Proc. of the European Safety and Reliability Conference (ESREL 2026)
Edited by José C. Matos, Paulo B. Lourenço, Michael Beer, Daniel Oliveira, Edoardo Patelli, Hélder S. Sousa, Rui A. Silva and Monica Santamaria

doi: 10.3850/ESREL2026061419_esrel26-p26323-cd

Impact of modeling tool choice on ERPG safety zones for ammonia releases

Lorenza Saturnino
NOVOTEC - Process Safety and Major Accidents Department. Cerdanyola del Vallès, 08193, Spain.
E-mail: lorenza.saturnino@novotec.es

Alba Àgueda
Centre for Technological Risk Studies (CERTEC), Universitat Politècnica de Catalunya - Barcelona Tech, Barcelona East School of Engineering (EEBE). Barcelona, 08019, Spain. E-mail: alba.agueda@upc.edu

Miguel Muñoz
NOVOTEC - Process Safety and Major Accidents Department. Cerdanyola del Vallès, 08193, Spain.
E-mail: miguel.munoz.m@novotec.es

Eulàlia Planas
Centre for Technological Risk Studies (CERTEC), Universitat Politècnica de Catalunya - Barcelona Tech, Barcelona East School of Engineering (EEBE). Barcelona, 08019, Spain. E-mail: eulalia.planas@upc.edu

Accurate prediction of the consequences of accidental releases of hazardous substances is essential for defining safety distances and emergency planning zones. Several software tools formally recognized by regulatory frameworks are available for use in safety studies. However, they often rely on different theoretical formulations and simplifying assumptions, which may lead to discrepancies in the predicted hazard footprints for identical release scenarios. This study compares toxic dispersion predictions from PHAST v9.0 and EFFECTS v12.5 for representative release scenarios under 1.5F and 5D meteorological conditions. The analysis focused on discrepancies between Emergency Response Planning Guidelines (ERPGs) concentration footprints at 1 m above the ground level, evaluated using similarity and distance metrics, including Jaccard and Dice indices, as well as Hausdorff and Chamfer distances. Results showed substantial differences under default software configurations in both source-term representation and predicted footprints, with Jaccard indices ranging from 0.35 to 0.69 and maximum downwind discrepancies exceeding 1.7 km in some cases. Harmonizing the representative release rate reduced part of this discrepancy, increasing footprint similarity by up to 10% for the 5D condition. The findings emphasize the need to raise awareness among practitioners and regulators regarding variability arising from tool choice and underlying modeling assumptions. Quantifying these differences supports transparency, consistency in safety studies, accurate safety zone definition, and the harmonization of risk assessment practices.

Keywords: Toxic gas dispersion, Integral models, PHAST, EFFECTS, Similarity metrics

1. Introduction

Ammonia is a hazardous substance widely used in the chemical industry and is increasingly considered a potential energy vector in low-carbon energy systems due to its high energy density and the feasibility of its storage and transportation in compressed or refrigerated form (Cardoso et al., 2021). However, its significant toxicity requires careful evaluation of the consequences of accidental releases. When ammonia is stored and handled as a pressurized liquefied gas, estimating the source term associated with accidental releases remains one of the most challenging tasks in dispersion modeling (Hanna et al., 2022), as releases often involve flashing two-phase jets and transitions from dense-gas to passive dispersion.

Industrial practitioners and regulatory authorities frequently employ specialized commercial software tools for land-use planning and quantitative risk assessment (QRA) in the oil, gas, and chemical sectors (Gant et al., 2025). The dispersion modules in these tools most commonly rely on integral models, which enable rapid computation but assume simplified terrain representations, such as flat topography or homogeneous surface roughness. These models solve scalar conservation equations over the control volume defined by the cloud boundary, achieving computational efficiency at the expense of detailed physical representation (Lacome et al., 2023). Within the software frameworks, the dispersion models are coupled with submodels for release characterization, flashing behavior, droplet formation, rainout, pool evaporation, and transitions between jet, heavy-gas, and passive dispersion regimes.

Despite their widespread use, the choice of modeling tool may influence the outcomes of consequence analyses (Park et al., 2025). Differences in theoretical formulations, submodel implementations, thermodynamic assumptions, and default parameterizations lead to discrepancies in predicted hazard footprints, potentially resulting in either conservative or non-conservative estimates for toxic exposure.

Previous studies have largely focused on benchmarking dispersion models against experimental datasets, with the aim of quantifying predictive accuracy for specific release and dispersion conditions (Gant et al., 2025; Ruiz Pérez, 2017; Witlox et al., 2023). These validation efforts typically rely on point-wise concentration comparisons using well-established statistical performance indicators, including Fractional Bias, Geometric Mean Bias, Normalized Mean Square Error, and Geometric Variance. However, such metrics do not capture differences in the shape and extent of concentration footprints.

Concentration footprints define the geographic area where contaminant concentration exceeds regulatory thresholds, such as Emergency Response Planning Guidelines (ERPGs), Acute Exposure Guideline Levels (AEGLs), or Protective Action Criteria (PACs). Comparing two-dimensional footprints therefore requires alternative classes of metrics, including similarity indices (e.g., Jaccard index and Dice score) and distance-based measures (e.g., Hausdorff and Chamfer distances), which quantify both spatial overlap and geometric differences between contours.

This study investigates the discrepancy in ammonia dispersion footprints predicted by two widely used consequence modeling tools, PHAST v9.0 and EFFECTS v12.5. ERPG concentration contours obtained under representative release scenarios and meteorological conditions are compared using geometric similarity and distance metrics to quantify model-driven variability relevant to consequence assessment and emergency planning. The analysis evaluates consistency between model outputs under comparable conditions, without addressing experimental validation.

2. Methodology

The simulated system was based on the configuration presented in Orozco et al. (2019) and consisted of a pressurized vessel storing 120 tons of liquefied ammonia at 12.75 bar and 15 °C. Two meteorological conditions were analyzed: stability class F with 1.5 m/s wind and class D with 5 m/s wind. In both cases, the ambient temperature was 15 °C and relative humidity 70 %.

Release scenarios were selected according to the Purple Book recommendations for pressurized vessels (RIVM, 2005a). Two cases were analyzed: a continuous release through a 10 mm diameter orifice and a catastrophic release of the entire vessel inventory over 10 minutes. In all simulations, the release point was located 1 m above ground level (a.g.l.).

The toxic thresholds used for contour extraction were the ERPG levels for ammonia shown in Table 1. ERPG-3 is particularly relevant because it approximates lethality thresholds (Currie et al., 2024) and can therefore be related to risk metrics. PHAST and EFFECTS use different default ERPG-3 values (750 ppm and 1500 ppm, respectively). For consistency, PHAST ERPG thresholds were updated to the latest recommended values prior to simulation, while EFFECTS already uses these values by default.

The first simulation run ($S1$) used the default modeling assumptions implemented in each software tool. For the 10 mm leak, PHAST represents the release as a steady-state discharge at the peak mass flow rate, whereas EFFECTS models a time-dependent release and applies the second-20% approximation recommended in the Purple Book (RIVM, 2005a). As a result, the representa-

Table 1. ERPG values for ammonia.

Threshold	Value [ppm]
ERPG-1	25
ERPG-2	150
ERPG-3	1500

Source: AHIA (2016)

tive release rate for the 10 mm scenario differed between the tools, with PHAST using 2.02 kg/s and EFFECTS using 1.38 kg/s.

Because release rate strongly affects dispersion outcomes, a second simulation run ($S2$) was conducted to isolate differences attributable to the dispersion models themselves. In $S2$, EFFECTS was configured to use the peak mass flow rate for the 10 mm leak, thereby aligning the source term with PHAST. No changes were applied to the 10-minute release scenario, as both tools already model this case as a prescribed mass release. The key differences in configurations for both runs are summarized in Table 2.

Contour coordinates were exported from each software package prior to metric evaluation. PHAST contours were extracted from Excel output files (Cartesian x-y coordinates), while EFFECTS contours were exported as KML files and converted from geographic to planar coordinates. To eliminate discretization bias, all contours were resampled to a uniform 10 m spatial resolution using linear interpolation. All analyses were performed in MATLAB R2024b.

Discrepancies between PHAST and EFFECTS concentration footprints were quantified using the Dice score, Jaccard index, Hausdorff distance, and Chamfer distance. These metrics evaluate spatial overlap and contour geometry, with the areas enclosed by the ERPG contours denoted as A and B, and their boundary points as a and b (Table 3).

3. Results

In this section, results obtained under default software settings are labeled $S1$, whereas those using harmonized release rates are labeled $S2$. The results are presented for both release scenarios and meteorological conditions.

Figure 1 and Figure 2, together with Table 4 and Table 5, present the maximum ERPG footprints at 1 m a.g.l. and the corresponding similarity metrics for all scenarios and both software tools in the $S1$ run. Figure 3 and Table 6 refer instead to the $S2$ run.

3.1. *Results Using Default Software Settings (S1)*

Under the 1.5F condition, EFFECTS consistently predicted longer maximum downwind distances than PHAST for both release scenarios. Conversely, under 5D, EFFECTS generally produced shorter downwind distances, except for the ERPG-2 and ERPG-3 contours of the 10-minute release scenario.

With respect to the footprint area, PHAST generally produced larger total affected zones. This behavior is attributable to the broader crosswind spread of the maximum footprints generated by PHAST, even when its predicted downwind distances were shorter than those obtained with EFFECTS (Figures 1 and 2).

The similarity metrics accounting for the area of the footprints (D and J in Tables 4 and 5) indicate moderate agreement between the two software tools. The highest similarity values were consistently observed for the ERPG-3 contours, which correspond to concentrations closest to lethality thresholds, implying that the results of the two software tend to converge when predicting the most hazardous regions. Jaccard index values ranged from approximately 0.35 to 0.69, with higher similarity under the more stable 1.5F condition. This indicates that the degree of agreement between PHAST and EFFECTS depends also on the meteorological conditions considered.

Regarding distance metrics, Hausdorff Distance (HD) corresponded in nearly all cases to the difference in maximum downwind distances between PHAST and EFFECTS contours. Consequently, in this case, HD represents the discrepancy that will be obtained when defining the emergency zones. The Chamfer Distance (CD) was consistently smaller, indicating that although localized deviations in plume reach were large, the average spatial separation between the contours was substantially lower.

Table 2. Configuration differences between the $S1$ and $S2$ simulation runs.

Scenario	*S1*		*S2*	
	EFFECTS	**PHAST**	**EFFECTS**	**PHAST**
10 mm leak	Second-20% approximation; default ERPG-3	Peak mass flow; standardized ERPG-3	Peak mass flow; default ERPG-3	No changes
10 min release	Second-20% approximation; default ERPG-3	Peak mass flow; standardized ERPG-3	No changes	No changes

Note: 'No changes' indicates that the same configuration as in $S1$ was used.

Table 3. Similarity and distance metrics used for comparing ERPG contours.

Metric	Formula	Description
Dice score (D)[a]	$D = \frac{2\|A \cap B\|}{\|A\| + \|B\|}$	Measures spatial overlap between two areas; ranges from 0 to 1.
Jaccard index (J)[a]	$J = \frac{\|A \cap B\|}{\|A \cup B\|}$	Stricter overlap metric; penalizes mismatches more strongly.
Hausdorff Distance (HD)[b]	$HD(A,B) = \max_{a \in A}\{\min_{b \in B} \|\|a-b\|\|\}$	Captures maximum point wise deviation; sensitive to outliers.
Chamfer Distance (CD)[c]	$CD(A,B) = \frac{1}{2}\left(\frac{1}{\|A\|}\sum_{a \in A}\min_{b \in B}\|\|a-b\|\| + \frac{1}{\|B\|}\sum_{b \in B}\min_{a \in A}\|\|b-a\|\|\right)$	Average minimum distance between contour points; less sensitive than HD.

Source: (a)Bertels et al. (2019); (b)Taha and Hanbury (2015); (c)Fan et al. (2017)

3.2. *Results Using Harmonized Release Parameters (S2)*

The $S2$ simulations were performed in order to isolate differences attributable to the dispersion models by defining the same release rate for the two software, as described in Section 2.

Because EFFECTS used a higher representative release rate for the 10 mm scenario in $S2$, the resulting ERPG concentration footprints expanded compared to those obtained in $S1$ (Figure 3). Under both meteorological conditions, EFFECTS predicted longer maximum downwind distances than PHAST for the 10 mm release, consistent with the behavior observed for the 10-minute release in the $S1$ simulations.

Aligning the source term improved the agreement between the ERPG footprints predicted by the two software tools. For the 5D condition, the Jaccard index increased by approximately 10% compared to $S1$ (Table 6), indicating greater similarity in the affected areas once differences in release magnitude were removed. Under the 1.5F condition, the improvement was smaller and was mainly observed for the ERPG-3 contour. The only case showing a reduction in similarity was the ERPG-1 contour under 1.5F, where both indices decreased by about 3%. This occurred because, in EFFECTS, the maximum downwind ex-

Table 4. Similarity and distance metrics computed for the release from a 10 mm diameter hole in the S1 simulation run. E stands for ERPG.

S1	10 mm and 1.5F			10 mm and 5D		
	E1	**E2**	**E3**	**E1**	**E2**	**E3**
D [-]	0.73	0.69	0.75	0.58	0.59	0.61
J [-]	0.57	0.52	0.61	0.41	0.42	0.44
HD [m]	1909	539	42	161	52	31
CD [m]	252	104	20	62	26	10

Table 5. Similarity and distance metrics for the release of the entire content in 10 min in the S1 simulation run. E stands for ERPG.

S1	10 min and 1.5F			10 min and 5D		
	E1	**E2**	**E3**	**E1**	**E2**	**E3**
D [-]	0.77	0.51	0.82	0.69	0.68	0.76
J [-]	0.62	0.35	0.69	0.52	0.52	0.61
HD [m]	1697	11311	1101	816	681	344
CD [m]	693	2204	195	411	232	72

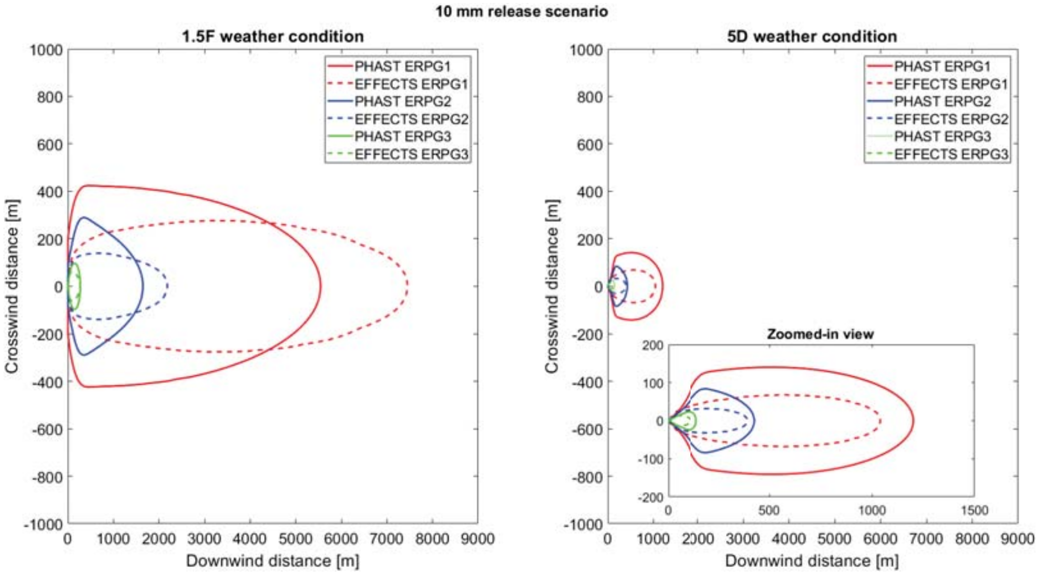

Fig. 1. ERPGs maximum footprints for S1 run of 10 mm release scenario.

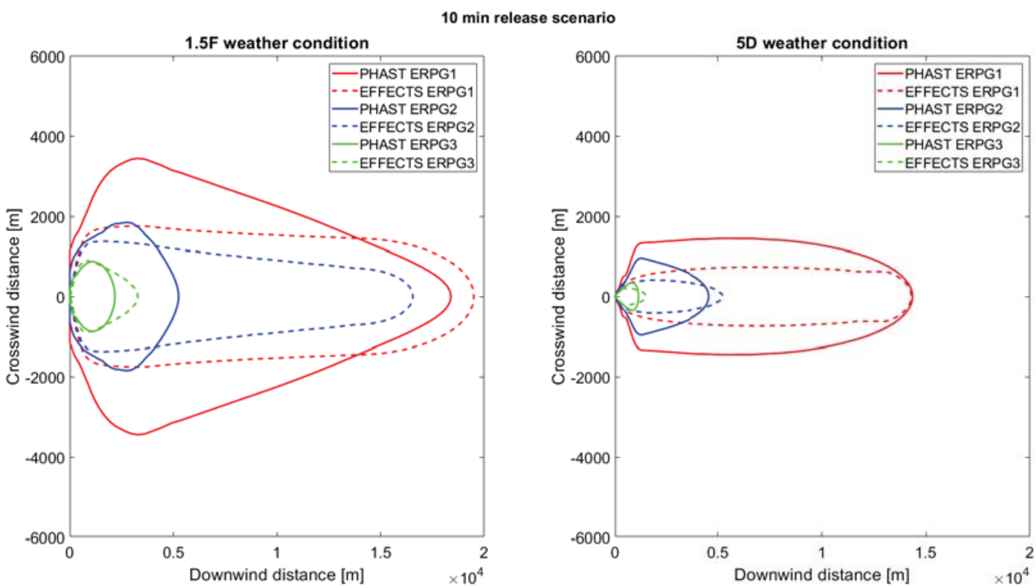

Fig. 2. ERPGs maximum footprints for S1 run of 10 min release scenario.

tent of the contour increased more than the crosswind width. As a result, the overall footprint area expanded, but the overlap with PHAST did not increase accordingly, leading to lower Dice and Jaccard values.

Greater footprint overlap in $S2$ also reduced the distance based discrepancies. For the 5D condition, HD decreased substantially compared to $S1$ and no longer reflected differences in maximum downwind extent. Instead, it highlighted residual discrepancies in crosswind plume spread. Under the 1.5F condition, the Hausdorff Distance remained closely associated with differences in maximum downwind distance, indicating that plume elongation along the wind direction continued to dominate the discrepancies between the two model outputs.

The Chamfer Distance exhibited a similar trend. Compared to $S1$, CD decreased under the 5D condition, reflecting a reduction in the average spatial separation between the contours after alignment of the release rate. For the 1.5F condition, reductions in CD were more limited and were primarily evident for the ERPG-3 contour.

Table 6. Similarity and distance metrics computed for the release from a 10 mm diameter hole in the S1 simulation run. E stands for ERPG.

S2	10 mm and 1.5F			10 mm and 5D		
	E1	E2	E3	E1	E2	E3
D [-]	0.70	0.69	0.84	0.72	0.70	0.74
J [-]	0.54	0.53	0.72	0.56	0.54	0.59
HD [m]	3225	965	46	73	47	14
CD [m]	427	153	15	37	19	6

4. Discussion

The comparison between PHAST and EFFECTS shows that assumptions related to both release modeling and dispersion modeling strongly influence the predicted ERPG concentration footprints. The $S1$ results indicate that default software configurations can lead to substantial differences in maximum downwind distances and total affected areas. For small-leak scenarios, discrepancies in the predicted downwind extent of the lethality zone ranged from approximately 15% to 30% of the footprint length, while for catastrophic releases the relative differences increased to between 30% and 50%, depending on meteorological conditions. These differences arise from variations in source-term representation, which were reduced in the $S2$ run. However, model-specific differences, such as droplet formation and rainout models, and the transitions between jet, dense-gas, and passive dispersion regimes remain and continue to influence the discrepancies observed between the contours.

For the release of the full inventory over 10 minutes, the two software tools adopt markedly different approaches to rainout and pool formation. PHAST predicts partial rainout followed by the development of an evaporating liquid pool. As a result, not all released ammonia contributes immediately to the airborne cloud, and the effective emission duration extends beyond the prescribed 10 minutes. EFFECTS, in contrast, applies the Yellow Book spray model (RIVM, 2005b), which results in no rainout and no pool formation, causing the full inventory to enter the atmosphere directly. This divergence in source-term treatment likely explains the differences observed in the ERPG footprints for the 10-minute scenario (Figure 2).

For this catastrophic release scenario, the predicted maximum downwind distances exceeded 10 km, which corresponds to the upper validity limit of the dispersion model implemented in EFFECTS (GEXCON, 2024). Under such conditions, the ERPG-1 and ERPG-2 contours are subject to increasing uncertainty, and the associated footprints should therefore be interpreted with caution.

The $S2$ run shows that aligning the representative release rate between the two tools improves agreement between the predicted footprints, particularly under the 5D meteorological condition. This confirms that differences in release modeling are a significant contributor to the discrepancies observed in $S1$, while the remaining variability arises from differences in dispersion formulations and embedded model parameterizations.

Differences in maximum downwind distances, as reflected by the Hausdorff Distance in the $S1$ simulations, directly affect how emergency planning zones are delineated. Because these distances are often used as a geometric basis for defining protective action areas, variations that arise solely from model selection and configuration can lead to different spatial extents for the same release scenario. Recent studies have emphasized the importance of cross-validating outputs from different modeling tools to strengthen the robustness of consequence assessments (Park et al., 2025). This

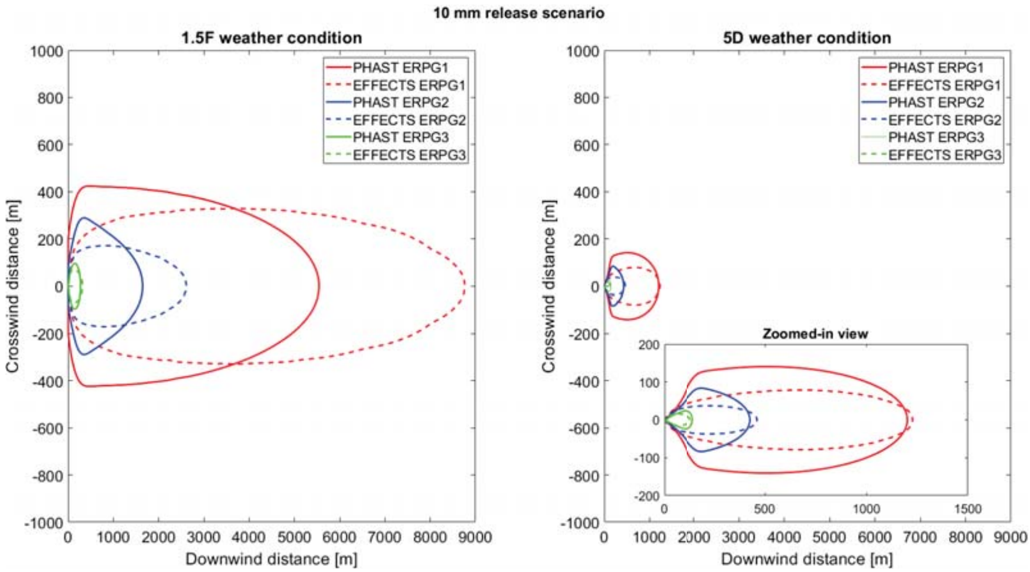

Fig. 3. ERPGs maximum footprints for S2 run of 10 mm release scenario.

highlights the need for transparency in modeling assumptions and for practitioners to interpret software-specific behaviors carefully when model results are used to support emergency preparedness and land-use planning.

5. Conclusion

This study assessed the variability in ammonia toxic dispersion predictions produced by two widely used consequence-analysis tools, PHAST and EFFECTS. Representative accidental release scenarios were simulated under different meteorological conditions, and the resulting ERPG concentration footprints were compared using geometric similarity and distance-based metrics. The analysis quantitatively demonstrated how the choice of modeling software can lead to divergent hazard predictions.

Under default assumptions ($S1$), substantial discrepancies emerged: EFFECTS consistently predicted longer downwind distances under stable conditions, while PHAST produced broader crosswind spreads and larger total affected areas. These differences were most pronounced under neutral atmospheric conditions (5D), where spatial overlap between the two models decreased significantly.

Harmonizing the release rates ($S2$) improved agreement between the predicted footprints, particularly under 5D, where the Jaccard index increased by approximately 10%, yet notable discrepancies persisted in both downwind and crosswind dispersion. These residual differences confirm that the variability cannot be attributed solely to source-term assumptions but also reflects differences in dispersion formulations, submodel implementations, and embedded parameterizations.

Overall, the results show that model-driven variability can significantly influence the delineation of emergency planning zones and the characterization of toxic hazard footprints. Practitioners, safety analysts, and regulators should be aware of these differences, especially when model outputs inform land-use planning or emergency preparedness decisions.

This study provides a reproducible framework for comparing two-dimensional model outputs using geometric similarity measures and offers a quantitative characterization of discrepancies between two major industrial tools across representative release scenarios. The results illustrate that part of the variability originates from default modeling choices but also that complete harmonization is inherently difficult, as each tool incorporates numerous assumptions that affect predicted footprints. These findings reinforce the need for transparent documentation and careful interpretation of software-specific behaviors when model outputs support safety assessments.

Future work will extend this analysis to additional release scenarios and include a sensitivity study of key modeling parameters to quantify their influence on contour geometry and inter-model variability. Furthermore, expanding the comparison to other consequence analysis tools would further support the development of more harmonized modeling practices within the industry.

Acknowledgement

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 Húnter fellow. This work was also partly funded by Grant PID2023-150607OB-I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF/EU.

References

AHIA (2016). Emergency Response Planning Guidelines (ERPGs) Values. Technical report, American Industrial Hygiene Association.

Bertels, J., T. Eelbode, M. Berman, D. Vandermeulen, F. Maes, R. Bisschops, and M. B. Blaschko (2019). Optimizing the Dice Score and Jaccard Index for Medical Image Segmentation: Theory and Practice. In D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap, and A. Khan (Eds.), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, Volume 11765, pp. 92–100. Cham: Springer International Publishing. Series Title: Lecture Notes in Computer Science.

Cardoso, J. S., V. Silva, R. C. Rocha, M. J. Hall, M. Costa, and D. Eusébio (2021, May). Ammonia as an energy vector: Current and future prospects for low-carbon fuel applications in internal combustion engines. *Journal of Cleaner Production 296*, 126562.

Currie, J., K. Study, P. Delanoy, and K. First (2024, December). Advanced consequence modeling and risk analysis. *Process Safety Progress 43*(4), 678–685.

Fan, H., H. Su, and L. Guibas (2017, July). A Point Set Generation Network for 3D Object Reconstruction from a Single Image. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, pp. 2463–2471. IEEE.

Gant, S., J. Chang, R. Hetherington, S. Hanna, G. Tickle, T. Spicer, S. McMasters, S. Fox, R. Meris, S. Bradley, S. Miner, M. King, S. Simpson, T. Mazzola, A. McGillivray, H. Tucker, O. Björnham, B. Carissimo, L. Fabbri, M. Wood, K. Habib, M. Harper, F. Hart, T. Vik, A. Helgeland, J. Howard, L. Mauri, S. Mackie, A. Mack, J.-M. Lacome, S. Puttick, A. Ibrahim, D. Miller, S. Dharmavaram, A. Shen, A. Cunningham, D. Beverly, D. M. O'Neal, L. Verdier, S. Burkhart, C. Dixon, S. Nilsen, R. Bradley, H. L. Skarsvåg, E. H. Fyhn, and A. Aasen (2025, December). Pressure-liquefied ammonia jet dispersion: Multi-model intercomparison using Desert Tortoise and FLADIS field data. *Atmospheric Environment: X 28*, 100389.

GEXCON (2024). EFFECTS v12 - User and reference manual.

Hanna, S., T. Mazzola, J. Chang, T. Spicer, S. Gant, and R. Batt (2022, March). Gaps in toxic industrial chemical model systems: Improvements and changes over past 10 years. *Process Safety Progress 41*(1), 151–166.

Lacome, J.-M., G. Leroy, L. Joubert, and B. Truchot (2023, October). Harmonisation in Atmospheric Dispersion Modelling Approaches to Assess Toxic Consequences in the Neighbourhood of Industrial Facilities. *Atmosphere 14*(11), 1605.

Orozco, J., J. Van Caneghem, L. Hens, L. González, R. Lugo, S. Díaz, and I. Pedroso (2019, June). Assessment of an ammonia incident in the industrial area of Matanzas. *Journal of Cleaner Production 222*, 934–941.

Park, H. K., J. H. Kim, and M. C. Lee (2025, October). Sensitivity and regression analysis-based risk assessment of ammonia as fuel for combined cycle power plants. *International Journal of Hydrogen Energy 179*, 151566.

RIVM (2005a). *Guidelines for quantitative risk assessment (Purple Book), CPR 18e*. National Institute of Public Health and the Environment.

RIVM (2005b). *Methods for the calculation of Physical Effects (Yellow Book), CPR 14e*. National Institute of Public Health and the Environment.

Ruiz Pérez, S. (2017). TNO dense gas model verification and validation.

Taha, A. A. and A. Hanbury (2015, November). An Efficient Algorithm for Calculating the Exact Hausdorff Distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence 37*(11), 2153–2163.

Witlox, H., M. Fernandez, M. Harper, A. Oke, Y. Xu, and J. Stene (2023). Consequence models for accidental releases of toxic or flammable chemicals to the atmosphere - Verification and Validation.