IFRLUP Product P5 – Report on the Review of HSE Models and Methodologies Used to Set Land-Use Planning Consultation Zones

Report by:
John Allen
Matt Lea
Richard Thomas

Date: 06-12-04

Version: Final

Contents

Executive Summary

Introduction

1. The primary purpose of the IFRLUP project is to implement recommendations (relevant to HID) from the Safety Policy Directorate (now PG) led Fundamental Review of Land Use Planning. A key recommendation of the review was:

"The criteria and methodology used for setting planning zones and for "calling in" planning applications should be reviewed and, if necessary, revised and then published."

2. Product P5 addresses the methodology element by carrying out an appropriate and proportionate review of the models and methodologies most commonly used by HSE's Methodology and Standards Development Unit (MSDU) in setting Consultation Distances (CDs). The key aim of the review being to prioritise the models and methodologies by development need.

The Assessment Methodology

3. To meet the needs of P5 and ensure that the wide range of models and methodologies used by MSDU were assessed in a fair, robust and comparable way an assessment methodology was developed by a multidisciplinary team which included representatives from:

   - The IFRLUP project team
   - MSDU
   - WS Atkins –(Independent external expert)
   - HSL
   - HID OPU

4. The methodology developed was a judgmental process relying on a common team of people carrying out the assessments to achieve a suitable degree of consistency.

5. The assessment team assessed each selected model/methodology against a set of 3 key criteria and 3 supplementary criteria.

   - Fitness for purpose - How well does the model or methodology meet HSE's risk assessment needs?
   - Risk to HSE - Is the model/methodology likely to be challenged?
   - Wider influences - Are there any external drivers likely to require changes?

   } Key Criteria

- Extent of use - How much need is there for this model/methodology?
- Likelihood of successful development. - How easy is it for development of the model/methodology to be carried out?
- Snapshot information - Where site-specific data is used, how valid is it?

} Supplementary Criteria

6. Each criterion was assessed against a number of sub criteria and scored to assist the later analysis stage.
7. To validate/verify the findings of the assessment team an independent specialist team from HSL assessed a small selection of models and methodologies and compared their results with those of the P5 team to ensure that the assessments had been thorough, complete and consistent.
8. Also, MSDU Topic Specialists play a key role in advising on the use of the models and methodologies and were interviewed independently by HSL to provide supporting information regarding where further development work may be required.

Prioritising Models and Methodologies for the Assessment

9. With approximately 80 models and over 20 methodologies in the MSDU LUP portfolio it was impractical to assess all of them. Consequently the models/methodologies were prioritised to review those most commonly used. These are shown in the table below:

MODELS AND METHODOLOGIES (RISK ASSESSMENT TOOLS)
STREAM
TOXIC RISKAT – including assessments of DENZ and CRUNCH (and as applied to chlorine, ammonia and hydrogen fluoride
PASKER
MEM
FLAMCALC – as applied to LPG
GASP
DRIFT
FIREPEST
YELLEVAP
SANDRA
RASP (as applied to sulphur trioxide)
TOEM (as applied to 20% oleum)
REACTPOOL
PRAM PSR
PIPERS
PIPIN
MISHAP
TRAM
POOLFIRE 6
LPG RISKAT – as applied to LPG

Results

10. There were four elements to the results of this review;

- Model assessments
- Methodology assessments
- Topic Specialist interviews
- HSL independent review

11. The results of these elements were analysed in a thorough multistage approach, resulting in the selection of the models and methodologies considered most appropriate for development or replacement. These are shown in the table below:

MODELS AND METHODOLOGIES (RISK ASSESSMENT TOOLS)
REACTPOOL
SANDRA
TOXIC RISKAT
MISHAP
DRIFT
PASKER

Note:

- Although no methodologies, as categorised in the assessment, have been individually marked out for development, it is assumed that the model development process will drive that of the methodologies. Any methodology involving a model chosen for development will itself have to be at least revised.
- The MSDU documentation of methodologies, including PCAG, is in general need of updating – this process is at its early stages and is supported by the P5 findings

Conclusions on Assessment Process

12. Overall, it is concluded that the recommendations for development are based on a rigorous analysis of reliable data obtained from the P5 programme, and that those recommendations are justified.
13. The reliability of the assessment procedure used to obtain the data, and therefore the data generated by it, is fully supported by the conclusions of the independent HSL comparative study.

Main Report

Introduction

14. The primary purpose of the IFRLUP project is to implement those recommendations relevant to the Hazardous Installations Directorate (HID), of the Safety Policy Directorate (SPD) led Fundamental Review of Land-Use Planning.

15. A key recommendation of the review was:

"The criteria and methodology used for setting planning zones and for "calling in" planning applications should be reviewed and, if necessary, revised and then published."

16. This product (P5) addresses the methodology review element by way of a prioritised review of the key models and methodologies used in land-use planning (LUP).

17. The aim of P5 was to carry out an appropriate and proportionate review of the models and methodologies (M&Ms) used by HSE's Methodology and Standards Development Unit (MSDU) in setting CDs in order to:

    - Obtain a robust picture of the relative strengths and weaknesses of the M&Ms
    - Enable prioritisation of the M&Ms for development. Note: product P6 will deal with the actual development of the M&Ms on a needs basis.
    - Provide evidence-based data to inform the debate with stakeholders on CD setting.

Developing the Assessment Methodology

18. In order to meet the needs of P5 and ensure the wide range of M&Ms used by MSDU were assessed in a fair, robust and comparable way an assessment methodology was required to be developed.

19. The methodology was developed by a multidisciplinary team which included representatives from:

    - The IFRLUP project team
    - MSDU
    - WS Atkins –(Independent external expert)
    - HSL
    - HID OPU

20. The team took an iterative approach to developing the methodology, keeping the IFRLUP Project Management Board informed and taking direction from them as the work progressed. From this process they identified the following key issues that should form the basis of the assessment process.

    - How well does the model/methodology meet HSE's risk assessment needs? - Fitness for purpose

- Is the model/methodology likely to be challenged? e.g. in a planning appeal - Risk to HSE
- Are there any external drivers likely to require changes? e.g. from Europe - Wider influences
- How much need is there for this model/methodology by HSE? - Extent of use
- How possible/easy is it for development of the model/methodology to be carried out? - Suitability for development
- Where site-specific data (snapshot information) is used how valid is this data?

21. The review had to ensure that it gathered information on the models and methodologies from relevant specialists and this was validated independently to ensure the review was robust.

The Methodology

22. The methodology developed was a judgmental process relying on a common team of people (the assessment team + relevant Topic Specialist) carrying out the assessments to achieve a suitable degree of consistency.

23. The assessment team consisted of representatives from MSDU, IFRLUP team, WS Atkins (independent contractor input) and industry (independent challenge function). The team discussed each model/methodology against a set of agreed criteria (3 key criteria + supplementary criteria).

- Fitness for purpose
- Risk to HSE
- Wider influences

} Key criteria

- Extent of use
- Likelihood of success (for development)
- Snapshot information (only relevant for methodologies)

} Supplementary criteria

24. Each criterion was assessed against a number of sub criteria in the form of statements, which were scored on a 4-point scale (0=strongly disagree, 1 = disagree, 2 = agree, 3 = strongly agree). The scoring was devised to assist the later analysis stage. (See Annex 1 for the detailed criteria and statements used)

25. To validate/verify the findings of the assessment team an independent specialist team from HSL assessed a small selection of models and methodologies and compared results with those of the P5 team to ensure the assessments had been thorough, complete and consistent.

26. The MSDU Planning Case Assessment Guide (PCAG) frequently refers to the use of Topic Specialists, indicating there are a significant number of uncodified methodologies with heavy reliance on individual specialists. Topic Specialists were interviewed independently by HSL to identify where further development work may be required and, in particular, where codification of advice could be carried out. The question framework used for these interviews is available at Annex 2

Methodology Pilot

27. The assessment methodology was piloted to determine how comprehensive and workable it was, and how fair and robust a view of the models or methodologies it generated.

28. The Fitness-for-purpose (FFP) Criterion was chosen for the pilot as it involved only people with a detailed knowledge and experience of the processes being assessed. This thus maximised the validity of the results and conclusions of the pilot. The independent HSL validation process was also included in the pilot so as to ensure that a clear unbiased view of the technical issues was obtained.

Pilot approach

29. For the pilot a model and a methodology that are in regular use and fairly representative of the range of models and methodologies available were selected for assessment by the FFP team and HSL. The poolfire model and poolfire methodology were selected

30. The assessment team and HSL carried out their assessments independently, followed by a joint review of the process, criteria used and the outcome to identify any areas of concern and consider any required changes.

Findings of pilot

31. The pilot demonstrated that the approach taken to assess the FFP criteria was comprehensive, workable, and fundamentally sound. Fine-tuning of the process, summarised below was recommended and carried out.

Summary of changes

- Assessment of sub criteria (a) “Is the tool scientifically credible?” needs to be recorded in sufficient detail to enable HSL to validate on the same basis.
- Sub criteria (d), second part, “Do the outputs bear any resemblance to observations?” needs to be clarified to ensure any available evidence of validation of model predictions against real life incidents is captured.
- Information regarding the “authority” a model/methodology has (e.g. how a model was developed, what expertise it is based on, any peer review process etc) should be recorded as a new sub criteria under key criteria 2 “risk to HSE”
- To aid consistency of scoring sub criteria develop descriptors for the 4-point scoring scale.
- For security of data and ease of use an access database was developed to replace the pilot excel spreadsheet.

32. In support of the FFP assessment the Topic Specialist interview process was also tested with interviews of the poolfire and bund/tank overtopping specialists.

Prioritising Models and Methodologies for Assessment

33. With approximately 80 models and over 20 methodologies in the MSDU LUP portfolio it was impractical to assess all of them. Consequently the models/methodologies were prioritised so that those most commonly used were assessed.

34. Earlier work had been done by MSDU where the sites subject to most planning applications had been identified. This was used as the basis for prioritisation identifying which models were relevant to the top 40 sites. The prioritisation was carried out by the following process.

    a. For each of the substances identified for the top 40 sites the models used for them were identified, they were then ranked by the number of times the model occurred in the top 40.

    b. Each MSDU methodology was reviewed to identify how many times individual models were used within a methodology and ranked by number of occurrences.

    c. The two ranked lists from a. and b. were then compared and models ranked on the basis of occurrence in both lists and the scores within both lists.

    d. A reality check of the ranked list from c was undertaken with MSDU. This aimed to identify:

        - Key models that should be a priority
        - Models that performed utility functions (e.g. routine plotting operations) only and could be excluded from the assessment process at this stage.
        - Models that were no longer used.

35. The M&Ms selected for assessment are listed below

MODELS	METHODOLOGIES
STREAM	**Chlorine**
TOXIC RISKAT – including assessments of DENZ and CRUNCH	**Toxic gas/vapour release this to be assessed for two substances -Ammonia and Hydrogen fluoride**
PASKER	**LPG**
MEM	**Sulphur trioxide/Oleum**
FLAMCALC	**TRAM**
GASP	
DRIFT	
FIREPEST	
YELLEVAP	
SANDRA	
RASP	
TOEM	
REACTPOOL	
PRAM PSR	
PIPERS	
PIPIN	
MISHAP	
POOLFIRE6	

The Assessment Process

36. The assessment process was carried out in a series of 8 meetings between 10-2-04 and 5-5-04. During these meetings the assessment team and relevant Topic Specialist, for the Model or Methodology being reviewed, systematically worked through each element of the criteria. The team agreed the comments and score for each element and recorded them directly into the P5 database. Following this process ensured that a detailed view on each M&M being assessed was agreed and recorded in real time, eliminating potential transcription errors that could have resulted if documenting was done after the meetings.

37. The Topic Specialist interviews were carried out by HSL roughly in parallel with the M&M assessments. HSL recorded the outcomes of these interviews in a proforma (see Annex 2) and these were input to the P5 database once all interviews had been completed.

38. The final stage of the process was the HSL M&M validation review. This was carried our by an HSL team once the assessment reviews had been completed.

Results

There are four elements to the results of this review;

- Model assessments
- Methodology assessments
- Topic Specialist interviews
- HSL independent review

39. An analysis of these results is provided below. The detailed results of the M&M assessments and Topic Specialist interviews run to tens of pages and aren't included in this document but are available from the IFRLUP team. A copy of the HSL assessment validation report is at Annex 4.

Analysis of Results

40. This section aims to present a brief description of the data analysis process and the recommendations stemming from it. A detailed and thorough presentation is available at Annex 3.

Overall Aims of the Review Process

41. The purpose of this review was to prioritise the models and methodologies chosen for assessment in terms of their development needs. This prioritisation has been undertaken by analysis of the quantitative and qualitative data obtained by the assessment processes detailed previously.

42. It should be noted that it was neither the aim nor the responsibility of this review to determine absolutely the models and methodologies which would be developed. Ultimately, development, and the allocation of funding for such, will be made by others. These decisions will be influenced by the availability of finances and other strategic priorities and considerations, both internal and external to MSDU.

43. It should be borne in mind that the models and methodologies recommended for development from this review have, by necessity, been chosen from a pre-selected subset of all of MSDU's LUP models and methodologies currently in use. Whilst this subset was carefully chosen there may, or may not, be models and methodologies outside of it that would have also been considered appropriate for development had they been included.

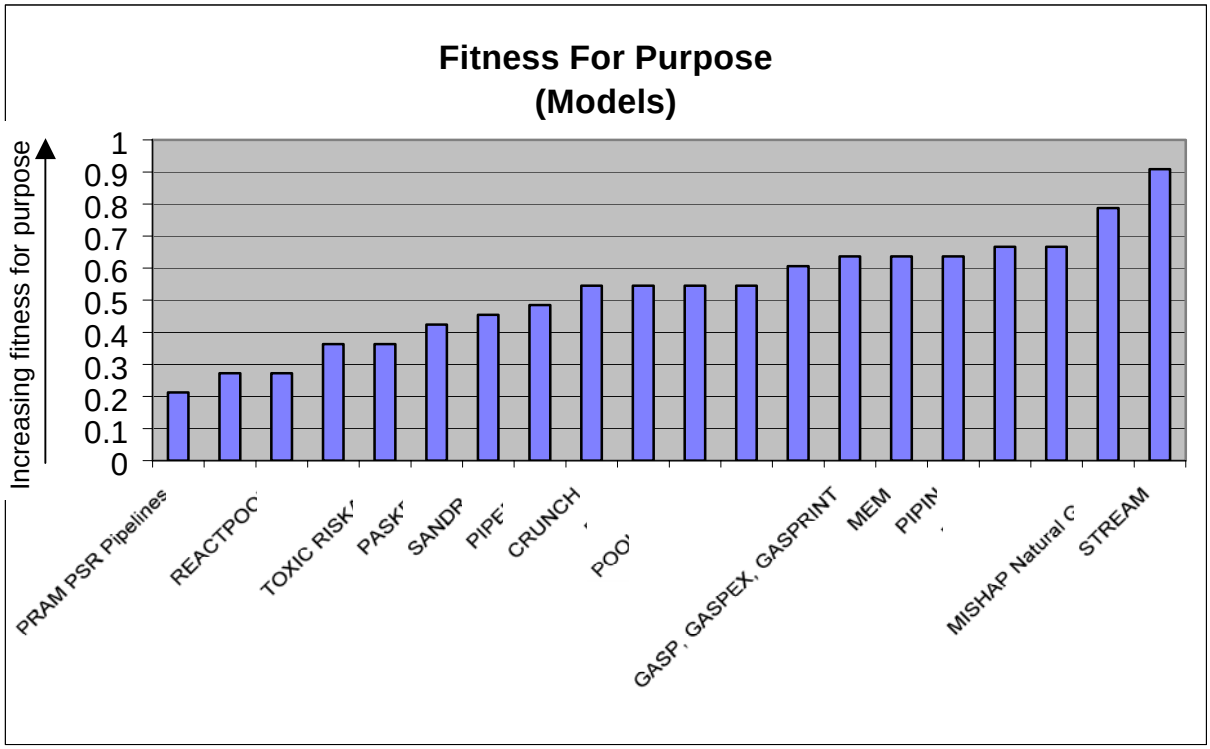
44. The analysis and conclusions contained in this document (including Annex 3) have been subject to peer review within HSE. This review included Topic Specialists and other members of MSDU.

Analysis Structure

45. The analyses of the model and methodology data were performed in several stages, leading finally to the selection of models and methodologies considered most appropriate for development (or replacement if necessary). Briefly those stages were:

- Preliminary Analysis – the assessment scores and qualitative responses were analysed in detail. This was done to serve as quality assurance, to ensure that the score attributed to a given qualitative response was consistent with that response, and to inform later analysis stages. Where necessary scores were revised, in consultation with appropriate parties.

- Criterion Ranking - each Criterion was individually analysed, and the models ranked within it to form a ranking profile. For this purpose the scores alone were utilised. An example of a ranking profile thus generated is shown below:



- Categorisation – the Risk and FFP Criteria ranking profiles obtained were used to assign the models and methodologies to a number of assessment performance categories. The categorisation system employed was similar, but not identical, for the models and methodologies. Both assigned a two-letter combination to indicate the combined risk-FFP performance, ranging from AA for low risk, high FFP through to EE (models) or CC (methodologies) for high risk, low FFP.

- Initial Selection – a number of models were selected based on their Risk-FFP categorisation. This selection was then cross-referenced with the ranking profile obtained for the Wider Influences Criterion. From this process an initial selection of models was determined for individual assessment. Qualitative responses were also considered at this point to supplement the selection process, leading to consideration of some models not automatically selected as a result of categorisation.

- Individual Analysis – the models chosen in the initial selection were subjected to a more detailed analysis. Both the Criterion rankings and the information obtained from the preliminary analysis were used to inform the overall assessment. Inter-dependencies were also assessed and used to determine the final selection.

- Final Selection

46. Throughout each stage, where appropriate, further analysis of the sub-criteria was undertaken and supplementary evidence or clarification sought from Topic Specialists and other relevant parties.

47. The Topic Specialist Interview data underwent the same process of Preliminary Analysis, although the main analysis was conducted differently. This was primarily due to the fact that the assessment utilised a unique set of questions rather than the statement set used for the general models and methodologies assessment. A similar approach to scoring the responses was however used.

48. The stages of the Topic Specialist data analysis were:

    - Preliminary Analysis – as with the models and methodologies the data from the interviews was checked, cross-referenced and quality assured.

    - Main Analysis – the information relating to models and methodologies was separated out and the scores analysed in order to generate ranking profiles. A ranking profile was generated for both the models and the methodologies involved in the Topic Specialist Interviews.

    - Comparative Analysis – the ranking profiles obtained from the Topic Specialist Interview data was compared with its general assessment counterpart.

    - Conclusions

Development Recommendations

49. From the analysis process the following conclusions were drawn:

Models

50. It is recommended, subject to the caveats presented in the detailed consideration in Annex 3, that the following models or model replacements are suitable for development consideration. No importance should be attached to the order in which the list is presented.

    - REACTPOOL
    - SANDRA replacement process
    - TOXIC RISKAT or replacement process
    - MISHAP

51. Dependent upon the outcome of other development processes detailed above or already in progress, limited development should be considered for:

    - DRIFT as part of TOXIC RISKAT
    - PASKER replacement process (VIEWRISK)

52. This is covered in greater detail in Annex 3.

Methodologies

53. It is recommended, for the reasons presented in Annex 3, that consideration should be given to reviewing, and revising where appropriate, the documentation of all methodologies.

54. It is also recommended that the model development priorities be used to determine the methodology development requirements. Whilst TRAM and AMMONIA performed worst in the assessment they are not specifically recommended for development. The reasons for this are presented in the detailed analysis in Annex 3.

55. LPG-RISKAT must be considered to be strongly recommended if HSE decides that further detailed risk assessments are required in this area.

56. Other methodologies would be recommended, but they will automatically have to be reviewed as a result of the models employed within them and their recommendation for development. For example the SULPHUR TRIOXIDE/OLEUM methodology would be revised in the light of the suggested development of REACTPOOL. Whilst not singly identified in the methodology assessment, it did perform very poorly in the FFP Criterion

Topic Specialist Interviews

57. The data obtained from these assessments was considered to be best employed as a source of supplementary information to inform the individual model and methodology assessments.

58. Where it was appropriate to use the rankings generated by this analysis they were found to further support the selection of models and methodologies detailed previously.

Assessment Conclusions

Models

59. Very good agreement was found between the Risk/Fit For Purpose/Wider Influences Criteria rankings and the proposed improvement/replacement strategies currently in progress within HSE. All models intended for replacement in the near term fell into the high risk, low FFP (AA) category, as did their replacements if already in use. In the former case this can be explained by their being ‘life expired’ and in the latter due to them requiring further refinement to be fully effective.

60. TOXIC RISKAT was the only highlighted ‘model’ which does not have a successor determined. This is principally down to the effort and complexity necessary in its replacement. It should also be noted that, whilst TOXIC RISKAT has been evaluated as a model, it is more of a process involving a sequence of models such as CRUNCH and DENZ and not dissimilar to LPG-RISKAT (assessed as a methodology.)

Methodologies

61. It is believed that the results of the model assessment will primarily determine the methodologies requiring development.

62. Any methodology involving a model considered appropriate for development would itself have to be at least reviewed.

63. Not surprisingly the outcomes of the two assessments are interlinked, with all of the worst performing methodologies being connected to models deemed worthy of development or replacement.

64. It is also accepted that many methodologies as detailed in MSDU documentation, including PCAG, require reviewing and updating. This generic development process has already been started and it is recommended by this review that it is continued and supported.

Overall Conclusions on Assessment and Analysis Process

65. A number of areas for improvement in the assessment process were identified, along with possible minor biasing effects, and these are presented in Annex 3.

66. Overall, it is concluded that the recommendations for development are based on a rigorous analysis of reliable data obtained from the P5 programme, and that those recommendations are justified.

67. The reliability of the assessment procedure used to obtain the data, and therefore the data generated by it, is fully supported by the conclusions of the independent HSL comparative study and associated report (Annex 4).

Annex 1 – Assessment Criteria

The criteria statements were scored on a 4-point scale:

- 0=strongly disagree,
- 1 = disagree,
- 2 = agree,
- 3 = strongly agree

Key Criteria

Fitness For (Current) Purpose

a. The tool is scientifically credible

b. The tool has no recognised significant weaknesses

c. The tool gives broadly acceptable results/performs as expected

d. There is confidence in the tool when correctly applied. Consider: Do the outputs bear resemblance to observations

e. Assumptions are considered sensible/realistic

f. Sensitivity - There are no inputs that significantly/disproportionately affect the outputs

g. There are no existing plans to revise/replace this tool
Consider: The grounds for replacement

h. The user is rarely directed to consult with a Topic Specialist (indicative of shortcomings in the method/model). Consider: If this is purely a shortcoming of the model/method or a shortcoming of the software failing to identify that reference should be made

i. The tool is quality assured for use

j. The tool has been validated against “real events”. Consider: If it has been possible to compare model predictions with accident experience and whether this gives any confidence that the model predictions resembled reality?

k. The model has been thoroughly validated in experiments.
Consider: What size were the experiments in relation to actual loss of control accidents?

Risk To HSE

a. There is no controversy associated with this tool. Consider:

    i. Has it been used where there is disagreement between LPAs and HSE?

    ii. Has it been challenged at an external enquiry or call-in? (and if so, what was the outcome?)

b. The tool has been or could be used to assist judgements where the approximate risk integral (ARI) indicates societal risk of concern.

c. The tool has been or could be used in assessments for ‘sensitive’ hazardous installations – e.g. where there is a significant build up of development including sensitive developments

d. The tool has been quality assured for development

e. The tool has been rigorously peer reviewed

f. Development to address any fitness for purpose shortcomings identified would not impact significantly on the size of CD’s or otherwise have the potential to affect HSE credibility. Consider: Would we get results that are inconsistent with HSE policy?

Wider Influences

a. There are no drivers for change. Consider:

   - Any recent challenges
   - Developing thinking
   - Possible moves to harmonise UK and European methodologies /assumptions
   - TWG5, in particular objective 2 (database of models etc.)

b. The tool supports HSE’s credibility and is sufficiently leading edge (Note: We need to be informed by any existing justifications for HSE to be leading edge – only where an identified need exists should we consider, if it is not already, whether a tool has the potential to be leading edge. P5 is not required to make a case for leading edge tools)

c. The tool can cope with constraints on maximum ranges to accommodate political/policy/pragmatic restraints. Consider: Are outputs flexible enough to enable limits to be imposed?

d. There is a commercially available package that could replace this method/model

<u>Supplementary Criteria</u>

Additional criteria will not feature in the overall priority rating score; but will provide the P6 team with useful additional information when allocating resources to deliver improvements. They cover issues such as extent of use of the tool, is the technology available to deliver any required improvements to the tool.

Extent of use

a. The tool is widely applicable

b. The tool is used often

c. The tool was used recently

d. The tool is likely to be used again in the near future

Likelihood of success

a. The technology is available to deliver the desired improvement/s

b. A minimal amount of effort is required for any proposed improvements

c. Proposed developments would have a minimal impact on the extent of CDs, leading to minimal costs (e.g. minimal working of 3 zone maps)

Snapshot information (Methodologies only)

a. There is clear guidance to ensure snapshot info is accurate (a one-off consideration)

b. The tool is not sensitive to snapshot information

c. The snapshot information for this tool is current

Annex 2 – Topic Specialist Interview Question Set

Topic Specialist: **Methodology:** **Interviewees:**

Topic: **Model:**

Interview date:

Question	**Comments**	**Driver for Codification/ Development/ analysis** **0 – Strong** **1 – Moderate** **2 - Weak** **3 – None**
1) How often is the Topic Specialist consulted?		
2) How variable is the advice given?		
3) How often have issues relating to the advice given by the Topic Specialist been taken to panel?		
4) How sensitive are the assessment outputs to the advice given?		
5) Have there been any developments in the field since the code/methodology was last reviewed?		
6) Have there been any appeals against the assessments that resulted from the advice of the Topic Specialist?		
7) How sensitive is the advice to the questions asked (or information provided)?		
8) Are the questions asked of the Topic Specialist reasonable?		
9) Are there any areas where questions are often asked where there are no easy answers?		
10) Are the reasons for referrals to the specialist due to failings of the model or the fact that the software is good at flagging where referrals are required?		
Other comments:		

Annex 3 – Model and Methodology Data Analysis

Overall Aims of the Review Process

A3-1. The purpose of this review was to prioritise the models and methodologies chosen for assessment in terms of their development needs. This prioritisation has been undertaken by analysis of the quantitative and qualitative data obtained by the processes detailed in the main body of this paper.

A3-2. It should be noted that it was neither the aim nor the responsibility of this review to determine absolutely the models and methodologies which would be developed. Ultimately, development decisions, and the allocation of funding for such, will be made by others. These decisions will be influenced by the availability of finances and other strategic priorities and considerations, both internal and external to HSE's Methodology and Standards Development Unit (MSDU).

A3-3. It should be borne in mind that the models and methodologies recommended for development from this review have, by necessity, been chosen from a pre-selected subset of all of MSDU's land-use planning (LUP) models and methodologies currently in use. Whilst this subset was carefully chosen there may, or may not, be models and methodologies outside of it that would have been considered appropriate for development had they been included.

A3-4. The analysis and conclusions contained in this document have been subject to peer review within HSE, including Topic Specialists and other members of MSDU.

Analysis of Models

Analysis Structure

A3-5. The analysis of the data was performed in several stages, leading finally to the selection of a number of models considered most appropriate for development (or replacement if necessary). Briefly, the stages were:

- Preliminary Analysis – the assessment data was analysed in detail to ensure that the scores used in the main analysis were correct and appropriate. This analysis also provided detailed information to qualitatively inform the decisions of later stages.
- Criterion Ranking - each Criterion was individually analysed, and the models ranked within it. For this purpose the scores alone were utilised.
- Categorisation – the Risk and FFP Criteria ranking profiles obtained were used to assign the models to a number of assessment performance categories.
- Initial Selection – a number of models were selected based on their Risk-FFP categorisation. This selection was cross-referenced with the ranking profile obtained for the Wider Influences Criterion and an initial selection of models undertaken for individual assessment. Qualitative responses were also considered at this point to supplement the

selection process, leading to consideration of some models not automatically selected as a result of categorisation.

- Individual Analysis – the models chosen in the initial selection were subjected to a more detailed analysis. The various Criterion rankings and a detailed analysis of the sub-criteria qualitative responses, gained from the preliminary analysis, were used to inform the overall assessment. Inter-dependencies were also assessed and used to determine the final selection.
- Final Selection

A3-6. Throughout each stage, where appropriate, further analysis of the sub-criteria was undertaken and supplementary evidence or clarification sought from Topic Specialists and other relevant parties.

Preliminary Analysis

A3-7. Prior to undertaking the actual ranking analysis the scores and qualitative responses obtained during the assessment process were analysed in detail. This was done to serve as quality assurance, to ensure that the score attributed to a given qualitative response was consistent with that response, and to inform later analysis processes.

A3-8. This preliminary analysis showed that some apparently inappropriate score/response combinations were present in the data set. The scores were therefore revised in line with the qualitative response given, subject to further information or clarification being sought in order to confirm and justify that revision. Examples of this include questions where a “disagree” qualitative response has been allocated an “agree” score.

A3-9. It was also found that, in certain circumstances, a default score of 0 had been attributed to ‘not applicable’ responses. In these cases the problem was addressed by a ‘no score’ being fed into the assessment process, and this being taken into account in the overall analysis.

A3-10. Questions were found where the response indicated that further information was pending and an ‘interim’ score had been attributed. After consultation with the appropriate respondent, and others where appropriate, additional information was obtained allowing a final score to be determined.

A3-11. The number of sub-criteria scores revised in these ways was small. Throughout this review the use of the word ‘scores’ should be taken to mean those values determined after the preliminary analysis, therefore including revised scores where appropriate.

A3-12. As part of this preliminary analysis it was also decided to omit certain questions, and their response scores, from the direct assessment process. The qualitative responses to these questions were used as supplementary evidence to support aspects of the model under analysis and to inform its overall assessment.

The questions omitted from the model analysis were:

- Criterion 2, sub-criterion 2b
- Criterion 2, sub-criterion 2c
- Criterion 3, sub-criterion 3d

A3-13. Criterion 2 sub-criteria 2b and 2c were omitted as they were not applicable or relevant to all models and their inclusion would therefore have biased the assessment.

A3-14. Criterion3 sub-criterion 3d related to the availability, or otherwise, of a replacement commercial software package. It was felt that this was irrelevant in terms of determining the performance of the currently employed HSE model. Should a particular model be deemed in need of development or replacement as a result of this assessment then the responses to this sub-criterion would be taken into consideration in determining the best way forward.

Criterion Ranking

A3-15. Within each Criterion the sub-criteria scores were combined, where appropriate, for each model under assessment. The combined score was then normalised relative to the maximum possible score for that model, resulting in a performance rating from 0 (poor) to 1 (excellent). This process thus generated a ranking profile for every Criterion. These are shown in Figures 1 to 5.

Categorisation

A3-16. As detailed earlier, the assessment for development priority was primarily based on the Fit For Purpose (FFP), Risk and Wider Influence Criteria ranking profiles. The Likelihood of Success (Figure 4) and Frequency of Use (Figure 5) Criterion profiles were not used explicitly in this process.

A3-17. The Likelihood of Success Criterion was omitted on the basis that the purpose of the review was solely to determine the need for development/replacement of models and not whether such a process was prohibitively costly. This would be dealt with at a later stage, where the information gathered here would be utilised. Where relevant the qualitative responses from this Criterion were used to inform and support decisions based on responses in the other Criteria.

A3-18. The Frequency of Use Criterion was not considered as it was artificially biased. One of the factors determining the inclusion of a model in the HSE review process was that it had to be frequently used (with one exception - PRAM PSR). Thus all models, except PRAM PSR, would belong to the high usage category. Such information will, however, be taken into account in the next stage of the project which will determine the models and methodologies to be developed or replaced, based on consideration of a range of issues such as funding availability, policy issues etc.

A3-19. Analysis of the ranking profile for the Risk and FFP Criteria allowed the models to be categorised in those terms, resulting in models being assigned from high risk, low FFP through to low risk, high FFP.

A3-20. The correlation between these two Criteria proved to be high, with no models being determined as opposites (low risk, low FFP and vice versa). Each model was then assigned to a defined risk-FFP category. These categories were designated by a two letter combination, representing the relative standing in the Risk and FFP Criterion rankings respectively. The

categories ranged from AA (very high risk, very low FFP) through to EE (very low risk, very high FFP). At this stage no distinction was drawn between risk and FFP category assignment i.e. BC was considered to be the same as CB

A3-21. The full list of models under consideration and their deemed category is presented in Table 1.

Worst ⟶ Best

AA	AB/BA	BB	BC/CB	CC	CD/DC	CE & DD	DE/ED	EE
PRAM PSR RASP REACTPOOL	none	PASKER SANDRA PIPERS	TOEM TOXIC RISKAT	CRUNCH DENZ	PIPIN YELLEVAP	POOLFIRE6 (CE) FIREPEST (DD) MEM (DD)	DRIFT GASP etc	FLAMCALC MISHAP STREAM

Table 1 – Categorisation of Models

Initial Selection

A3-22. From a study of the distribution within the categories a number of the models were identified as being candidates for further detailed assessment. These were:

- PRAM PSR (AA)
- RASP (AA)
- REACTPOOL (AA)
- PASKER (BB)
- SANDRA (BB)
- PIPERS (BB)
- TOEM (CB)
- TOXIC RISKAT (CB)

A3-23. Similar analysis on the Wider Influences ranking profile showed that the same eight models occupied the lowest, worst, eight places in that Criterion. Thus the choice of models detailed above was further supported, and deemed conclusive.

A3-24. As a result of qualitative responses CRUNCH (CC) and DENZ (CC) were also considered. Inter-dependencies led to consideration being given to MISHAP (EE), PIPIN (CD) and DRIFT (DE).

Individual Analysis

A3-25. Further detailed analysis of the eight models determined by the ranking profile assessment highlighted several inter-dependencies amongst them. The detailed analysis is presented below, including the other models where those dependencies exist.

REACTPOOL

A3-26. Whilst this was itself classed as a very high risk, very low fit for purpose model (AA) it was found to be intended as a replacement for RASP and TOEM, subject to further development. As both RASP and TOEM are

included in the eight models for further consideration, (AA) and (CB) respectively, it can be seen that suitable development of this model would be advisable. The development of REACTPOOL would strengthen its own position and remove the two other weak models from service.

A3-27. It should be noted that REACTPOOL's low performance in the assessment is primarily due to it being a replacement model and not having been fully coded up or validated at this time.

SANDRA

A3-28. This was deemed to be a high risk, low fit-for-purpose model (BB). It is understood that a replacement for SANDRA is already 80% complete and, therefore, support is recommended for completion of this work. This should result in a replacement model that has high fitness-for-purpose and poses low risk to HSE.

TOXIC RISKAT

A3-29. This was categorised as a low fit-for-purpose model posing medium risk to HSE. As it is a fundamental part of HSE's LUP etc risk assessment process it is suggested that funding should be found for improvement or, more realistically, replacement. It should be noted that TOXIC RISKAT is not an individual model, despite being assessed as one, but is a process involving a number of models. It is the inclusion of some of these models, CRUNCH and DENZ, which leads to its poor overall performance. These two models were independently assessed as part of this review and were borderline in relation to inclusion in the development priority list. They are no longer utilised outside the TOXIC RISKAT system and their performance in the assessment was artificially increased by this fact, e.g. they were scored as low risk by virtue of not being used independently. A more thorough review of their scores would have led to explicit inclusion in the list of models for consideration, but this was deemed unnecessary due to the fact that they were only employed as an integral part of TOXIC RISKAT which was to be included.

A3-30. In the independent HSL Review, discrepancies found between the HSL and HSE assessment of TOXIC RISKAT were solely down to the interpretation of how poor the CRUNCH and DENZ models were, and how this would impact on the performance of TOXIC RISKAT. HSL considered the two models to pose a much higher risk and be considerably less fit-for-purpose and scientifically credible than did the HSE assessment. This may have arisen partly due to the fact that some HSE scores reflected the situation that CRUNCH and DENZ were no longer used as stand-alone models and had been replaced for such uses by DRIFT. As already explained such scores were not dealt with under the preliminary analysis and revision process as their only usage was within TOXIC RISKAT, which was being assessed itself. It is debatable, given this knowledge, whether CRUNCH and DENZ should have been included in the assessment in their own rights at all.

DRIFT

A3-31. The recommendation for DRIFT is principally based on it being the replacement stand-alone model for DENZ and CRUNCH, integral and outdated components of TOXIC RISKAT. The assumption is made that it

would be used in their stead in any revised/replacement TOXIC RISKAT system. Should an entirely new system be developed that did not incorporate DRIFT then its development would be less of a priority.

PASKER

A3-32. This was deemed as a high risk, low fit-for-purpose model (BB). The intended replacement of this model by the introduction of VIEWRISK should remove it from posing any risk to HSE. Some development effort, however, is likely to be required to ensure that the replacement process is fully effective. This replacement development is supported.

MISHAP

A3-33. This model was one of the best performing of those assessed, being deemed to be of very low risk and very high fit-for-purpose (EE). It is included as being appropriate for development due to the fact that it is considered possible to develop this model and use it as the replacement for the very high risk, very low FFP (AA) PRAM PSR model. The PRAM PSR model has been entirely discredited and is not even able to operate on current HSE IT technology.

PIPERS

A3-34. This model was categorised as high risk and low FFP (BB) by the assessment and analysis procedures. The recommendation for development of this model was based ostensibly on the fact that it would be used to determine the CDs of pipelines that are not currently classed as Major Hazards. Its performance would therefore be under close scrutiny and its risk to HSE significant. Whether the model will be used to calculate currently undetermined CDs depends entirely on the introduction of new legislation, and the benefits and timing of any development is dependent upon this introduction. As the legislation to bring these pipelines into the Major Hazards classification is currently ‘on hold’, there is no justification for further development at this time. It is important, however, that this potential need should not be forgotten.

A3-35. There is also a desire to introduce a ‘new’ PIPELINE RISKAT that would incorporate MISHAP along with PIPIN. This would further strengthen the case for looking at MISHAP whilst also possibly making minor cost-effective changes to PIPIN. It should be noted that PIPIN was assessed to be of average risk to HSE, and of high FFP (CD), so would not be recommended purely on its own merits. The possible future need to incorporate PIPERS into this RISKAT should also not be overlooked in its development.

Development Recommendations

A3-36. It is recommended, subject to the caveats presented in the previous section, that the following models or model replacements are suitable for development consideration. No importance should be attached to the order in which the list is presented.

- REACTPOOL
- SANDRA replacement process
- TOXIC RISKAT or replacement process

- MISHAP

A3-37. Limited development support, subject to the caveats already presented, should be considered for:

- DRIFT as part of TOXIC RISKAT
- PASKER replacement programme (VIEWRISK)

Model Analysis Conclusions

A3-38. Very good agreement was found between the Risk/Fit For Purpose/Wider Influences Criteria rankings and the proposed improvement/replacement strategies currently in train within HSE. All models intended for replacement in the near term fell into the high risk, low FFP (AA) category, as did their replacements if already in use. In the former case this can be explained by their being ‘life expired’ and in the latter due to them requiring further refinement to be fully effective.

A3-39. TOXIC RISKAT was the only highlighted ‘model’ which does not have a successor determined. This is principally down to the effort and complexity necessary in its replacement. It should also be noted that, whilst TOXIC RISKAT has been evaluated as a model, it is more of a process involving a sequence of models such as CRUNCH and DENZ and not dissimilar to LPG-RISKAT, assessed as a methodology.

A3-40. For the reasons outlined in this review, it is concluded that the recommendations for consideration for development are based on reliable data obtained from the P5 programme. Those models recommended, in no particular order, are:

- REACTPOOL,
- MISHAP,
- TOXIC RISKAT or replacement process,
- SANDRA replacement process,
- PASKER replacement process,
- DRIFT

Figure 1 - Fitness For Purpose (Models)

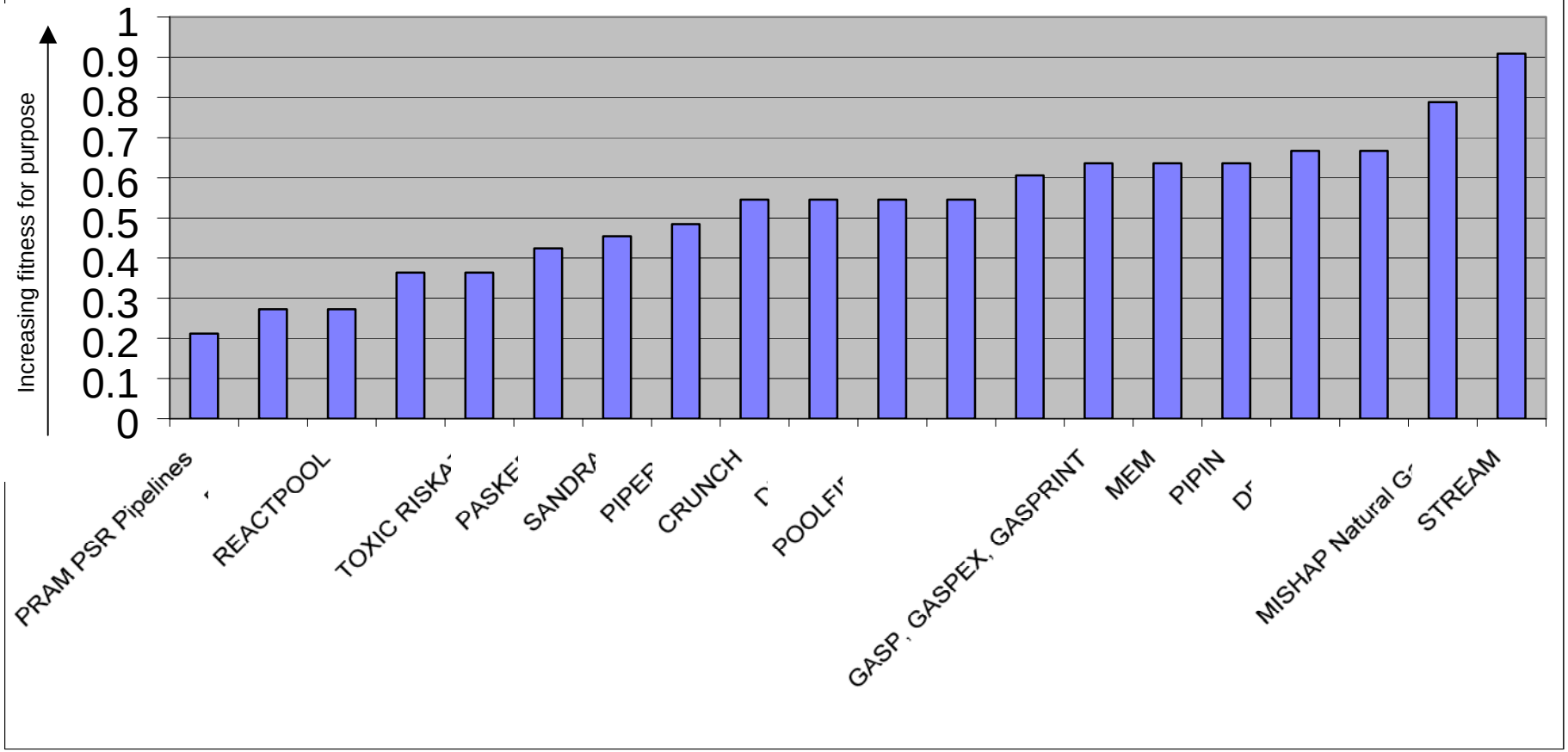


Figure 2 - Risk to HSE (Models)

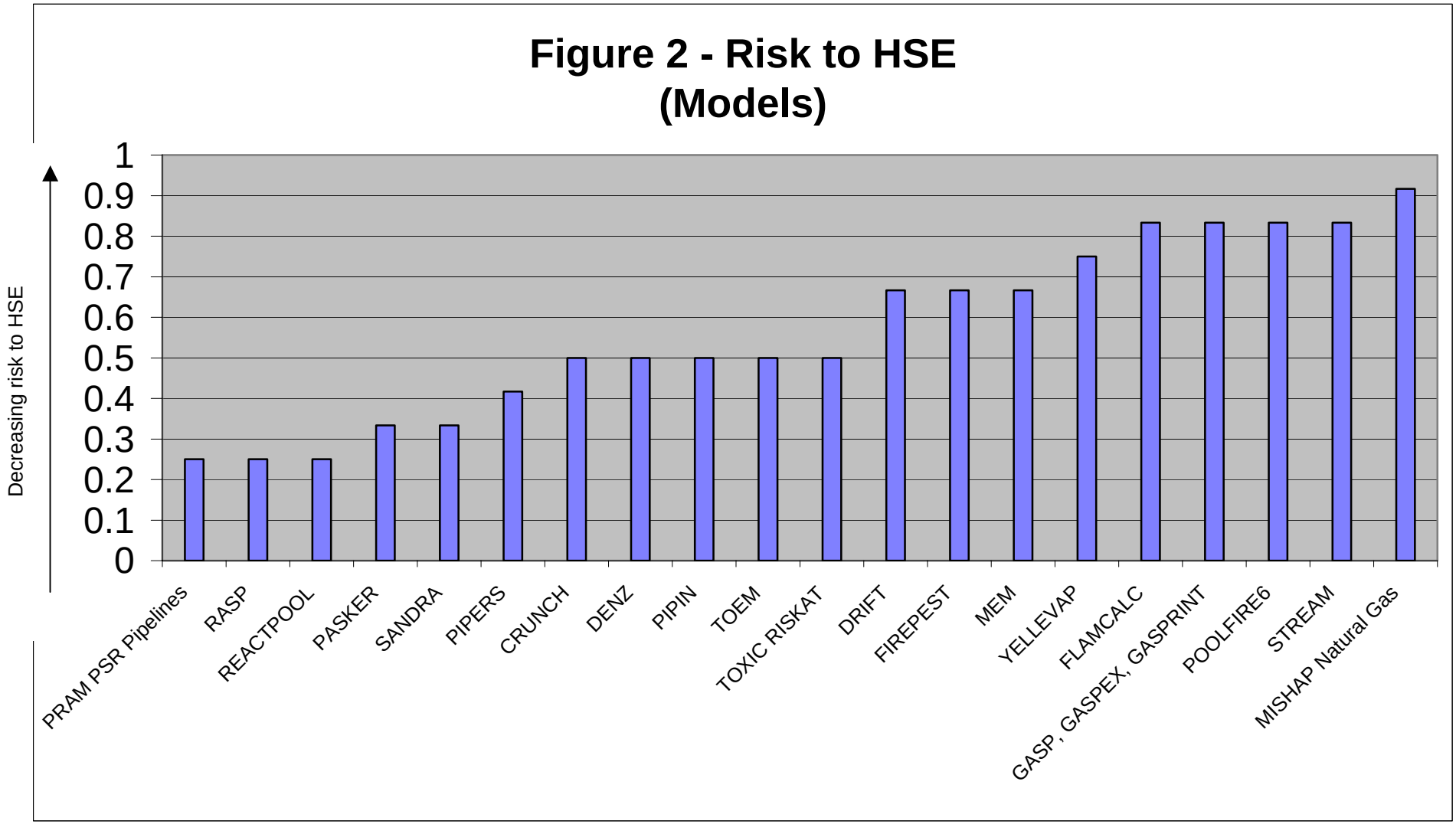


Figure 3 - Wider Influences (Models)

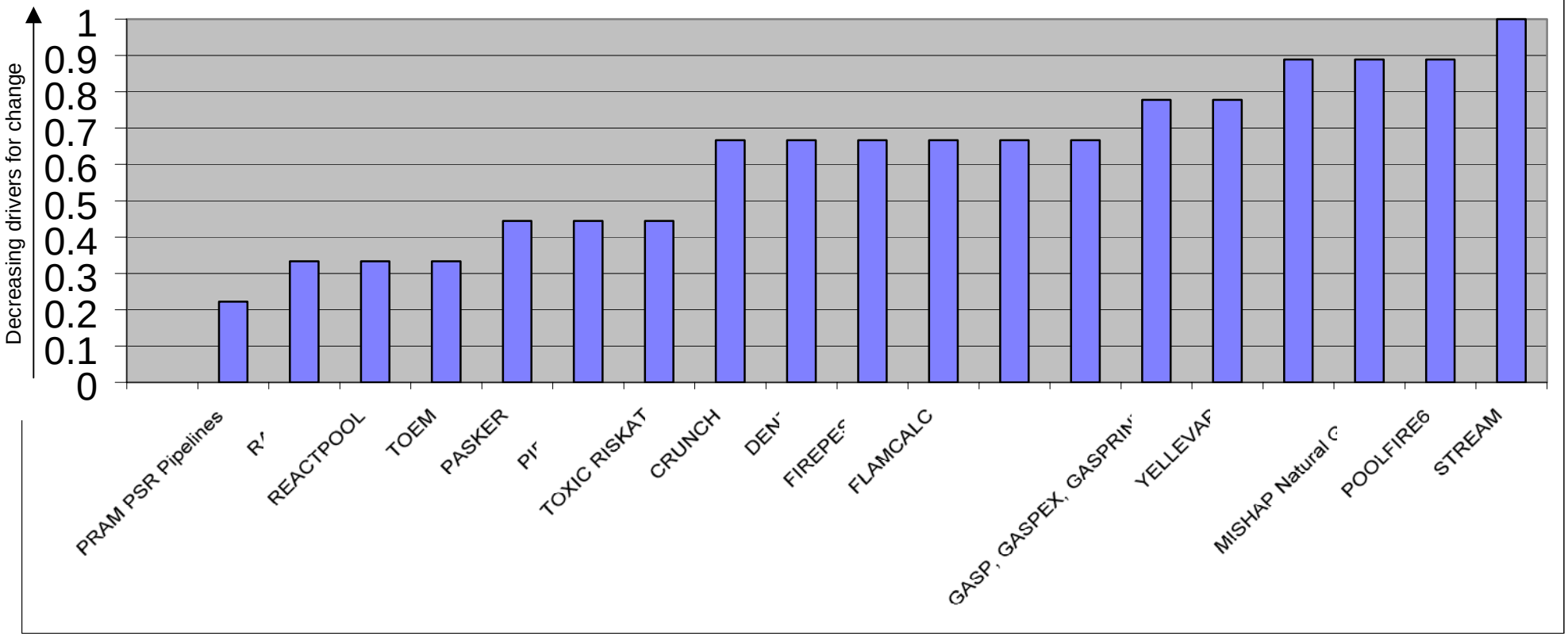


Figure 4 - Likelihood of Success (Models)

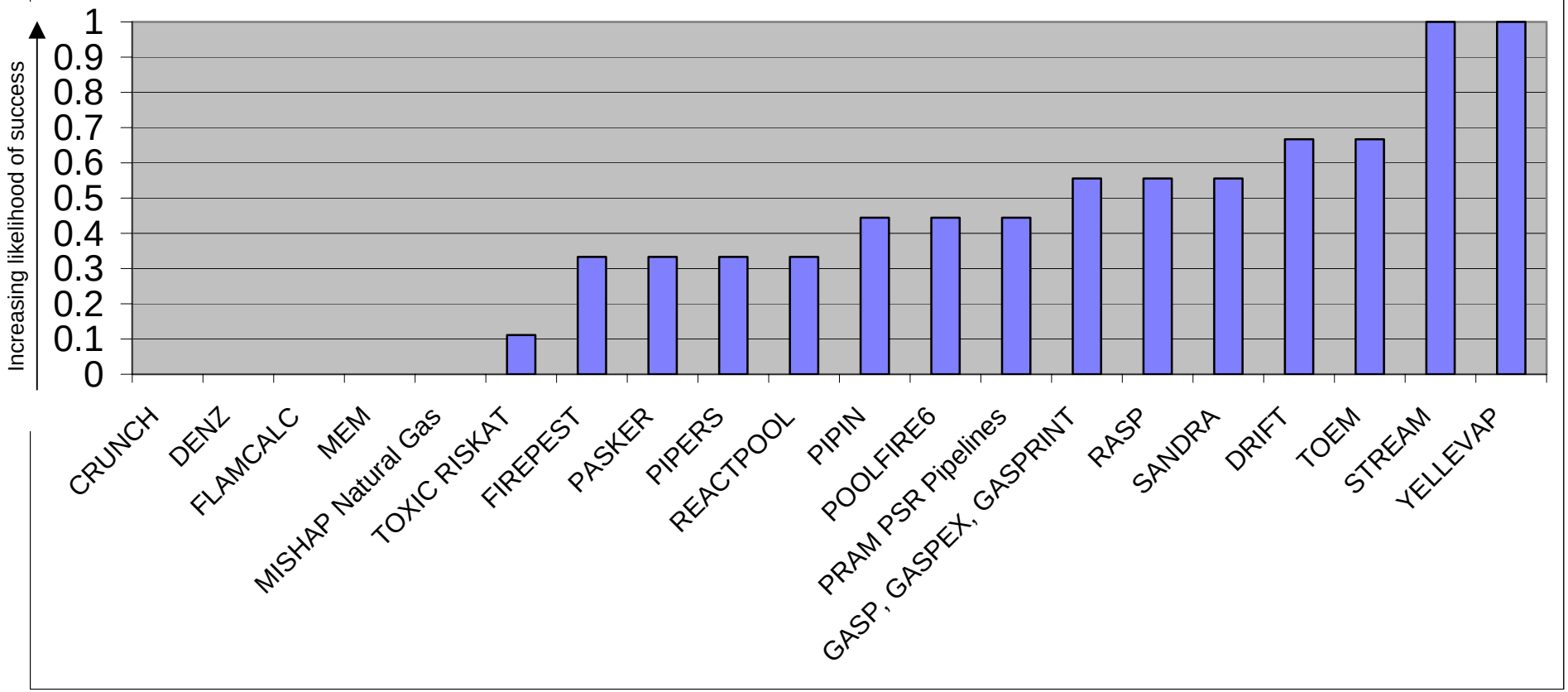
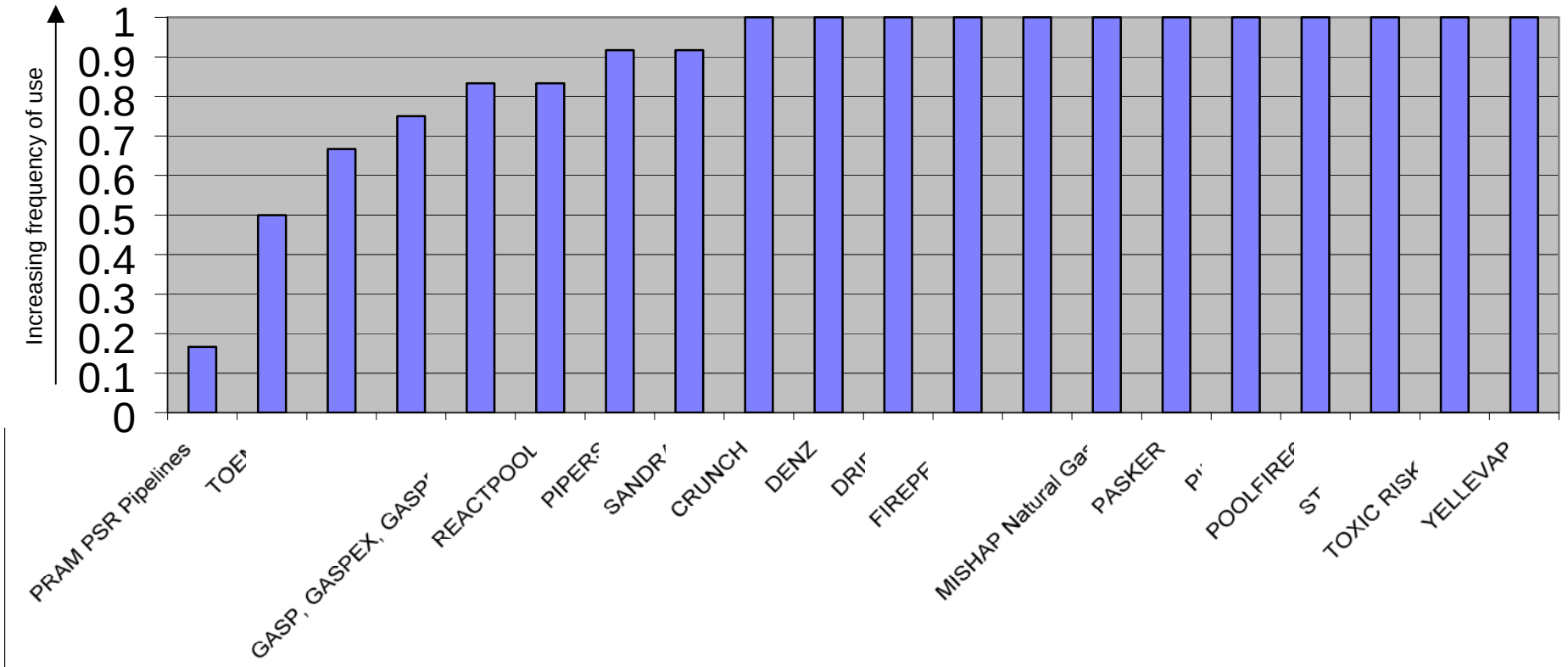


Figure 5 - Frequency of Use (Models)



Analysis of Methodologies

Analysis Structure

A3-41. As with the models, a multi-stage analysis process was undertaken. This process consisted of:

- Preliminary Analysis
- Criterion Ranking
- Categorisation
- Initial Selection
- Individual Analysis
- Final Selection

Preliminary Analysis

A3-42. Again, prior to the main analysis a preliminary analysis was performed to, effectively, quality check the data being used.

A3-43. A number of sub-criteria were deemed inappropriate for use with the methodologies and they, and their response scores, were omitted from the direct assessment process. The qualitative responses to these questions were used as supplementary evidence to support aspects of the model under analysis and to inform its overall assessment where applicable.

A3-44. The questions omitted from the model analysis were:

- Criterion 1, sub-criterion 1j
- Criterion 1, sub-criterion 1k
- Criterion 2, sub-criterion 2b
- Criterion 2, sub-criterion 2c
- Criterion 3, sub-criterion 3d

A3-45. It was decided to remove Criterion1, sub-criteria 1j and 1k from the ranking process as, in the vast number of instances, a response of ‘not applicable’ was returned for them.

A3-46. Criterion 2 sub-criteria 2b and 2c were omitted as they were not applicable or relevant to all methodologies and their inclusion would therefore have biased the assessment.

A3-47. Criterion3 sub-criterion 3d related to the availability, or otherwise, of a replacement commercial software package. This had no relevance to the methodology assessment and was therefore removed.

A3-48. A number of other sub-criteria received ‘not applicable’ responses but, as there was no consistency between methodologies, they were not removed. This also highlighted the fact that, as the same sub-criteria had been used for the models and methodologies, some confusion appears to have been generated as a result of their phrasing. For example the use of the word ‘tool’ in some sub-criteria led some to respond in relation to the models involved in the methodology as opposed to the methodology itself. The confusion was reflected to some extent in the scores attributed. This was not widespread. No revision of scores was undertaken.

A3-49. From the qualitative responses and additional consultation outside the assessment there was an opinion formed that the documentation of the methodologies, for example in the form of the PCAG, had deficiencies. In some cases it had not been rigorously kept up to date, and in a small number did not actually exist. This appeared mainly to be due to pressure of other work. It did however mean that some methodologies as actually employed were different to those in document form.

A3-50. The preliminary analysis seemed to suggest that a general review of the methodologies, or rather their documentation, would be beneficial. It also suggested that the model assessment conclusions were likely to be more robust.

Criterion Ranking

A3-51. Within each Criterion the sub-criteria scores were combined, where appropriate, for each methodology under assessment. The combined score was then normalised relative to the maximum possible score for that methodology, resulting in a performance rating from 0 (poor) to 1 (excellent). This process thus generated a ranking profile for every Criterion. These are shown in Figures 6 to 11.

Categorisation

A3-52. It was decided not to use the AA/EE categorisation method employed previously for the model assessment. This decision was taken principally due to there being fewer methodologies under scrutiny than models, with a narrower spread of relative scores. This made the distinction between the performance of different methodologies less marked.

A3-53. Instead, the ranked methodologies were analysed and assigned to three simple categories of A, B and C for both Risk and FFP. Again, AA equates to the worst performing (high risk, low FFP) and CC the best (low risk, high FFP). No distinction is made between the risk and FFP categories at this stage, i.e. BC and CB are treated as the same.

The full list of methodologies and their deemed category is presented in Table 2.

Worst ⟶ Best

AA	AB/BA	BB	BC/CB	CC
LPG-RISKAT TRAM	AMMONIA SULPHUR TRIOXIDE /OLEUM	HYDROGEN FLUORIDE POOL FIRES LPG:REFRIGERATION	none	CHLORINE LPG:CYLINDERS LPG:PRESSURISED BULK

Table 2 – Categorisation of Methodologies

Initial Selection

A3-54. The methodologies exhibited a similar degree of correlation between the Risk and FFP Criteria as the models, although this is less decisive due to the smaller number of categories. For example, both AMMONIA and SULPHUR TRIOXIDE/OLEUM fell into the AB/BA category but their performance from the ranking profiles is markedly different: AMMONIA can

be easily separated as the worst performer of the two. The initial selection was therefore made from analysis of the categorisation process but with reference to the ranking profiles.

A3-55. Some methodologies marked themselves out as obviously being the weakest performing, with TRAM (AA) and LPG RISKAT (AA) falling in the bottom three for both the Risk and FFP Criteria. The next weakest methodology was AMMONIA (AB) showing the highest risk to HSE with a relatively poor FFP score. Inversely LPG:CYLINDERS (CC), LPG:PRESSURISED BULK (CC) and CHLORINE (CC) methodologies showed themselves to be the best performing, being the three highest ranked in both Risk and FFP.

A3-56. The Wider Influences Criterion further supported the choice of the poorest performing methodologies with TRAM (AA), LPG-RISKAT (AA) and AMMONIA (AB) being three of the worst four performing, although CHLORINE (CC) scored badly here.

A3-57. In the light of the findings of the preliminary analysis, with respect to the qualitative responses, and further discussions with relevant parties it is suggested that the selection of methodologies for development is not as straightforward as for the models. As indicated however, some do perform noticeably worse in the assessment than the majority.

A3-58. Overall it is recommended that the development priority of the methodologies be driven by the conclusions of the model assessment.

Individual Analysis

A3-59. A more detailed analysis of the three worst performing methodologies served to illustrate the model-methodology dependency. The detailed analysis is presented below:

TRAM

A3-60. This methodology was deemed to be of very high risk and very low FFP (AA) and, based on the methodology review alone, would be a primary candidate for development. However, further investigation has led to the knowledge that, for various reasons, it has been very infrequently employed. This reinforced the position that the recommended development of the MISHAP model should be undertaken as part of a new PIPELINES RISKAT methodology. There is, therefore, no case for supporting further development of TRAM.

LPG RISKAT (also known as FLAMMABLE RISKAT)

A3-61. Was assessed as being of very high risk and very low FFP (AA) although it is currently not used except for specialist circumstances. FLAMCALC is generally used instead, but is not a replacement for LPG RISKAT as they work to differing levels of detail.

A3-62. LPG RISKAT makes use of the CRUNCH and DENZ models and, as a result, is subject to the same problems and comments detailed for TOXIC RISKAT in the model assessment.

A3-63. Due to the fact that this methodology has limited current applicability it is not recommended for development. Should HSE decide that detailed risk calculations for the areas covered by LPG RISKAT are routinely

required then development would be necessary and considered to be a high priority. As with TOXIC RISKAT it is more likely that the optimum solution would be to develop a replacement methodology.

AMMONIA

A3-64. Was deemed by the data analysis process to be very high risk and low FFP (AB) and would be expected to be automatically selected on this basis. A detailed study of the data, and further discussion with MSDU staff, showed however that the data was flawed.

A3-65. The phrase ‘Ammonia Methodology’ had resulted in the assessment process eliciting responses for both refrigerated and non-refrigerated ammonia methodologies to varying questions. These are two very different methodologies. Thus the overall assessment data was a combination of the two, and the data analysis had effectively highlighted the worst aspects of both.

A3-66. Non-refrigerated ammonia methodology is used frequently, but has few problems, whereas refrigerated ammonia methodology is problematic but used very infrequently. The combined data gave the impression of a frequently used problematic methodology, which is simply not the case.

A3-67. This confusion may have stemmed from the fact that there is no formalised methodology for ammonia in PCAG; a variation of the CHLORINE methodology is employed for non-refrigerated cases and ad-hoc assessments undertaken for refrigerated. The documentation aspect should be addressed, as part of an overall documentation review, but priority development of a methodology in either individual ammonia case cannot be justified.

Development Recommendations

A3-68. It is recommended, for the reasons presented previously, that consideration should be given to reviewing all methodologies and revising and/or documenting those found to be in need of it.

A3-69. It is also recommended that the model development priorities are used to determine the methodology development requirements.

A3-70. LPG-RISKAT must be considered to be strongly recommended if HSE decides that detailed risk assessments are required in this area.

A3-71. Other methodologies would be recommended, but they will automatically have to be reviewed as a result of the models employed within them and their recommendation for development. For example the SULPHUR TRIOXIDE/OLEUM methodology would be revised in the light of the suggested development of REACTPOOL. Whilst not singly identified in the methodology assessment, it did perform very poorly in the FFP Criterion.

Methodology Analysis Conclusions

A3-72. It is believed that the results of the model assessment will primarily determine the methodologies requiring development. Any methodology involving a model considered appropriate for development would itself have to be at least reviewed. Not surprisingly the outcomes of the two

assessments are interlinked, with all of the worst performing methodologies being connected to models considered appropriate for development or replacement.

A3-73. It is also accepted that the documentation of many methodologies, including PCAG, requires reviewing and updating. This generic development process has already been started and it is recommended by this review that it is continued and supported.

Figure 6 - Fitness for Purpose (Methodologies)

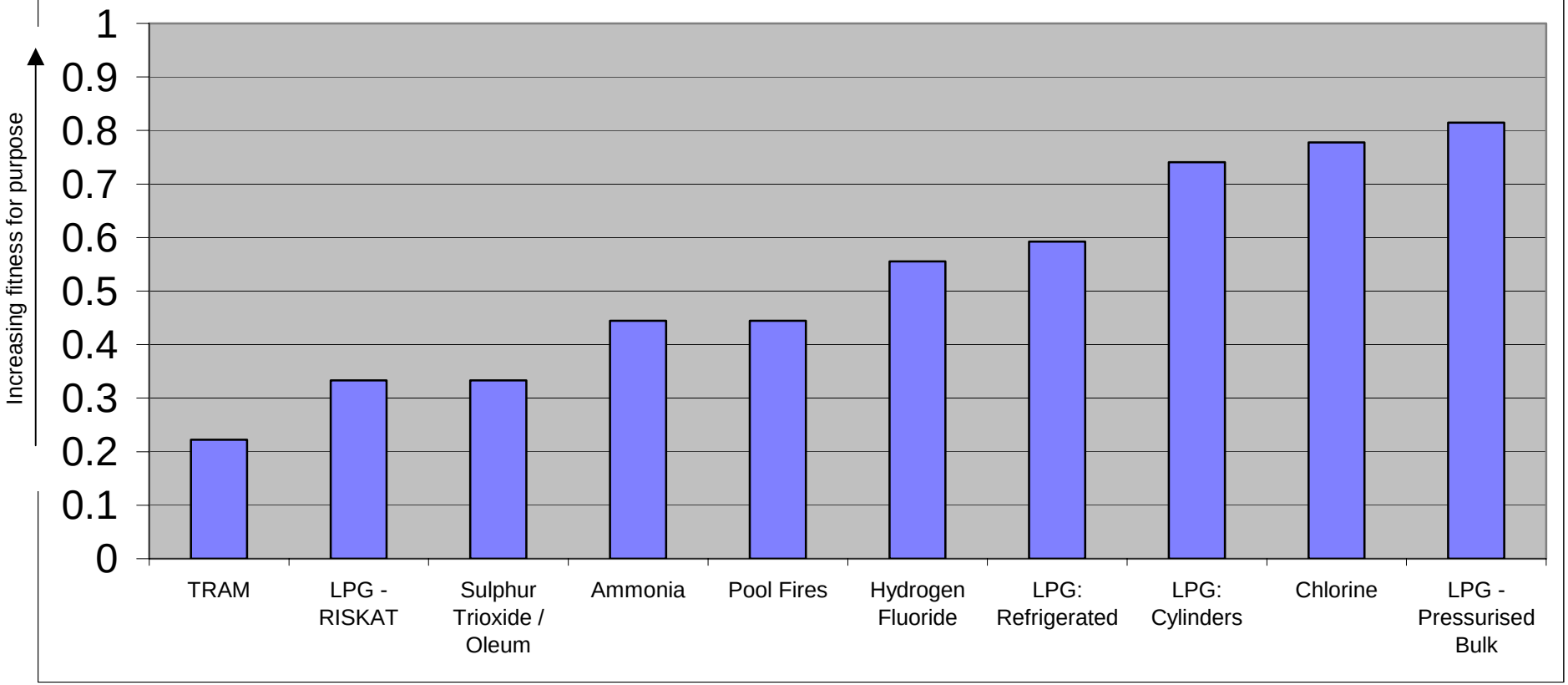


Figure 7 - Risk to HSE (Methodologies)

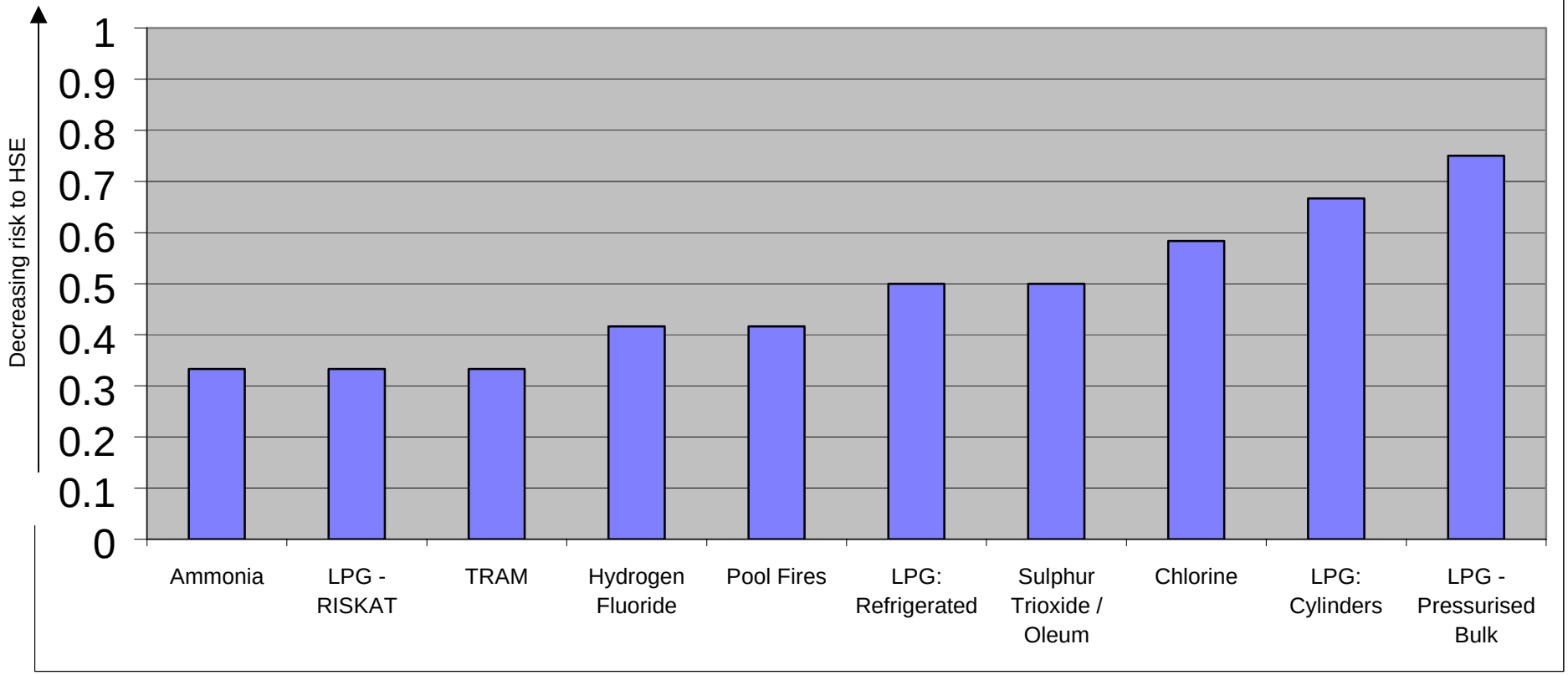


Figure 8 - Wider Influences (Methodologies)

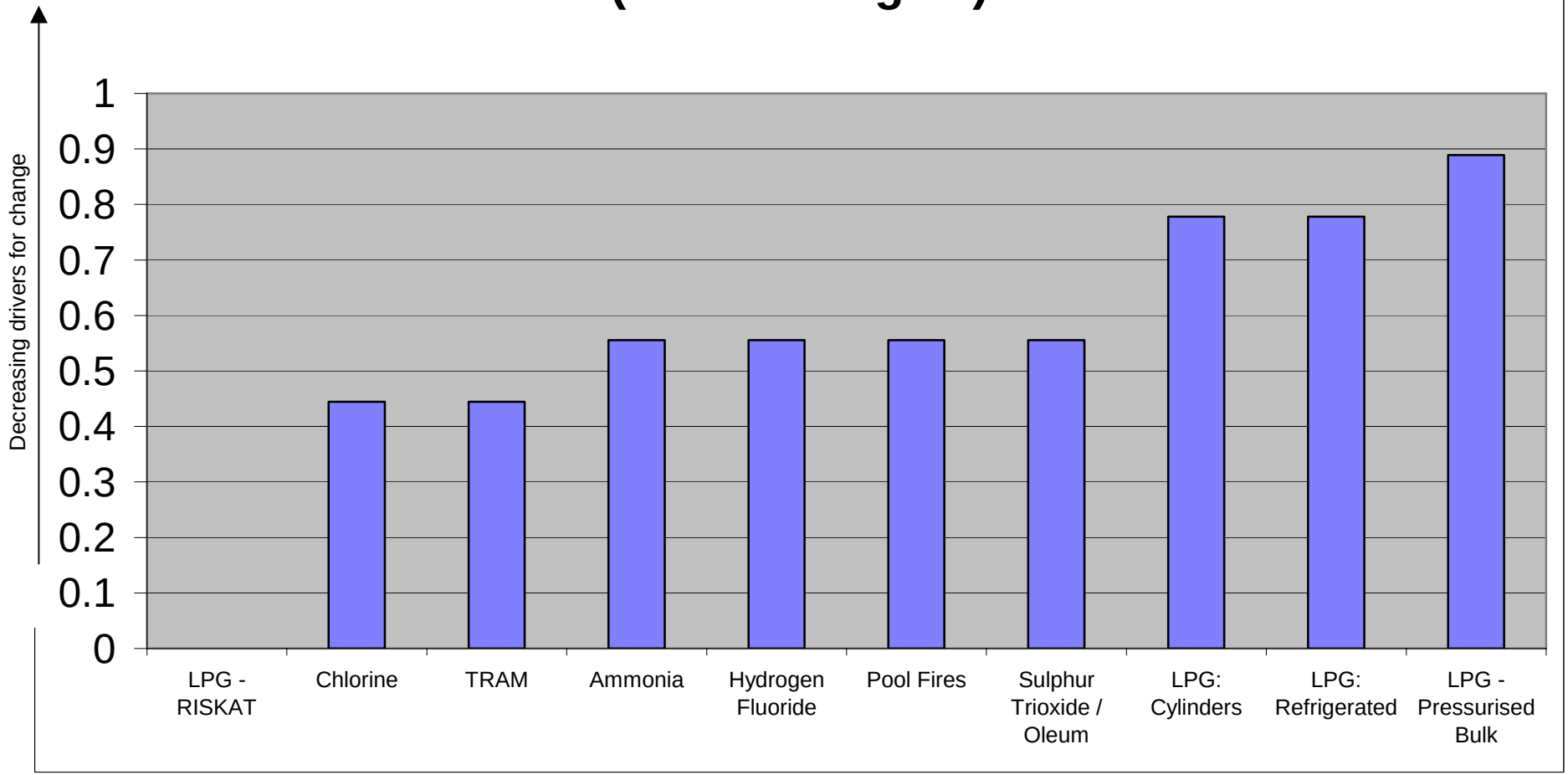


Figure 9 - Frequency of Use (Methodologies)

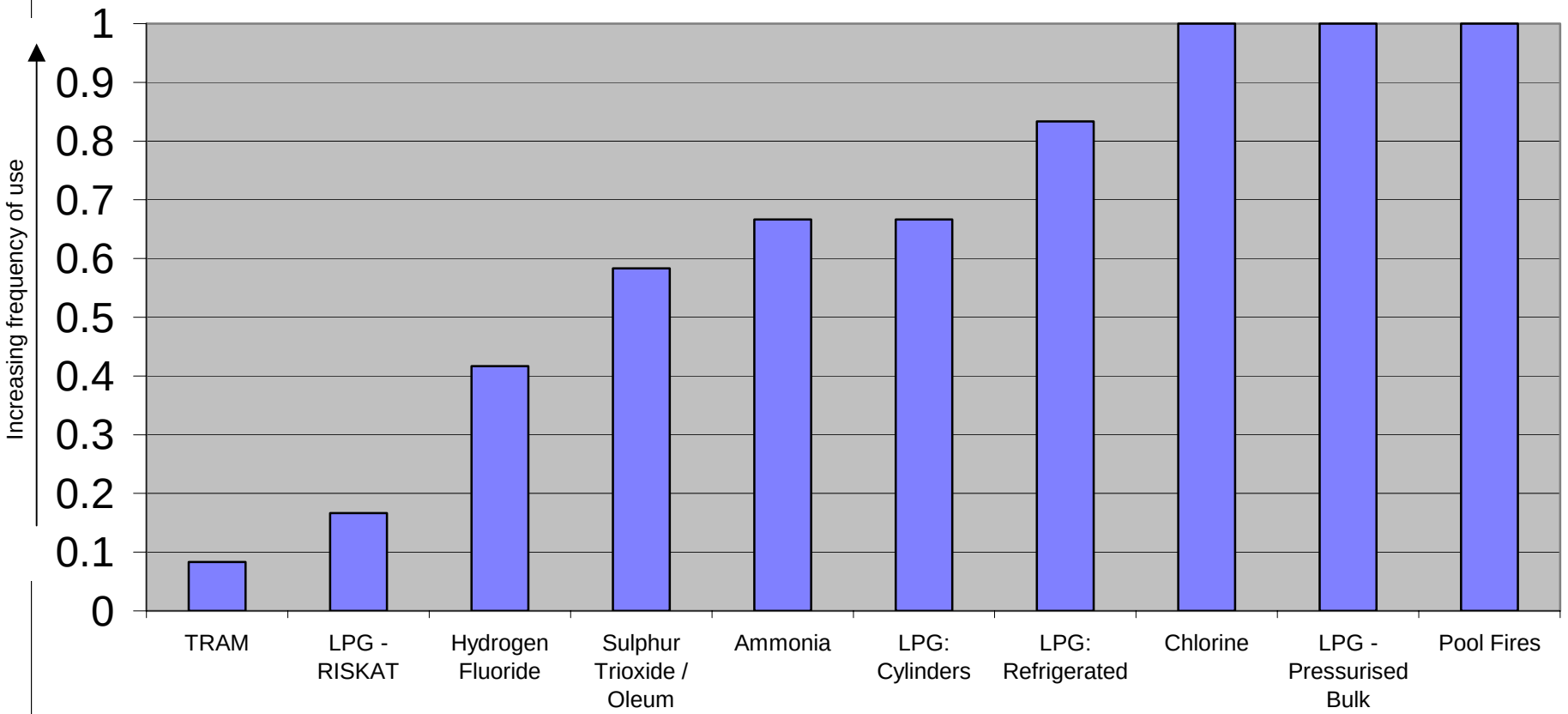


Figure 10 - Likelihood of Success (Methodologies)

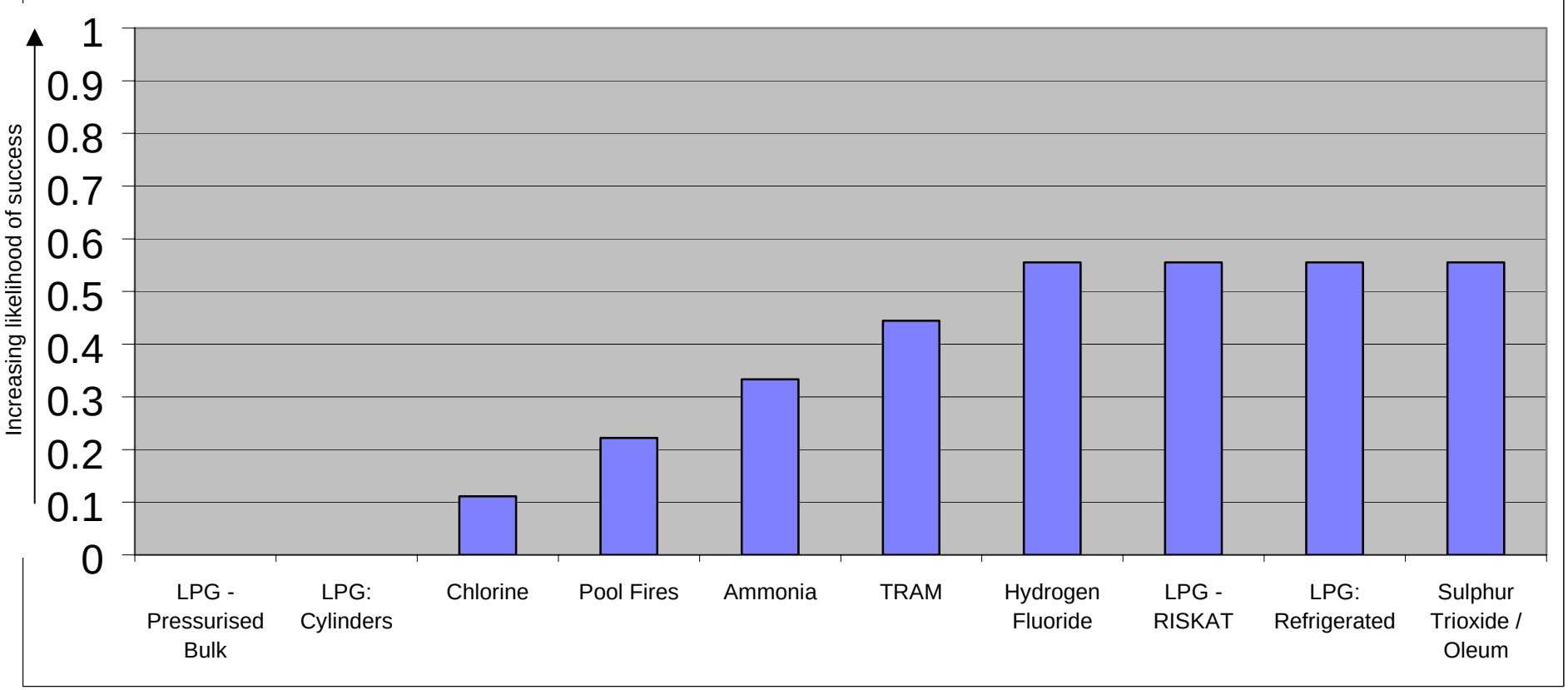
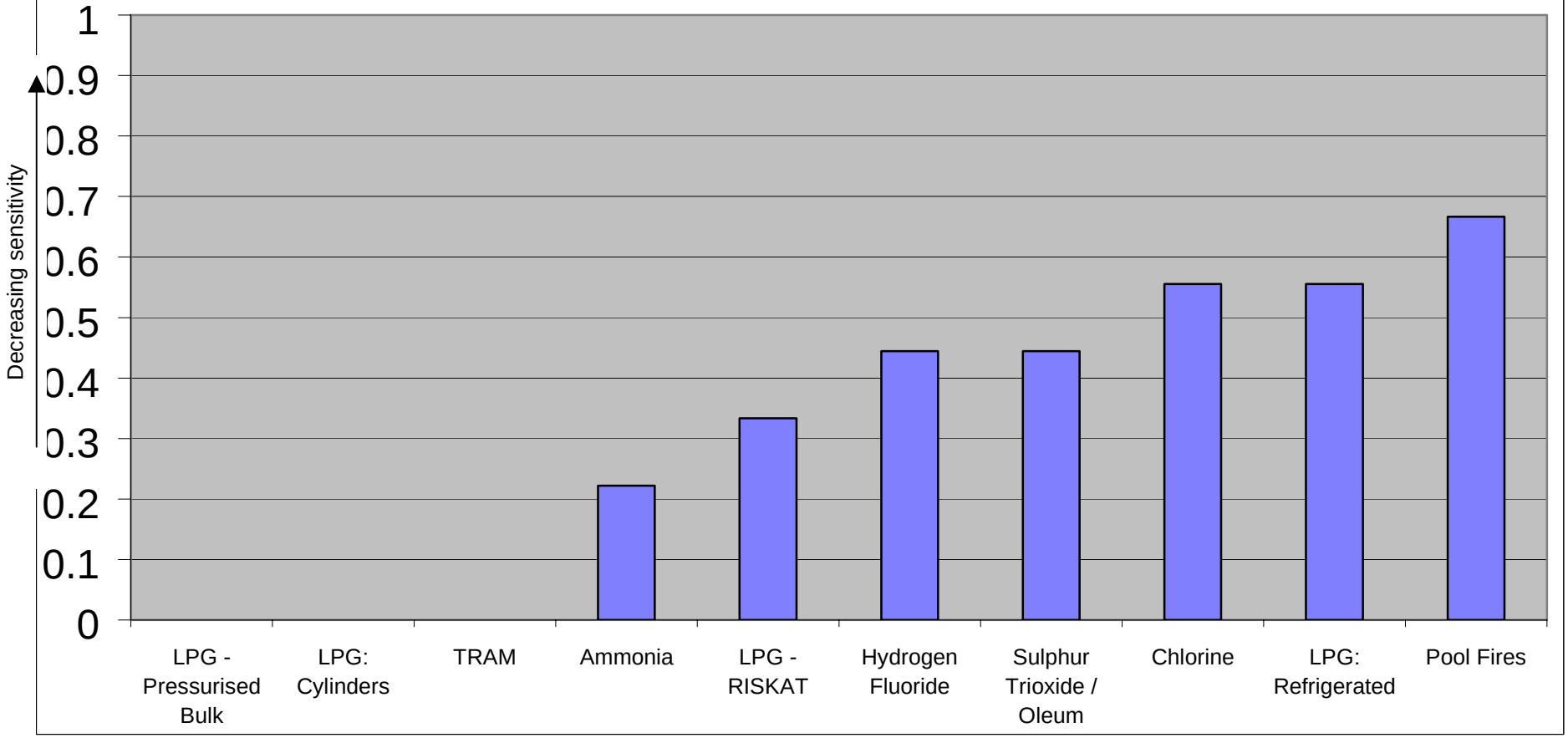


Figure 11 - Snapshot Information (Methodologies)



Topic Specialist Assessments

A3-74. A number of Topic Specialists were interviewed in relation to a selection of models and methodologies from their specialist area.

A3-75. The interviews took the form of a series of ten questions; as opposed to the Criterion sub-criteria statements applied in the general model and methodology assessments. A similar scoring system was however applied, as detailed in Annex 2. As with the previous assessments, the lower the score achieved then the worse the performance and greater the driver for development or change.

Analysis of Interview Data

Preliminary Analysis

A3-76. The scores and qualitative responses were analysed in detail prior to the ranking process taking place. As with the general assessments this was done in order to serve as quality assurance, to ensure that the score attributed to a given qualitative response was consistent with that response, and to inform the general assessment analysis and ranking processes.

A3-77. This preliminary analysis indicated a number of inappropriate score/response combinations. These were investigated, with further advice and clarification being sought where necessary, and revised. The revisions made a minor difference to three models in their attributed scores, and a major difference in the case of two other models.

A3-78. The minor revisions occurred as a result of scores being attributed, which were inconsistent with the response given.

A3-79. The major revisions occurred as a result of the Topic Specialist responding that the models in question (TOEM and RASP) were “rarely consulted” and therefore “this tool not considered further” but scores of 3 were assigned by default to all questions. Hence, despite having only one response each (out of ten questions) the two models attained very high performance scores. The scores for these models were revised to 0 on both counts.

A3-80. In some instances a response of not applicable, or no response at all, was recorded yet was assigned a default score. These scores were replaced with ‘no score’ and this was taken into account in the overall analysis process.

Analysis

A3-81. The models and methodologies were separated, and their scores were used to rank them. The ranking was done in a similar manner to that employed in the general assessments with the scores being normalised relative to the maximum possible for that model/methodology. The ranking profiles for the models and methodologies are shown in Figures 12 and 13 respectively.

Comparative Analysis

A3-82. Although, in the case of the models, there was some similarity in the rankings obtained by the Topic Specialist and general assessments (RASP/TOEM and STREAM) good correlation was not obtained. The Topic Specialist ranking profile did not exhibit the same wide range of performance as seen with the general model assessment, and differences were less marked.

A3-83. With the methodologies the range of performance was greater, but with only a small number of them assessed in this way any ranking has little worth.

A3-84. The fact that certain models and methodologies are only present in either the Topic Specialist assessment or the general assessment (but not both) further complicates any comparative analysis.

Conclusions

A3-85. It should not be surprising that there is a lack of consistency between the Criterion based assessments and the Topic Specialist ones as both the approach taken and the focus of the ‘questions’ was not the same.

A3-86. Primarily the Topic Specialist Interviews were best employed as sources of supplementary information to inform the general assessments performed under the Criterion statements system. In many instances the information here further supported the rankings obtained in the Criterion based analyses and assisted in determining the models and methodologies for the final selection.

Figure 12 - Topic Specialist Responses (Models)

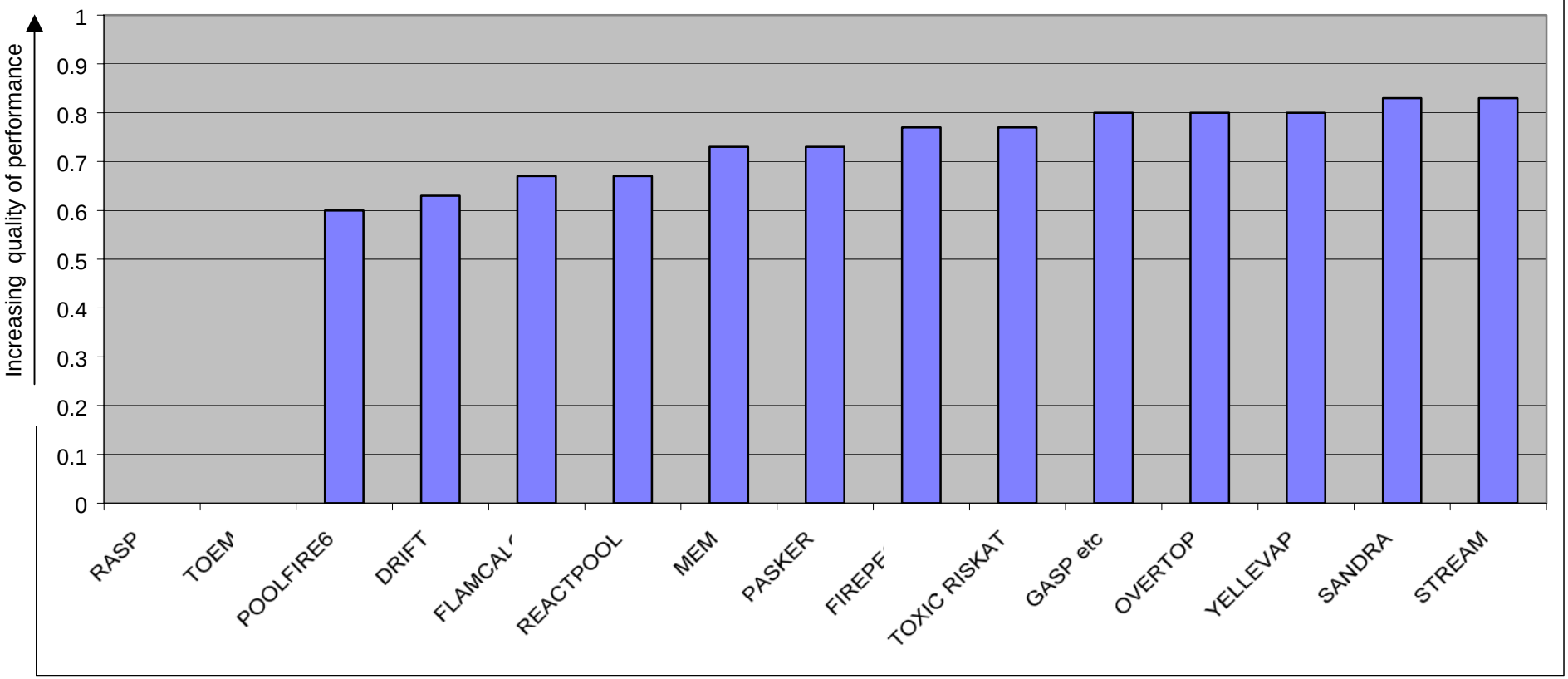
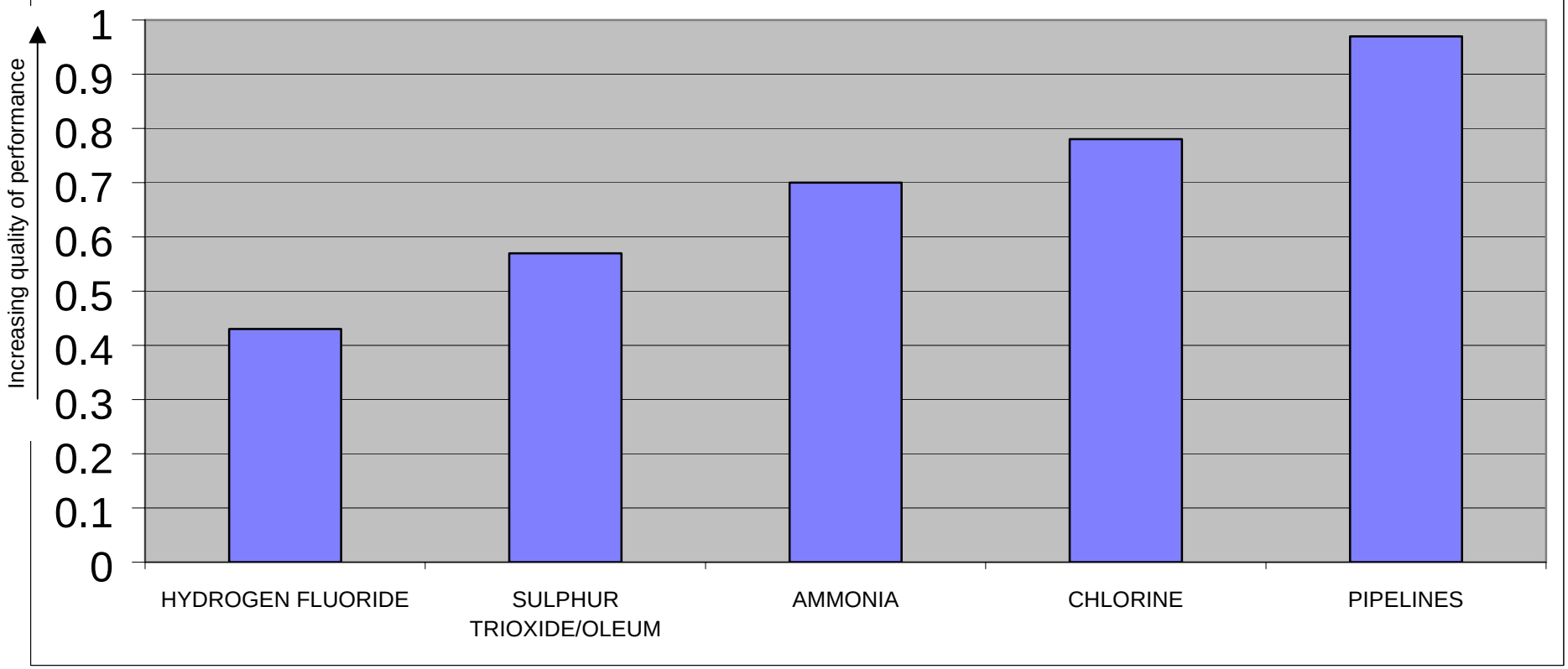


Figure 13 - Topic Specialist Responses (Methodologies)



Critical Review of Assessment Procedure

A3-87. The responses to the statements in the model assessment seemed to be indicative that the questions were appropriate and easily assimilated by both questioner and respondent. This view is borne out by the independent HSL review. Some statements were slightly ambiguous or unclear and required a minor degree of interpretation, although did not result in any fundamental problems.

A3-88. With respect to the methodology assessment however there was a greater degree of ambiguity, with the identical set of statements being applied across the Criteria. In hindsight the statements should have been rephrased to be more appropriate to the methodology approach. In some cases this led to responses being supplied that commented on the models' performance within the methodology rather than the methodology itself. This, in part, explains the differences seen in the HSL and HSE methodology assessments; HSL interpreted the statements relative to the methodology in some instances whereas they were taken at face value in the HSE assessment.

A3-89. In addition, I believe the fact that the HSE methodology assessments were undertaken with the Topic Specialists, whereas the HSL ones were not, may have had an effect. Being very familiar with the systems, the Topic Specialists may have based their assessment on how well the methodologies performed as they applied them (with the benefit of their experience), which may not be identical to the way that they would be employed by someone less knowledgeable and being solely directed by the relevant PCAG (where in existence).

A3-90. When such reviews are carried out involving the people intimately associated with, or even responsible for producing, the product or processes under review an element of bias is very difficult to avoid. This review is no exception.

A3-91. It is believed however that, through the inclusion of external experts in the data generation process, this bias has at least been minimised. The results of the independent HSL review give credence to the view that such biases, if any, and minor flaws in the review process have not undermined the credibility of the process, its outcomes and conclusions.

A3-92. The Topic Specialist interviews were conducted by the use of questions, as opposed to statements, but a similar relative scoring format was employed. This obviously required some degree of interpretation in allocating a score and may have contributed to the apparent variations between the outcomes of the model/methodology assessments and the Topic Specialist interviews.

A3-93. The transfer of responsibility for this product through several members of project staff may also have contributed to some errors. The principal example of this is the assessment of the distinct refrigerated and non-refrigerated ammonia 'methodologies' under a single generic AMMONIA heading. The lack of a clear definition between models and methodologies

(e.g. TOXIC RISKAT-assessed as a model, LPG RISKAT assessed as a methodology) is a further example.

A3-94. As stated in the Overall Aims section of this Annex, a model or methodology could only be considered for development if it was included in the assessment process. Consequently this review is a relative, not absolute, determination of development priorities. There may, or may not, be other models and methodologies excluded from the review that would have been considered appropriate for development had they been included. The likelihood of this is greatly reduced due to the factors used to determine those considered, but cannot be entirely dismissed.

A3-95. Overall, it is concluded that the recommendations for development are based on a rigorous analysis of reliable data obtained from the P5 programme, and that those recommendations are justified. The reliability of the data and the assessment process is fully supported by the conclusions of the independent HSL comparative study.

Annex 4 – HSL Independent Review

Introduction

As part of product P5 of the Implementation of the Fundamental Review of Land Use Planning (IFRLUP) project HSE reviewed a number of models and methodologies used for land use planning (review criteria included in Annex 1). The output from this review was to rank all models and methodologies to determine priorities for further development, if required. An integral part of the IFRLUP project was for an independent and impartial review of a sample of the models and methodologies to be carried out; this was undertaken by HSL. The aim of this independent review was to give confidence in the HSE review process and review outputs. The same approach and criteria were used as used by HSE in their review.

Four models and two methodologies were selected by HSL to be reviewed from those reviewed by HSE. Two criteria were used to aid selection. The first was to select models and methodologies that covered the range in rankings determined by HSE. The second was to select models and methodologies that HSL had sufficient expertise and knowledge.

The HSL review utilised a team approach, involving people with experience from using the models and methodologies, and in developing them. The team was made up of three permanent members (one of which was the chairman who led the review and recorded the results, and the other two had experience in carrying out Land Use Planning assessments) whose numbers were augmented by additional members with specialist knowledge of particular models. The review and scoring of the models and methodologies were performed blind from the detailed results obtained by HSE. However, HSE's overall scores were used so that models and methodologies with a range of scores could be selected; this information was only utilised by the chairman.

The following models were selected:

- Toxic Riskat;
- Drift;
- Sandra; and
- Firepest.

The following methodologies were selected:

- Oleum/Sulphur trioxide; and
- Chlorine.

It is noted however that there was less expertise available for the review of the methodologies and therefore this review was mainly based on the information present in PCAG.

This report presents the scores assigned by HSL to each sub criteria, for the four models and two methodologies reviewed, and compares these to those generated by HSE's review. Large differences (two points or more) between the scoring of the sub criteria of the HSE and HSL reviews are highlighted.

Small differences in scoring (only one point) have not been highlighted as these have an insignificant effect on the overall scoring, and ranking, of the tools. Where there were large differences, these were identified and the comments made during the review were re-examined to identify if there was any obvious reason for the difference. In this report a difference of two points has been termed a significant difference while for a difference of three points this has been termed an opposite response.

Review of the models

The results of the four models reviewed by HSL, with a comparison against the results of the review by HSE, are included in this section. Results are shown for each of the five assessment criteria, including the sub criteria, and the overall score.

Each sub criteria statement has, where possible, been scored as follows:

- 0 – Strongly disagree
- 1 – Disagree
- 2 – Agree
- 3 – Strongly agree

Some of the sub criteria have not been scored during the HSL review, and have been left blank in the tables below. This corresponds to where the HSL review team had insufficient knowledge to come to any conclusions. Where this has been the case HSE's overall criteria scores have been adjusted to remove the contribution of these sub criteria (original score is shown in parentheses in the tables below).

From a review of the sub criteria carried out by HSE it has been identified by them that sub criteria 2b and 2c are not applicable/relevant to all models, and sub criterion 3d does not have a bearing on the performance of the model. Therefore, these have not formed part of the ranking exercise, although the information has been used qualitatively. Comparative scores for these sub criteria are therefore presented separately.

Criterion 1: Fitness for current purpose

The following sub criteria were used in the assessment:

1a The tool is scientifically credible;
1b The tool has no recognised significant weaknesses;
1c The tool gives broadly acceptable results/performs as expected;
1d There is confidence in the tool when correctly applied;
1e Assumptions are considered sensible/ realistic;
1f Sensitivity - there are no inputs that significantly/ disproportionately affect the outputs;
1g There are no existing plans to revise/replace this tool;
1h The user is rarely directed to consult with a Topic Specialist;
1i The tool is quality assured for use;
1j The tool has been validated against "real events"; and
1k The tool has been thoroughly validated in experiments.

Table 1 compares the sub criteria scores from HSE's review with those from HSL's review for the four models reviewed by HSL. From Table 1 it can be seen that there are no 'opposite responses' for any of the sub criteria, but there are a number where there are significant differences, and these are discussed in greater detail below.

Table 1 Criterion 1 comparison for the models

Sub criteria	Score							
	Drift		Firepest		Sandra		Toxic Riskat	
	HSE	HSL	HSE	HSL	HSE	HSL	HSE	HSL
1a	3	3	2	2	1	0	2	0
1b	2	3	1	2	1	0	1	0
1c	3	3	2	2	1	2	2	1
1d	3	3	3	3	1	1	2	0
1e	3	3	2	2	1	1	2	1
1f	2	3	1	1	3	3	1	-
1g	1	-	2	3	0	1	1	0
1h	1	2	3	3	3	3	1	2
1i	2	0	1	0	2	0	0	0
1j	0	-	1	0	2	0	N/A	-
1k	2	3	2	1	0	0	N/A	-
Total	21 (22)	23	20	19	15	11	11 (12)	4

For the model Drift there is a significant difference for sub criterion 1i. Both reviews stated that the model was produced before QA procedures were in place but different scores have been assigned. Further HSE statements are that it has been submitted to SMEDIS and that a QA system is now available for modifications. This appears to be the basis of the difference.

For the model Sandra there are significant differences for sub criteria 1i and 1j. Again, for 1i both review teams state that there were no QA procedures in place when the model was produced but a further comment by HSE that appropriate records exist has produced a higher score. The difference for sub criterion 1j is believed to be due to HSL not knowing of additional work that has been performed to validate the model.

For the model Toxic Riskat there are significant differences for sub criteria 1a and 1d. For both 1a and 1d the lower scores in the HSL review were due to the use of DENZ and CRUNCH within RISKAT, which are both considered by the HSL team to be no longer scientifically credible.

Criterion 2: Risk to HSE

The following sub criteria were used in the assessment:

2a There is no controversy associated with this tool;
2b The tool has been used or could be used to assist judgements where ARI indicates societal risk of concern;

2c The tool has been used or could be used in assessments for 'sensitive' hazardous installations – e.g. where there is a significant build up of development including sensitive developments;
2d The tool has been quality assured for development;
2e The tool has been rigorously peer reviewed; and
2f Development to address any fitness for purpose shortcomings identified would not impact significantly on the size of the CDs or otherwise have the potential to affect HSE credibility.

Table 2 compares the sub criteria scores from HSE's review with those from HSL's review for the four models reviewed by HSL. From Table 2 it can be seen that there are a number sub criteria where there are significant differences; these are discussed in greater detail below.

Table 2 Criterion 2 comparison for the models

Sub criteria	Score							
	Drift		Firepest		Sandra		Toxic Riskat	
	HSE	HSL	HSE	HSL	HSE	HSL	HSE	HSL
2a	1	3	3	3	1	1	2	0
2d	2	-	2	-	0	0	0	0
2e	2	3	2	1	2	0	3	3
2f	3	3	1	-	1	3	1	0
Total	6 (8)	9	5 (8)	4	4	4	6	3

For the model Drift there is a significant difference for sub criterion 2a. For 2a the difference is due to HSL providing a higher score as there was considered to be no controversy as Drift has been used in several EU joint-funded projects with wide agreement and that it is appropriate for its job. HSE has given it a lower score as it will be used for sensitive cases in the near future.

For the model Sandra there are significant differences for sub criteria 2e and 2f. For 2e the difference is due to HSE being able to identify papers where the dispersion and explosion modelling have been published. For 2f HSE has identified this as not applicable as they are going to replace the model not develop it further, while the higher value from HSL is due to the knowledge that new work has not produced significantly different consultation distances.

For the Toxic Riskat model there is a significant difference for sub criterion 2a. For 2a the issue with using DENZ and CRUNCH, as discussed previously, has caused a lower score to be assigned by HSL.

Criterion 3: Wider influences
The following sub criteria were used in the assessment:

3a There are no drivers for change;
3b The tool supports HSE's credibility and is sufficiently leading edge;
3c The tool can cope with constraints on maximum ranges to accommodate political/policy/pragmatic restraints; and
3d There is a commercially available package that could replace this tool.

Table 3 compares the sub criteria scores from HSE's review with those from HSL's review for the four models reviewed by HSL. From Table 3 it can be seen that there are no opposite responses for any of the sub criteria, but there is one with a significant difference, and this is discussed in greater detail below.

Table 3 Criterion 3 comparison for the models

Sub criteria	Score							
	Drift		Firepest		Sandra		Toxic Riskat	
	HSE	HSL	HSE	HSL	HSE	HSL	HSE	HSL
3a	2	2	1	2	0	1	1	0
3b	3	3	2	3	0	1	2	0
3c	3	-	3	-	0	-	1	-
Total	5 (8)	5	3 (6)	5	0 (0)	2	3 (4)	0

For the Toxic Riskat model there is a significant difference for sub criterion 3b. This difference is again due to HSL's view that the use of DENZ and CRUNCH is a significant weakness in the model and therefore does not support HSE's credibility and is not sufficiently leading edge.

Criterion 4: Extent of use

The following sub criteria were used in the assessment:

4a The tool is widely applicable;
4b The tool is used often;
4c The tool was used recently; and
4d The tool is likely to be used again in the near future.

Table 4 compares the sub criteria scores from HSE's review with those from HSL's review for the four models reviewed by HSL. From Table 4 it can be seen that there are no opposite responses for any of the sub criteria, but there is one with a significant difference, and this is discussed in greater detail below.

Table 4 Criterion 4 comparison for the models

Sub criteria	Score							
	Drift		Firepest		Sandra		Toxic Riskat	
	HSE	HSL	HSE	HSL	HSE	HSL	HSE	HSL
4a	3	3	3	3	3	1	3	2
4b	3	-	3	3	2	3	3	2
4c	3	-	3	3	3	3	3	3
4d	3	-	3	3	3	3	3	3
Total	3 (12)	3	12	12	11	10	12	10

For the Sandra model there is a significant difference for sub criterion 4a. For this criterion HSE has stated that it can be applied to all ammonium nitrate assessments. HSL identified that it was only appropriate for 300 tonne stacks of ammonium nitrate stored indoors, so a lower value was assigned.

Criterion 5: Likelihood of success
The following sub criteria were used in the assessment:

5a The technology is available to deliver the desired improvement(s);
5b A minimal amount of effort is required for any proposed improvements; and
5c Proposed developments would have a minimal impact on the extent of CDs, leading to minimal costs.

Table 5 compares the sub criteria scores from HSE's review with those from HSL's review for the four models reviewed by HSL. From Table 5 it can be seen that there are no opposite responses for any of the sub criteria, but there are two with significant differences, and these are discussed in greater detail below.

Table 5 Criterion 5 comparison for the models

Sub criteria	Score							
	Drift		Firepest		Sandra		Toxic Riskat	
	HSE	HSL	HSE	HSL	HSE	HSL	HSE	HSL
5a	3	3	1	3	3	3	1	3
5b	1	0	1	-	1	0	0	0
5c	2	2	1	-	1	2	0	0
Total	6	5	1 (3)	3	5	5	1	3

For the Firepest model there is a significant difference for sub criterion 5a. HSL believe that the technology is available to deliver the desired improvements. This relates to stage one of the improvement process identified by HSE, and if the second phase identified by HSE had been considered by HSL it is likely that a different score would have been assigned.

For the Toxic Riskat model there is a significant difference for sub criterion 5a. HSL indicated that dispersion models are available that could be used in Toxic Riskat instead of DENZ and CRUNCH. HSE state that the architecture would make the changeover difficult and have therefore given it a lower value.

Overall scoring
Table 6 compares both the total scores for each criterion and the overall score from HSE's review with those from HSL's review for the four models reviewed by HSL.

Table 6 Overall comparison for the models

Criteria	Score							
	Drift		Firepest		Sandra		Toxic Riskat	
	HSE	HSL	HSE	HSL	HSE	HSL	HSE	HSL
1	21 (22)	23	20	19	15	11	11 (12)	4
2	6 (8)	9	5 (8)	4	4	4	6	3
3	5 (8)	5	3 (6)	5	0 (0)	2	3 (4)	0
4	3 (12)	3	12	12	11	10	12	10
5	6	5	1 (3)	3	5	5	1	3
Total	41 (56)	45	41 (49)	43	35 (35)	32	33 (35)	20
Number of sub criteria not scored	0	7 2 in cr1 1 in cr2 1 in cr3 3 in cr4	0	5 2 in cr2 1 in cr3 2 in cr5	0	1 1 in cr3	2 2 in cr1	4 3 in cr1 1 in cr3

For most of the criteria the results of the HSE review and the HSL review are relatively close. The overall ranking of the four models is also consistent between the two reviews. The major differences in results are for Toxic Riskat, which as discussed earlier is due to the HSL team assigning lower scores for various sub criteria due to the incorporation of DENZ and CRUNCH into Toxic Riskat as the HSL team believed these not to be fit for purpose. However, overall there is sufficient correlation between the HSE and HSL reviews to give confidence in HSE's review of the other models.

Sub criteria not used to rank the models

As stated previously sub criteria 2b, 2c, and 3d have not been used by HSE for ranking the various models so have been separated from the information above. For completeness they are included in Table 7.

Table 7 Sub criteria not used for ranking

Sub criteria	Score							
	Drift		Firepest		Sandra		Toxic Riskat	
	HSE	HSL	HSE	HSL	HSE	HSL	HSE	HSL
2b	2	3	2	3	2	2	1	3
2c	3	3	2	3	2	2	3	3
3d	3	2	1	0	0	0	2	3

From Table 7 it can be seen that there is only one significant difference. For the model Toxic Riskat, sub criterion 2b the HSL team believed that that while the Toxic Riskat file itself could not be used in the ARI calculation an ISO file is produced that would provide information appropriate to perform an ARI calculation.

Review of methodologies

The results of the two methodologies reviewed by HSL, with a comparison against the results of the review by HSE, are included in this section. Results are shown for each of the six assessment criteria, including the sub criteria, and the overall score.

Each sub criteria statement has, where possible, been scored as follows:

- 0 – Strongly disagree
- 1 – Disagree
- 2 – Agree
- 3 – Strongly agree

Again, some of the sub criteria have not been scored during the HSL review, and have been left blank in Tables 8 to 13 below. This corresponds to where the HSL review team had insufficient knowledge to come to any conclusions. Where this has been the case HSE's overall criteria scores have been adjusted to remove the contribution of these sub criteria (original score is shown in parentheses in the tables below).

From a review of the sub criteria carried out by HSE it has been identified by them that sub criteria 2b and 2c are not applicable/relevant to all models, and sub criterion 3d does not have a bearing on the performance of the model. Therefore, these have not formed part of the ranking exercise, although the information has been used qualitatively. Comparative scores for these sub criteria are therefore presented separately.

Unlike with the models a detailed assessment of the scores has not been performed for the methodologies. This was not deemed appropriate as the same issues would have been raised in each instance. A higher-level examination of the differences has been performed instead; this is shown in the overall scoring section below.

Overall scoring

Table 8 compares the total criterion scores and the overall score from HSE's review with those from HSL's review for the two methodologies reviewed by HSL.

Table 8 Overall comparison for the methodologies

Criteria	Score			
	SO3/Oleum		Chlorine	
	HSE	HSL	HSE	HSL
1	9 (9)	4	16 (21)	3
2	6	0	4 (7)	1
3	2 (5)	0	3 (4)	0
4	3 (7)	0	6 (12)	3
5	5	3	1	9
6	1 (4)	0	3 (5)	0
Total	26 (36)	7	33 (50)	16
Number of sub criteria not scored	2 2 in cr1	8 3 in cr1 1 in cr3 2 in cr4 2 in cr6	2 2 in cr1	9 4 in cr1 1 in cr2 1 in cr3 2 in cr4 1 in cr6

As can be seen in Table 8 there are significant differences between HSL's review of the methodologies and HSE's. There are 2 main reasons for the differences in the scoring.

The first reason is that unlike the review of the models there was not the level of expertise in the use of the methodologies within the HSL review team. HSL did not have use of the Topic Specialists within their review, as did HSE. To try and counteract this limitation the relevant chapters in PCAG were used to help with the review. However, PCAG is not completely up-to-date and it has limitations. The scoring by HSL relates to these issues and produced such comments as PCAG was 'not clear' or 'not in a logical order'. It was also the belief of the HSL review team that a relatively small amount of effort to improve PCAG would make a significant improvement to the overall methodology. Possible improvements such as incorporating flowcharts of the assessment process would significantly help, and would aid in a consistent approach in Land Use Planning assessment.

The second reason is that during the HSL review of the methodologies it was considered that if the methodology stated that a model should be used that is known to have weaknesses then that is a weakness in the methodology itself. For both methodologies reviewed by HSL, models that are known to be limited are indicated to be used. For the sulphur trioxide / oleum methodology TOEM and RASP are used, with HSL commenting that 'The choice of TOEM and RASP is not appropriate due to problems.' For the chlorine methodology, ultimately DENZ and CRUNCH are to be used within the Toxic Riskat model,

with HSL commenting ‘Directs you to use inappropriate models.’ This has also been taken into account in the HSE review but possibly this has affected their scoring to a different extent.

The HSL team have reviewed the scores assigned by HSE. Overall these scores look appropriate given the evidence recorded. Therefore, the differences observed appear to be due to the HSL review team having limited knowledge, limitations in what is recorded in PCAG and the HSL review team not having input from the Topic Specialists who hold most of the relevant knowledge. The differences are not indicative of problems with the HSE review. Overall, the HSE review of methodologies would appear to be fit for purpose.

Criterion 1: Fitness for current purpose

The following sub criteria were used in the assessment:

1a The tool is scientifically credible;
1b The tool has no recognised significant weaknesses;
1c The tool gives broadly acceptable results/performs as expected;
1d There is confidence in the tool when correctly applied;
1e Assumptions are considered sensible/ realistic;
1f Sensitivity – there are no inputs that significantly/ disproportionately affect the outputs;
1g There are no existing plans to revise/replace this tool;
1h The user is rarely directed to consult with a Topic Specialist;
1i The tool is quality assured for use;
1j The tool has been validated against “real events”; and
1k The tool has been thoroughly validated in experiments.

Table 9 compares the sub criteria scores from HSE’s review with those from HSL’s review for the two methodologies reviewed by HSL.

Table 9 Criterion 1 comparison for the methodologies

Sub criteria	Score			
	SO3/Oleum		Chlorine	
	HSE	HSL	HSE	HSL
1a	1	0	2	0
1b	1	0	1	0
1c	1	0	3	0
1d	1	0	3	0
1e	1	0	2	1
1f	1	1	2	-
1g	0	-	3	-
1h	1	3	3	2
1i	2	0	2	0
1j	N/A	N/A	N/A	N/A
1k	N/A	N/A	N/A	N/A
Total	9 (9)	4	16 (21)	3

Criterion 2: Risk to HSE

The following sub criteria were used in the assessment:

2a There is no controversy associated with this tool;
2b The tool has been used or could be used to assist judgements where ARI indicates societal risk of concern;
2c The tool has been used or could be used in assessments for 'sensitive' hazardous installations – e.g. where there is a significant build up of development including sensitive developments;
2d The tool has been quality assured for development;
2e The tool has been rigorously peer reviewed; and
2f Development to address any fitness for purpose shortcomings identified would not impact significantly on the size of the CDs or otherwise have the potential to affect HSE credibility.

Table 10 compares the sub criteria scores from HSE's review with those from HSL's review for the two methodologies reviewed by HSL.

Table 10 Criterion 2 comparison for the methodologies

Sub criteria	Score			
	SO3/Oleum		Chlorine	
	HSE	HSL	HSE	HSL
2a	1	0	1	0
2d	1	0	3	-
2e	2	0	3	0
2f	2	0	0	1
Total	6	0	4 (7)	1

Criterion 3: Wider influences

The following sub criteria were used in the assessment:

3a There are no drivers for change;
3b The tool supports HSE's credibility and is sufficiently leading edge;
3c The tool can cope with constraints on maximum ranges to accommodate political/policy/pragmatic restraints; and
3d There is a commercially available package that could replace this tool.

Table 11 compares the sub criteria scores from HSE's review with those from HSL's review for the two methodologies reviewed by HSL.

Table 11 Criterion 3 comparison for the methodologies

Sub criteria	Score			
	SO3/Oleum		Chlorine	
	HSE	HSL	HSE	HSL
3a	1	0	1	0
3b	1	0	2	0
3c	3	-	1	-
Total	2 (5)	0	3 (4)	0

Criterion 4: Extent of use
The following sub criteria were used in the assessment:

4a The tool is widely applicable;
4b The tool is used often;
4c The tool was used recently; and
4d The tool is likely to be used again in the near future.

Table 12 compares the sub criteria scores from HSE's review with those from HSL's review for the two methodologies reviewed by HSL.

Table 12 Criterion 4 comparison for the methodologies

Sub criteria	Score			
	SO3/Oleum		Chlorine	
	HSE	HSL	HSE	HSL
4a	2	-	3	3
4b	1	0	3	0
4c	2	0	3	-
4d	2	-	3	-
Total	3 (7)	0	6 (12)	3

Criterion 5: Likelihood of success
The following sub criteria were used in the assessment:

5a The technology is available to deliver the desired improvement(s);
5b A minimal amount of effort is required for any proposed improvements; and
5c Proposed developments would have a minimal impact on the extent of CDs, leading to minimal costs.

Table 13 compares the sub criteria scores from HSE's review with those from HSL's review for the two methodologies reviewed by HSL.

Table 13 Criterion 5 comparison for the methodologies

Sub criteria	Score			
	SO3/Oleum		Chlorine	
	HSE	HSL	HSE	HSL
5a	3	3	1	3
5b	1	0	0	3
5c	1	0	0	3
Total	5	3	1	9

Criterion 6: Snapshot information
The following sub criteria were used in the assessment:

6a There is clear guidance to ensure snapshot information is accurate (a one off consideration);

6b The tool is not sensitive to snapshot information; and

6c The snapshot information for this tool is current.

Table 14 compares the sub criteria scores from HSE's review with those from HSL's review for the two methodologies reviewed by HSL.

Table 14 Criterion 6 comparison for the methodologies

Sub criteria	Score			
	SO3/Oleum		Chlorine	
	HSE	HSL	HSE	HSL
6a	2	-	2	0
6b	1	0	1	0
6c	1	-	2	-
Total	1 (4)	0	3 (5)	0

Sub criteria not used to rank the methodologies
As stated previously, sub criteria 2b, 2c, and 3d will not be used for ranking the various methodologies so have been separated from the information above. For completeness of review they are included in Table 15.

Table 15 Sub criteria not used for ranking

Sub criteria	Score			
	SO3/Oleum		Chlorine	
	HSE	HSL	HSE	HSL
2b	2	-	3	0
2c	2	-	3	3
3d	0	0	0	-

From Table 15 it can be seen that there is only one opposite response. For the methodology Chlorine, sub criterion 2b HSL believed that it could not be

used to perform an ARI calculation, while HSE indicated that it could be used in ARI calculations.

Overall conclusions of the HSL independent review
It can be seen in Table 6 that for the models Drift, Firepest and Sandra there is a good correlation between the results from the HSE review and those from the independent review by HSL. However, for the model Toxic Riskat there is a significant difference in the total score; this is mainly from criterion 1. The lower score in the HSL review was due to the use of Denz and Crunch in Toxic Riskat, which were considered to be no longer credible, and therefore many of the sub criteria were scored as 0. The order of ranking though is consistent.

It can be seen from Table 14 that for the methodologies there are significant differences in the overall scores, but the order of ranking is consistent. The reasons behind the large differences between the HSE and HSL review scores is that in the independent review by HSL the methodology described in PCAG was used as the basis of the review as well as if the methodology used a model that is considered not to be credible it was assumed to be a weakness in the methodology.

Overall, notwithstanding the differences discussed above, HSL have confidence in the HSE review, and in particular have confidence in the ranking of the models and methodologies.