Exploring discrepancies between consequence analysis software: from validation to common QRA initiating events

Lorenza Saturnino1 , Alba Àgueda2 , Miguel Muñoz1 , Eulàlia Planas2,\*

1NOVOTEC - Process Safety and Major Accidents Department. Cerdanyola del Vallès, 08193, Spain;

2Centre for Technological Risk Studies (CERTEC), Universitat Politècnica de Catalunya ⋅ Barcelona Tech, Barcelona East School of Engineering (EEBE). Avinguda Eduard Maristany, 16, 08019 Barcelona, Spain

\*eulalia.planas@upc.edu

1. Introduction

Quantitative risk analysis (QRA) consists of a set of methodologies for estimating the risk posed by a given system in terms of human loss or, in some cases, economic loss (Casal, 2008). Throughout QRA, varying degrees of uncertainty influence the reliability of the results.

Consequence analysis, which models different loss-of-containment scenarios, is especially uncertain due to the complexity of physical phenomena like dispersion, fires, and explosions. These processes are highly nonlinear and governed by multiple interacting variables, necessitating the use of mathematical models that approximate these complex systems. The approximations introduced during modelling represent a form of knowledge-based uncertainty (Abrahassom, 2002), which arises from the inherent limitations in our understanding and representation of these hazardous events.

Among the various modelling approaches, integral models are the most widely used in QRA due to their low computational demands. These models are typically based on pseudo-one-dimensional representations, in which the system is analysed along a primary spatial axis. Unlike one-dimensional models, three-dimensional effects are incorporated in a simplified way, often through correction factors or semi-empirical correlations calibrated using field test data.

Several available software tools —such as PHAST, EFFECTS, and ALOHA— implement these integral models, each with its own algorithms, tuning parameters, and assumptions.

Numerous studies have compared the predictive capabilities of different modelling tools across a variety of release scenarios (Bernatik et al., 2011; Bubbico & Mazzarotta, 2008; Hanna et al., 1993; Mazzola et al., 2021). In some instances (Hanna et al., 1993; Mazzola et al., 2021), simulation results were benchmarked against experimental data that were independent of those used in model development—providing a basis for model validation and enabling an assessment of predictive accuracy. These validation exercises revealed discrepancies between tools, although such differences generally fell within an acceptable range, supporting their practical use in QRA. However, in other studies (Bernatik et al., 2011; Bubbico & Mazzarotta, 2008), suitable experimental datasets were unavailable, and comparisons were limited to the outputs generated by different software. While this approach offers insight on the range of possible predictions, it does not allow for conclusions about accuracy. Moreover, a critical concern remains: it is unclear whether the acceptable discrepancy between software tools observed during validation persists when models are applied to QRA scenarios that deviate from the controlled conditions under which they were originally validated. This issue raises the possibility that discrepancies between software predictions may increase in real-world applications, thereby amplifying uncertainty in consequence estimation.

This study investigates whether inter-software discrepancies become more pronounced as scenarios move from controlled validation cases to initiating events commonly used in QRA, potentially leading to an increased divergence in predictions and a loss of accuracy in consequence estimation. By comparing ammonia dispersion predictions from PHAST, EFFECTS, and ALOHA across both validated and common QRA scenarios, this work seeks to evaluate the stability of inter-software variability and its implications for the reliability of QRA outcomes.

2. Methods

This study compares ammonia dispersion predictions from three software tools—PHAST v9.0, EFFECTS v12.3, and ALOHA v5.4.7—across both validation and QRA scenarios.

2.1 Definition of the baseline discrepancy and the QRA scenario-related discrepancy

This study investigates whether inter-software discrepancies become more pronounced when moving from controlled experimental scenario used for validation to the initiating events typically considered in QRA. The core objective is to assess whether the relative agreement observed during validation holds under more generalized conditions, or whether divergence increases—potentially compromising the reliability of consequence estimation.

To quantify this effect, a baseline discrepancy is defined as the relative difference between software outputs when modelling a scenario for which the experimental data are available. This baseline reflects the expected level of variability between tools that is considered sufficiently accurate for that specific case due to prior validation. A comparable metric, referred to as the QRA scenario-related discrepancy, is calculated for typical QRA scenarios for which experimental data are not available.

The discrepancy $D\_{ij}$ between two software tools *i* and *j* is calculated as represented in equation (1).

|  |  |
| --- | --- |
| $$D\_{ij}=\frac{(R\_{i}-R\_{j})}{(R\_{i}+R\_{j})}$$ | (1) |

Where $R\_{j}$ and $R\_{i}$ are the results predicted by the software *i* and *j*, respectively. This formulation is applied both to the experimental validation case (yielding the baseline discrepancy) and to QRA scenarios (yielding the QRA scenario-related discrepancy).

The discrepancy $D\_{ij}$, as defined in this work, ranges from –1 to 1. A value of 0 indicates perfect agreement between the software compared, while positive or negative values reflect which software produces higher results. Values close to zero suggest high consistency, whereas larger absolute values (e.g., above 0.2) highlight significant divergence.

By comparing the QRA scenario-related discrepancy to the baseline discrepancy, it is possible to assess whether the predictive agreement between software tools deteriorates when simulating QRA initiating events. If the discrepancy observed in QRA scenarios exceeds the baseline one, this suggests that the acceptable variability established during validation is no longer maintained. Such an increase in discrepancy may indicate a loss of predictive reliability when the models are applied outside the controlled conditions under which their accuracy was originally verified.

2.2 Description of the experimental and QRA scenarios simulated

As reference case for the baseline discrepancy, the Desert Tortoise field experiment was selected. This experiment involved the horizontal release of pressurized liquid ammonia, with the release direction nearly aligned with the prevailing wind direction (Goldwire et al., 1985). The experimental campaign consisted of four tests at different spill rates. The test chosen to be the reference case in this study is Desert Tortoise 1 (DT1). The data describing the release conditions, weather parameters, and measured values were obtained from the SMEDIS database (UK Health and Safety Executive et al., 2001). This database was developed during the European project with the same name, which aimed to establish a methodology for evaluating dense gas atmospheric dispersion models. The main parameters defining the DT1 test conditions are presented in Table 1. The measured variable was the concentration of ammonia, recorded at 29 different locations within the test field.

Table 1: Desert Tortoise 1 test parameters (Goldwire et al., 1985; UK Health and Safety Executive et al., 2001).

|  |  |
| --- | --- |
| Parameter  | Value |
| Substance | Ammonia |
| Pressure [barg] | 12.6 |
| Temperature [ºC] | 21.5 |
| Spill mass [kg] | 10200 |
| Release point (x,y,z) [m] | (0, 0, 0) |
| Release direction | Horizontal, 45º from N |
| Release duration [s] | 126 |
| Ambient pressure [bara] | 0.909 |
| Ambient temperature [ºC] | 29.3 |
| Soil temperature [ºC] | 31.7 |
| Relative humidity  | 0.132 |
| Surface roughness [m] | 0.003 |
| Average wind speed [m/s] | 7.42 |
| Average wind direction | 223.7º from N |
| Reference height for wind [m] | 2 |
| Stability class | D |
| Cloud cover | 0.01 |

The other two simulated scenarios are initiating events defined in the BEVI guideline (Bevi, 2009), the Dutch standard for conducting QRA. For pressurized vessels, the two continuous release scenarios considered as initiating events in risk assessments are:

* A continuous release through a 10 mm diameter orifice, with the release duration defined either by the time required to fully empty the tank or a maximum of 1800 seconds.
* A continuous release of the entire tank contents over a fixed duration of 10 minutes.

The storage parameters used in the simulation software were selected based on the study by Orozco et al. (2019), in which ALOHA was employed to model an actual accidental release from ammonia storage tanks in an industrial setting. Furthermore, each scenario was modelled two times, with varying wind and atmospheric stability class parameters. Table 2 summarizes the storage parameters and meteorological conditions adopted to simulate the BEVI initiating events. For consistency with the DT1 case, concentration levels at various locations were compared across the three software tools.

Table 2: BEVI scenarios simulation parameters (Orozco et al., 2019).

|  |  |
| --- | --- |
| Parameter  | Value |
| Substance | Ammonia |
| Pressure [barg] | 11.75 |
| Temperature [ºC] | 15 |
| Mass in the tank [kg] | 120,000 |
| Release point (x,y,z) [m] | (0, 0, 0) |
| Release direction | Horizontal |
| Ambient pressure [atm] | 1 |
| Ambient temperature [ºC] | 15 |
| Soil temperature [ºC] | 15 |
| Relative humidity  | 0.7 |
| Surface roughness [m] | 0.25 |
| Average wind speed [m/s] and stability class | 1.5F-5D |
| Reference height for wind [m] | 10 |

2.3 Models adopted in the different software

To model the baseline and QRA scenario-related discrepancies, different software tools were employed, each utilizing its own modelling approach. The selection of discharge and dispersion models in PHAST, EFFECTS, and ALOHA was based on the technical recommendations provided by the software developers. This section outlines the selection and application of these models within each software, emphasizing the key choices and assumptions made for each scenario.

In PHAST, the discharge process is modelled using the DISC module, which simulates the flow from stagnation conditions to the orifice and subsequently from the orifice to atmospheric conditions (DNV, 2023a). Following this, the Unified Dispersion Model (UDM) utilizes the output from DISC to simulate atmospheric dispersion. For the discharge phase, PHAST allows users to choose among several modelling options for the expansion from stagnation to the orifice, from the orifice to the atmosphere, and for droplet formation.

During the validation of PHAST, a specific combination of these models was identified as the most accurate for simulating the DT1 scenario (DNV, 2023b). It is important to note that this combination is not the default configuration used by the software for similar release scenarios. Therefore, DT1 was simulated using the model combination recommended in PHAST validation study, whereas the BEVI scenarios were modelled using the software's default assumptions.

The models used in EFFECTS to simulate the scenarios under investigation were also selected based on the recommendations provided in the user manual (GEXCON, 2024). No specific adjustments or parameter settings were proposed for the DT1 scenario, as validation of the software against this particular experiment —performed using EFFECTS v10.1— showed an underprediction of the concentrations of approximately 50% (Ruiz Pérez, 2017). A more recent validation, introduced in EFFECTS v12.5, is still very recent and has not yet been made publicly available.

In ALOHA, all models employed were the default ones, as the software does not provide the user with the option to select or modify modelling approaches.

3. Results and discussion

This section focuses on the comparison of results using PHAST as the reference software. However, Figure 1 also provides a visualization of both the baseline discrepancy and the QRA scenario-related discrepancies between EFFECTS and ALOHA.

In the DT1 case, ALOHA and EFFECTS showed absolute baseline discrepancies of approximately 0.3 and 0.5, respectively, relative to PHAST, generally underpredicting ammonia concentrations.

In the BEVI scenarios, the discrepancies varied depending on the type of release and meteorological conditions. For the BEVI scenario involving a 10 mm orifice release, the discrepancy between PHAST and ALOHA remained similar to the baseline level. However, in the 10-minute release scenario under 1.5F weather conditions, the discrepancy between PHAST and ALOHA increased significantly, reaching values around 0.7.

The discrepancies between PHAST and EFFECTS across both BEVI scenarios ranged between 0.5 and 0.6, comparable to the baseline discrepancy but slightly higher.

While ALOHA predicted higher concentrations than PHAST in the BEVI scenarios, EFFECTS maintained the trend observed in the DT1 case, continuing to underpredict concentrations relative to PHAST.

The differences suggest a change in predictive agreement when moving from controlled validation conditions to more generalized QRA scenarios. According to the framework outlined in the methodology, discrepancies exceeding the baseline imply that the acceptable variability established during validation is no longer maintained. The observed increases in discrepancy indicate a potential loss of predictive reliability.



Figure 1: Comparison of baseline discrepancy with QRA scenario-related discrepancy.

4. Conclusions

This study examined the consistency of ammonia dispersion predictions from three widely used tools implementing integral models —PHAST v9.0, EFFECTS v12.3, and ALOHA v5.4.7— across both controlled validation scenarios and initiating events used in QRA. While the tools demonstrated relatively consistent performance under validation conditions, notable differences emerged when applied to typical QRA scenarios.

The results suggest that model agreement observed under validation conditions may not consistently extend to broader QRA applications, particularly where input data are less detailed or more uncertain. The discrepancy observed among tools in QRA contexts may represent a source of uncertainty not fully addressed in typical validation exercises. While these findings point to the potential value of considering inter-model differences within uncertainty assessments, further investigation is needed. This study represents a preliminary step toward a more comprehensive understanding of model behaviour in QRA settings. Future work will explore additional substances, a wider range of scenarios, and a more detailed examination of model structures and assumptions.

Acknowledgments

L. Saturnino works within the framework of Project 101119358, ‘PROSAFE’, funded by the Marie Skłodowska-Curie Actions programme, HORIZON-MSCA-2022-DN-01. A. Àgueda is a Serra Hunter fellow. This work was also partly funded by Grant PID2023-150607OB-I00 funded by MICIU/AEI/ 10.13039/501100011033 and by ERDF/EU.

References

Abrahassom, M. (2002). *Uncertainty in Quantitative Risk Analysis—Characterisation and Methods of Treatment.* Fire Safety Engineering and Systems Safety.

Bernatik, A., Senovsky, P., & Pitt, M. (2011). LNG as a potential alternative fuel – Safety and security of storage facilities. *Journal of Loss Prevention in the Process Industries*, *24*(1), 19–24. https://doi.org/10.1016/j.jlp.2010.08.003

Bevi. (2009). Reference Manual Bevi Risk Assessments, v. 3.2. *National Institute of Public Health and the Environment (RIVM)*, 189.

Bubbico, R., & Mazzarotta, B. (2008). Accidental release of toxic chemicals: Influence of the main input parameters on consequence calculation. *Journal of Hazardous Materials*, *151*(2–3), 394–406. https://doi.org/10.1016/j.jhazmat.2007.06.002

Casal, J. (2008). *Evaluation of the Effects and Consequences of Major Accidents in Industrial Plants. Industrial Safety Series* (Vol. 8). Elsevier.

DNV. (2023a). *DISC Discharge Scenarios Theory—PHAST 9.0 Technical Documentation*.

DNV. (2023b). *UNIFIED DISPERSION MODEL Validation—PHAST 9.0 Technical Documentation*.

GEXCON. (2024). *EFFECTS v12—User and reference manual*.

Goldwire, H. C., McRae, T. G., Johnson, G. W., Hipple, D. L., Koopman, R. P., McClure, J. W., Morris, L. K., & Cederwall, R. T. (1985). *Desert Tortoise Series Data Report 1983 Pressurized Ammonia Spills*. Lawrence Livermore National Laboratory.

Hanna, S. R., Chang, J. C., & Strimaitis, D. G. (1993). Hazardous gas model evaluation with field observations. *Atmospheric Environment. Part A. General Topics*, *27*(15), 2265–2285. https://doi.org/10.1016/0960-1686(93)90397-H

Mazzola, T., Hanna, S., Chang, J., Bradley, S., Meris, R., Simpson, S., Miner, S., Gant, S., Weil, J., Harper, M., Nikmo, J., Kukkonen, J., Lacome, J.-M., Nibart, M., Björnham, O., Khajehnajafi, S., Habib, K., Armand, P., Bauer, T., … Ek, N. (2021). Results of comparisons of the predictions of 17 dense gas dispersion models with observations from the Jack Rabbit II chlorine field experiment. *Atmospheric Environment*, *244*, 117887. https://doi.org/10.1016/j.atmosenv.2020.117887

Orozco, J. L., Van Caneghem, J., Hens, L., González, L., Lugo, R., Díaz, S., & Pedroso, I. (2019). Assessment of an ammonia incident in the industrial area of Matanzas. *Journal of Cleaner Production*, *222*, 934–941. https://doi.org/10.1016/j.jclepro.2019.03.024

Ruiz Pérez, S. (2017). *TNO Dense Gas Model Verification and Validation*.

UK Health and Safety Executive, Cambridge Environmental Research Consultants, & Electricité de France. (2001). *Scientific Model Evaluation of dense gas DISpersion models (SMEDIS) database—Excel spreadsheets version 2.4*.