CAAP Quarterly Report

June 19, 2024

Project Name: Determination of Potential Impact Radius for CO_2 Pipelines using Machine Learning Approach

Contract Number: 693JK32250011CAAP

Prime University: Texas A&M University

Prepared By: Sam Wang, qwang@tamu.edu, 979-845-9803

Reporting Period: 3/27/2024 – 6/26/2024

Project Activities for Reporting Period:

The following relevant tasks in the proposal have been completed:

- Built initial machine learning models to predict PIRs with CO_2 concentrations of 1%, 4%, and 9% for different geography. More details are provided in the appendix.
- Studied the near-field simulation with User-Defined Functions (UDFs) and User-Defined Real Gas Models (UDRGMs) in Ansys Fluent. This research is prepared for the upcoming Skylark project. The model could be applied to the flow behavior in wind tunnel experiments, which are planned at University of Arkansas. More details are provided in the appendix.
- Studied the influences of wind direction and pipeline location on sloped geometries. Then, planned the simulation accordingly. More details are provided in the appendix.
- Used Ansys Fluent to continue to conduct CFD simulations based on the results of the calculation for the near field.

Project Financial Activities Incurred during the Reporting Period:

Based on the proposed budget, the cost is broken down into two parts:

- Efforts from the PI Dr. Wang for about 0.25 month.
- Efforts and work by graduate students, Chi-Yang Li and Jazmine Aiya D. Marquez, totally for about 3 months for each of them.

Project Activities with Cost Share Partners:

Dr. Wang's time and efforts (0.25 month) in this quarterly period are used as cost share. He devoted his time to supervise the graduate students on research, review all paperwork, and prepare the progress report.

Project Activities with External Partners:

Dr. Wang attended the technical discussion session on June 19 about the Skylark Joint Industry Project (JIP), led by Simon Gant at UK HSE and Dan Allason at DNV. The technical proposal was also sent to PHMSA for review. DNV will host a contractual discussion session on June 27. A final consultation period to end-July will allow for final changes with an aim for signatures to be applied on August 1. The team plans to kick off the Skylark JIP on September 1, 2024. Ashley Kroon was also in the technical discussion session. We will set up a time to discuss the CFD modeling work and details to participate in the Skylark JIP.

Potential Project Risks:

As we expect, performing hundreds of CFD simulations does take a significant amount of time. With two PhD students working on the project, we can finish the simulations by the next quarter progress report.

Future Project Work:

- The future work is to continue to conduct ongoing parametric studies at TAMU HPRC for various dispersion scenarios using Ansys Fluent, incorporating the numerical simulation setup. Simulation of Flat geometry is already finished. Simulations on the four other terrain categories (VM, VB, SH, BH) will be finished, coupled with pipeline characteristics and weather conditions as outlined in Table 1.

Table 1. The variables for pipeline characteristics and weather conditions.

	Variable	High	Medium	Low
Pipeline characteristics	pressure (MPa)	20	10	1
	diameter (inch)	30	16	4
	flow rate (MMcfd)	1300	590	30
Weather conditions	wind speed (mph)	25	14	3
	temperature (°F)	100	60	0

- Continuously enhance the PIR database for CO_2 pipeline dispersion using the simulation results obtained with the aforementioned setup.

- Conduct parametric studies to identify appropriate machine learning techniques and their corresponding hyperparameters for the machine learning model.
- Study near-field simulations with the application of UDFs and UDRGMs in Ansys Fluent.

Potential Impacts to Pipeline Safety:

- The variables for pipeline characteristics and weather conditions cover the upper limits and lower limits of the current industrial practices; therefore, the machine-learning model is believed to have accurate predictions for other CO_2 pipelines in the range.

Appendix

1. Machine learning models to predict PIRs

The simulation results from the Ansys Fluent for the Flat geometry are shown in Table 2. Because the distributions of the distances for the three different concentrations are quite divergent (Figure 1), we built three distinct models for each of them. The machine learning models applied for searching for the best model are multiple linear regression (MLR), Support Vector Regression (SVR), K nearest neighbors (KNN), random forest (RF), extreme gradient boosting regression (XGBoost), gradient boosting regression (GBR), and Bootstrap Aggregating (Bagging). Mean square error (MSE) and R^2 with the 10-fold cross validation are used to evaluate the performance of the models. In each model, the inputs (features) for the models are gauge pressure, diameter of pipeline, flow rate of CO_2, wind speed, and ambient temperature, and the output (response) is the corresponding distances from simulation. With the fine-tuning of hyperparameters for each model, the best version of each machine learning model is demonstrated in Table 3, and the hyperparameters for the best model is displayed in Table 4. The R^2 for models for the each are, respectively, 0.9243, 0.9201, and 0.9580, which represents highly accurate on predictions. The predictions for 10-fold cross validations results are as Figure 2.

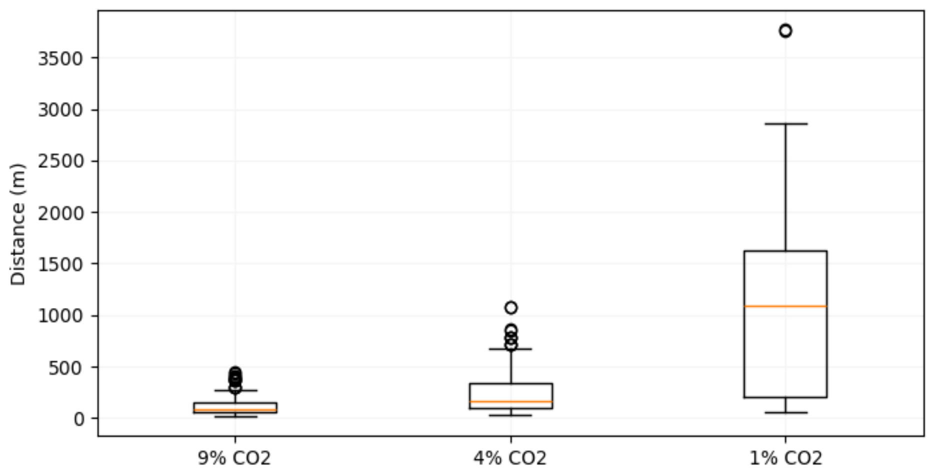

Figure 1. Boxplots for the distances for the three different concentrations.

Table 2. Simulation results from Ansys Fluent for Flat geometry.

Gauge pressure (MPa)	Diameter (inch)	Flow rate (mmcfd)	Wind speed (mph)	Ambient temperature (°F)	Distance for 9% (m)	Distance for 4% (m)	Distance for 1% (m)
1	4	30	3	0	15	34	102
1	4	30	3	60	15	34	102
1	4	30	3	100	15	34	102
1	4	30	14	0	15	31	80
1	4	30	14	60	15	31	80
1	4	30	14	100	15	31	80
1	4	30	25	0	15	29	57
1	4	30	25	60	15	29	57
1	4	30	25	100	15	29	57
10	30	30	14	0	90	152	403
10	30	30	14	60	85	146	241
10	30	30	14	100	85	127	266
10	30	30	25	0	97	162	836
10	30	30	25	60	96	164	597
10	30	30	25	100	96	170	622
1	16	30	3	0	42	72	184
1	16	590	3	0	60	148	1107
1	16	30	3	60	42	71	182
1	16	590	3	60	60	148	1090
1	16	30	3	100	42	72	184
1	16	590	3	100	60	148	1086
1	16	30	14	0	47	68	122
1	16	590	14	0	60	135	1040
1	16	30	14	60	47	68	122
1	16	590	14	60	60	135	1040
1	16	30	14	100	47	68	122

1	16	590	14	100	60	135	1040
1	16	30	25	0	43	61	121
1	16	590	25	0	59	118	928
1	16	30	25	60	43	61	121
1	16	590	25	60	59	118	928
1	16	30	25	100	43	61	121
1	16	590	25	100	59	119	928
1	16	1300	3	0	60	153	1764
1	16	1300	3	60	60	153	1706
1	16	1300	3	100	60	154	1683
1	16	1300	14	0	60	148	1537
1	16	1300	14	60	60	147	1516
1	16	1300	14	100	60	147	1511
1	16	1300	25	0	60	140	1152
1	16	1300	25	60	60	138	1148
1	16	1300	25	100	60	140	1148
10	4	30	3	0	39	60	202
10	4	590	3	0	96	227	1189
10	4	30	3	60	39	60	199
10	4	590	3	60	96	227	1193
10	4	30	3	100	39	61	199
10	4	590	3	100	97	227	1188
10	4	30	14	0	56	89	202
10	4	590	14	0	85	122	1055
10	4	30	14	60	56	89	202
10	4	590	14	60	85	122	1055
10	4	30	14	100	56	89	203
10	4	590	14	100	85	122	1055
10	4	30	25	0	47	85	155
10	4	590	25	0	80	122	1058

10	4	30	25	60	47	85	155
10	4	590	25	60	80	122	1056
10	4	30	25	100	47	85	155
10	4	590	25	100	80	122	1057
10	4	1300	3	0	99	320	1638
10	4	1300	3	60	98	320	1626
10	4	1300	3	100	98	317	1626
10	4	1300	14	0	91	234	1746
10	4	1300	14	60	91	233	1737
10	4	1300	14	100	91	233	1731
10	4	1300	25	0	85	178	1613
10	4	1300	25	60	85	178	1589
10	4	1300	25	100	85	177	1581
20	4	30	3	0	40	61	198
20	4	590	3	0	99	252	1188
20	4	30	3	60	40	61	199
20	4	590	3	60	100	249	1188
20	4	30	3	100	40	61	200
20	4	590	3	100	100	249	1188
20	4	30	14	0	61	90	252
20	4	590	14	0	87	177	1062
20	4	30	14	60	61	91	223
20	4	590	14	60	87	177	1061
20	4	30	14	100	61	91	223
20	4	590	14	100	87	177	1061
20	4	30	25	0	60	86	199
20	4	590	25	0	62	170	1073
20	4	30	25	60	60	86	199
20	4	590	25	60	62	170	1073
20	4	30	25	100	60	86	199

20	4	590	25	100	62	170	1076
20	4	1300	3	0	101	332	1624
20	4	1300	3	60	101	332	1611
20	4	1300	3	100	101	331	1609
20	4	1300	14	0	92	274	1775
20	4	1300	14	60	94	271	1767
20	4	1300	14	100	94	271	1761
20	4	1300	25	0	63	189	1689
20	4	1300	25	60	63	189	1725
20	4	1300	25	100	63	189	1727
10	16	30	3	0	46	63	140
10	16	30	14	0	82	92	192
10	16	30	14	60	82	92	192
10	16	30	14	100	82	92	192
10	16	30	25	0	87	114	192
10	16	30	25	60	87	114	192
10	16	30	25	100	87	114	192
20	16	30	3	0	57	77	157
20	16	30	3	60	55	70	157
20	16	30	3	100	57	72	160
20	16	30	14	0	91	113	274
20	16	30	14	60	90	113	278
20	16	30	14	100	90	113	278
20	16	30	25	0	87	121	256
20	16	30	25	60	87	121	256
20	16	30	25	100	86	121	256
1	30	30	3	0	43	72	228
1	30	590	3	0	122	256	1175
1	30	1300	3	0	127	375	1620
1	30	30	3	60	43	72	228

1	30	590	3	60	122	256	1175
1	30	1300	3	60	125	370	1623
1	30	30	3	100	43	72	225
1	30	590	3	100	122	256	1176
1	30	1300	3	100	128	360	1621
1	30	30	14	0	79	116	309
1	30	590	14	0	114	173	1053
1	30	1300	14	0	116	334	1776
1	30	30	14	60	79	116	309
1	30	590	14	60	114	173	1051
1	30	1300	14	60	116	333	1767
1	30	30	14	100	79	116	309
1	30	590	14	100	114	173	1053
1	30	1300	14	100	116	334	1762
1	30	30	25	0	61	92	212
1	30	590	25	0	102	149	1092
1	30	1300	25	0	109	157	1913
1	30	30	25	60	61	92	212
1	30	590	25	60	102	149	1098
1	30	1300	25	60	109	157	1891
1	30	30	25	100	61	92	213
1	30	590	25	100	102	149	1095
1	30	1300	25	100	109	157	1891
20	30	30	3	0	32	53	130
20	30	30	3	60	32	52	128
20	30	30	3	100	32	52	130
20	30	30	14	0	73	96	186
20	30	30	14	60	73	96	186
20	30	30	14	100	73	96	186
20	30	30	25	0	61	92	184

20	30	30	25	60	61	91	183
20	30	30	25	100	61	92	184
10	16	30	3	60	62	99	185
10	16	30	3	100	62	98	192
10	16	590	14	0	121	160	1055
10	16	590	14	60	121	160	1055
10	16	590	14	100	121	160	1055
10	16	590	25	0	142	160	1086
10	16	1300	25	0	148	173	1366
10	16	590	25	60	142	160	1086
10	16	590	25	100	142	160	1086
10	30	30	3	0	80	122	344
10	30	30	3	60	80	122	345
10	30	30	3	100	80	122	343
10	30	590	14	0	137	238	963
10	30	590	14	60	137	238	963
10	30	590	14	100	128	233	963
10	30	590	25	0	157	349	1288
10	30	1300	25	0	157	378	1439
10	30	590	25	60	157	349	1355
10	30	590	25	100	157	351	1286
20	30	590	14	0	187	371	1190
20	30	1300	14	0	197	386	1348
20	30	590	14	60	187	371	1186
20	30	590	14	100	186	370	1142
20	30	590	25	0	231	426	1824
20	30	1300	25	0	249	456	1926
20	30	590	25	60	232	427	1821
20	30	1300	25	60	250	457	1913
20	30	590	25	100	231	427	1818

20	30	1300	25	100	249	456	1907
10	16	590	3	0	110	242	1210
10	16	1300	3	0	157	388	1614
10	16	590	3	60	110	262	1212
10	16	1300	3	60	157	386	1627
10	16	590	3	100	109	240	1213
10	16	1300	3	100	157	387	1617
10	16	1300	14	0	191	551	2174
10	16	1300	14	60	191	546	2170
10	16	1300	14	100	192	546	2165
10	16	1300	25	60	276	681	2610
10	16	1300	25	100	276	682	2605
20	16	590	3	0	115	290	1295
20	16	1300	3	0	157	460	1658
20	16	590	3	60	117	294	1300
20	16	1300	3	60	157	447	1638
20	16	590	3	100	116	287	1301
20	16	1300	3	100	157	458	1649
20	16	590	14	0	197	434	1893
20	16	1300	14	0	221	546	2296
20	16	590	14	60	200	435	1882
20	16	1300	14	60	221	547	2289
20	16	590	14	100	200	435	1878
20	16	1300	14	100	221	549	2287
20	16	590	25	0	296	630	2421
20	16	1300	25	0	366	718	2859
20	16	590	25	60	296	630	2402
20	16	1300	25	60	367	718	2821
20	16	590	25	100	296	628	2390
20	16	1300	25	100	367	718	2821

10	30	590	3	0	155	368	1439
10	30	1300	3	0	175	525	1690
10	30	590	3	60	156	333	1380
10	30	1300	3	60	167	541	1685
10	30	590	3	100	156	333	1350
10	30	1300	3	100	172	552	1712
10	30	1300	14	0	389	780	2640
10	30	1300	14	60	386	778	2622
10	30	1300	14	100	386	778	2624
10	30	1300	25	60	428	1073	3767
10	30	1300	25	100	445	1073	3752
20	30	590	3	0	162	307	1395
20	30	1300	3	0	191	438	1615
20	30	590	3	60	162	342	1393
20	30	1300	3	60	191	433	1603
20	30	590	3	100	162	338	1385
20	30	1300	3	100	191	425	1591
20	30	1300	14	60	403	857	2707
20	30	1300	14	100	403	851	2702

Table 3. Performance for each fine-tuned machine learning model.

CO_2 concentration (%)	Model	R^2	MSE
9	Gradient Boosting	0.9243	484.31
	Bagging	0.9234	579.21
	Random Forest	0.9227	585.02
	XGBoost	0.9129	672.75
	K nearest neighbors	0.5816	2799.65
	Multiple Linear Regression	0.5065	3255.94
	Support Vector Regression	0.2545	5250.17
4	Bagging	0.9201	3706.08
	Random Forest	0.9136	3909.26
	XGBoost	0.9038	4296.44
	Gradient Boosting	0.9007	4110.81
	K nearest neighbors	0.6126	14733.75
	Multiple Linear Regression	0.5612	16177.29
	Support Vector Regression	0.3787	24006.72
1	Gradient Boosting	0.9580	27563.63
	XGBoost	0.9487	34112.93
	Random Forest	0.9419	40308.39
	Bagging	0.9419	40364.34
	K nearest neighbors	0.8458	99575.83
	Multiple Linear Regression	0.7967	124326.65
	Support Vector Regression	0.6114	245217.11

Table 4. The hyperparameters for the best model for each scenario.

CO_2 concentration (%)	Machine learning approach	Setup
9	Gradient Boosting	Learning rate: 0.01 Max depth of trees: 9 Number of estimators: 250
4	Bagging	Number of estimators: 50
1	Gradient Boosting	Learning rate: 0.11 Max depth of trees: 8 Number of estimators: 75

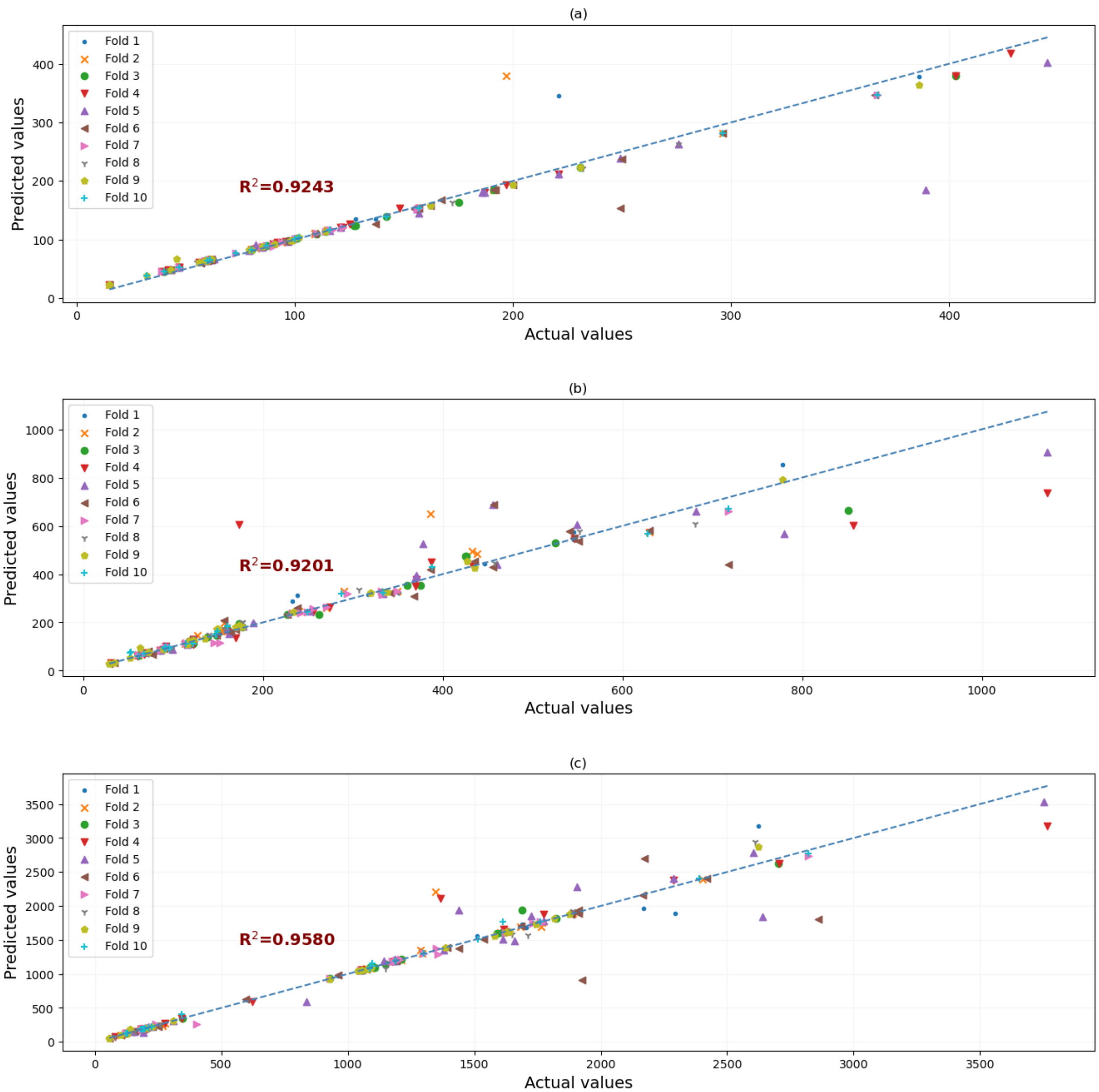

Figure 2. Actual vs. Predicted Values (10-fold cross validation) for (a) Distance for 9% CO_2, (b) Distance for 4% CO_2, and (c) Distance for 1% CO_2.

2. User-Defined Functions (UDFs) and User-Defined Real Gas Models (UDRGMs)

Because the scenarios of CO_2 pipeline release experience extreme pressure and temperature transition, the behavior of CO_2 in the process is complicated. Therefore, accurate formulas to represent the CO_2 behavior in extreme conditions are critical. The Span-Wagner equation of state and database covers the fluid region from triple-point temperature to 1100K at pressure up to 800 MPa. The database includes information such as density, enthalpy, entropy, and specific heat capacity, making it valuable for researchers and engineers working on processes involving CO_2, such as carbon capture and storage, refrigeration, and industrial applications. Additionally, NIST also established a website for accessing the data for further research and development.

In Ansys Fluent, UDF is a feature that allows users to customize or extend Fluent's capabilities by writing their own code in C or C++ to define specific functions or boundary conditions that are not available through Fluent's built-in options. While UDRGM allows users to create their own customized gas-phase reaction models in Fluent using User-Defined Functions (UDFs). Both UDFs and UDRGMs are powerful tools in Fluent that enable users to tailor their simulations to specific requirements and simulate a wide range of complex phenomena.

Therefore, we accessed the database from NIST and studied the application of UDFs and UDRGMs in Ansys Fluent. Therefore, we could have better simulation for CO_2 under extreme conditions.

3. Studied the influences of wind direction and pipeline location on sloped geometries

Simulating worst case scenario for CO_2 dispersion along a sloped terrain required us to explore different pipeline locations and wind direction. Depending on the pipeline location, it is possible that a terrain can slope down in more than one direction. We have observed this for the VB terrain (Figure 3). It can be observed from the contour that the slope is directed diagonally, which can influence the direction of the dispersion. Hence, we first investigated the effect of wind direction if the pipeline is situated at the middle of the lower slope. Figure 4 illustrates the effect of wind direction on CO_2 dispersion at the same atmospheric and pipeline conditions. It was observed that when the wind direction is parallel to the pipe (Figure 4A) the dispersion of CO_2 was farthest. If the wind direction is perpendicular to the pipe (Figure 4B and Figure 4C), the wind tends to induce mixing between air and CO_2. This prevents CO_2 from dispersing sideward and becomes diluted instead, especially if the direction of the wind is towards the downward slope (see Figure 4B). That being said, we also explored the effects of wind if it was coming from two directions. The aim is to mimic the wind blowing diagonally – following the downward slope of the terrain. For this, we have assigned wind to blow parallel to the pipe and another perpendicular to the pipe. Figure 4D depicts the effect of using wind blowing parallel to the pipe and wind blowing from the right side of the terrain. It was observed that CO_2 did not immediately mix with air (contrast to Figure 4B). Instead, CO_2 dispersion curved towards the left. Although this is a possibility, there is little accumulation at the base of the slope which indicates that CO_2 will not disperse and accumulate further from what is shown. Hence, it will not be considered a worst-case scenario. The same can be seen if the perpendicular wind direction came from the left (Figure 4E). Hence, we retain the original terrain, pipeline, and wind direction configuration for VB.

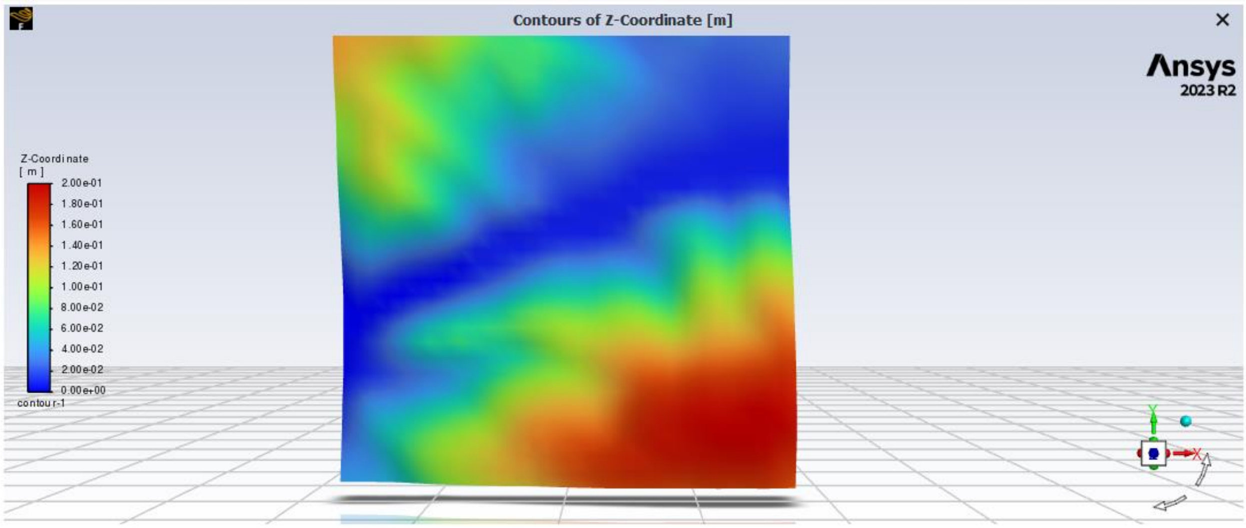

Figure 3. Contour of terrain VB depicting the highest part of the slope (red) and the lowest part of the valley (blue).

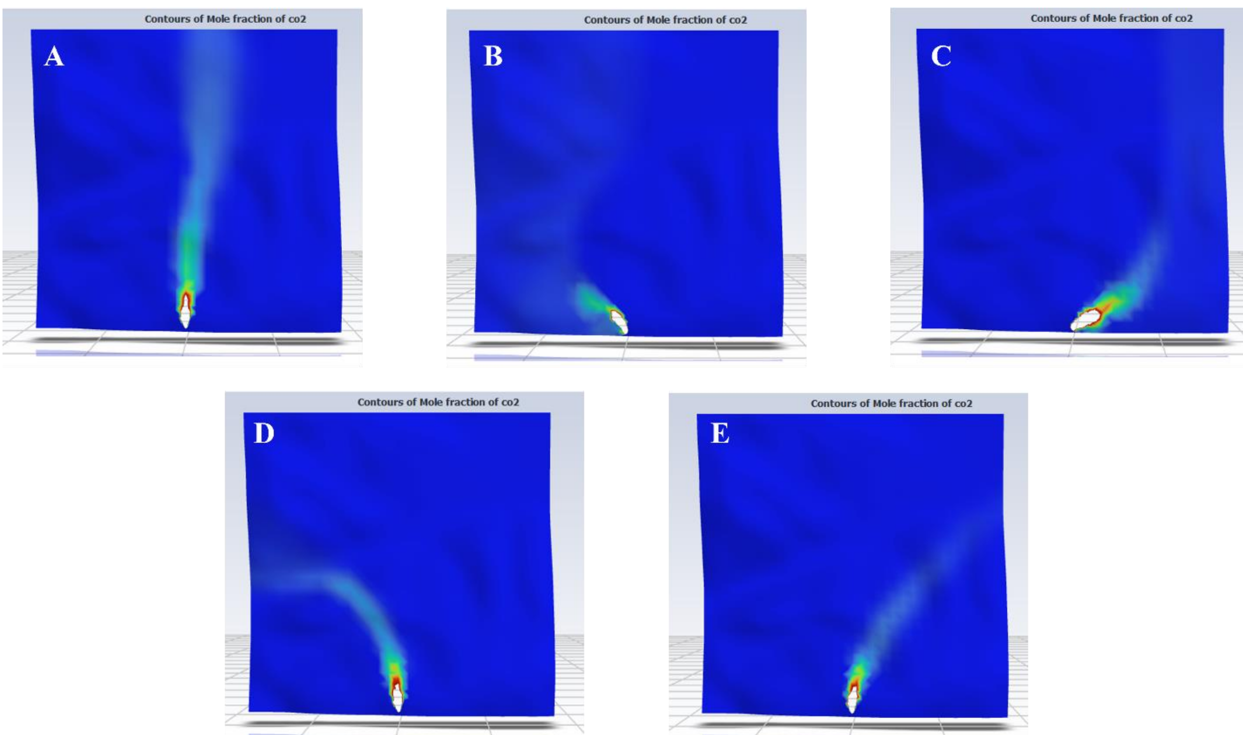

Figure 4. Dispersion of CO_2 with different wind direction configurations: A) wind is parallel to pipe, B) wind blows from the right side of the pipe, C) wind blows from the left side of the pipe, D) wind blows parallel to the pipe and from the right side of the pipe, E) wind blows parallel to the pipe and from the left side of the pipe.

From the dispersion behavior that was observed from VM, it was determined that positioning the pipeline parallel to the direction of the lowest part of the valley may also be considered a worst-case scenario. Thus, we plan to also perform simulations considering a secondary position of the pipe – pipe parallel to lowest part of the valley. In addition, the initial results for the small hill (SH) have shown that the direction of the slope has naturally allowed CO_2 to disperse diagonally. This could potentially indicate that having the wind direction follow the slope natural downward slope of the hill is also a worst-case scenario. Hence, this will also be investigated further. Table 3 outlines the additional cases that will be considered for simulations.

Table 3. Wind direction and pipeline position configuration.

Terrain	Wind Direction	Pipeline position	Notes
VB	1) Parallel to pipe 2) Perpendicular to pipe (right) 3) Perpendicular to pipe (left) 4) Parallel and perpendicular (right) to pipe 5) Parallel and perpendicular (left) to pipe	Middle of slope	Only perform parallel to pipe
	1) Parallel to pipe	Parallel to valley lowest point	
SH	1) Parallel to pipe 2) Parallel and perpendicular (right) to pipe	Middle of slope	