conducted, the results should be accompanied by**
517 **deterministic outputs.** This will allow for comparison of techniques and outputs and
518 will also help regulators and other end-users to become familiar with the new
519 methodology whilst still working with the *status quo*. Another possibility is to consider
520 using distributions of toxicity-exposure ratios (TERs) as an output of probabilistic
521 assessments (see Section 3.2.2). However, case studies have shown that if
522 distributions of TERs are used, it is important to accompany them with a narrative
523 explaining their interpretation. In fact, this is a general requirement that is essential
524 for all types of probabilistic output: **the quantitative outputs of probabilistic**
525 **assessments should always be accompanied by narrative text explaining their**
526 **interpretation.**



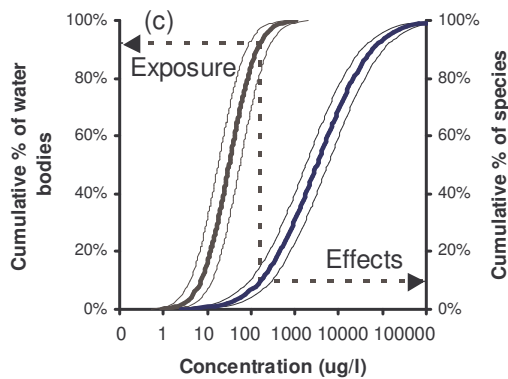
16% of species have no-effect concentrations at or below the predicted environmental concentration of 316 $\mu\text{g/l}$

i.e. this PEC will affect 16% of species



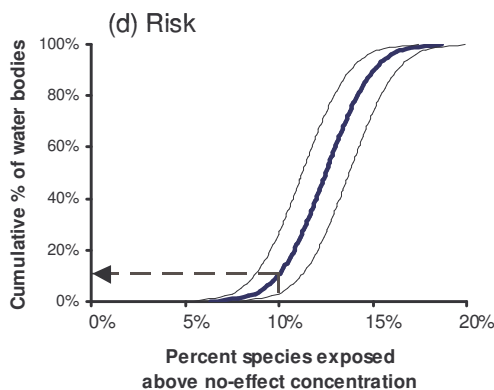
84% of water bodies have predicted environmental concentrations equal to or less than the no-effect level of 100 $\mu\text{g/l}$

So, by subtraction, 16% of water bodies have PECs greater than the NOEC



92% of water bodies have equal to or less than 10% of species affected (i.e. exposed above their NOEC)

So, 8% of water bodies have more than 10% of species affected



10% of water bodies have equal to or less than 10% of species affected (i.e. exposed above their NOEC)

So, 90% of water bodies have more than 10% of species affected

527

528 *Figure 5. Four main types of graphical output proposed for probabilistic assessments*
 529 *of ecological risks of pesticides, together with examples of risk statements that they*
 530 *could be used to make. The measures of magnitude and frequency (x and y axes) will*
 531 *vary according to the needs of the assessment. Note: these examples are illustrations*
 532 *and do not relate to real data.*

533 In summary, EUFRAM provisionally concludes that:

534 **PC3. Probabilistic results for ecological risks of pesticides should generally be**
535 **presented as cumulative distribution functions (CDFs), together with**
536 **confidence intervals if required.**

537 **PC4. Probabilistic assessments for ecological risks of pesticides should**
538 **produce one or more of the following types of cumulative distribution as**
539 **output:**

- 540 • **Distributions for effects (e.g. a species sensitivity distribution), to be**
541 **compared with point estimates for exposure,**
- 542 • **Distributions for exposure, to be compared with point estimates for effects,**
- 543 • **Distributions for exposure and effects, plotted together on the same graph**
544 **to show the extent of overlap between them,**
- 545 • **Distributions for risk or impact (e.g. toxicity-exposure ratio or % species**
546 **affected), in which exposure and effects are combined.**

547 **PC5. Probabilistic assessments should include a list of unquantified sources of**
548 **variability and uncertainty, and a qualitative evaluation of how they might affect**
549 **the assessment outcome.**

550 **PC6. Probabilistic risk assessment should be accompanied whenever possible**
551 **by a deterministic assessment, for comparison.**

552 **PC7. The quantitative outputs of probabilistic assessments should always be**
553 **accompanied by narrative text explaining their interpretation and ecological**
554 **implications.**

555 **2.5 Validation of probabilistic methods**

556 Validation is often cited as an important requirement for the acceptance of new
557 approaches for regulatory assessment. Modelling approaches such as those used in
558 probabilistic risk assessment can be subjected to several different types of validation
559 (European Commission, 2003), including:

- 560 • **Conceptual validation** – does the model structure accurately represent the
561 system?
- 562 • **Validation of algorithms and software code** – do the model equations properly
563 represent the conceptual model?
- 564 • **Functional validation** – does output from the model agree with independently-
565 obtained observations and measurements?

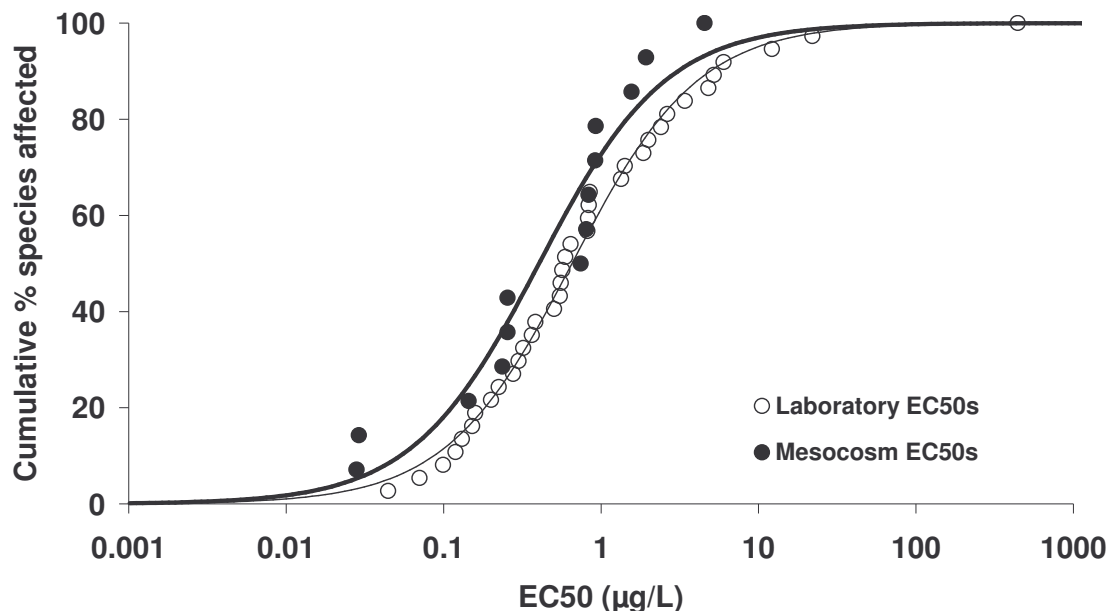
566 The first two types of validation rely largely on peer review by relevant experts,
567 whereas functional validation requires comparisons with independent data. Complete
568 functional validation of models is not possible, as they are necessarily simplifications
569 of the real world. Validation is especially challenging for probabilistic methods, as
570 they aim to predict variation in effects over a range of conditions, and also quantify
571 uncertainty. It would be necessary to obtain independent data for a wide range of
572 conditions, and repeat this for a range of different pesticides, in order to test the
573 predictions comprehensively. In practice, only partial functional validation is feasible,
574 for specific model components applied to particular combinations of pesticides and
575 environmental conditions.

576 Existing examples of partial functional validation of probabilistic approaches for
577 pesticides have been reviewed by EUFRAM (Volume 2, Work Package 9). Existing
578 examples relate primarily to comparisons between species sensitivity distributions
579 (SSDs) for effects on aquatic organisms in laboratory and semi-field experiments
580 (mesocosms). One of the earliest examples is illustrated in Figure 6. In this case,
581 there was good agreement between laboratory and semi-field SSDs when both are
582 restricted to the same taxonomic group, however comparisons of this type exist for
583 only a limited number of pesticides.

584 The limited feasibility and high cost of functional validation increase the importance of
585 comprehensive peer review in evaluating probabilistic approaches for regulatory use.
586 The importance of peer review as means of establishing credibility of approaches for
587 regulatory risk assessment has also been emphasised by Suter *et al.* (1993).

588 **PC8. Comprehensive peer review, together with functional validation where**
589 **possible, are important requirements for the acceptance of probabilistic**
590 **approaches.**

591



592

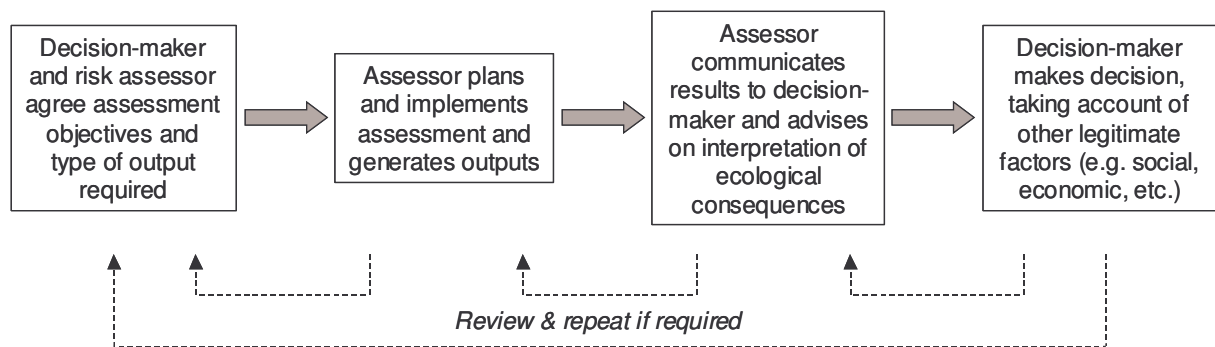
593 *Figure 6. Example of partial validation of an element of probabilistic assessment:*
594 *comparison between species sensitivity distributions for arthropod species based on*
595 *laboratory and mesocosm (semi-field) experiments with chlorpyrifos (adapted from*
596 *van den Brink et al., 2002).*

597 3 FRAMEWORK FOR PROBABILISTIC RISK ASSESSMENT

598 3.1 Interactions between risk assessors and decision-makers

599 Effective interaction and discussion between risk assessors, evaluators and decision-
600 makers is vital to ensure each probabilistic assessment meets the regulatory need
601 and to minimise waste of resources due to revision or rejection. Interaction is
602 especially important in the initial problem definition phase (section 3.2), and should

603 be repeated if it becomes apparent later that the problem definition needs revision
 604 (Figure 7). Interaction between technical experts and decision-makers is also
 605 important at the end of the assessment, to facilitate interpretation and use of the
 606 outputs.



607
608

609 *Figure 7. Simplified representation of the risk assessment/risk management process.*
 610 *Interaction between decision-makers and risk assessors may occur at any point, but*
 611 *is especially important in the first and third stages illustrated¹².*

612 3.2 Problem definition

613 Each assessment should start with a process of problem definition to ensure it meets
 614 the regulatory need. Problem definition should first define the assessment objectives,
 615 then specify what form of probabilistic output will meet those objectives.

616 Problem definition is a vital step that may require substantial effort if the assessment
 617 scenario is new. Effort will be saved if a similar assessment has been done
 618 previously, or if a standardised problem definition has been established, but a check
 619 should always be made to confirm that the problem definition is truly appropriate for
 620 the issue under consideration.

621 3.2.1 Assessment objectives

622 The first step is to define the objectives and scope of the assessment, starting with
 623 the basic elements that should already have been identified for the first-tier
 624 assessment:

- 625 • Define the **pesticide uses** to be considered (crops, application rates, etc.)
- 626 • Define the **non-target organisms** to be considered.
- 627 • Define the **types of effect** to be considered (e.g. lethal, sublethal, etc.)
- 628 • Define the **timescale** of the assessment (often implied by the type of effect, e.g.
 629 acute/chronic).
- 630 • Define the **spatial boundaries** of the assessment (e.g. regional, national, EU).

631 These choices should be made to reflect as closely as possible the **protection goal**
 632 – what the decision-maker wants to protect, and what types of effect they wish to limit
 633 or prevent. If it is too difficult to assess the types of effect that ultimately concern the
 634 decision-maker, then the assessment can be directed at a simpler measure of effect
 635 (e.g. acute toxicity-exposure ratio instead of population effects). However, if this is

¹² Note: in practice the situation is often more complex, e.g. there are multiple assessors and decision-makers, and some individuals may undertake both roles, but the sequence of activities shown in Figure 6 is still applicable.

636 done then it will be necessary to consider the relation between these two measures
637 of effect when interpreting the results (see section 3.4).

638 The purpose of probabilistic assessment is to quantify variation and/or uncertainty. It
639 is important to specify at the outset **what types of variation are of interest**. Possible
640 examples include variation in space or time, between individual organisms, or
641 between species. This decision will have a major impact on the structure of the
642 analysis and the types of data required. For example, data on variation in exposure
643 over time are required to quantify variation in risk over time.

644 It is important to specify in advance whether **confidence intervals** are required, as
645 this has substantial implications for assessment methodology (see section 3.4.2). It is
646 also helpful to agree in advance what level of confidence to show in the outputs¹³,
647 and how the confidence intervals will be used in decision-making.

648 If the decision-maker is concerned about particular **sources of variability and**
649 **uncertainty** in the assessment inputs (e.g. species differences in toxicity), then it is
650 helpful to identify these as part of the problem definition. However, this should not
651 preclude additional sources being examined later, if it becomes apparent during the
652 assessment that they are important.

653 It is helpful if the decision-maker is able to specify in advance the **decision criteria**
654 that they might use in decision-making, e.g. critical levels for the toxicity-exposure
655 ratio or the percentage of species affected. This will enable the assessor to produce
656 results focussed on those criteria (e.g. probability of TER<10). However, this is not
657 essential and decision-makers often prefer not to commit to particular criteria.

658 Finally, any **limitations** on the scope of the assessment should be specified, e.g.
659 avian assessments are often limited to dietary routes of exposure.

660 **3.2.2 Assessment endpoint**

661 The next step is to define a form for the key probabilistic output that will address the
662 assessment objectives: we refer to this key output as the **assessment endpoint**. If
663 the assessment endpoint is to quantify variation in effects, exposure or risk, three
664 essential components must be defined¹⁴:

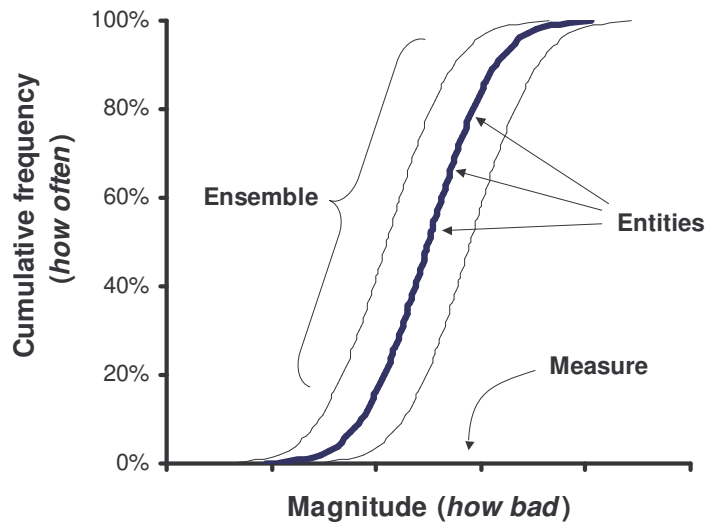
- 665 • What **measure** to use for the magnitude of effects, exposure or risk?
- 666 • For what individual **entities**¹⁵ will the magnitudes be estimated?
- 667 • For what **ensemble** or group of entities will frequencies be estimated?

668 The measure and entity together define the x-axis of a cumulative distribution for the
669 output (magnitude), while the entities and ensemble together define the y-axis
670 (frequency). This is illustrated in Figure 8. The curve simply shows the variation in
671 effects, exposure or risk between the entities in the ensemble. It can be thought of as
672 being formed by stacking up all of the entities in the ensemble, ordered by their
673 individual magnitudes.

¹³ In biological research, 95% confidence intervals are used as a convention. Different levels might be preferred for pesticide assessments, depending on the degree of conservatism that is required.

¹⁴ This is compatible with existing uses of the term assessment endpoint (e.g. US EPA, 1998).

¹⁵ The terms “entity” and “ensemble” are used because they are flexible in meaning and can refer to both biological entities (e.g. species) and physical entities (e.g. water bodies).



674

675 *Figure 8. Representation of the 3 elements of an assessment endpoint in a cumulative*
 676 *distribution: measure, entities and ensemble.*

677 **The most important consideration is to ensure that the measure, entity and**
 678 **ensemble are meaningful and relevant to the decision-maker.** To achieve this, it
 679 may help to convert the assessment objectives into a specific question for the
 680 assessment. Some examples are shown in Table 1. Note that the entity can take
 681 many forms, either physical (e.g. a water body) or biological (e.g. an individual,
 682 population, species or community), but must be compatible with the measure of
 683 magnitude. The ensemble is simply a group of entities that is meaningful and
 684 relevant to the decision-maker.

685 At its simplest, the cumulative distribution represents one type of variation (e.g. in
 686 space, or between species). If two types of variation are important to the decision-
 687 maker, one of them can be represented by a suitably-defined measure of magnitude.
 688 For example, a decision-maker may be interested in both the proportion of species
 689 that are affected and the proportion of water bodies in which this will happen. In this
 690 case, one option is to define the entity as a water body, the measure of effect as the
 691 proportion of species affected, and the ensemble as the population of water bodies
 692 relevant to the assessment (e.g. water bodies of a particular type, or in a particular
 693 region). Another good option is to produce several different cumulative distributions,
 694 each quantifying just one type of variation in the output. Other options exist, e.g.
 695 combining both types of variation on the frequency axis (ensemble = all species/ditch
 696 combinations) or adding a third axis to the output graph, but these are much more
 697 difficult to interpret and communicate.

698 Often, more than one combination of measure, entity and ensemble could be
 699 appropriate, and it may be helpful to use several different combinations that reflect
 700 alternative ways of forming the assessment question. For example, the decision-
 701 maker may be interested in the proportion of all water bodies that are affected, and
 702 also in the proportion of exposed water bodies that are affected (same measure and
 703 entity, different ensembles). As a more complex example, the decision maker might
 704 be interested in what proportion of ditches have effects on more than 10% of species,
 705 and also in what proportion of species are affected in more than 10% of ditches
 706 (different effects, entities and ensembles – see last two rows in Table 1).

707 *Table 1. Examples of assessment endpoints that could be used to answer different*
 708 *types of assessment question¹⁶.*

Assessment question	Assessment endpoint		
	Measure of magnitude	Entity	Ensemble
What proportion of species will be affected at the predicted environmental concentration?	No effect concentration, NOEC ¹⁷	One species	Those species relevant to the assessment (e.g. present in water bodies adjacent to treated fields)
What proportion of water bodies will have concentrations exceeding the no-effect level?	Predicted environmental concentration, PEC	One water body	Water bodies adjacent to treated fields, in a specified geographic region
How many species will suffer significant mortality?	Percent mortality of individuals	One species	Those species relevant to the assessment
What proportion of aquatic species will have a TER less than 10?	Toxicity-exposure ratio, TER	One aquatic species	Those species present in water bodies adjacent to treated fields
What proportion of birds will have a TER of less than 10?	TER	One individual bird of the relevant indicator species	Those individuals of the indicator species that visit treated fields
In what proportion of ditches are more than 10% of species affected?	Percentage of species affected	One ditch	Ditches adjacent to treated fields
What proportion of species are affected in more than 10% of ditches?	Percentage of ditches in which species is affected	One species	Species relevant to the assessment

709
 710 Often, the entity used in the conventional first tier assessment is abstract, for
 711 example a standard ditch or a generic insectivorous bird. If an abstract entity is used
 712 in a probabilistic assessment then the ensemble will also be abstract, for example an
 713 ensemble of standard ditches. Abstract entities and ensembles can still be useful
 714 provided they can be interpreted in relation to the real world – for example, if they are
 715 conservative.

716 **It is not always necessary to use a distribution to quantify variation.** For
 717 example, if the measure is binary (e.g. mortality of individuals), then variation in effect
 718 can be quantified by a simple percentage (% mortality). In such cases the
 719 assessment endpoint can be expressed simply as an estimate of the frequency of
 720 effects (e.g. % mortality), together with confidence intervals if required. This
 721 simplification may also be possible when the measure of magnitude is continuous, if

¹⁶ These examples are not definitive: other choices are possible for the same assessment questions.

¹⁷ In this case, the CDF is a species sensitivity distribution (SSD) and the measure of magnitude is sensitivity to the toxicant.

722 there are critical values of particular interest to the decision-maker. For example, if
723 the measure of effect is a TER, the decision-maker may be interested primarily in the
724 proportion of entities with $TER < 10$. In this case, the primary assessment endpoint
725 could be simply the percentage of entities with $TER < 10$, together with confidence
726 intervals if required. Focussing on critical values in this way simplifies the output and
727 may aid communication (see section 3.6).

728 Finally, we emphasise that **it is important to define the output ensemble first, and**
729 **then use that to infer what the input ensembles should be**. It is tempting to do the
730 opposite – i.e. allow the available input datasets to determine the output ensemble.
731 This can cause under- or over-estimation of variability or produce output ensembles
732 that have no real-world interpretation¹⁸.

733 **3.3 Planning the assessment model**

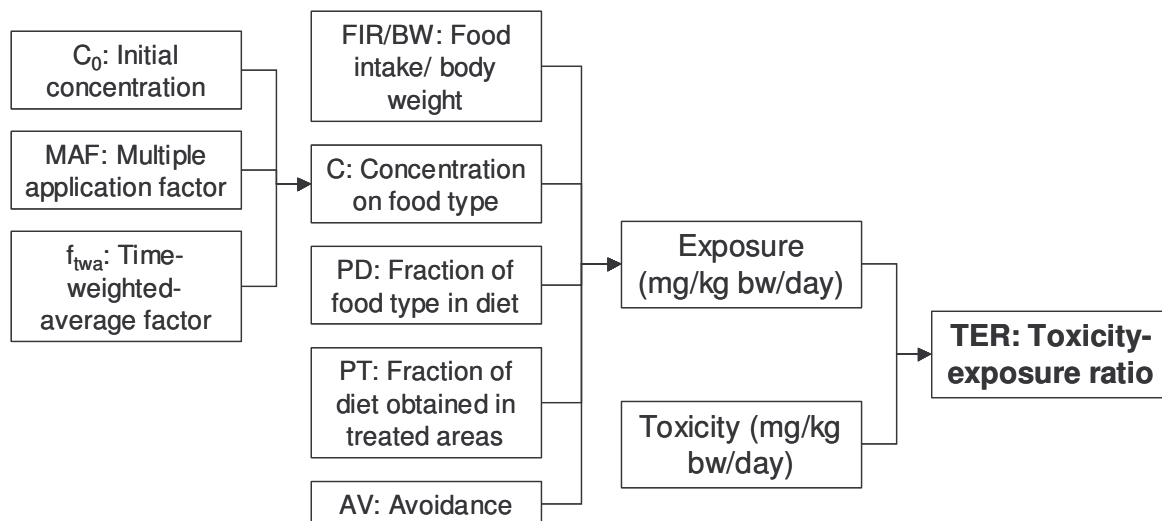
734 Having defined the output that is required, the next step is to plan the assessment
735 model that will generate the output from the available data. If the assessment model
736 is inappropriate, the output will be invalid or misleading.

737 Planning the assessment may require substantial effort if the assessment scenario is
738 new. Effort will be saved if a similar assessment has been done previously, or if a
739 standardised assessment has been established, but a check should always be made
740 to confirm that the assessment is truly appropriate for the issue under consideration.

741 The key steps in planning the assessment model are as follows:

- 742 1. **Identify the key factors and mechanisms that influence the effect that is to**
743 **be assessed, and how they interact**. Generally, these will be the same as in the
744 conventional first tier assessment, with additional refinements if required to
745 address the objectives agreed for the probabilistic assessments. It may help to
746 draw a diagram representing the various factors and mechanisms and showing
747 their interactions, e.g. as a flow chart (this is sometimes referred to as a
748 **conceptual model**). An example is shown in Figure 9.
- 749 2. **Consider each part of the model in turn, and decide which factors and**
750 **mechanisms might contribute significantly to variability in the assessment**
751 **endpoint**. These will need to be represented by distributions in the assessment
752 model.
- 753 3. **For each variable in the assessment, identify the appropriate ensemble**.
754 This depends on the ensemble of the assessment endpoint and the model
755 structure. For example, if the output is effects in an ensemble of water bodies,
756 then the appropriate ensemble for exposure is PECs for different water bodies.
- 757 4. **Identify what data are available** that can help in quantifying each factor and
758 mechanism, including distributions for those that will be treated as variables.

¹⁸ For example, using a distribution of average values from different studies could cause either over- or under-estimation of variability within studies, depending on which is more relevant to the assessment endpoint and whether between- or within-study variation is greater. Combining a within-species variation for one parameter with between-species variation for another would create an output ensemble of entities that are all one species with respect to the first parameter but represent multiple species with respect to the second parameter.



759

760 *Figure 9. Example of a conceptual model, depicted in the form of a flow chart. This*
 761 *example shows the current model for first tier assessment of risks to birds and*
 762 *mammals (European Commission, 2002c).*

763 5. **Decide whether any extrapolations or adjustments are needed to model the**
 764 **key factors and mechanisms from the available data.** Extrapolation or
 765 adjustment may be required to deal with various issues including:

- 766
- 767 • Lab-to-field extrapolation.
 - 768 • Non-random sampling. For example, species tested in the laboratory may not
 769 be representative of the ensemble of species considered in the assessment.
 - 770 • Surrogacy – where one type of data is used as a surrogate for another type.
 771 For example, in the first tier avian assessment, the estimate used for
 772 pesticide residues on insects is actually derived from measurements of
 773 residues on plants (European Commission, 2002c).
 - 774 • Incompatible ensembles. For example, if the required exposure ensemble is
 775 variation in pesticide concentrations between water bodies, using a series of
 776 concentrations measured in one water body would be invalid.
 - 777 • Inappropriate averaging. Distributions based on average results from different
 778 studies will underestimate variation in the underlying values.
 - 779 • Differences in spatial and temporal scale. For example, peak exposures in
 780 water bodies adjacent to treated fields differ from concentrations from routine
 781 sampling of large rivers.

781 If issues of this type do require the use of adjustment or extrapolation factors,
 782 then they become, in effect, part of the assessment model and should be
 783 identified as such in the model plan. Also, consideration should be given to
 784 whether these factors should be represented by point values or by distributions (if
 785 they are subject to substantial variability or uncertainty).

786 6. **Consider each part of the model in turn to identify possible sources of**
 787 **uncertainty**¹⁹. Decide which uncertainties might contribute significantly to
 788 uncertainty in the assessment endpoint, and which of these will be quantified
 789 using distributions (if any). Decide what to do about important uncertainties that

¹⁹ This is essential if uncertainty and variability are to be separated, to provide confidence bounds on the outputs, or if a distribution representing only variability is required. However, it is desirable even in other cases, to enable the assessor to communicate the meaning of the output distribution (e.g. does it represent mostly uncertainty or mostly variability).

790 cannot or will not be quantified (e.g. use best estimates or conservative values).
791 Different types of uncertainty, and methods for dealing with them, are discussed
792 in section 3.4.4.

793 7. **Consider carefully for each input distribution whether it contributes**
794 **uncertainty or variability to the output.** This again depends partly on the
795 ensembles. For example, a species sensitivity distribution contributes variability if
796 the output ensemble comprises multiple species, but uncertainty if the output
797 ensemble comprises individuals of a single species. If a distribution includes both
798 variability and uncertainty and they cannot be separated, it is better to treat the
799 distribution as uncertainty²⁰ as this reflects the state of knowledge more
800 accurately than if it were treated as variability.

801 8. **Identify potential dependencies affecting the assessment.** Dependencies
802 occur where the value of one variable depends upon the value of another variable
803 (e.g. food intake may be positively related to body weight). It is important to
804 consider dependencies carefully, as they can have a major impact on the
805 assessment outcome. Different types of dependency, and methods for dealing
806 with them, are briefly discussed in section 3.4.4.

807 9. **Express the entire assessment model as a set of mathematical equations,**
808 so that they can be used for calculations.

809 It is essential to **document** the justification for each choice made in planning the
810 assessment model, and each component of the resulting equations, to support the
811 credibility of the assessment and facilitate peer review (see section 3.5).

812 **3.4 Key methods for probabilistic assessment**

813 **3.4.1 Overview of alternative methods for combining distributions**

814 Distributions for assessment inputs must be combined to obtain a distribution for the
815 assessment output. A large number of different methods exist for combining
816 distributions. EUFRAM has examined some that were thought promising for pesticide
817 risk assessments. Descriptions and comparisons of these methods are presented in
818 Volume 2, Work Package 4, together with references to more detailed publications.
819 Additional investigations into methods for use with limited datasets are presented in
820 Work Package 5.

821 This section briefly outlines the essence of the methods considered by EUFRAM,
822 and the next section offers provisional recommendations on which methods can
823 generate which types of output, and how to choose between them.

824 **Scenario analysis.** This method is used in current first-tier assessments for most of
825 the taxonomic groups (e.g. use of 90th percentile for spray drift). Distributions are
826 used to quantify one or more sources of variability or uncertainty, and a single
827 percentile is chosen from each distribution. These point values are then combined by
828 simple algebra to derive a point estimate for the assessment endpoint, which
829 represents a scenario defined by the choice of the percentiles for the various inputs.
830 This method is very simple to use but cannot quantify the overall conservatism of the
831 assessment when more than one or two inputs are derived from distributions.

²⁰ If this is done then it is important to take account of the resulting confidence bounds in decision-making, as using only the central estimate would under-represent variability.

832 **Combining one distribution with point estimates.** If only one source of variation or
833 uncertainty is represented by a distribution, each percentile of this can be combined
834 with point estimates for other parts of the assessment model to obtain corresponding
835 percentiles for the assessment endpoint. This requires only simple algebra but is
836 limited to a single input distribution, and assumes it is independent of other model
837 components.

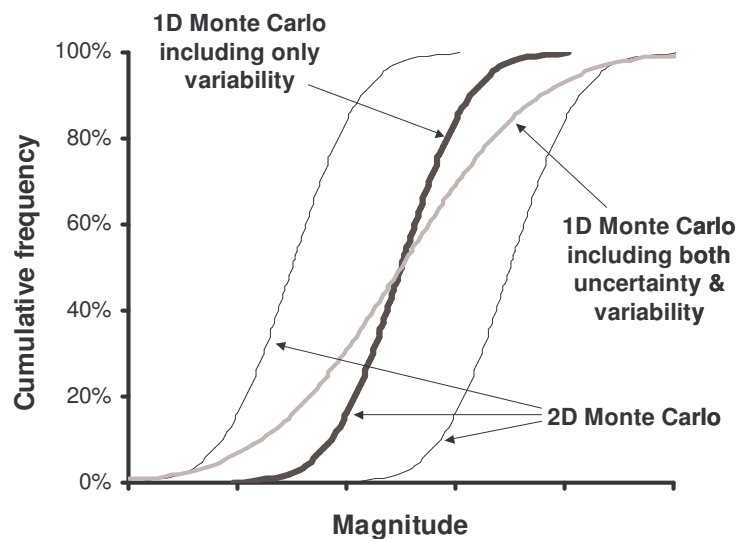
838 **Plotting two distributions on the same graph.** Distributions for exposure and
839 effects can be plotted on the same graph if they have the same x-axis (e.g. Figure
840 5c). The overlap between the two curves provides a visual impression of the degree
841 of risk, and for any point on the horizontal axis, one can read the frequency of
842 exposure from one vertical axis and the frequency of effects on the other. This
843 method is very simple but limited to use with pairs of distributions sharing the same x-
844 axis, and interpretation is difficult if the distributions are not independent.

845 **Method of joint probabilities.** Two distributions can be combined into one without
846 calculations by the method of joint probabilities, if they have the same x-axis. The
847 result is often called a “joint probability curve” or JPC and can be plotted in several
848 formats, one of which is equivalent to a cumulative distribution for the frequency of
849 effects. The JPC is simple to construct without any calculations. It is limited to use
850 with pairs of distributions sharing the same x-axis, and is invalid if the distributions
851 are not independent. The joint probability curve can also be used to estimate the
852 average frequency of effects, but this is harder to compute.

853 **Lookup tables.** Risk assessment statistics can be provided in lookup tables, of a
854 similar type to those commonly used for logarithms and statistics such as Student’s t
855 and chi-square. These allow the user to obtain best estimates and confidence
856 intervals for assessment endpoints without having to do complex calculations.
857 However, they are only available for a few specific types of assessment model.
858 Aldenberg and Jaworska (2000) provide tabulated extrapolation factors for
859 calculating selected percentiles of a single Normal distribution together with
860 confidence intervals for sampling uncertainty. Aldenberg et al. (2002, Table 5.3)
861 provide tabulated factors that can be used to calculate the average percent of
862 species affected from the means and standard deviations of an exposure distribution
863 and species sensitivity distribution, both of which must be Normal.

864 **1D Monte Carlo.** Monte Carlo simulation combines distributions by taking large
865 numbers of samples from each one at random. In 1D (1-dimensional) Monte Carlo,
866 all the input distributions are sampled together, and produce a single distribution for
867 the output. When all the input distributions represent *only variability*, the output
868 distribution represents the best estimate of variation in the assessment endpoint.
869 This can be used to estimate specific percentiles of the output ensemble (e.g. the 5th
870 percentile or HC5), but provides no confidence intervals and may give a false
871 impression of certainty. When input distributions representing *variability and*
872 *uncertainty* are combined by 1D Monte Carlo, the output distribution represents
873 uncertainty about the assessment endpoint for a *single ensemble member selected*
874 *at random* (e.g. a species selected at random, rather than a particular percentile
875 species). The difference between 1D Monte Carlo including and excluding
876 uncertainty is illustrated in Figure 10. The distribution excluding uncertainty
877 represents a best estimate of the CDF, whereas the 1D distribution combining both
878 variability and uncertainty spans the full range of possible values for the assessment
879 endpoint.

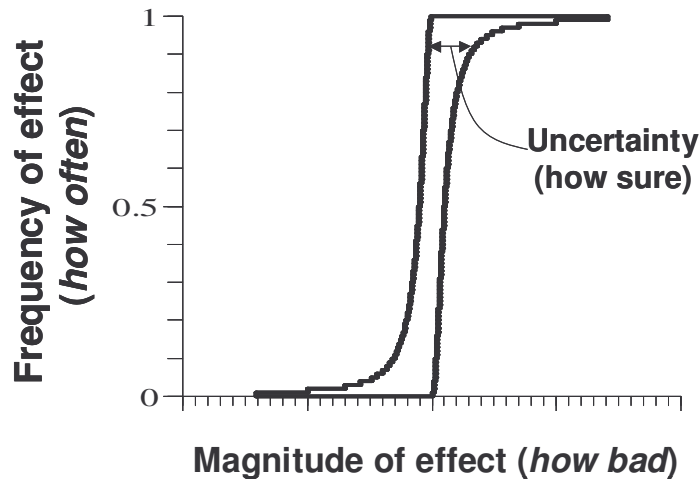
880 Monte Carlo methods can take account of dependencies between input variables, if
881 information is available to characterise them, but requires strong assumptions about
882 distribution shape. A number of software packages exist for 1D Monte Carlo and are
883 fairly easy to use.



884
885 *Figure 10. Diagrammatic comparison between the output of 1D and 2D Monte Carlo for*
886 *the same input distributions. 2D Monte Carlo produces the cumulative distribution*
887 *with confidence intervals, whereas 1D Monte Carlo produces a single distribution*
888 *representing uncertainty about a single member of the ensemble selected at random.*

889 **2D Monte Carlo.** Two dimensional (2D) Monte Carlo is very similar to 1D Monte
890 Carlo except that it separates uncertainty and variability. In 2D Monte Carlo, the
891 distributions for uncertainty and variability are sampled separately, so that the
892 combined effect of the uncertainties can be shown as a confidence interval around
893 the output distribution (Figure 10). The ability to show the effect of uncertainties
894 separately as confidence intervals makes 2D Monte Carlo especially useful for limited
895 datasets, but at present appropriate techniques are available to quantify only some
896 types of uncertainty (see section 3.3.4). 2D options are included in some Monte Carlo
897 software. Like 1D Monte Carlo, 2D methods can take account of dependencies
898 between input variables but require strong assumptions about distribution shape.

899 **Probability bounds.** Probability bounds is a method of combining uncertain
900 distributions that does not require any data or assumptions about either distribution
901 shape or dependencies among variables, although such information can be used to
902 some extent if available. It first produces absolute (outer) bounds for the cumulative
903 distribution of each input variable, and uses these to derive absolute (outer) bounds
904 for the output distribution. However, an important limitation is that this method does
905 not provide any information on where the true distribution might lie between the
906 bounds. An example is shown in Figure 11. The method can work with many types of
907 information, and each piece of information can itself be uncertain. Probability bounds
908 copes well with most types of uncertainty but has difficulties with sampling
909 uncertainty, which is important for the small sample sizes that are common in risk
910 assessment (see Volume 2, Work Package 5). Probability bounds may be very wide,
911 but this is appropriate if there is genuinely little information about the inputs.



912

913 *Figure 11. Example of output from probability bounds analysis. This method provides*
 914 *absolute bounds, within which the cumulative distribution of the assessment endpoint*
 915 *must lie.*

916 **Bayesian methods.** Bayesian methods use a subjective concept of probability and
 917 can make use of subjective information (e.g. expert judgements, which are often
 918 used in risk assessment), as well as objective data. Bayes theorem is used to
 919 combine existing (“prior”) knowledge with new data, a process sometimes called
 920 “updating”. This produces “posterior” distributions for the assessment endpoint.
 921 Confidence intervals can be generated for different percentiles of the assessment
 922 endpoint, if required. The concept of updating fits neatly with the process of tiered
 923 risk assessment, and can also be used to combine different types of information
 924 (e.g., lab and field). Some of the theory and computations involved in Bayesian
 925 methods are complex and require specialist expertise. However, this need not be an
 926 obstacle to their use if they can be provided in a pre-packaged form that is known to
 927 be appropriate to the problem in hand (e.g. customised software or lookup tables).

928 **3.4.2 Choosing between methods for combining distributions**

929 **NOTE:** this section is provisional and will be revised as the project progresses,
 930 depending on further evaluation of the methods and feedback from participants.

931 Preferences between the different methods for combining distributions will depend on
 932 a number of factors. At this stage the following factors appear important:

- 933 • **What type of output you require from the assessment** (specified in problem
 934 definition, section 3.2)
 - 935 - *Point estimates for specified input scenario* – can be provided by scenario
 936 analysis, 1 distribution, overlay graph, joint probability method
 - 937 - *Best estimate of cumulative distribution* – any method except scenario
 938 analysis
 - 939 - *Confidence intervals for the cumulative distribution* – lookup tables, 2D Monte
 940 Carlo or Bayesian methods.
 - 941 - *Absolute outer bounds for the cumulative distribution* – probability bounds

- 942 • **How many input distributions you have**²¹ (specified in the assessment model,
 943 section 3.3)
- 944 - *1 input distribution* – any method except overlay graph and joint probability
- 945 - *2 input distributions* – any method except combining 1 distribution with point
 946 values
- 947 - *3 or more input distributions* – scenario analysis, 1D or 2D Monte Carlo,
 948 probability bounds or Bayesian methods
- 949 • **How you want to handle uncertainty about distribution parameters,
 950 distribution shape and dependencies** (discussed in more detail in sections
 951 3.4.4 and 3.4.5). Only 2D Monte Carlo, probability bounds, and Bayesian
 952 methods can provide confidence intervals on output distributions, while 1D Monte
 953 Carlo can include uncertainty but not separate it from variability (see Figure 10).
 954 These methods can also take account of known or assumed dependencies.
 955 Probability bounds can produce bounds that enclose all possible output
 956 distributions, without requiring any knowledge or assumptions on dependencies.
 957 The simpler methods (scenario analysis, combining one distribution with point
 958 estimates, overlaying two distributions and the joint probability method) can
 959 represent uncertainty by using alternative assumptions (e.g. best estimate vs.
 960 worst case) but do not provide confidence intervals and generally assume that
 961 inputs are independent of one another.
- 962 • **Ease of use** is an important practical consideration. The current view of EUFRAM
 963 project partners may be summarised as follows:
- 964 - Scenario analysis, combining single distributions with point values, overlay
 965 graphs of two distributions, the joint probability method, lookup tables and 1D
 966 Monte Carlo are relatively easy to use, and seem most likely to gain regulatory
 967 acceptance in the short-term.
- 968 - 2D Monte Carlo is of intermediate difficulty and has received only limited
 969 evaluation for pesticide assessments.
- 970 - Probability bounds and Bayesian methods are (or are perceived as) more
 971 difficult to use and have received least evaluation.

972 Consideration should also be given to whether different methods could serve
 973 complementary functions, either sequentially (e.g. as successive steps in a tiered
 974 approach) or concurrently (e.g. overlaying probability bounds on Monte Carlo results,
 975 to show the effect of relaxing assumptions about distribution shape and
 976 dependencies, Regan et al. 2002 and Volume 2, Work Package 4).

977 **QUESTION FOR FEEDBACK:**

978 **Q1. Do you have any comments on the strengths and weaknesses of different
 979 methods for combining distributions, or suggestions for how to choose
 980 between them? Which methods do you prefer, and why?**

²¹ In addition it should be borne in mind that the joint probability method and overlay graph are applicable only when 2 input distributions share a common x-axis, and that at present lookup tables are available only for normal/lognormal distributions.

981 **3.4.3 Specifying distributions for input variables²²**

982 **The specification of input distributions is critical, as inappropriate choices will**
983 **give misleading or invalid outputs** (e.g. negative exposures). There is a wide
984 range of options for specifying distributions and an extensive literature, including
985 substantial chapters in publications on probabilistic risk assessment (e.g. Vose 2000,
986 Cullen & Frey 1999, US EPA 1997, 2001). Specifying appropriate distributions
987 requires **expertise** in these statistical approaches, combined with a clear
988 understanding of the assessment model, and expert knowledge of the parameter in
989 question (e.g. ecology, chemistry etc.).

990 The measure, entities and ensemble required for each input distribution should be
991 identified from the structure of the assessment model and the entities and ensemble
992 of the eventual assessment endpoint. Data available for specifying the input
993 distribution often deviate from these requirements in various ways: any necessary
994 adjustments or extrapolations to allow for these deviations should be defined as part
995 of the assessment model (see section 3.3). Alternatively, it may be appropriate to
996 select certain subsets of the data or give different weights to different points within
997 the dataset, e.g. to correct over-representation of certain taxonomic groups in toxicity
998 data.

999 A wide range of approaches can be used to specify distributions, depending on the
1000 amount and type of data available. For large datasets it may be best to use
1001 **empirical distributions** or **empirical bootstrapping** methods, which directly reflect
1002 the sample data.

1003 For smaller datasets, or where there are reasons to expect a particular form of
1004 distribution, it may be preferable to use **parametric distributions** (e.g. the Normal
1005 distribution). Parametric distributions are defined mathematically by parameters (e.g.
1006 mean and standard deviation) that may be estimated from the sample data in various
1007 ways. The first consideration in choosing between the many possible parametric
1008 distributions should be which are logically compatible with the nature of the variable
1009 in question, e.g. whether it is discrete or continuous, the existence of absolute
1010 minima or maxima, and what type of process or mechanism is responsible for the
1011 variation. Various graphical and statistical approaches are available for assessing
1012 **goodness of fit**, and some computer programs allow semi-automated testing of
1013 multiple distribution types to find the best fit (e.g. Crystal Ball® and @Risk®).
1014 However, goodness of fit tests can be misleading and it is generally recommended to
1015 select distributions using both statistical and graphical methods in conjunction with
1016 mechanistic criteria. Sometimes it may be appropriate to **transform** the sample data
1017 (e.g. to logarithms) to improve goodness of fit, or to represent a variable using a
1018 **mixture of distributions** instead of only one. Sometimes it may be appropriate to
1019 **truncate** a parametric distribution, to prevent sampling of extreme values that cannot
1020 occur in reality.

1021 Where data are very limited, or only summary statistics or subjective information are
1022 available, different approaches may be appropriate. **Maximum entropy** approaches
1023 are intended to be conservative, selecting the type of parametric distribution that
1024 maximises variability, given the data. **Probability bounds** and related methods use
1025 “probability boxes” (e.g. Figure 11 above) to define absolute bounds within which the

²² More discussion of some of the approaches identified in this section is presented in Volume 2, Work Package 4.

1026 cumulative distribution for the variable is certain to lie, and can be defined from many
1027 types of limited input information. Finally, there are a variety of formal methods of
1028 **expert elicitation**, that can specify distributions using only subjective expert
1029 knowledge and opinion.

1030 Whatever methods are used, special attention should be paid to ensuring the **tails** of
1031 distributions are reasonable, as they are frequently critical to the outcome of risk
1032 assessment but poorly represented by data except in very large samples.

1033 Finally, it is emphasised that users should refer to detailed texts, such as those cited
1034 at the start of this section, and consider seeking expert assistance for the
1035 specification of input distributions.

1036 **3.4.4 Methods for dealing with uncertainties**

1037 One of the purposes of probabilistic assessment is to quantify uncertainty. It is
1038 important to specify at the outset whether confidence intervals are required on output
1039 distributions, as this has substantial implications for assessment methodology (see
1040 section 3.4.2). If the decision-maker is concerned about particular sources of
1041 uncertainty, these should be specified as part of the problem definition. In addition,
1042 when planning the assessment model, each component should be considered in turn
1043 to identify further potentially important sources of uncertainty (section 3.3).

1044 Uncertainties can be classified in various ways (e.g. Morgan & Henrion 1990, Cullen
1045 & Frey 1999). Important types of uncertainty include:

- 1046 • **Uncertainty about distribution shape** – often, several different distributions may
1047 be plausible for the same model input, and may show similar goodness of fit to
1048 sample data.
- 1049 • **Sampling uncertainty** – when a sample is used to estimate distribution
1050 parameters or to derive an empirical distribution, there is uncertainty about their
1051 relationship to the true parameters or distribution for the larger population from
1052 which the sample was drawn.
- 1053 • **Measurement uncertainty** – various factors may cause random errors or bias in
1054 measurements or experimental data.
- 1055 • **Extrapolation uncertainty** – when it is necessary to extrapolate beyond the
1056 range of a dataset, or from one type of data to another (surrogacy), there is
1057 uncertainty about how closely the extrapolated values match the true values that
1058 are being estimated.
- 1059 • **Model uncertainty** – there is often uncertainty about which of several alternative
1060 model structures best represents real mechanisms and processes.
- 1061 • **Uncertain dependencies** – there may be uncertainty about the nature, direction
1062 and magnitude of dependencies in the assessment (see section 3.4.5).

1063 In general, uncertainties are greater when datasets are small.

1064 Objective methods exist for **quantifying** some types of uncertainties. Exact analytical
1065 expressions are available for sampling uncertainty of a few parametric distributions,
1066 notably the Normal distribution (e.g. Vose 2000, Cullen & Frey 1999). For other types
1067 of distributions, it may be possible to estimate sampling uncertainty by Bayesian or
1068 bootstrapping approaches (e.g. Beta distribution, Vose 2000, pp. 255-262)²³. Some

²³ An alternative to quantifying sampling uncertainty may be to collect enough data to make sampling uncertainty negligible for the purposes of the assessment. This approach is implied when people propose minimum sample sizes, e.g. for SSDs, but implies a risk management judgement about how much certainty is required.

1069 types of measurement uncertainty can be quantified using the distribution of repeated
1070 measurements.

1071 A key strength of probability bounds is that it can show the effects of uncertainty
1072 about distribution shape, by producing bounds for the output that make no
1073 assumption about distribution shape for some or all of the inputs (Ferson, 2002).

1074 It may be possible to quantify some types of uncertainty subjectively, e.g. use expert
1075 judgement to specify a triangular distribution based on estimates of the minimum,
1076 maximum and most likely value for a quantity. This is probably best done using
1077 formal expert elicitation methods (e.g. Morgan & Henrion 1990, Vose 2000). It is
1078 currently more common to use fixed values (either best estimate or worst case) for
1079 subjective uncertainties. This is simpler but makes it difficult to characterise the
1080 conservatism of the assessment, especially when best estimates or worst-case
1081 assumptions have been used for multiple inputs.

1082 Uncertainty about distribution shape and model structure may be explored by
1083 repeating the assessment with a limited number of credible alternative models, and
1084 comparing the results. It is possible to include alternative models within a single
1085 Monte Carlo assessment, weighted by some estimate of their relative likelihood, but
1086 this in effect averages the models and may be misleading (e.g. produce intermediate
1087 results that are not compatible with either model). Probability bounds can however
1088 combine alternative models without incurring this problem (Ferson 2002).

1089 Annex VI of Directive 91/414 includes decision criteria for first tier assessments, e.g.
1090 critical TERs, below which further assessment is required. These critical values are
1091 intended, in part, to allow for variability and uncertainties affecting the assessment.
1092 There is no published documentation of the quantitative basis for the specified
1093 values, nor is it clear what types of uncertainty they were originally intended to cover,
1094 although post-hoc interpretations are given in the EU Guidance Documents
1095 (European Commission, 2002a, 2002c). The Guidance Documents indicate that in
1096 principle these critical values could be reduced when uncertainty is reduced by
1097 provision of additional data, but provide little detailed guidance on this.

1098 Ultimately it is never possible to quantify all uncertainties affecting an assessment.
1099 This implies it should be clearly stated in the assessment (a) which uncertainties
1100 have been quantified and (b) what uncertainties remain unquantified and their
1101 potential implications for the assessment. The following approaches may assist in
1102 **accounting for unquantified uncertainties:**

- 1103 1. Systematically list or tabulate all unquantified sources of uncertainty
- 1104 2. Qualitatively evaluate the direction and relative magnitude of each unquantified
1105 uncertainty in terms of its impact on the assessment outcome
- 1106 3. Qualitatively evaluate the possible combined effect of all the quantified and
1107 unquantified uncertainties on the assessment outcome.

1108 **3.4.5 Methods for dealing with dependencies**

1109 Dependencies occur when the value of one quantity in the model depends upon the
1110 value of another. They can occur between variables (e.g. age and weight of exposed
1111 individuals), between uncertainties (e.g. the slope and intercept of a regression
1112 relationship), or between variables and uncertainties. They can arise from a direct
1113 mechanistic or logical relationship between two quantities, or indirectly when both are
1114 dependent on a third. Potentially important dependencies should be identified when
1115 planning the assessment model (section 3.3).

1116 The most familiar form of dependency is linear correlation, but there are many others.
1117 Sometimes they may be detected and quantified from data, by examining scatterplots
1118 of two variables measured on the same sample. More usually, it is necessary to
1119 deduce them from the structure of the model and the mechanisms it represents:
1120 examine different parts of the model in turn and consider subjectively whether they
1121 might be inter-dependent, including the possibility of dependencies in time and
1122 space.

1123 **Failing to account for important dependencies can cause major over- or under-**
1124 **estimation of risk.** Methods to account for dependencies in Monte Carlo
1125 assessments are discussed by Cullen and Frey (1999) and US EPA (2001) and in
1126 more detail by Vose (2000). Cullen and Frey (1999) prefer to incorporate
1127 dependencies into the model structure by making one quantity a function of the other
1128 (e.g. using regression relationships). The most familiar alternative is to simulate linear
1129 correlations by restricted sampling of input distributions in Monte Carlo, and is
1130 implemented in commercial software including Crystal Ball® and @Risk®.

1131 **Dependencies are often affected by substantial uncertainty.** Vose (2000)
1132 discusses how to quantify sampling uncertainty for correlations estimated from data.
1133 When a dependency is suspected but cannot be estimated from data, its potential
1134 importance can be investigated by trying a range of different assumed dependencies
1135 (Vose, 2000; US EPA, 1997).

1136 One of the important advantages of probability bounds analysis is that it can
1137 accommodate total uncertainty about dependencies, by calculating bounds that
1138 enclose all possible output distributions regardless of the dependencies between the
1139 inputs (Ferson, 2002). Probability bounds can be narrowed if some variables are
1140 known to be totally independent or perfectly correlated, but cannot take account of
1141 intermediate degrees of dependency.

1142 **3.4.6 Sensitivity analysis**

1143 Sensitivity analysis evaluates the influence of alternative inputs or assumptions on
1144 the output of an assessment. At its simplest, it could examine the effect of changing a
1145 single deterministic input from one value to another. In probabilistic assessments, it
1146 often examines correlations between input distributions and output distributions.

1147 Sensitivity analysis can provide valuable assistance at various stages, from initial
1148 planning of the assessment model to communicating the results. It can help to:

- 1149 • Understand how inputs affect outputs and check for unrealistic model behaviour
- 1150 • Screen the potential influence of different sources of variability and uncertainty, to
1151 determine which are most important to quantify
- 1152 • Show the potential effects of uncertainties that cannot be satisfactorily
1153 represented with distributions (e.g. different choices for distribution shape, or
1154 different model structures)
- 1155 • Identify key sources of uncertainty, as options for further data collection if required
- 1156 • Identify key sources of controllable variability, as options for risk reduction if
1157 required (e.g. adjusting the use pattern of the pesticide)
- 1158 • Identify the main factors driving the assessment outcome, so that they can be
1159 used in risk communication to help explain the causes of risk.

1160 Many different methods are available for sensitivity analysis. Vose (2000) describes
1161 some examples of analysis outputs that can be used for sensitivity analysis. US EPA

1162 (2001) and Cullen and Frey (1999) discuss a wider range of approaches to sensitivity
1163 analysis, and offer summary tables to assist in choosing between them. There is
1164 also a substantial general literature on sensitivity analysis (e.g. Saltelli *et al.*, 2000).

1165 **3.4.7 Software**

1166 **NOTE:** This section will be revised as the project progresses to take account of
1167 feedback received on available software and databases.

1168 The EUPRA workshop, which preceded EUFRAM, recommended that standard
1169 software tools for probabilistic assessment should be adopted, at a level of
1170 complexity appropriate for users in all parts of the EU regulatory arena (Hart, 2001).
1171 EUFRAM Work Package 10 endorsed this view, and considered the characteristics
1172 that would be desirable in software and databases for probabilistic assessment. An
1173 ideal system for probabilistic assessment would:

- 1174 1. provide a comprehensive set of models for assessing the ecological risks of
1175 pesticides and their metabolites, including both exposure and effects.
- 1176 2. provide comprehensive, appropriate, referenced, quality-controlled data for all
1177 model inputs including pesticide use, physico-chemical properties, toxicity,
1178 ecological and landscape factors, and field data.
- 1179 3. include dynamic links between data and models.
- 1180 4. include appropriate methods for taking account of uncertainty, variability and
1181 dependencies, and for incorporating spatial and temporal variation.
- 1182 5. offer different modes for different users and purposes, including standardised
1183 models and data for lower tier assessments and more flexibility for higher tier
1184 assessments.
- 1185 6. be fully tested and documented, approved by appropriate expert panels,
1186 published in peer-reviewed literature and, as far as possible, empirically validated.
- 1187 7. run efficiently on computer systems typically used by regulatory assessors, or be
1188 usable remotely via the internet.
- 1189 8. have an intuitive user interface including automatic error-checking, extensive help
1190 facilities, and convenient methods for input and output of data and results.
- 1191 9. be provided with long-term maintenance and user support services.
- 1192 10. be suitable for regulatory use at both national and European level.

1193 It is recognised that some users will require more flexibility and specialist tools, that
1194 would be difficult to incorporate in a standard system, but they would still benefit from
1195 the other characteristics mentioned.

1196 A detailed evaluation of existing software and databases is presented in Volume 2,
1197 Work Package 10. The main conclusion is that existing tools meet the criteria only to
1198 a limited extent, covering only particular types of assessment²⁴, and providing limited

²⁴ For example the ETX software combines one distribution for exposure with one for effects, both of which must be lognormal; and the US EPA models for acute avian and aquatic risks are tailored to US regulatory requirements and scenarios.

1199 functionality for handling uncertainties and dependencies²⁵. Although pesticide
1200 assessments can be produced with the existing tools, further development is highly
1201 desirable.

1202 **QUESTION FOR FEEDBACK:**

1203 **Q2. What software and databases have you used for probabilistic ecological**
1204 **risk assessments? Do you have any comments on their strengths and**
1205 **weaknesses? Are you developing any software or databases yourself?**

1206 **3.5 Interpreting the results of probabilistic assessments²⁶**

1207 Risk assessment is not just about doing calculations. Quantitative risk assessment
1208 should be seen as one of several lines of evidence, each of which requires critical
1209 interpretation and evaluation to arrive at an overall characterisation of risk.

1210 **3.5.1 Interpreting the quantitative results**

1211 In general, the interpretability of the quantitative results will depend on how clearly
1212 the assessment objectives were defined, and whether the assessment endpoint was
1213 suitably designed to address those objectives. If the assessment endpoint does not
1214 or cannot relate closely to the assessment objectives (e.g. if the assessment relates
1215 to individual effects but the real interest is in population effects), then considerable
1216 interpretation may be required to bridge the gap. Such extrapolation should be
1217 supported by other lines of evidence (see 3.5.3).

1218 One of the challenges when interpreting probabilistic results is the lack of established
1219 criteria for using them in decision-making, e.g. what percentage of species may be
1220 affected? (European Commission, 2002a). This is a very important issue but requires
1221 balancing ecological risk against other relevant factors that are outside the scope of
1222 EUFRAM, such as the benefits of pesticide use and social and aesthetic
1223 considerations. EUFRAM can contribute by assisting the development of probabilistic
1224 outputs that express risk in terms that are meaningful to decision-makers (see
1225 sections 3.2 and 3.3), but if decision criteria are to be established this needs to be
1226 done by the responsible authorities in consultation with other interested parties as
1227 appropriate. So long as such criteria are lacking, decisions will have to be considered
1228 case-by-case: assessors should use the quantitative results to characterise the risk,
1229 and decision-makers should then consider its acceptability. This is the same as for
1230 most other types of refined assessment, which also lack established decision criteria.

1231 It is important always to ask “Do I believe my results?”. For example:

- 1232 • Compare the upper tails from the probabilistic assessment with the conventional
1233 first tier assessment, and compare the central estimate with the results of
1234 deterministic calculations using the mean or median values of the input
1235 distributions (US EPA, 2001). Are differences consistent with the differences in
1236 inputs, or do they suggest problems?

²⁵ Commercial risk analysis software packages (e.g. Crystal Ball®, @Risk®, Risk Calc®) provide more functionality but are general tools not tailored to pesticide assessment, so the user has to specify the assessment model and provide all the data.

²⁶ See Volume 2, Work Package 6 for further discussion.

- 1237 • For assessments using Monte Carlo, bootstrap or other iterative computations,
1238 check whether sufficient iterations have been done to obtain stable and
1239 repeatable results.
- 1240 • Compare the probabilistic results with other lines of evidence (see 3.5.3)
- 1241 If the central estimate or variability of the output seem too high or too low:
- 1242 • Check whether the assessment endpoint relates to the objectives in the way that
1243 was intended. Are the output measure and its entities and ensemble as intended,
1244 or have they been inadvertently altered by the choice of input distributions, or by
1245 the way the distributions are combined?
- 1246 • Check for errors in data units (e.g. mg to kg)
- 1247 • Consider whether important variables or relationships have been omitted
- 1248 • Identify which input variables are driving the result (using sensitivity analysis, see
1249 above) and consider whether they have been modelled correctly (e.g. whether
1250 distributions relate to the wrong ensemble, or extend beyond feasible limits). If the
1251 output is heavily driven by the tail of an input distribution, check how well the
1252 shape of that tail is supported by data or other evidence. Also consider whether
1253 the model is being driven by impossible combinations of input variables, that may
1254 occur if important dependencies are omitted.
- 1255 • Check the assumptions used for any fixed (deterministic) inputs.
- 1256 If uncertainty intervals seem too wide or too narrow to be credible:
- 1257 • Check whether you have handled them correctly (consult a statistician)
- 1258 • Consider what reasons you have for thinking they are too wide, or too narrow. If it
1259 is based on other sources of knowledge or uncertainty that are not yet
1260 incorporated in the model, either try to incorporate them in the model, or
1261 incorporate them in the final risk characterisation as other lines of evidence (see
1262 3.5.3).
- 1263 • Consider whether the intervals are being inflated by impossible combinations of
1264 input uncertainties, which may occur if important dependencies are omitted.

1265 **3.5.2 Considering additional uncertainties and dependencies**

- 1266 It is never possible to quantify all the uncertainties and dependencies that might
1267 affect an assessment. The potential influence of unquantified uncertainties and
1268 dependencies should therefore be considered:
- 1269 • Systematically list or tabulate all unquantified uncertainties and dependencies
 - 1270 • Qualitatively evaluate the direction and relative magnitude of each unquantified
1271 uncertainty and dependency in terms of its impact on the assessment outcome
 - 1272 • Consider re-running the assessment with different assumptions to assess the
1273 influence of potentially important uncertainties or dependencies (a form of
1274 sensitivity analysis)
 - 1275 • Consider using probability bounds to explore the impact of potentially important
1276 uncertainties about distribution shape and dependencies (in effect, putting
1277 conservative outer bounds around the assessment result)

- 1278 • Qualitatively evaluate the possible combined effect of all the quantified and
1279 unquantified uncertainties and dependencies on the assessment outcome (how
1280 different could the “true” outcome be?).

1281 **3.5.3 Considering other lines of evidence**

1282 Risk characterisation should seek to provide an overall assessment, combining the
1283 outcome of the probabilistic assessment with any other lines of evidence that are
1284 relevant to the assessment objectives. Examples of other lines of evidence might
1285 include:

- 1286 • higher-tier laboratory, semi-field or field studies
- 1287 • monitoring data on exposure or impacts
- 1288 • data on similar compounds
- 1289 • the context and scope of the assessment – how worst-case it is, how generally
1290 applicable it is.

1291 Sometimes, different lines of evidence can be compared quantitatively: for example if
1292 they represent independent estimates of the same quantity, or if one is an
1293 independent estimate of a component of the other (e.g. measured concentrations vs.
1294 predicted concentrations within an aquatic assessment). It may even be possible to
1295 combine them, e.g. by Bayesian updating. Often, however, the lines of evidence will
1296 be sufficiently different in nature that only a qualitative comparison is possible.

1297 If other lines of evidence are to influence the characterisation of risk, then they
1298 should be subject to the same standards of objective assessment as the probabilistic
1299 assessment. This will involve many of the same considerations, such as
1300 measurement and sampling uncertainties, bias, and any extrapolation required
1301 between the measure or ensemble of the data and those of the assessment.

1302 Two lines of evidence may give very different central estimates, if one or both are
1303 very uncertain. If the disagreement is larger than can be accounted for, then the
1304 basis of each line of evidence should be re-examined. For example, disagreement
1305 between a probabilistic assessment and monitoring data might result either from
1306 invalid model assumptions or biases in monitoring. If both appear sound, then the
1307 difference between them should be regarded as a form of uncertainty in the overall
1308 risk characterisation. No line of evidence should be discounted unless there is
1309 objective evidence that it is invalid.

1310 **3.5.4 Considering the wider ecological consequences**

1311 Risk characterisation may also include consideration of the wider ecological
1312 consequences of effects indicated by the assessment. This is often necessary
1313 because the quantitative assessment is constrained (for reasons of practicality) to
1314 simple measures of risk that are only indirect or partial measures of the assessment
1315 objectives or protection goal. For example, quantitative estimates often relate to
1316 individual effects, but decision-makers may be more interested in impacts at the level
1317 of the population or community.

1318 Interpreting wider ecological consequences involves:

- 1319 • Identifying which types of consequences are of interest, and to whom. Ideally, this
1320 should already have been considered when specifying the assessment objectives
1321 (section 3.2). It is a risk management judgement, informed partly by scientific
1322 understanding of ecological structure and function but also by social and aesthetic
1323 values.

1324 • Using the quantitative assessment together with any other relevant lines of
1325 evidence to assess those consequences. This should be an objective scientific
1326 assessment, applying the same standards of enquiry to each of the relevant lines
1327 of evidence. For example, arguments about the ability of the ecological systems
1328 to absorb or recover from pesticide impacts should be supported by evidence
1329 (e.g. relevant field studies or modelling studies), and the effect of those processes
1330 in the specific scenario under consideration should be evaluated.

1331 Risk management decisions may legitimately be influenced by other considerations,
1332 including economic, social and environmental benefits that the chemical may confer.
1333 However, while these factors may alter the balance of risk and benefit they are part
1334 of risk management and should not influence the scientific characterisation of risk.

1335 **3.6 Reporting the results of probabilistic assessments**

1336 This section summarises provisional EUFRAM recommendations on how to report
1337 the results of probabilistic assessments. For full details, see Volume 2, Work
1338 Package 6.

1339 The formal assessment report should comprise 3 main parts, serving complementary
1340 purposes:

- 1341 • **Summary** – to **communicate the main points** required for decision-making
- 1342 • **Main report** – to **document and justify the methods, results and conclusions**
1343 clearly and concisely but in sufficient detail to enable critical evaluation of all
1344 stages by other specialists (e.g. peer reviewers)
- 1345 • **Appendices** – to provide any additional background information required for
1346 others to **duplicate the assessment** if required.

1347 Each section should be drafted with the needs of the intended audience in mind.
1348 Specialists conducting detailed evaluation or peer review of the assessment (e.g.
1349 Rapporteur Member State scientists in current EU procedures) are the primary
1350 audience for the main report. Decision-makers, and technical specialists who are not
1351 themselves required to review the assessment in detail, will probably rely mainly on
1352 the summary and only refer to the main report for clarification or detail on specific
1353 points.

1354 The **summary** should therefore provide a clear and concise summary of the
1355 assessment objectives, the main results of the quantitative assessment, and the
1356 main points of the overall risk characterisation and ecological interpretation including
1357 unquantified variability, uncertainties and dependencies and other lines of evidence.
1358 It should also give a very brief indication of the methods used, and state whether they
1359 are novel or follow an established approach.

1360 The **main report** should comprise a series of sections documenting and justifying the
1361 main components of the assessment:

- 1362 • **Problem definition.** This should provide full documentation of the assessment
1363 objectives (section 3.2) and describe and justify the chosen assessment
1364 endpoints (3.3).
- 1365 • **Exposure assessment.** This should document and justify the exposure model
1366 and its inputs, report the results of any sensitivity analysis, describe the

1367 computational methods used, evaluate the model output, and explain the use
1368 made of environmental measurements (if any).

1369 • **Effects assessment.** This should summarise the effects data and justify any
1370 selection criteria or weightings, describe and justify the computational methods,
1371 and present and evaluate the results.

1372 • **Risk characterisation.** This should describe the methods used to combine the
1373 effects and exposure distributions and derive the assessment endpoint, provide a
1374 description of the organisms likely to be at risk, and characterise the probability of
1375 exceeding any defined effects thresholds and the frequency, magnitude and
1376 duration of effects. It should also list the quantified and unquantified sources of
1377 variability, uncertainty and dependency and discuss their potential influence on
1378 the assessment outcome. The overall risk characterisation should also include a
1379 comparison with deterministic results and evaluate any other lines of evidence
1380 bearing on the endpoint of the probabilistic assessment.

1381 • **Ecological interpretation.** This should discuss which wider ecological
1382 consequences are of concern, and use the quantitative assessment together with
1383 any other relevant lines of evidence to assess those consequences.

1384 The sections on exposure and effects assessment and risk characterisation should
1385 document and justify all aspects of the **methodology** including the assessment
1386 model (section 3.3), the specification of input distributions (section 3.4.3), methods
1387 used for combining distributions (3.4.1 and 3.4.2), methods for dealing with
1388 uncertainties and dependencies (3.4.4 and 3.4.5), sensitivity analysis (3.4.6), and
1389 software (3.4.7). Details necessary to repeat the assessment (e.g. software version
1390 numbers and random number seeds) should be included either in these sections or
1391 in appendices.

1392 The choice of graphical, tabular and narrative formats for communicating **output**
1393 **distributions** is discussed in section 3.7. Suitable graphs should also be presented
1394 for **input distributions**, to support decisions made in distribution selection.

1395 The **appendices** should list sufficient of the input data, methodological details and
1396 output data to enable other specialists to duplicate the assessment if required.

1397 **Technical terms** used in the report should follow established convention where
1398 possible. Consideration should be given to including a glossary in the appendices,
1399 especially if less familiar terms are used.

1400 **3.7 Communicating results of probabilistic assessments**

1401 Probabilistic assessments are more difficult to communicate than conventional
1402 deterministic ones, and this could be a major obstacle to the acceptance of
1403 probabilistic approaches by end-users. The development of effective communication
1404 strategies is therefore a **key objective** of EUFRAM. This section summarises the
1405 provisional findings of EUFRAM Work Package 7 on this topic, which are presented
1406 in detail in Volume 2. Evaluation of different communication strategies is an ongoing
1407 activity within the project, and feedback via the EUFRAM website and workshops will
1408 be used to update this section as the project progresses.

1409

QUESTION FOR FEEDBACK:

1410

Q3. Do you have any suggestions for improving the communication of probabilistic assessments and their outputs?

1411

1412

3.7.1 Communicating with different types of audience

1413

The immediate audiences for probabilistic assessments are technical specialists (e.g. peer reviewers, evaluators) and decision-makers or policy staff in government, industry and NGOs (non-governmental organisations). It is also important to consider other interested parties, including the general public, particularly in the context of increased transparency associated with the practice of risk analysis. Some members of the public may take a special interest in pesticide risks and, more generally, it is important to build societal trust in the process of risk assessment and risk management. Increasing *public* trust in scientific risk assessments is contingent on effective communication about uncertainties regarding risks, about what is not known as well as what is known, as otherwise the public will not perceive that there is an active attempt on the part of the scientific community to explain what is really happening within the risk assessment process.

1414

1415

1416

1417

1418

1419

1420

1421

1422

1423

1424

1425

Different audiences have different communication needs, which must be considered as part of the process of developing an effective risk communication strategy. Technical specialists who are peer-reviewing an assessment, or attempting to repeat or extend it, will need detailed reports and appendices containing comprehensive information on every aspect. Other technical specialists and decision-makers may require only summary information (Section 3.6). Summary information will also be important for public audiences, and this may need presenting using a different approach e.g. avoiding unfamiliar technical terms and using analogies to explain unfamiliar concepts. Although public audiences will rarely want detailed information, public trust may be improved if they know it is available on request (Frewer et al, 2004).

1426

1427

1428

1429

1430

1431

1432

1433

1434

1435

1436

Despite the different needs of these audiences, many of basic strategies for effective communication are generic. The following section considers how these strategies may apply to the communication of results from probabilistic risk assessments.

1437

1438

1439

3.7.2 Effective formats for communicating probabilistic results

1440

The following list of interim findings has been developed from discussions within the EUFRAM project, based in part on experience gained when presenting some of the EUFRAM case studies to regulatory authorities and committees and on feedback from participants at the first EUFRAM workshop.

1441

1442

1443

1444

1. EUFRAM provisionally recommends the use of cumulative distributions in preference to other graphical formats for the immediate technical outputs of probabilistic assessments (see Section 2.4.5). It is hoped that standardising on these will facilitate communication amongst technical specialists, but further evaluation is required to confirm this.

1445

1446

1447

1448

1449

2. Efforts by EUFRAM partners to use cumulative distributions with less technical audiences have been discouraging up to now. However, Morgan & Henrion (1990) report experiments showing that non-technical audiences find cumulative distributions more intuitive than other kinds of display. Therefore it would be desirable to attempt improvements in the detailed format of cumulative graphs

1450

1451

1452

1453

- 1454 and then evaluate again whether they can contribute to communication with
1455 decision-makers, in combination with other techniques.
- 1456 3. Further work is required to optimise the detailed formatting of cumulative graphs.
1457 In particular, we have not yet found an optimal solution to labelling the axes of
1458 cumulative graphs to ensure their cumulative nature is readily understood (e.g.
1459 putting “proportion <X” on the y-axis and “X” on the x-axis is confusing). If we
1460 cannot find an adequate solution for this, we may need to reconsider whether
1461 other formats (including exceedance plots) might be more readily understood.
- 1462 4. EUFRAM workshop participants expressed a preference for graphs showing fitted
1463 cumulative distributions together with datapoints, rather than empirical
1464 distributions.
- 1465 5. It may be preferable to avoid, or provide additional explanation for, graphs with
1466 frequencies or percentages on both axes (e.g. “joint probability curves”), as they
1467 seem generally more difficult to understand and communicate.
- 1468 6. US EPA (1997) recommend showing both the probability density function and
1469 cumulative distribution together on same page with axes aligned, for inputs as
1470 well as outputs. We should consider whether this would be helpful.
- 1471 7. Graphical formats other than distributions should also be considered further,
1472 including bar charts, which have been liked by some regulatory audiences.
1473 Publications focussing on the design of scientific graphics (e.g. Tuffe, 2001) may
1474 be helpful in refining our preferred formats.
- 1475 8. Care is required in choice of symbols, colours and line styles, to ensure that
1476 displays are still understandable when photocopied.
- 1477 9. Some people like distributions of TERs, probably because the TER is a familiar
1478 assessment output. However, others don’t like distributions of TERs, because the
1479 established decision criteria don’t apply to them in the normal way.
- 1480 10. Although graphical formats may be more effective in communicating an overall
1481 impression of the results, tabular formats are more effective when numeric results
1482 for specific statistics (e.g. the average or 95th percentile) are required. EUFRAM
1483 has given less attention so far to evaluating tabular formats.
- 1484 11. Different individuals within any audience have different preferences. It may
1485 therefore be good practice to present and explain the same information in both
1486 graphical and tabular formats, as well as text.
- 1487 12. Responses at the EUFRAM workshop indicated conflicting desires for simple and
1488 uncluttered graphs and tables on the one hand, and on the other hand for
1489 supporting information (e.g. context, assumptions, identification of species on
1490 SSDs etc.). The appropriate balance between these needs may vary from case to
1491 case depending on the context and audience.
- 1492 13. Good legends can do a lot to improve understanding of graphs and tables and we
1493 need to make better use of this.
- 1494 14. One firm conclusion, for all types of audiences, is the need always to express and
1495 explain the results in words alongside any graphical or tabular presentation. Use
1496 everyday language as far as possible, even for technical audiences; where
1497 technical terms and acronyms are necessary they should be explained. Consider

- 1498 providing a glossary. Bullets and text boxes containing key messages may be
1499 helpful.
- 1500 15. Insights from psychometrics and evolutionary psychology suggest that natural
1501 frequencies (e.g. 1/10, 2/5) are more readily understood than percentages
1502 (Gigerenzer, 2002). How far this applies to graphical, tabular and narrative
1503 outputs of probabilistic pesticide assessments will be evaluated further as the
1504 EUFRAM project continues.
- 1505 16. It is important to provide a balanced picture of what is known and what is
1506 uncertain. Where uncertainties have been incorporated into the analysis, they can
1507 be shown as a range or confidence interval for the result. The potential effect of
1508 other, unquantified, uncertainties on the assessment outcome should also be
1509 described (e.g. how different could the 'true' risk be?). However, the strengths of
1510 the analysis (what is known) should also be made clear.
- 1511 17. If the decision-maker is likely to consider options for risk mitigation (e.g. buffer
1512 zones), then it will be helpful to include comparative assessment results for the
1513 different mitigation options. Ideally, the need to assess such options should be
1514 identified at the outset and included in the assessment objectives (Section 3.2).
- 1515 18. Inclusion of a "*positive control*", the effects of which are already well understood
1516 by those involved in the risk analysis process, facilitates communication about
1517 new methods. Although this should not be a requirement within individual risk
1518 assessments, it could be very useful for:
- 1519 ○ *Demonstration/teaching purposes* (provided it is not interpreted as a
1520 benchmark for what effects are acceptable, because this may vary between
1521 pesticides/scenarios),
 - 1522 ○ Providing a *test case* for evaluation of new tools (to help build confidence and
1523 check consistency between tools)
- 1524 19. The ecological implications (the "so what" question) should be explained, bringing
1525 in relevant expertise or supporting evidence, in order to understand which
1526 organisms are affected, the implications for vulnerable or protected habitats, etc.
- 1527 20. A substantial part of the difficulty in communicating probabilistic results is due to
1528 the lack of established criteria for decision-making. Even if the results address the
1529 assessment objectives and are clearly communicated, the lack of decision criteria
1530 makes them hard to interpret. It is unlikely that standard criteria will be established
1531 soon so, for the time being, results will have to be evaluated case by case. This
1532 further increases the importance of providing good interpretative text with results.
- 1533 21. Familiar formats and concepts (e.g. TER) are more easily understood than novel
1534 ones. Explanations should be started with familiar assessment methodologies
1535 and subsequently move to unfamiliar assessment approaches (e.g. from
1536 deterministic to probabilistic assessments). Novel ideas should be introduced one
1537 at a time rather than all at once.
- 1538 22. Although regulatory processes focus heavily on documents and reports, the
1539 importance of effective face-to-face interactions should not be overlooked.
1540 Consideration should be given to providing joint training exercises for risk
1541 assessors and risk managers, to promote effective interaction during the risk
1542 assessment process and train them in the use and interpretation of relevant
1543 communication formats.

1544 **3.8 Potential for harmonised approaches**

1545 Harmonisation of approaches in regulatory assessment can help to ensure quality
1546 and improve cost-effectiveness, but its applicability to probabilistic risk assessment
1547 requires careful consideration. US EPA (1997) noted that quantitative risk
1548 assessment approaches are developing rapidly and stated that EPA guidance is not
1549 intended to constrain the use of new or innovative approaches where scientifically
1550 defensible. Similarly, in a report on microbial risk assessment, the former EU
1551 Scientific Steering Committee (European Commission, 2003) stressed the need for
1552 further evolution of the methodologies for quantitative risk assessment and
1553 concluded that a quick harmonisation at the present state-of-the-art should be
1554 avoided. These arguments seem equally valid for the assessment of pesticide risks.

1555 Although flexibility is essential, it may not be necessary to develop every probabilistic
1556 assessment completely afresh, because similar pesticide uses require similar
1557 assessments. For example, assessing the same type of risk (e.g. acute aquatic) for
1558 two insecticides with similar use patterns would probably imply similar assessment
1559 objectives, endpoints, and models, could use similar methods and software to
1560 combine distributions and handle uncertainty and dependencies, and similar formats
1561 for communicating results. Only things specific to each pesticide and their specific
1562 use scenarios need differ: e.g. their toxicities and physico-chemical properties.
1563 Therefore, if a probabilistic assessment has been conducted for one such pesticide
1564 and accepted by the relevant authorities, then much of the approach may be directly
1565 transferable to other, similar pesticides. This implies increasing potential for resource
1566 savings as examples of “accepted” assessments accumulate.

1567 This transferability of approaches opens the possibility of establishing generic
1568 probabilistic approaches for some of the scenarios that most frequently require
1569 refined assessment under 91/414/EEC. This would increase efficiency both for
1570 people conducting probabilistic assessments, and also for people evaluating them.
1571 An essential pre-condition for the initial establishment of such generic approaches
1572 will be thorough peer review by relevant technical experts (including ecotoxicology,
1573 environmental fate, probabilistic risk analysis and statistics). Once established, they
1574 could then be used in subsequent assessments without the need for detailed model
1575 development or peer review on every occasion. Furthermore, consideration could be
1576 given to developing customised tools (e.g. software and databases) for such generic
1577 scenarios, similar to the FOCUS models for exposure assessment. Of course, the
1578 suitability of generic approaches and tools should be checked for each new
1579 assessment, and it should always remain open for assessors and decision-makers to
1580 select new or non-standard approaches where appropriate.

1581 **PC9. EUFRAM provisionally recommends that consideration should be given to**
1582 **establishing generic, peer-reviewed probabilistic approaches and tools for**
1583 **scenarios that frequently require refined assessment under 91/414/EEC.**

1584 **3.9 Summary checklist**

1585 **NOTE:** When preparing future revisions of this document, consideration will be given
1586 to including a 1-2 page summary checklist of the approaches recommended by
1587 EUFRAM.

1588 4 CASE STUDIES OF PROBABILISTIC ASSESSMENT

1589 **NOTE:** It is intended in later versions to include summaries of 2 (or possibly more)
1590 case studies at this point in the document.

1591 5 LIST OF CONTENTS FOR VOLUME 2

1592 Work package 3: Role and outputs of probabilistic assessments

1593 Work package 4: Methods of uncertainty analysis

1594 Work package 5: Probabilistic approaches for use with typical datasets

1595 Work package 6: Reporting probabilistic assessments

1596 Work package 7: Communicating the results of probabilistic assessments

1597 Work package 9: Practical approaches for validation

1598 Work package 10: Software and databases for probabilistic assessment

1599 Work package 11: Pooling data for probabilistic approaches

1600 6 LIST OF CONTENTS FOR VOLUME 3

1601 Work Package 8. Case studies:

1602 • Case study 1 - Aquatic risk assessment for carbaryl

1603 • Case study 2 - Atrazine aquatic risk assessment

1604 • Case study 3 - Risk to aquatic organisms arising from exposure to 'Herburon' via
1605 spray drift

1606 • Case study 4 - Acute risks to skylarks arising from a summer cereal spray

1607 • Case study 5 - Risk assessment for an organophosphate insecticide under
1608 Mediterranean conditions

1609 7 GLOSSARY AND ABBREVIATIONS

Assessment endpoint	The output required from a risk assessment. For a probabilistic risk assessment, it is defined by a measure of the magnitude of effects, the entities for which those effects are to be estimated, and the ensemble of entities for which the frequency of effects is to be estimated.
Bayesian statistics	One of two main schools of statistics. Bayesian approaches define the probability of an event as the degree of belief that a person has, given some state of knowledge, that the event will occur. Also called the subjective view of probability, in contrast to the frequentist view.
Bias	Systematic error, e.g. a measurement or statistical method is biased if it produces mean results that are consistently different from the "true" mean.
Bootstrapping	A computational method that estimates uncertainty in population parameters such as the mean by random sampling from observed data, generating a series of samples of equal size to the original dataset.
Classical statistics	See Frequentist.

Conceptual model	A written and/or graphical representation (e.g. flow chart) of the model to be used for a risk assessment, showing the causal mechanisms leading from pesticide use to effects on the assessment endpoint. Can also be used to show the relationships between these mechanisms and the input data and assumptions used in the assessment.
Confidence interval	A range of numbers believed to include an unknown quantity with a specified degree of confidence (e.g. 95%).
Correlation	A quantitative relationship between two or more variables. The most familiar form is a linear positive (or negative) relationship, such that high values of one variable are associated with higher (or lower) values of the other.
Cumulative distribution function (CDF)	A function or graph expressing the probability that a random variable is less than or equal to a certain value. The CDF is obtained by integration of the PDF for a continuous random variable, or summation of the PDF in the case of a discrete random variable.
Dependency	Any departure from statistical independence between two or more quantities, such that the expected value of each varies depending on the value of the other.
Deterministic	Methods that use point estimates to represent one or more factors in a risk assessment and treat them as if they were fixed and precisely known.
Distribution	A graph or mathematical function defining the relative probabilities or frequencies of alternative values for a variable or uncertain quantity.
EC50	Effective concentration affecting 50% of tested population.
Empirical distribution	A distribution of actual data, in some cases smoothed with interpolation techniques. Data are not fit to a particular parametric distribution, but are described by percentiles of the dataset.
Ensemble	A group of entities or population of individuals. EUFRAM uses ensemble to refer to the group of entities to which a distribution refers. The ensemble for the assessment endpoint is the group of entities, or population of individuals, for which the frequency of effects is estimated.
Entity	EUFRAM uses entity to refer to the individual organisms or things to which the values in a distribution relate. The entity for the assessment endpoint is the organism, population, species, community, or ecological structure or function, for which the magnitude and frequency of effects are estimated.
Exceedance distribution	A function expressing the probability that a random variable is greater than a certain value. The probability at each point on an exceedance distribution is equal to one minus the probability at the same point on the cumulative distribution function for the same variable.
Extrapolation uncertainty	The uncertainty which results when it is necessary to extrapolate beyond the range of a dataset, or from one type of data to another (surrogacy). Uncertainty about how closely the extrapolated values match the true values that are being estimated.
FOCUS	FORum for the Co-ordination of pesticide fate models and their Use.
Frequentist statistics	One of two main schools of statistics. Frequentist approaches define the probability of an event occurring as the frequency of occurrence measured in a series of repeated trials. Sometimes called classical statistics. See also Bayesian statistics.
Goodness of fit	Assessment of how well (or poorly) a sample of data can be described by a hypothesised parametric distribution. Methods for assessing goodness of fit include statistical tests and various graphical comparisons (e.g. US EPA 2001, Cullen & Frey 1999).
HC _p (e.g. HC ₅)	Hazardous Concentration for p% (e.g. 5%) of the species. The 5 th percentile of a species sensitivity distribution for a toxicity endpoint expressed as concentration (e.g. LC50 or NOEC).

Histogram	A graph that groups the data into intervals and displays the count of observations within each interval.
Independence	Two variables are independent if the value assumed by each variable is not related to or influenced by the value of the other. Two events are independent if the probability of each event occurring is not related in any way to whether the other occurs.
Joint probability method/curve (JPC)	Graphical method of combining exposure and effects distributions that share the same x-axis (e.g. concentration) into a single output distribution or joint probability curve. Can be plotted in several formats, one of which is equivalent to a cumulative distribution for the frequency of effects.
LC50/LD50	Lethal concentration/dose affecting 50% of tested population.
Maximum entropy	A method for selecting parametric distributions, which is conservative in that it chooses the broadest distribution that is compatible with the information available.
Measurement uncertainty	Uncertainty in measurements due to factors that cause random errors or bias in the measurement process. Uncertainty about how closely measured values match the true values of the quantities that are being measured.
Mixture distribution	Use of a combination of several different parametric distributions to describe a single dataset or theoretical population.
Model uncertainty	The component of the uncertainty concerning an estimated value that is due to possible misspecification of a model used for the estimation. It may be due to the choice of the form of the model, its parameters, or its bounds.
Monte Carlo	A probabilistic technique which repeatedly samples distributions for input variables to estimate the distribution of a model's output variable. 1D Monte Carlo – all input distributions are sampled together in every repetition of the calculation, producing a single output distribution with no confidence intervals. 2D Monte Carlo – distributions representing variability and uncertainty are sampled separately, so that the combined effect of the uncertainties can be shown as a confidence interval around the output distribution.
NOEC	No observed effect concentration
Normal distribution	Familiar symmetrical bell-shaped parametric distribution, defined by its mean and variance (or standard deviation).
Parametric distribution	A theoretical distribution that is specified by one or more parameters, e.g. the normal distribution, which is specified by its mean and variance.
PDF	Probability density function.
PEC	Predicted environmental concentration.
Percentile	The p -th percentile of a distribution is the value such that p percent of observations fall at or below it. Percentiles (more properly called centiles) are expressed as percentages (0-100%). Equivalent to quantiles or fractiles, which are expressed as proportions (0-1).
Point values or estimates	Representation of a measured or estimated quantity by a single number, instead of a distribution.
Probabilistic Risk Assessment	Risk assessments that use probabilities or distributions to quantify one or more sources of variability and/or uncertainty in exposure and/or effects and the resulting risk.
Probability	Probabilities are used to quantify variability and/or uncertainty. The probability of something can be defined as its frequency in repeated independent trials, or as the degree of belief that it will occur, expressed as a proportion (i.e. a number between 0 and 1).
Probability bounds	A method for representing and combining uncertain distributions that does not require any data or assumptions about either distribution shape or dependencies among variables (see Ferson, 2002).

Probability Density Function (PDF)	For a continuous random variable, the PDF expresses the relative probability of different values. For a discrete random variable, the PDF expresses the probability that the random variable is equal to a specific (discrete) value.
Problem definition	The phase in a risk assessment in which the assessment objectives and required output (assessment endpoint) are defined.
Regression	Relationship between two variables such that one is determined at least partly by the other.
Risk	The magnitude, frequency and uncertainty of a specified effect, event or consequence.
Sampling uncertainty	Uncertainty due to estimating a population distribution or statistic from a sample of data representing the larger population. Uncertainty about how closely the estimated distribution or statistic matches the true distribution or statistic for the larger population from which the sample was drawn.
Sensitivity analysis	Analysis of the influence of alternative inputs or assumptions on the output of an assessment.
Spaghetti plot	Graph showing multiple PDFs or CDFs plotted together, to represent uncertainty about the true PDF or CDF. Each plotted distribution represents one possible realisation of the true distribution.
Species sensitivity distribution (SSD)	A PDF or CDF of the toxicity of a certain substance or mixture to a set of species which may be defined as a taxon, assemblage, or community. Empirically, a PDF or CDF is estimated from a sample of toxicity data for the specified species set.
Surrogacy	Where data or measurements on one quantity is used as a substitute for a different quantity, e.g. because it is too difficult or costly to measure directly.
Tails	Lower and upper ends of a distribution. Often important in risk assessment, and often poorly estimated by data.
TER	Toxicity-exposure ratio.
Transformation	Mathematical modification or conversion of data, e.g. by taking logarithms. Often used to improve the goodness of fit between data and a parametric distribution.
Truncation	The process of setting lower and upper limits on the range of a distribution, in order to avoid unrealistic or impossible values.
Uncertainty	Limitations in knowledge, e.g. about factors that influence risk.
Variability	Real variation, e.g. in factors that influence risk. Heterogeneity over time, space or different members of a population.

1610

1611 8 REFERENCES

1612 Aldenberg T, Jaworska JS & Traas TP. 2002. Normal species sensitivity distributions
1613 and probabilistic ecological risk assessment. In: The use of Species Sensitivity
1614 Distributions in Ecotoxicology, Posthuma L., T.P. Traas and G.W. Suter (eds.) Lewis
1615 Publishers, Boca Raton, FL, USA. Pp. 49-102.

1616 Aldenberg T, Jaworska J. 2000. Uncertainty of the hazardous concentration and
1617 fraction affected for Normal species sensitivity distributions. *Ecotoxicology and
1618 Environmental Safety*, 46: 1-18.

1619 Cullen AC and Frey HC, 1999. Probabilistic techniques in exposure assessment. A
1620 handbook for dealing with variability and uncertainty in models and inputs. Plenum
1621 Press, NY.

- 1622 European Commission, 2002a. Guidance document on aquatic ecotoxicology in the
1623 context of Directive 91/414/EEC. Sanco/3268/2001 rev.4 (final), 17 October 2002,
1624 Brussels.
- 1625 European Commission, 2002b. Guidance document on terrestrial ecotoxicology
1626 under Directive 91/414/EEC. SANCO/10329/2002 rev 2 final, 17 October 2002,
1627 Brussels.
- 1628 European Commission, 2002c. Guidance document on risk assessment for birds and
1629 mammals under Council Directive 91/414/EEC. Sanco/4145/2000, 25 September
1630 2002, Brussels.
- 1631 European Commission, 2003. Risk assessment of food borne bacterial pathogens:
1632 Quantitative methodology relevant for human exposure assessment. Appendix 3 in:
1633 The future of risk assessment in the European Union. The second report on the
1634 harmonisation of risk assessment procedures. Scientific Steering Committee.
1635 European Commission, Brussels, Belgium.
- 1636 Ferson, S. 2002. *RAMAS Risk Calc 4.0 Software: Risk Assessment with Uncertain*
1637 *Numbers*. Lewis Publishers, Boca Raton, Florida.
- 1638 L. Frewer, J. Lassen, B. Kettlitz, J. Scholderer, V. Beekman, K.G. Berdal (2004)
1639 Societal aspects of genetically modified foods. *Food and Chemical Toxicology* 42(7,
1640 special issue ENTRANSFOOD): 1181-1193.
- 1641 Gigerenzer, G. 2002. *Calculated risks: how to know when numbers deceive you*.
1642 Simon & Schuster, New York.
- 1643 Hart, A. 2001. Probabilistic risk assessment for pesticides in Europe: implementation
1644 and research needs. Report of the European workshop on Probabilistic Risk
1645 Assessment for the Environmental Impacts of Plant Protection Products (EUPRA).
1646 Central Science Laboratory, Sand Hutton, UK. 109pp. Electronic copies can be
1647 downloaded via links at www.eufram.com.
- 1648 Morgan MG and Henrion M, 1990. *Uncertainty*. Cambridge University Press.
- 1649 Posthuma L, Suter GW III, Traas TP. 2002. *Species sensitivity distributions in*
1650 *ecotoxicology*. Lewis Publishers, Boca Raton, FL.
- 1651 Regan HM, Hope BK, Ferson S. 2002. Analysis and portrayal of uncertainty in a
1652 food-web exposure model. *Human and Ecological Risk Assessment*, 8: 1757-1777.
- 1653 Saltelli A, Chan K, EM Scott. 2000. *Sensitivity analysis*. John Wiley & Sons Ltd,
1654 Chichester.
- 1655 Suter, G.W. II., L.W. Barnthouse, S.M. Bartell, T. Mill, D. Mackay and S. Patterson.
1656 1993. *Ecological Risk Assessment*. Lewis Publishers, Boca Raton, FL, USA. 538 p.
- 1657 Tufte ER, 2001. *The visual display of quantitative information*. Graphics Press, USA.
- 1658 Van den Brink, P.J., L. Posthuma and T.C.M. Brock (2002). The value of the Species
1659 Sensitivity Distribution concept for predicting field effects: (non-)confirmation of the
1660 concept using semi-field experiments. In: *The use of Species Sensitivity Distributions*
1661 *in Ecotoxicology*, Posthuma L., T.P. Traas and G.W. Suter (eds.) Lewis Publishers,
1662 Boca Raton, Fl, USA. Pp. 155-193.
- 1663 US EPA. 1997. *Guiding principles for Monte Carlo analysis*. US Environmental
1664 Protection Agency, Risk Assessment Forum, Washington DC. Available for
1665 downloading at www.epa.gov.

- 1666 US EPA. 1998. Guidelines for ecological risk assessment. Washington, DC: US
1667 Environmental Protection Agency.
- 1668 US EPA. 2001. Risk assessment guidance for Superfund. Volume III – Part A,
1669 Process for conducting probabilistic risk assessment. US Environmental Protection
1670 Agency, Office of Emergency and Remedial Response, Washington DC. Available
1671 for downloading at www.epa.gov.
- 1672 Vose, D. 2000. Risk analysis: a quantitative guide. 2nd edition. John Wiley & Sons
1673 Ltd.
- 1674

1674 **ANNEX 1. MEMBERS OF THE EUFRAM PROJECT CONSORTIUM.**

- 1675 Central Science Laboratory, UK (Coordinator)
- 1676 ALTERRA Green World Research, NL
- 1677 Applied Biomathematics, NY, USA
- 1678 Bayer CropScience AG, Germany
- 1679 Bayer CropScience GmbH, Germany
- 1680 Board for the Authorisation of Pesticides, The Netherlands
- 1681 Centre for EcoChemistry, York University, UK
- 1682 Centre for Environmental Research, Germany
- 1683 Centre for Toxicology, University of Guelph, Canada
- 1684 Danish Environmental Protection Agency, Denmark
- 1685 Ecotox Limited, UK
- 1686 Federal Office for Consumer Protection and Food Safety, Germany
- 1687 Federal Environmental Agency, Germany
- 1688 Fraunhofer Institute for Molecular Biology and Applied Ecology, Germany
- 1689 General Direction Of Crop Protection, Portugal
- 1690 Institute "Mario Negri", Milan, Italy
- 1691 MAPP, The Aarhus School of Business, Denmark
- 1692 National Chemicals Inspectorate, Sweden
- 1693 National Institute for Agriculture and Food Research and Technology, Spain
- 1694 National Institute for Public Health and the Environment, NL
- 1695 Norwegian Agricultural Inspection Service
- 1696 Pesticide Control Service, Dept. of Agriculture, Food & Rural Development, Ireland
- 1697 Pesticides Safety Directorate, UK
- 1698 Procter & Gamble Eurocor SA, Belgium
- 1699 Scientific Consulting Company GmbH, Germany
- 1700 Springborn Laboratories (Europe) AG, Switzerland
- 1701 Syngenta, UK
- 1702 The National Environmental Research Institute, Denmark
- 1703 Wageningen University, The Netherlands