A Benchmark Comparison of Level 3 Probabilistic Safety Assessment Codes: Preliminary Results from First Calculations

G.I.L. van Oudenaren[a*], J.B. Grupa[a], S. Kim[b],
C. Suzuki[c], K. Wadayama[c], A. Nagakubo[d]
[a] NRG, Petten, The Netherlands
[b] KAERI, Daejeon, Republic of Korea
[c] NRA, Tokyo, Japan
[d] JAEA, Ibaraki, Japan

Abstract: An ongoing task of the Working Group on Risk Assessment (WGRISK), Organisation for Economic Co-operation and Development (OECD), Nuclear Energy Agency (NEA) Committee on the Safety of Nuclear Installations (CSNI) is a Level 3 Probabilistic Safety Assessment (PSA) benchmark study. Six independent institutions from various countries are participating in this task, marking the first large international probabilistic accident consequence analysis benchmark study in the last 30 years.

This paper presents the first interim results of the ongoing Level 3 PSA code comparison, specifically focused on the atmospheric dispersion modelling and the radiological consequence analyses. The study is centred around various atmospheric dispersion models applied by several international parties with expertise in performing a Level 3 PSA. The preliminary results of this study encompass the radiological consequences of a severe accident involving a light-water reactor (LWR) considering a diverse array of meteorological conditions as a case study, in which a set of representative radionuclides is released into the environment. While focused on the case of a LWR, the outcomes can be deemed representative for other types of nuclear reactors and installations with similar types of radiological release characteristics.

In addition to investigating the atmospheric dispersion, several radiological health impact endpoints are assessed. These include the effective dose accumulated over the lifetime of the reference person (as defined by the International Commission on Radiological Protection (ICRP)), its constituent dose pathways (groundshine, cloudshine, inhalation and ingestion) and the conditional risk probability associated with this dose – an important metric in assessing individual risk. In addition, special interest is given to calculating a deterministic health effect, which is commonly used in determining the group risk for a large population body that has been exposed.

The benchmark study will explore how different users approach the implementation and customisation of atmospheric dispersion models, considering factors such as meteorological data inputs, release modelling and parameter choices. Additionally, the activity will address the implications of model choices within the framework of their intended application, which is the determination of radiological risks associated with the release of large amounts of radionuclides during severe accident conditions at nuclear installations.

The outcomes of this benchmark exercise are expected to contribute significantly to the refinement and standardisation of atmospheric dispersion modelling practices within a Level 3 PSA. By identifying best practices and areas for improvement, this study seeks to enhance the reliability and accuracy of radiological consequence analyses, ultimately fostering improved insight into the applicability of risk assessment. The findings, which will be presented in the final benchmark study report, aim to serve as a starting point for a reference source (for reproducibility purposes) for the nuclear risk assessment community, guiding future developments and advancements in relevant Level 3 PSA methodologies.

Keywords: Atmospheric dispersion, benchmark, Level 3 PSA, probabilistic safety assessment (PSA), radiological consequence analysis.

---

1. INTRODUCTION

1.1 Level 3 PSA

A Level 3 PSA is a useful tool in modelling the off-site radiological consequences of severe accidents involving nuclear installations. It can be used to translate/interpret the outcomes of a Level 2 PSA, which typically provides information on the radiological releases and frequencies associated with a representative set of accidents, into metrics on detrimental health effects due to the exposure to radiotoxic materials. These metrics include the risk of radiological health damage, which can be tested against legal limits prescribed by (inter)national safety standards through regulatory assessment (mostly within the licensing context).

Not all countries in which PSA studies are performed make use of a Level 3 PSA. Even fewer countries, have a(implicit) legal requirement to do so. In a recent task report of the WGRISK published by the NEA [1],a survey was undertaken resulting in an overview of whether and how a Level 3 PSA is conducted on a country by country level. The results of this survey showed that seven of the ten surveyed countries use or plan to use a Level 3 PSA, while only two countries have a legal requirement to do so.

Although there are some generally accepted standards for a Level 3 PSA, there are various (national) differences. In any case, the analysis consists of a sequence of intermediate model results feeding into the subsequent model, as shown in Figure 1. However, due to these modelling differences, it is difficult to usefully compare the outcomes of two independent Level 3 PSA codes purely based on their calculated endpoints.

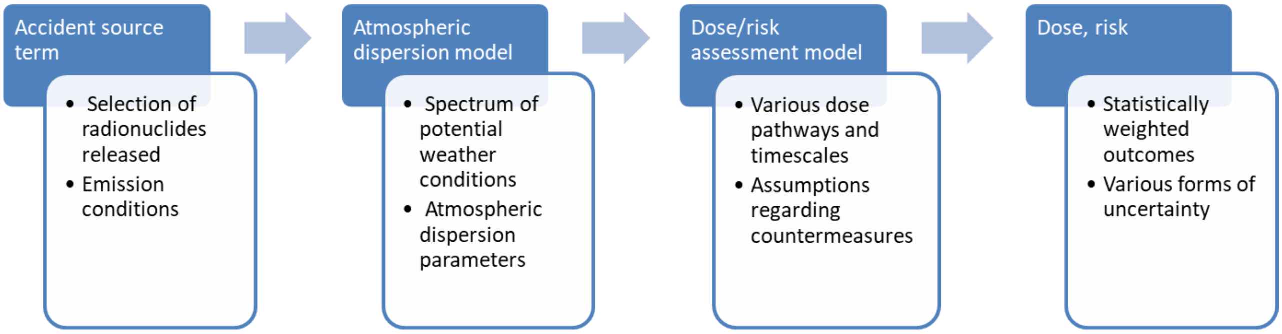

Figure 1. Main steps and dependencies in a Level 3 PSA calculation.

1.2 Purpose of the Level 3 PSA benchmark study

A Level 3 PSA benchmark exercise was established to better understand how different models applied within different Level 3 PSA codes yield dissimilar outcomes. The benchmark explores how the models, using an identical set of provided input parameters, result in a range of outcomes. For outputs with a larger variation in results, an attempt was made to provide an in-depth analysis of the causes. This knowledge will provide numerous benefits:

- An understanding of which models are used in existing Level 3 PSA methodologies;
- Insights on how intermediate outputs/results can deviate during the progression of various sub-models applied in a Level 3 PSA code;
- An understanding of which assumptions and parameter choices can be made and how these impact the outcomes.

The final result of the benchmark exercise will be a set of well-defined calculation cases with a range of expected outcomes, which can serve as a reference document for future practitioners of a Level 3 PSA which can be used to validate their results. This benchmark study will be the first of its kind after approximately 30 years, since a comparable study (also under the auspices of the OECD/NEA and the EU) was performed in 1994 [2]. Since then, many new models and codes have been developed, and insights have been gained which have been incorporated into modern Level 3 PSA methodologies. The ongoing benchmark exercise is significantly more detailed in delineating the relationship between the meteorological conditions and results. In the previous study, results were requested as complementary cumulative density functions (CCDF), a common statistical metric used to present radiological risks. The downside of this method is that large amounts

of information regarding the dependency on meteorological conditions and distance to the release point are lost, which is an explicit focus of the current study.
The participants of the ongoing benchmark study are members of the WGRISK, a forum for discussion and collaboration in the field of PSA for nuclear installations. Six members from WGRISK are participating in this benchmarking exercise: UKHSA (United Kingdom), BARC (India), KAERI (Republic of Korea), NRG (The Netherlands), JAEA and NRA (Japan).

2. Methodology

2.1 Initiation of the benchmark exercise

At the initial stage of the benchmark exercise, an assessment was made to establish how the different participants use the Level 3 PSA methodologies. This assessment consisted of a survey in which the participants were queried for the software used, which models were applied in the software, which parameter (set)s were used to supply these models with values and to establish their applicability range as well as understanding to what end their analyses are typically applied. By understanding the abilities and limitations of the different participants, the task group established a scope within which all parties would be able to operate. A concise list of the participating institutions and the codes applied by them is given below in Table 1 with a brief description of their atmospheric dispersion and stochastic risk models (although these are given as examples, many (sub)models were surveyed, which shows there is a large diversity of models in use).

Table 1. Codes used by the benchmark participants and their most important characteristics.

Participant	**KAERI**	**UKHSA**	**NRG**	**BARC**	**JAEA**	**NRA**
Software tool	MACCS 4.2	PACE 4.0	NUDOS2 (2022.3 ver.)	COSYMA (2001 ver.)	OSCAAR	MACCS2-NRA
Atmospheric dispersion model	Gaussian plume segment model	NAME 7.2	Bi-gaussian plume model	MUSEMET	3D Gaussian puff	Gaussian straight-line plume
Stochastic risk model	Linear-quadratic dose-response	ICRP103: risk factors & lethality fractions for various cancers, 1st and 2nd generation hereditary effects	Linear dose-response for solid cancers, no DDREF	Cancer incidence and mortality for various organs	Two types of models (linear and quadratic), depending on dose, dose rate, exposure time and pathway	Linear with DDREF at low dose (rate) (NUREG/CR-4691 [3])

2.2 Development of the benchmark case

Using the information gathered in Section 2.1, the task group was able to develop a suitable analysis case study. Some calculation iterations were made in order to converge on a case description that was as clearly and well-defined as possible (to prevent misinterpretation), and within the spectrum of capabilities for each participating code.

In the development of the case study, the benchmark participants strived to create a situation that was deemed representative of the expected application (effects of a large-scale early release in severe accidents involving an LWR) but which could also be widely applied to varying situations and applications. The scenario to be investigated is the following:

- A stack with a height of 20 m and a diameter of 1 m releases a plume of radionuclides at a constant rate over 1 hour, in which all meteorological conditions are constant. The emission contains no heat and is released without momentum (i.e. the stack exit velocity is approximately zero).
- The stack is present in a flat plane (other than a small defined surface roughness corresponding to light vegetation), with no surrounding buildings/structures. These choices were made to avoid modelling bias with respect to local geographic features or building effect models which could potentially affect a more realistic scenario of an existing nuclear installation.

- The release was chosen to be a set of some of the most important radionuclides in a severe LWR accident: Xe-133 (1.89 E18 Bq), I-131 (1.89 E17 Bq) and Cs-137 (1.57 E15 Bq), which would be an accident of Level 7 on the International Nuclear and Radiological Event Scale (INES). This release is inspired by the PWR-5 source term from the WASH-1400 report [4]. Xenon is released as a gas, and iodine and caesium are released as aerosols.

2.3 Calculation endpoints and input

For the scenario defined above, a spectrum of calculations is requested. The choice of calculation endpoints was based on their informative value (i.e. is this useful information to base meaningful conclusions with respect to the impact of different models used?) and the applicability (dose and risk assessment in a safety assessment context). The list of requested calculation endpoints is specified below:

- The time-integrated ground level activity concentration;
- The time-integrated deposition of radioactivity;
- The effective individual dose broken down into dose pathways;
- The lifetime stochastic risk, in particular, the increased risk of cancer;
- The deterministic fatality risk due to radiation related illness, in particular red bone marrow (RBM) syndrome.

Dose and risk quantities are calculated using a 50-year dose integration period (i.e. expected remaining lifetime) unless indicated otherwise. All requested quantities are to be calculated along the plume axis, i.e., the wind direction from the release at a fixed number of distances from the source (making it convenient to compare outcomes). Several conditions were specified to be taken into account when performing the calculations mentioned above:

- No shielding effects, either by buildings or surface roughness, for cloudshine or groundshine, should be accounted for (i.e., the reference person is permanently outside and exposed).
- No countermeasures are to be assumed during the progression and aftermath of the accident. This includes, but is not limited to, evacuation, relocation, decontamination and food bans.

One of the largest sources of variation in outcomes of the radiological consequences of an atmospheric release is the aleatoric uncertainty associated with the governing meteorological conditions at the time of the release. The source of uncertainty is eliminated by having the calculations performed for a fixed set of well-defined weather sequences. The outcomes can then be compared per weather sequence, which yields insight into how various atmospheric dispersion models behave differently under various circumstances. In addition, by not using the meteorological datasets of participating parties, potential modelling bias in atmospheric dispersion models with respect to geographically relevant or conservative weather conditions can be avoided.

Accompanying these variable parameters, there is a set of fixed parameters which are identical for each weather sequence:

- Surface roughness;
- Wind speed at 10 m height;
- Kinematic viscosity of air:
- Density of air:
- Density of dust particles in the release;
- Aerosol particle diameter;
- Dry deposition velocity;
- Scavenging coefficient (wet deposition):
- Air temperature.

Based on the variable parameters, twelve weather sequences were defined: Pasquill-Gifford stability class A – F with either 0 or 2 mm/h of rain (arbitrary choice). A schematic visualisation of the most important meteorological features, atmospheric dispersion effects and radiological consequences are shown in Figure 2.

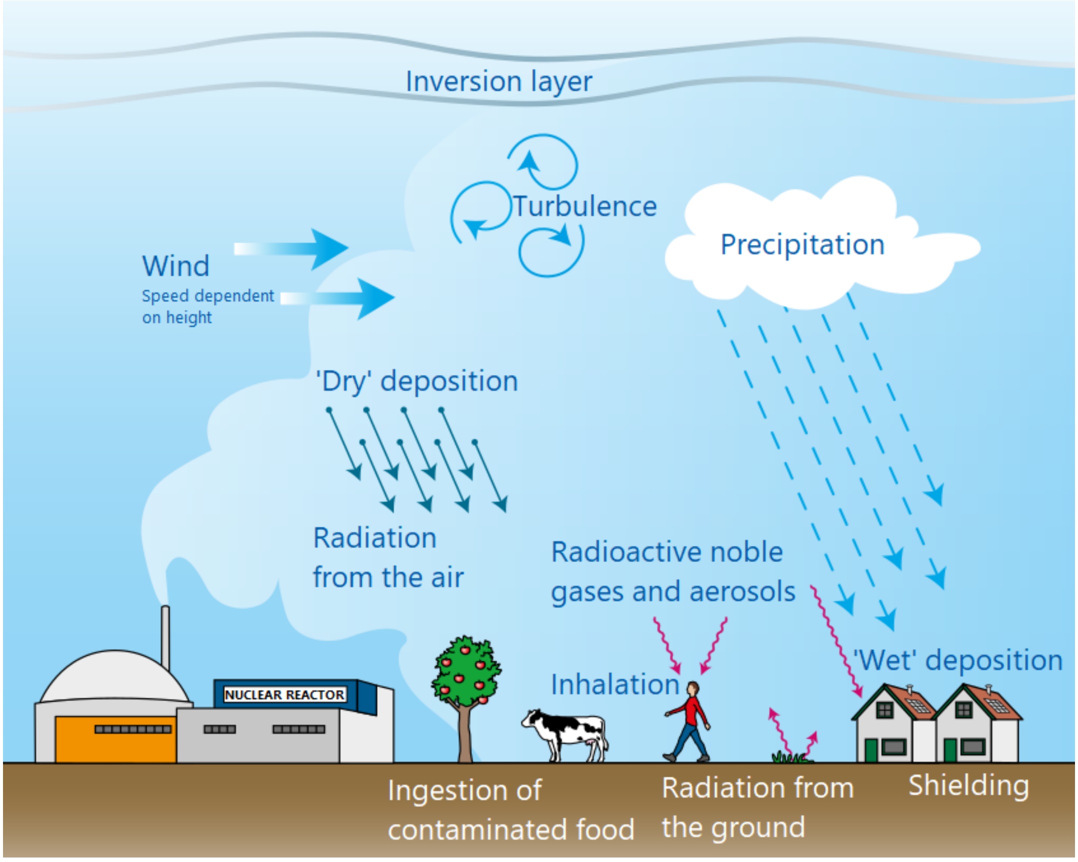

Figure 2. Schematic overview of some of the most relevant phenomena, properties and features within Level 3 PSA **[5]**.

A strict definition for each weather sequence is necessary. This entails a significant list of parameters to be defined since atmospheric dispersion models are highly complex and sensitive to many parameters (which are often implicitly assumed). The list of parameters defined within the benchmark are the following:

- Atmospheric stability class (using the Pasquill-Gifford definition), along with the Monin-Obukov length;
- Solar irradiance;
- Precipitation (rain) rate;
- Cloudiness;
- Friction velocity;
- Atmospheric boundary layer height.

2.4 Scope summary

A visual representation of how a Level 3 PSA is typically performed is given in Figure 3. In this flowchart, the information flow from one point to another is indicated by arrows. The boxes are colour-coded to indicate their status in the ongoing benchmark exercise: blue boxes are provided by the benchmark (in the case description), red boxes are out of scope (i.e. not relevant for this project), and green boxes are expected project outcomes.

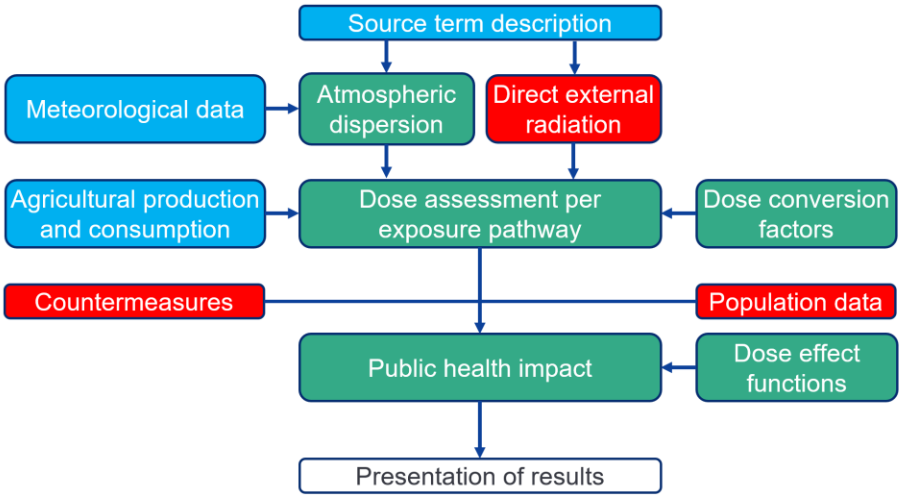

Figure 3. The steps of a Level 3 PSA and their relation to the scope of the WGRISK benchmarking exercise are indicated by their colours (blue: provided, red: excluded, green: result).

2.5 Analysis of the results

Each participant provides their calculation results to be analysed by the benchmark analysis team. The results are analysed in a sequential manner, per weather sequence. The analysis starts with the atmospheric dispersion results since these are the fundamentals on which subsequent outcomes of the calculations are based. For the entire data analysis, a comparison is made by inspecting log-log plots of the distance versus metric data as well as quantifying the spread (the largest value divided by the smallest value) at each distance.

3. Results

This section shows a selection of preliminary calculation results of the benchmark exercise. It should be noted that the outcomes are not based on the final iteration of the calculation rounds in the benchmark study. A subsequent iteration has commenced involving some slight refinements in the case description and the processing of the results of the various participants. Nevertheless, these interim results are representative enough to be suitable for a comparative analysis for the purposes of this work. Because of the preliminary nature of the results, they are anonymised (meaning they are not traceable to the corresponding participating parties).

As mentioned in Section 2, the results for atmospheric dispersion and dose/risk values are given as a function of distance to the release point along the plume axis (advection distance). In these figures, this distance is referred to as X [m].

3.1. Atmospheric dispersion results

Below, a selection of the results for the time-integrated ground level activity concentration and deposition are given for a selection of weather sequences. Generally, the results show that the spread in the results is highly dependent on the atmospheric stability class. The most unstable weather (A) shows a much smaller variation than the most stable weather (F). This is not a particularly surprising observation since the plume dimensions are highly sensitive to the stability class (more stable weather, which is typically relatively uncommon, means more compact dispersion), and different models possess a different interpretation of the stability classes.

Figure 4 shows the time-integrated ground level activity concentration as a function of distance to the release point along the plume axis for stability class A (dry) and F (wet). In class A, “dry weather”, the spread in the outcomes can be as low as 1.6, while for class F, “wet weather” it can be as high as 16 (more than an order of magnitude difference). Averaging over all data points for all weather conditions, the spread in concentration is 5.4. In the results for deposition (not shown in this work, since those outcomes are highly linked to those of the concentrations), a similar outcome is observed, albeit with smaller variations.

To note, most results show a dose reduction at a short distance in case 12. This is because the plume has not yet fully reached the ground due to the stack release height of 20 m. This effect is primarily important for more stable weather, which is why it can be observed in case 12 but not in case 1.

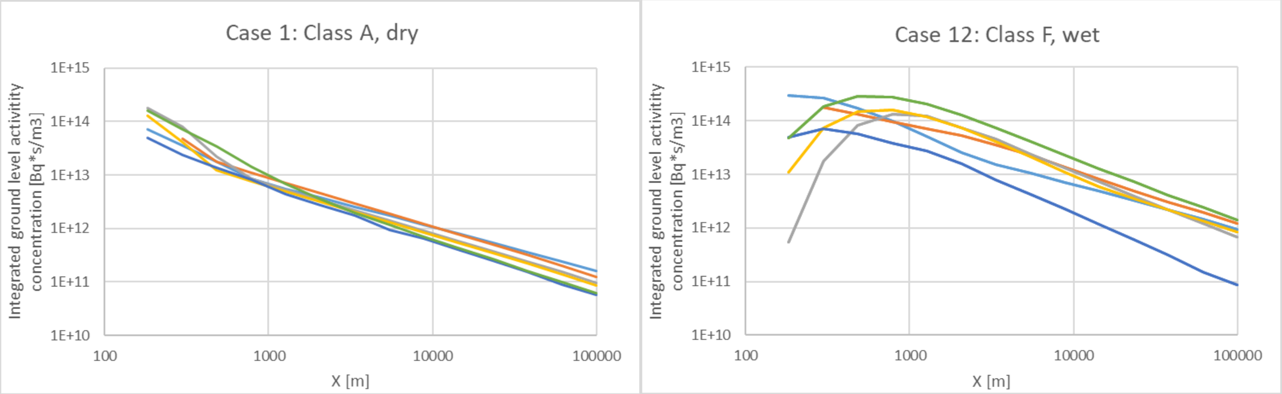

Figure 4. Time-integrated ground level activity concentration, as a function of distance to the release point along the plume axis, for stability class A (dry) and F (wet).

3.2. Dose assessment results

Figure 5 shows some of the total effective dose results. As can be observed, the dose is highly dependent on the stability class, and there is significant variation within stability classes. The first observation is an evident conclusion: the less stable the weather, the lower the concentrations, the lower the resulting doses are. The second observation has to do with two considerations: there is significant variation within the atmospheric dispersion (as can be seen in Figure 4), and the total effective dose (in this case, the sum of cloudshine, groundshine and inhalation) is an accumulation of intermediate results of individual dose paths, which exhibit varying degrees of divergence between models. It should be noted that since the total effective dose is affected by both the time-integrated ground level activity concentration and deposition, the distribution trend slightly differs from Figure 4, which only shows the concentration.

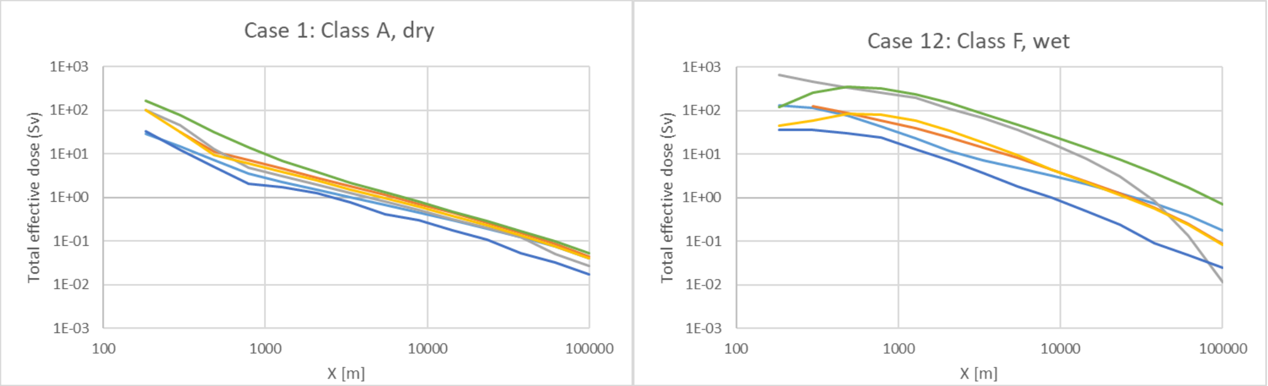

Figure 5. Total effective dose (being the sum of inhalation, groundshine and cloudshine), as a function of distance to the release point along the plume axis, for stability class A (dry) and F (wet).

The variation in individual dose pathways is exemplified in Figure 6, in which, for class A (dry) weather, the results for the inhalation and groundshine dose pathway are given. As is expected, the spread in the inhalation dose pathway (being one of the more straightforward and methodologically agreed upon doses to calculate) is very low, while for groundshine (which is, as mentioned in the previous section, prone to variation due to long-lived gamma emitters and weathering) this spread can exceed an order of magnitude.

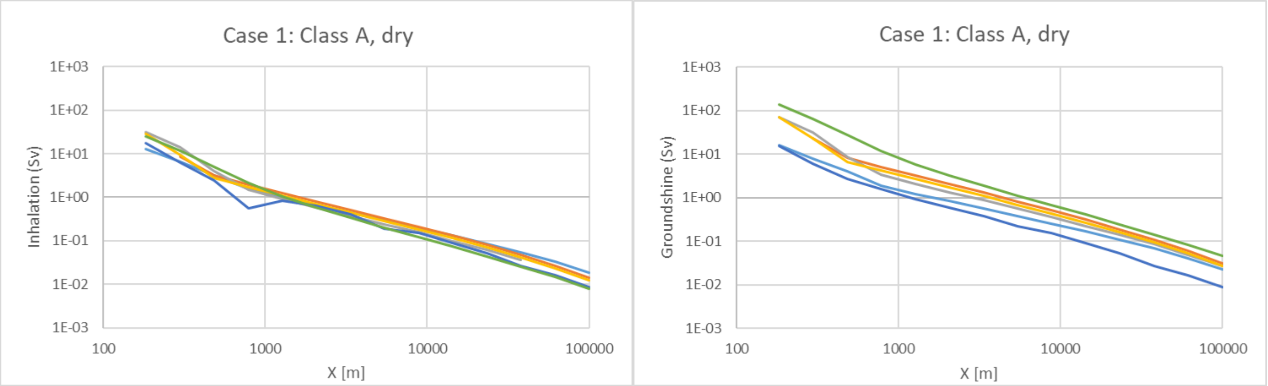

Figure 6. Inhalation and groundshine dose, as a function of distance to the release point along the plume axis, for stability class A (dry).

3.3. Risk assessment results

The progression through atmospheric dispersion and dose assessment eventually results in risk values, which are the primary figures of merit used to quantify the significance of a radiological release within the Level 3 PSA context. Figure 7 shows two exemplary results for the calculated stochastic and deterministic fatality risk (due to red bone marrow syndrome), for class F (wet) weather (note: risk cannot exceed 1 by definition). These results are interesting in several ways.

Although significant spread in the stochastic risk based on the variations in total effective dose (see Figure 4) can be expected to be observed, the contrast has increased even more (exceeding two orders of magnitude at times). This demonstrates that various models are applied to establish the relation between dose and stochastic risk, causing high degrees of inconsistency. The deterministic fatality risk value data points in Figure **7** are relatively scarce. This is because the existence of deterministic fatality risk is contingent on exceeding a threshold dose (typically in the order of multiple Gy-eq.), and the relatively small resolution of (logarithmically spaced) distance points, causing high jumps in dose and risk between subsequent steps.

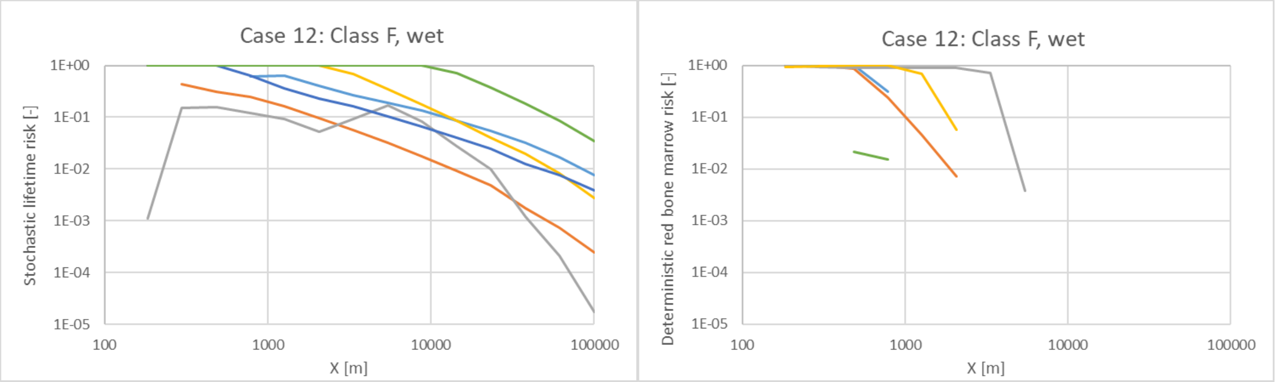

Figure 7. Stochastic and deterministic (red bone marrow) risk, as a function of distance to the release point along the plume axis, for stability class F (wet).

To understand the relation between the RBM dose and risk, a single scatter plot containing all the data points for the different weather sequence runs was generated, given in Figure 8. In this graph, it is evident that all participants calculate the risk similarly (using a two-parameter Weibull function); however, the parameter values (LD_{50} and threshold dose) describing the risk are different. Additionally, this also highlights the high spread in the calculated RBM dose values.

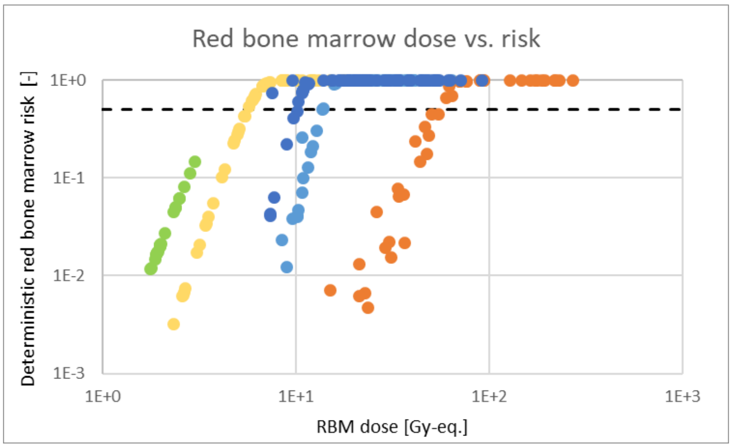

Figure 8. Relation between the red bone marrow dose and risk for the different participants. The dashed line is equal to 50% risk (corresponding to the LD_{50} dose).

Note: not all participants were able to calculate the red bone marrow dose in addition to the corresponding risk due to code limitations, so Figure 7 and Figure 8 *only contain 5 sets of data.*

4. Discussion: Refinement of the analysis results

Based on the initial analysis, it was observed that in some cases, a significant spread in the atmospheric dispersion results (around an order of magnitude) can be seen. As a result, the calculated values that are dependent on these outcomes (i.e., the dose depends on the time-integrated ground level activity concentrations and depositions) will also see significant changes. For that reason, the choice was made in the dose result analysis to look both at the outcomes "*as is*" and "normalised" by dividing them by the respective atmospheric dispersion metric most applicable (e.g., time-integrated deposition for groundshine, time-integrated ground level activity concentration for inhalation, etc.). By performing this comparison, the dose assessment model is 'isolated' from the atmospheric dispersion, making it possible to compare how differently a dose is calculated. A similar comparison is made for the stochastic and deterministic results in order to highlight the dose-risk relation.

Preliminary results of the first round of calculations show that some dose pathways are particularly prone to exhibiting large variations in the outcome between the different participating Level 3 PSA codes. Further investigation and discussion resulted in the conclusion that it is necessary to perform a separate calculation and analysis for two specific dose pathways under the given conditions:

- Groundshine, considering only a 24-hour dose integration period (compared to a 50-year period for the regular dose assessment). This dose pathway is prone to variation due to the fact that the benchmark task is dealing with a relatively long-lived gamma emitter (Cs-137), while also having to account for weathering effects (either mechanical or leaching effects occurring in the top-layer of the soil). The weathering effect difference could not be sufficiently excluded in the case description, so a compromise was made to perform an additional calculation on a time scale in which weathering is not a significant effect (24 h).
- Ingestion, considering exclusively the consumption of leafy greens and exclusively the consumption of milk. Radiological food chain models are highly complex and diverse and are rampant with assumptions, cultural biases and implicit parameter and modelling dependencies, which make their outcomes notoriously difficult to grasp and compare. By eliminating the variability of diet, consumption rate and production location (i.e., fixed yearly consumption of a single foodstuff grown/produced at the dose assessment reference location), the participants strived to converge to a result which should be as comparable as possible.

The aforementioned analysis refinement is a part of the ongoing work in the benchmark study. The results of this additional in-depth analysis will be reported extensively in an upcoming report covering the benchmark exercise comprehensively.

5. Conclusion

The current progress within this Level 3 PSA benchmark task has yielded some insights into model functionality, areas of agreement and disagreement among models, and the implications for their practical application.

For instance, the interim results of the ongoing task given in this paper (on the atmospheric dispersion, dose and risk assessment) show a high degree of variability in the agreement in calculation outcomes, depending on the calculation metric and the prevailing meteorological conditions. While variations were expected to occur, these results do prompt the need for a subsequent inquiry to understand the root causes better.

Further work will focus on refining the outcomes. This refinement involves the thorough analysis of a much broader spectrum of intermediate results than presented in the current work. The final product of the benchmark exercise will feature both a comprehensive set of well-defined meteorological sequences to be applied within our case definition as well as a vast set of reference data as provided by the benchmark participants and analysts.

In addition, specific efforts will be made to provide extensive analysis of the relevant metrics displaying the largest degrees of variation. Within this analysis, a quantification of the relationship between atmospheric and dose results, the dose-risk relation models as well as qualitative descriptions of the modelling specifications will be provided in order to supply future users of the benchmark study with adequate reference and guidance to understand the relation between their outcomes and the benchmark exercise. This relationship analysis, which is the starting point for qualitative conclusions on the modelling and parameter choices, will be one of the focus points of the final product.

Acknowledgements

The authors recognise the support and guidance of the OECD/NEA CSNI WGRISK under whose authority this task is being conducted.
Moreover, NRG acknowledges the financial support of the Dutch Ministry of Economic Affairs and Climate, sponsoring the PIONEER R&D program within parts of this work are performed.
The work part of KAERI was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT: Ministry of Science, ICT) (No. RS-2022-00144405).

References

[1] NEA, "Status of Practice for Level 3 Probabilistic Safety Assessments, NEA/CSNI/R(2018)1," OECD/NEA, Paris, 2018.

[2] EC, "Probabilistic Accident Consequence Assessment Codes, Second International Comparison. A Joint Report by the OECD Nuclear Energy Agency and the European Commission, EUR 15109 EN," European Comission, Paris, 1994.

[3] H.-N. Jow, J. L. Sprung, J. A. Rolistin and L. T.Ritchie, "MELCOR Accident Consequence Code System (MACCS) [NUREG/CR-4691]," Sandia National Laboratories, Albuquerque, 1990.

[4] U.S. NRC, "An Assessment of Accident Risks in U.S. Commercial Nuclear Power Plants [NUREG-75/014 (WASH-1400)]," Washington D.C., 1975.

[5] NRG, "Onderzoeksjaarverslag NRG 2021," NRG, Petten, 2022.