Comparison of Machine Learning Models for Prediction of Safety Distances in Toxic Dispersion Scenarios

Artemis Papadaki[a], Mercedes Belda[a], Stefania Bramato[b], Alba Àgueda[*a], Eulàlia Planas[a]

[a] CERTEC, Chemical Engineering Department, Polytechnic University of Catalonia – BarcelonaTech, Barcelona, 08019, Spain
[b] LISES – Department of Civil, Chemical, Environmental and Material Engineering, Alma Mater Studiorum - University of Bologna, Via Terracini 28, 40131, Bologna, 08019, Italy
alba.agueda@upc.edu

In the recent years, Machine Learning (ML) has demonstrated a significant potential in handling big amounts of diverse data and generating rapid and accurate predictions. These capabilities, applied in risk analysis studies, can be particularly valuable in emergency planning of industrial sites, providing timely and reasonably accurate information during and after accidents. This work focuses on the determination of safety distances, which are an indispensable part of the risk assessment of Seveso establishments. Specifically, ML models are implemented to predict safety distances in toxic dispersion scenarios of three substances, namely ammonia, chlorine and ethylene oxide. For the purpose of training the models, a database is constructed using ALOHA software, containing different scenarios over a wide range of atmospheric and process conditions that influence the release and the dispersion phenomena. Four distinct supervised ML models are deployed to perform classification and regression tasks, predicting distances for the intervention zone. The models are tuned and evaluated based on common statistical metrics, as well as a case-specific metric. The results indicate that classification is achieved in all the cases with similar degree of agreement, while regression varies significantly among the models.

1. Introduction

The Seveso III Directive obliges the Seveso establishments all over EU to develop and implement emergency plans (Directive 2012/18/EU1). Apart from the guidelines given in the Directive, state-level legislation determines the risk assessment approach (probabilistic, deterministic, quantitative, semi-quantitative, etc.) and provides further specifications to elaborate emergency and auto protection plans. Safety zones, commonly mentioned in the context of land-use planning (Laurent et al., 2021), can be used to measure the area potentially affected by a major accident. According to the Spanish and the Catalan laws (Real Decreto 840/2015, Instrucció 11/2010 SIE), upper-tier Seveso establishments are obliged to submit a safety report that explicitly includes the safety distances for accidental scenarios. In such cases, the pre-calculated distances provide a reference and facilitate the intervention of the civil protection and firefighters during an emergency. When it comes to accidents that occur in lower-tier or non-Seveso establishments, the safety distances should be estimated ad hoc or, in cases of scarce information, established according to pre-defined generic distances (EIC, 2023).

The demand for instant and even multiple simulations conducted by non-experts aligns with what Machine Learning (ML) can offer. Embracing the evolution of risk sciences towards data driven approaches, an increasing number of researchers experiment in applying ML tools in process safety (Hegde and Rokseth, 2020). While previous research (Jiao et al., 2023) has explored ML's application in establishing quantitative property-consequence relationships in the context of toxic dispersion, emergency planning demands a distinct problem formulation. This involves examining diverse atmospheric variables to address the unique challenges posed by emergency scenarios. This work delves into improving risk analysis through the prediction of safety distances, as the initial phase of implementing ML within emergency planning. Its primary objective is to demonstrate ML's

capability to reproduce the complex interdependencies among various variables involved in the dispersion. Once a reliable reproduction of the dispersion behavior by ML means is achieved, applicability improvements such as, data augmentation or uncertainty challenges can be addressed.

In the context of Spanish law, three different safety zones are defined, namely domino effect, intervention and alert zone. Each zone is delimited by threshold values for chemical, mechanical or thermal hazards. In this work chemical hazards were measured based on the AEGL criterion (Table 1). Most commonly, the computation of the zones is performed using a hazard modeling software, which employs empirical or semi-empirical equations. Complementary to this approach, a ML model can provide a reasonable prediction by processing an adequate number of accidental scenarios. In this way, it bypasses the empirical solution relying, instead, on logical and statistical schemes. The applicability of four supervised ML models was examined, namely, Decision Tree, Random Forest, Ada Boost and kNN.

Table 1: Safety zones' thresholds for chemical hazard

Accident type	Intervention zone	Alert zone	Domino zone
Toxic dispersion	AEGL - 2	AEGL - 1	-

2. Methodology

The building steps to develop a supervised ML model comprise three phases (Figure 1). Firstly, a reliable and sufficiently large database is needed. Database construction is achieved either by collecting information from existing sources or by generating new data. In both cases, the quality of the database is, probably, the most significant part of a data driven model, since the reliability of the database impacts directly on the reliability of the final model. In this work, ALOHA software is used, meaning that the database inherently contains the approximation error of the software. Once the database is obtained, the raw data is pre-processed so as to feed the ML model with the appropriate type and format of data. The modeling phase involves the essential steps to build an optimized ML model and assess its performance. In this work simulations and modeling were run on a desktop PC with 12th Gen Intel processor (6 cores x 4,900 MHz) using 16GB of RAM.
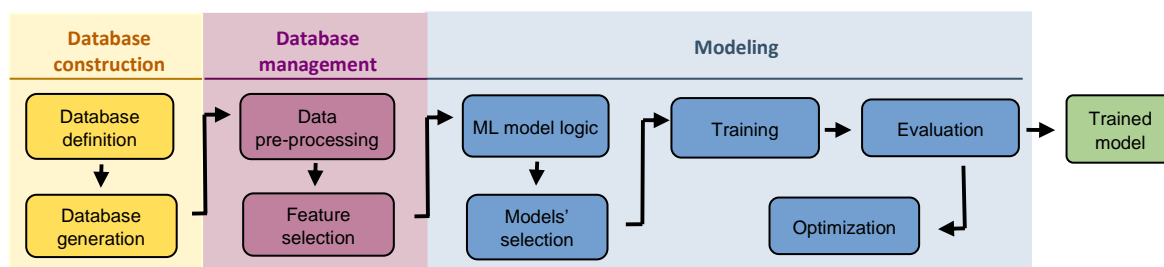

Figure 1: Model development methodology

3. Database construction

The database is constructed using ALOHA, which is a free software for hazard modelling developed for the CAMEO software suite by EPA and NOAA (EPA, 2016). The Spanish law indicates it as the software to be used for effect modelling in safety reports for toxic dispersion scenarios. The scenario under examination is the toxic dispersion of a substance in a storage tank, caused by the release from a hole of certain diameter. The substances included in the database are ammonia, chlorine and ethylene oxide due to their high toxicity and quantity registered in Catalan Seveso establishments (Comissió de Protecció Civil de Catalunya, 2023). Following an analysis on the ALOHA source and dispersion models, the type of variables influencing the phenomena are determined (Table 2). The stability class and wind speed combinations are chosen to provide a wide range of atmospheric turbulence conditions, more specifically to combine vertically stable atmospheric conditions with higher number of wind speeds. Ground roughness values include different situations considered plausible for storage installations (Casal, 2018). Air temperatures stem from 2022 Catalan climatology data for minimum, average and maximum registered temperatures (IDESCAT, 2023). Process variables are substance dependent, except for the hole diameter. The tank is horizontal because it is the most commonly used type for all the examined substances. Tank dimensions (diameter and length) are selected based on quantities that range from Seveso lower-tier to upper-tier thresholds for each substance (Directive 2012/18/EU). Two main storage conditions are included in the database: pressurized at ambient temperature, and cryogenic/refrigerated below the substance's boiling point at atmospheric pressure. Variables not included in Table 2 are kept constant,

e.g. leak height (assigned always at the bottom of the tank), and relative humidity and cloud cover, which are equal to 70 % and 50 %, respectively.

Table 2: Variables and considered values

Parameter	Considered values		
Stability class: [Wind speed (m/s)]	A: [1], B: [1, 2], C: [3, 4, 5], D(day): [4, 5, 6, 8] D(night): [3, 5, 7], E: [1, 2, 3, 5], F: [1, 1.5, 2, 2.5]		
Ground roughness (m)	[0.03, 0.1, 1]		
Air temp. (ºC)	[-5, 15, 35]		
Chemical substance	Ammonia	Chlorine	Ethylene oxide
Tank diameter (m): [Tank length (m)]	0.8: [2], 1.4: [6.2], 2.6: [7.6], 2.8: [8.8], 3.6: [13.6], 3.8: [16.29]	1.2: [3.98], 1.6: [4.4], 2: [5.65], 2.4: [8], 2.6: [10.36], 3: [10.25]	0.94: [2.12], 1.2: [3.98], 1.8: [5.25], 2.2: [7.1], 2.8: [7.11], 3: [10.25]
Process temp. (ºC) — Cryog. / refrig. storage	-33	-34	5
Pressurized storage	[-5, 15, 35]	[-5, 15, 35]	[-5, 15, 35]
Filling degree (%)	[5, 50, 85]		
Circular opening diameter (m)	[0.005, 0.015, 0.025, 0.035, 0.045]		

The number of scenarios simulated is in total 102,060, meaning 34,020 per substance, which are generated by automatizing ALOHA with the aid of a python script. From the calculated distances, which are intervention and alert zones, only the former are taken into account for the construction of the ML models.

4. Data base management

Data pre-processing covers multiple tasks, such as, scaling and conversion of categorical to numerical data and handling of missing values. These processes ensure that the features are transformed into a suitable format prior to the insertion to an ML model. In the current database two categorical variables exist, stability class and chemical substance. Instead of handling them as categorical data, a conversion to corresponding numerical data is preferred. Specifically, chemical substance was translated into molecular mass and for each stability class (A-F) a number was assigned (1-7), respectively, taking into account the differentiation between class D day and night. Moreover, the tank diameter and length were merged to a single feature, tank volume, in order to reduce the number of highly interrelated features. Finally, scaling by standardization was applied to Ada Boost and kNN models, for which scaling influences the quality of the predictions.

Feature selection refers to meticulously determining the features that contribute to the predictions. A sensitivity analysis helps distinguishing between sensitive features that impact significantly the output, and less important ones. A first sight of the features' importance is gained by analyzing the mathematical equations used in ALOHA, for instance the opening diameter has a squared relation with the release rate, while stability class shows a non-linear behavior. Then, a variance-based sensitivity analysis is run to quantify the interaction between 2 non-linearly correlated variables. More specifically, first-order Sobol indices (S1) measure the contribution of each single feature uncertainty to the total variance of the output (Sobol′, 2001). The S1 indices for the features in the database were generated by the SALib package (Herman and Usher, 2017) using random sampling in different subsets to ensure the robustness of the indices (Table 3). The database shows greater sensitivity to chemical substance followed by tank volume, process temperature and opening diameter (higher S1 indices). Moreover, ground roughness and air temperature are features that less influence the output (lower S1 values).

Table 3: Sensitivity analysis in the database features

Feature	Stab. class	Wind	Gr. rough.	Air T	Chem. sub.	Tank vol.	Proc. T	Fil. deg.	Circ. op. diam.
S1	0.034	0.012	0.003	0.005	0.41	0.28	0.28	0.15	0.18

5. Modeling

When calculating emergency planning zones in ALOHA, two possibilities exist, either the zone is above 10 km and the numerical value is not known, or it is below 10 km and it is expressed as an exact numerical value. Consequently, the objective of the ML is twofold, firstly to predict whether the zone is above 10 km or not with a classification task and secondly to predict its value performing a regression task. For these two tasks, four supervised ML models were built using the scikit-learn library (Pedregosa et al., 2011), namely Decision Tree, Ada Boost, Random Forest and kNN neighbors.

A Decision Tree algorithm starts from a set of features (root node) and progresses towards an output (leaf node) through a series of intermediate decisions (internal nodes). Every internal node is split into multiple subnodes,

creating paths which can lead to different leaf nodes. During optimization, the splitting process is controlled by the minimum number of samples split and minimum number of samples leaf. Additionally, the maximum number of features to consider in splitting is examined, as well as the maximum depth of the tree, which refers to the total number of internal nodes. Ada Boost algorithm is an ensemble technique, which improves the performance of a specific estimator, in this case the Decision Tree. Various instances of the estimator are created by training it in different subsets of the training set, called weak learners. The algorithm compares the weak learners and assigns different weights according to their performance. The hyperparameters used to tune the model are the number of weak learners and the learning rate, which determines the contribution of the weak learners to the overall performance. Random Forest is another ensemble technique which predicts an output by averaging the performance of multiple decision trees, a notion similar to the weak learners. The individual trees have equal weights, they are trained in parallel and independently, in contrast to Ada Boost which uses sequential training of the weak learners, with each new weak learner dependent on the performance of the previous ones. Finally, the kNN model functions by quantifying the proximity of a set of data to a neighboring one and the number of neighbors serves as a hyperparameter to the model. The definition of the hyperparameters can be found in Pedregosa et al. (2011).Once the models are defined, the database is divided into train and test set, the former used for the training of the models and the later for assessing its capacity to predict new unknown values. The test set contains ~ 20 % of the entire database (21,600 scenarios) and equal number of scenarios for every value of the features of chemical substance, stability class, internal temperature, tank volume, circular opening diameter and filling degree. The distribution of wind depends on stability class, and the features of ground roughness and air temperature were selected randomly, given their low Sobol indices.

The performance of the model is assessed depending on the task. The classification performance is measured with four metrics, Accuracy, Precision, Recall and F1 score, as defined by Pedregosa et al. (2011). Regression is quantified with two commonly used statistical metrics, namely R2 score and RMSE, and a case specific metric. The need for a case specific (CS) metric was due to the partial incapacity of common statistical metrics to depict the particularities of the safety zones. More specifically, a metric should treat the difference between a true output (IZ) and a prediction (IZ') as following:

- Higher penalization to underestimated predictions than overestimated. An error due to overprediction in safety distances, may lead to a more conservative result, while underprediction contributes to potential undervaluation of the risk level and mismanagement of the emergency. For this reason, the difference between IZ and IZ' is multiplied by a factor of 2 in case of underprediction and divided by a factor of 2 when overprediction takes place.

- Lower penalization of errors that fall out of the 95 % of the distances' distribution. This point can be realized by using the Mean Average Error (MAE) instead of the exact difference between IZ and IZ' when the distances are above 6,205 m, which corresponds to the 95 % of their distribution. In this way, large contributions in the mean error calculation, due to large differences in distances above 6,205 m, will be eliminated and substituted by the mean error value (MAE). These objectives can be mathematically expressed with the following loss function:

$$CS = \begin{cases} 2|IZ - IZ'|, if\, IZ < 6{,}205 \wedge IZ > IZ' \\ \dfrac{|IZ - IZ'|}{2}, if\, IZ < 6{,}205 \wedge IZ < IZ' \\ MAE, \quad if\, IZ > 6{,}205 \end{cases} \quad (1)$$

To optimize the performance of each model, tuning was performed on a wide range of their hyperparameters. The hyperparameters for each case along with their search space are summarized in Table 4.

Table 4: Hyperparameters' tuning

Model	Hyperparameters	Search space*	Classifier best set*	Regressor best set*
Decision Tree	Max depth [Max features]	1-150 [1-8]	67 [7]	45 [9]
	Min sample leaf [split]	2-20 [2-9]	2 [2]	2 [4]
Ada Boost	Learning rate	0.001 – 0.1, step=0.5	0.001	0.05
	N estimators	50-400	311	131
Random Forest	Bootstrap	False, True	False	False
	Max depth [Max features]	1-150 [1-8]	85 [7]	120 [6]
	Min sample leaf [split]	2-20 [2-9]	3 [7]	2 [3]
	N estimators	1-150	80	128
kNN	N neighbors	2-50	2	3
	Algorithm	Auto, ball tree, kd tree, brute	ball tree	ball tree

* The numbers in parenthesis correspond to the feature in parenthesis in the 1st column

Different sets of hyperparameters were tested using the k-fold cross validation method (Pedregosa et al., 2011). In the current modeling, the number of folds is equal to 4 and the 300 sets of hyperparameters for each model are used and defined by randomized selection within the search space. To avoid overfitting, the model is evaluated based on the absolute of the test loss (L) and its difference to the training loss (ΔL). The best hyperparameter set for each model is selected using the CS metric where L is minimized while ΔL remains in its 90 percentile.

6. Results

The distribution of the numerical values of the intervention zone in the database for each substance and in total is shown in Table 5. It can be noted that chlorine exhibits the highest distances, followed by ethylene oxide.

Table 5: Distribution of intervention zone values (m) in the database

Feature	Ammonia	Chlorine	Ethylene oxide	Total
Mean [Sd]	1,164 [1,042]	4,482 [2524]	711 [715]	1,996 [2,265]
Max. distance	9,600	10,000	7,500	10,000

In the classification task (Table 6), all models exhibit comparable and solid performance across the four metrics. kNN shows the worst performance in precision and F1 score, while for the rest of the models the differences can be considered non-significant. As for the regression (Table 6), Ada Boost and Random Forest show the best performance followed by Decision Tree. This behavior stems from the fact that both Ada Boost and Random Forest are ensemble techniques applied on a Decision Tree aiming to improve its performance. Similar to the classification task, kNN demonstrates the worst results being unable to capture the pattern of the training set. The run time of all the models for 21,600 scenarios ranged from 5 s to 60 s, rendering them significantly quicker comparing to ALOHA run time of a single scenario (5-10 min). For a better visual representation of the regression results, Random Forest is compared before and after tuning (2).

Table 6: Performance evaluation

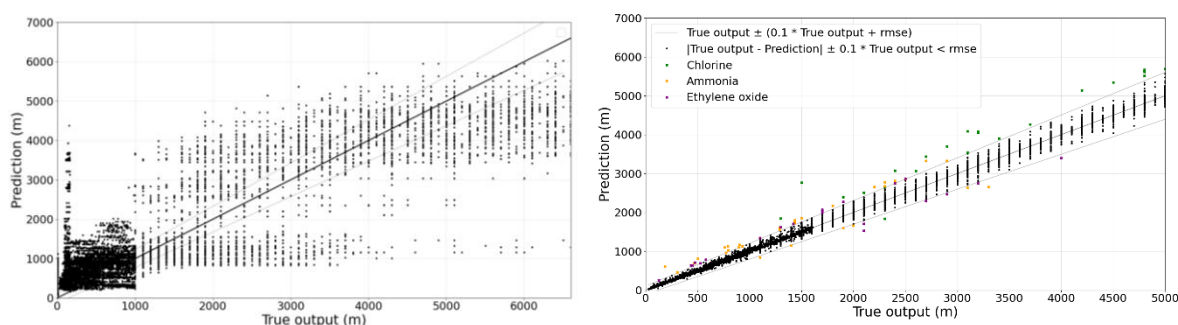| Model	Classification				Regression		
	Accuracy	Precision	Recall	F1	RMSE (m)	CS metric (m)	R2
Decision Tree	0.99	0.94	0.96	0.95	159	63	0.99
Ada Boost	0.99	0.97	0.98	0.97	103	42	0.99
Random Forest	0.99	0.96	0.96	0.96	102	43	0.99
kNN	0.98	0.76	0.94	0.84	443	240	0.96

Figure 2: Random Forest regression: (a) without tuning, (b) with tuning and outliers highlighted

In both plots, the prediction values are plotted against their true output for the test set, making evident the importance of optimization through hyperparameters' tuning. As calculated by ALOHA, values below 1,000 m are continuous (regions with dense points in plots), while above 1,000 m are calculated with precision of 100 m (intermittent region in plots). To further facilitate the analysis, the outliers are shown in 2b differentiating them based on the chemical substance, given that this was the only feature exhibiting significant differences in the distributions of the predicted values as well as the most significant feature (Table 3). The substances are depicted by 3 different colors for the outlier points. Chlorine exhibits outliers in large distances, while ammonia and ethylene oxide in lower values, which is expected given their distribution (Table 5).

7. Conclusions

This work is a contribution to the application of ML to the risk assessment in process safety, and more precisely to the prediction of safety distances for emergency planning purposes. Multiple toxic dispersion scenarios for ammonia, chlorine and ethylene oxide are examined. Sensitivity analysis is performed giving a better insight on the influence of the features upon the intervention zone. The performance of four supervised ML models is compared in classification and regression tasks, showing that classification obtained accurate results in all the models examined, while in the regression task, two of four models show noteworthy performance. Random Forest and Ada Boost are the best performing models in this case, showing a largest RMSE of 103 m in a range of distances from 10 to 10,000 m. As a result, certain ML trained models can reproduce toxic dispersion, as modelled by ALOHA. Future investigation will involve the study of the reduction of the number of input variables, so as to convert ML models in a handy tool for emergency response.

Acknowledgments

References

Comissió de Protecció Civil de Catalunya, 2023, Exterior Emergency Plan of the Chemical Sector of Catalonia (in Catalan). Departament d'Interior de la Generalitat de Catalunya.

Statistical Institute of Catalonia, 2023, Thermometry. Maximum temperatures. Counties, 2022. Ministry of Climate Action, Food and Rural Agenda.

Enginyers Indrustrials de Catalunya (EIC), 2023, Conferences - Emergency management in serious accidents: preparation and response (in Catalan), <www.eic.cat/jornades/1130637>.

Enviromental Protection Agency United States (EPA), 2016, ALOHA Software (Version 5.4.7) [Software] <www.epa.gov/cameo/aloha-software> accessed 15.04.2024.

Hegde, J., Rokseth, B., 2020, Applications of machine learning methods for engineering risk assessment – A Review. Safety Science, 122, 104492.

Herman, J., Usher, W., 2017, SALib: An open-source Python library for sensitivity analysis. Journal of Open Source Software, 2(9).

Instrucció 11/2010 SIE, 2010, Criteria for the preparation and evaluation of the safety report to be presented by the upper-tier establishments affected by the current legislation on major accidents (in Catalan). Generalitat de Catalunya. Departament d'Innovació, Universitats i Empresa.

Jiao, Z., Zhang, Z., Jung, S., Wang, Q., 2023, Machine learning based quantitative consequence prediction models for toxic dispersion casualty. Journal of Loss Prevention in the Process Industries, 81, 104952.

Laurent, A., Pey, A., Gurtel, P., Fabiano, B., 2021, A critical perspective on the implementation of the EU council seveso directives in France, Germany, Italy and Spain. Process Safety and Environmental Protection, 148, 47–74.

Ministerio de Economía, Industria y Competitividad, 2017, Royal Decree 656/2017, of June 23, which approves the Regulations for the Storage of Chemical Products and their Complementary Technical Instructions MIE APQ 0 to 10, Boletín Oficial del Estado.

Ministerio de Economía, Industria y Competitividad, 2015, Royal Decree 840/2015, September 21, which approves measures to control the inherent risks to major accidents involving dangerous substances, Boletín Oficial del Estado.

Pedregosa, F., Varoquaux, Gaël, Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al., 2011, Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 2825–2830.

Sobol′, I.M., 2001, Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Mathematics and Computers in Simulation, 55(1–3), 271–280.

The European Parliament & Council of the European Union, 2012, Directive 2012/18/EU of the European Parliament and of the Council of 4 July 2012 on the control of major-accident hazards involving dangerous substances, amending and subsequently repealing Council Directive 96/82/EC, Official Journal of the European Union.