

Exploratory Analysis of Packing Effective Area Data for Carbon Capture

Team 7

Benjamin Drewry, Jorge Martorell, Yuying Wang

Final report

SDS 384 Scientific Machine Learning, Spring 2023

April 26, 2023

1 Introduction and Literature Review

Climate change is an increasingly difficult problem to solve. Anthropogenic greenhouse gas emissions have been steadily increasing since the industrial revolution, with an excess of 50 Gt of CO₂-equivalents now being emitted each year. With these emissions having been proven to contribute to global climate change, including the warming of the oceans and increased extreme weather, the need for curbing emissions to stop irrevocable harm to the planet is more important than ever. Emissions will need to be reduced by around 70-80 Gt CO₂ equivalents per year in order to possibly keep global warming below 2 degrees Celsius. To accomplish this, Carbon Capture and Sequestration (CCS) technologies need to be implemented at power generation and heavy industry sites that emit the most CO₂ (IPCC 2014).

The most mature and ready-to-use capture technologies are amine-based processes. These processes utilize an aqueous amine solvent, which is water blended with some form of amine such as monoethanolamine (MEA), piperazine (PZ) or a blend of amines such as CESAR-1 (Vega et al. 2020). This solvent is fed to a column where it countercurrently contacts an inlet gas stream, for example the flue gas from a natural gas-fired power plant. The CO₂ in the flue gas stream reacts with the amine, absorbing the CO₂ from the gas phase to the liquid phase. The CO₂-rich solvent is then pumped to another column where the CO₂ is stripped into a CO₂-rich gas phase by heating with steam. The purified CO₂ gas is then compressed and sent off for another use or for geologic sequestration, while the CO₂-lean solvent is recirculated to the absorber column.

1.1 Column Packing Internals

A key part of these CCS systems is the packing in the internals of the absorption and stripping columns. Packing is typically either ceramic or metal rings randomly thrown into the column (referred to as random packing), or elements of sheet metal that are corrugated to create flow channels for both the liquid and gas (referred to as structured packing). Packing increases the mass transfer efficiency within a process column by increasing the likelihood of gas-liquid contact by directing the flow of gas and liquid into the void spaces between the packing elements. The wetting of the packing by the liquid also increases the surface area of vapor-liquid contact. These packings, shown in Figure 1, can be described by a number of physical variables, such as the void fraction of random packing (the fraction of void space within a given volume), the corrugation height, flow channel length/width, and corrugation angle of the structured packing element.

The characterization of packing types is typically performed at pilot-scale development units via wetted wall column experiments in order to define the most important variables: the effective wetted area of the packing (a_e), the gas-film mass transfer coefficient (k_G), and the liquid-film mass transfer coefficient (k_L). The effective wetted area is the actual amount of packing contributing to gas-liquid mass transfer, which is typically most strongly influenced by the geometry of the packing itself and the operating conditions. Both k_L and k_G are used to describe mass transfer at the film interfaces between the bulk gas and liquid phases. An example of gas and liquid concentration profiles predicted by two-film theory is shown in Figure 2. Similar to a_e , k_L and k_G are also affected by the column and packing geometries as well as the operating conditions. The operating conditions considered for the

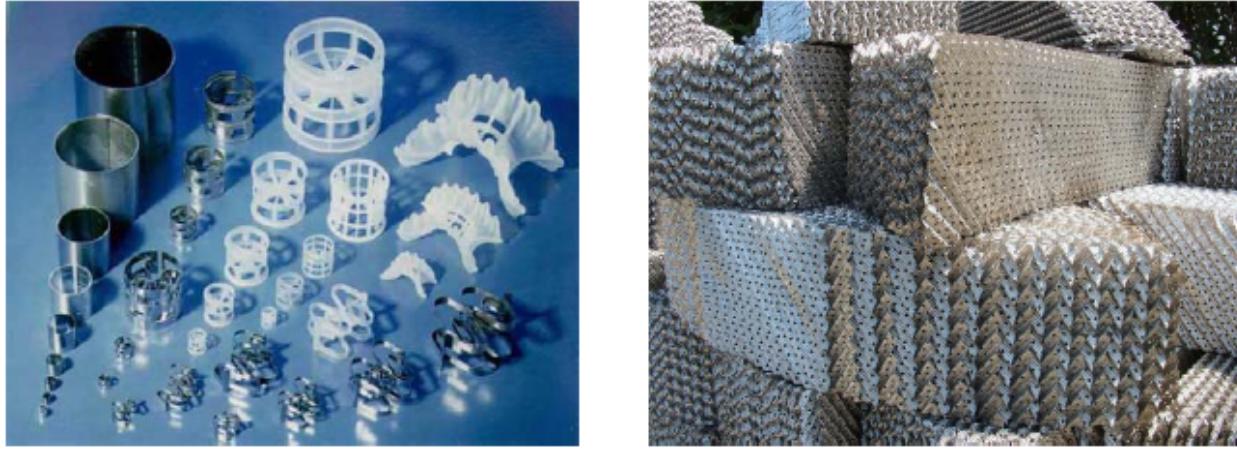


Figure 1: Examples of random (left) and structured (right) packings. Figure from ([Song 2017](#)).

characterization of these parameters typically are the gas and liquid flowrates, densities and viscosities, the temperature within the column, and the column pressure.

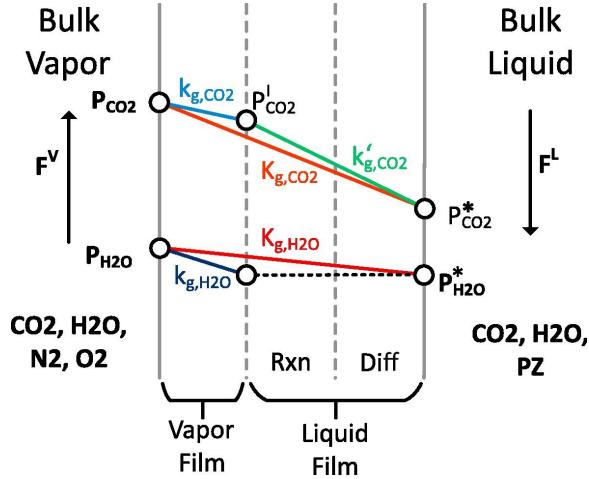


Figure 2: Two-film theory describing change in CO_2 and H_2O concentrations at the gas-liquid interface of a CO_2 absorber. Mass transfer is a function of mass transfer coefficients k_G and k_L (equivalent to k'_g in the figure) as well as contact area a_e . Figure from ([Tsay et al. 2019](#)).

The experiments that are typically performed to characterize these parameters can be expensive. Air-water column experiments are used to develop correlations for both a_e and k_L , while k_G is typically derived from experiments investigating the absorption and vaporization of solutes from carrier gases, such as SO_2 from air via NaOH ([Song 2017](#)).

1.2 Literature Review of Packing Correlations

Major hurdles to accurately characterizing any given packing are economic in nature: these are the cost associated with manufacturing and purchasing of the packing itself, the installation of said packing in a column, and the requirement of having a column built for the

packing to be tested in. This can also be very time-intensive to run hundreds of experiments over a wide variety of feasible operating conditions, and expensive to collect data with the proper instrumentation. Thus, it is preferable to develop generalized packing correlations for a wide variety of packings and operating conditions to produce universal correlations which can predict performance with reasonable accuracy.

Onda

Packing and mass transfer coefficient models have existed in some form since the 1960s, with the first significant model developed by Onda in 1968 for water-like random packing systems (Onda et al. 1968). This model predicted effective area as a function of liquid absorption, reaction rate, liquid concentration and diffusivity:

$$a_e = \frac{k_L^o a}{\sqrt{k_r C_B D_L}} \quad (1)$$

However, this correlation was developed using only random packings with relatively large surface areas, which led to the correlation underpredicting effective wetted area for smaller packings.

Bravo-Rocha-Fair, Billet-Schultes

Subsequent developments in packing correlations lead to the Bravo-Rocha-Fair (BRF) and Billet-Schultes correlations. The former correlation was developed by the Separations Research Program at the University of Texas at Austin (UT-SRP) between 1982 and 1996 (Rocha et al. 1993). Originally based on commercial packed distillation data, it was expanded to encompass data from an experimental database developed by Bolles and Fair (Bolles and Fair 1979) detailing a wide range of conditions, and gradually updated by changing the order of certain constants. A form of the updated effective area correlation is shown below:

$$\frac{a_e}{a_p} = F_S E \frac{29.12 u_L^{0.4} v_L^{0.2} S^{0.359}}{(1 - 0.93 \cos \gamma)(\sin \alpha)^{0.3} \varepsilon^{0.6}} \left(\frac{\rho_L}{\sigma g} \right)^{0.15} \quad (2)$$

Billet-Schultes, similar to BRF, was developed using experimental packing data from a wide range of chemical systems and packings (Billet and Schultes 1999). A similar correlation was developed from dimensional analysis of parameters.

Modern UT-SRP Models: Tsai, Wang, Song

The aforementioned models were not specifically developed for carbon capture applications. In 2010, Tsai measured nine structured packings and their ability to absorb CO₂ from air using 0.1 M NaOH, and developed an effective wetted area model as a function of gas and liquid flowrate, surface tension, and viscosity. This model represented the entire database within ±15% (Tsai 2010).

Wang built on this model in 2015, adding structured and random packings to the data gathered by Tsai and again measuring a_e via the absorption of CO₂ from air using an aqueous NaOH solution. A key improvement was the inclusion of packings with surface areas (a_p)

ranging from 125 to 500 [m^2/m^3] and different corrugation angles. Wang concluded that the specific packing area has a significant effect, while corrugation angle has a minimal effect on effective wetted area. This updated correlation predicts a_e within an error of $\pm 11\%$ (Wang 2015).

The current iteration of these models is the Song model (Song 2017). Additional structured and random packings were tested and added to the experimental database developed by Wang and Tsai. Fine tuning of the order of parameters resulted in the lowest error of $\pm 8.9\%$. The three forms of effective wetted area calculations developed by Tsai (3), Wang (4), and Song (5) are shown in the following equations:

$$a_{e,Tsai} = 1.34a_P \left[\left(\frac{\rho_L}{\sigma} \right) g^{1/3} \left(\frac{Q}{L_p} \right)^{4/3} \right]^{0.116} \quad (3)$$

$$a_{e,Wang} = 1.41a_P \left[\left(\frac{\rho_L}{\sigma} \right) g^{1/3} \left(\frac{u_L}{a_p} \right)^{4/3} \right]^{0.116} \quad (4)$$

$$a_{e,Song} = 1.16\eta a_P \left[\left(\frac{\rho_L}{\sigma} \right) g^{1/2} u_L a_p^{-3/2} \right]^{0.138} \quad (5)$$

1.3 Literature Review of Machine Learning for Area Correlation

A recent paper investigated the application of supervised machine learning algorithms for estimating the liquid mass transfer coefficient and effective interfacial area for bubble column reactors (Hazare et al. 2022). The authors extracted data for parameters such as gas velocity, column dimensions, a_e , k_{La} , and physiochemical system properties from plots in published papers. Independent variables included column diameter, liquid height, free area, gas and liquid velocities, pressure, temperature, densities and diffusion coefficients for the prediction of the output variables of k_{La} and a_e . Data was then scaled using the `MinMaxScaler` available within `sklearn` and split into 80/20 train/tests splits, and fed to Support Vector Regression (SVR), Random Forest (RF), and Artificial Neural Network (ANN) models. The tuning of each model's hyperparameters was performed via a grid search optimization method. Their optimized ML models were able to learn the predictor-prediction relationship with high accuracy, with the lowest R^2 value reported being 0.95 for the ANN model. Similar accuracy was achieved for the overall mass transfer coefficient models, with the worst model being the ANN, with an R^2 of 0.95 and an MSE of 1×10^{-3} . Although this paper did not specifically look at carbon capture applications, as bubble columns are multiphase reactors that can be used for applications such as N₂ sparging, the overall approach is very much applicable to the systems studied in this report.

2 Data

The dataset for this work was originally compiled as part of a review of k_{La} correlations (Abreu and Rochelle 2022). The majority of the measurements come from experiments conducted at the University of Texas Separations Research Program (UT-SRP) pilot facility (Tsai 2010; Wang 2015; Song 2017). Each of the three UT-SRP authors conducted bench-scale wetted wall column (WWC) and pilot-scale air-water column (AWC) experiments. The data of interest for this project comes from a_e measurements conducted by installing various different types of packings in the AWC and then measuring the rate of absorption of CO₂ into

aqueous NaOH solution. Each author collected slightly different types of measurements to suit different research goals (e.g. Song measured alkalinity instead of [OH-], Tsai measured ΔP). Therefore, there are a total of 23 possible features for each observation, but most observations are missing some values. Additionally, some packing geometry features like corrugation angle and characteristic dimensions are only applicable to structured packing and not random packing. The UT-SRP a_e dataset contains a total of 39 different packings. Most of these are standard corrugated metal structured packings, but a few random (i.e. dumped) and hybrid packings are also included as well as one grid (very coarse structured) and one gauze (very fine structured) packing.

Additional a_e data was collected from four other studies not affiliated with UT. One of these studies (Zakeri et al. 2011) used a similar methodology while the others instead used X-ray tomography to image cross sections of a packed air-water column and then applied image segmentation to calculate a_e (Aferka et al. 2010; Janzen et al. 2013; Bolenz et al. 2019). Due to the different methodology, these experiments did not measure the same set of features. Therefore, a reduced feature set (a_p, L, u_G, T) was used in fitting some of the models to allow for comparison to this dataset.

Figure 3 shows histograms of the numeric features of the complete dataset. There are a total of 983 observations from UT-SRP (orange) and 276 observations from the four other sources (blue). There are 21 numeric inputs (features) and 1 output (fractional wetted area, Φ). Fractional wetted area is defined as $\Phi \equiv a_e/a_p$ and can exceed 1.0 since it is possible to have $a_e > a_p$. There are also 2 non-numeric features, packing type and author name, which are used for visualization but not as part of model training.

3 Exploratory Analysis & Generating Hypotheses

The fractional area data appears to be approximately normally distributed over the range from 0 to 1.5. This means that the majority of the data points fall around the mean, with fewer and fewer points as we move away from the mean towards the edge. This uniform distribution is advantageous for model fitting, as there are no outlier of Y values to contend with. For most variables, the data points from non UT-SRP sources show a similar distribution to those from UT-SRP. However, the temperature variable was an exception to this trend, showing some variation in its distribution across the different sources. The similar distribution across most of the variables set the foundation to use data from both sources to train and test our developed models. However, it is important to note that there are 11 variables for which there are no data points from non UT-SRP sources, and these variables must be excluded from the combined dataset fitting. Figure 4 displays a correlation matrix heat map for all variables from the UT-SRP dataset. The focus of our investigation is on the relationship between the fractional area and other variables. Our analysis reveals that the specific area exhibits a strong negative correlation with the fractional area, as indicated by a correlation coefficient of -0.65. Similarly, the packing element height also shows a negative correlation with the fractional area, with a correlation coefficient of -0.62. On the other hand, a positive correlation with a coefficient of 0.68 is observed between the fractional area and the liquid flux rate. These findings provide valuable insights into the interplay between these variables and pave the way towards the development of predictive models for optimization.

Our research hypothesis posits that specific area, packing element height, and liquid

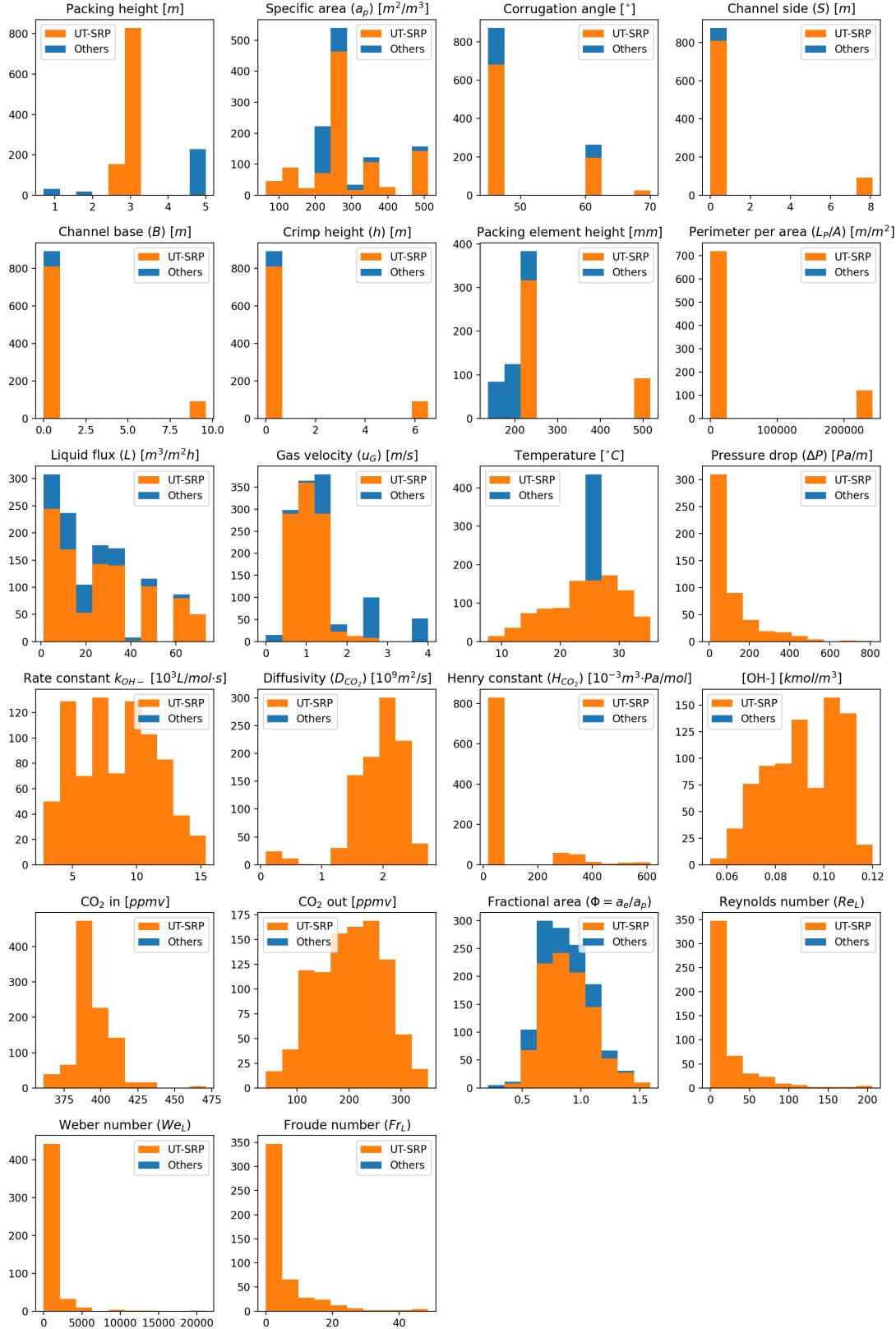


Figure 3: Histograms of numeric features in the complete dataset.

flux rate may be considered as potential influential features with a substantial impact on the target variable, fractional area Φ . Artificial Neural Network (ANN), Random Forest Regression and Support Vector Regression (SVR) are performed to determine the effective wetted area followed by feature importance analysis.

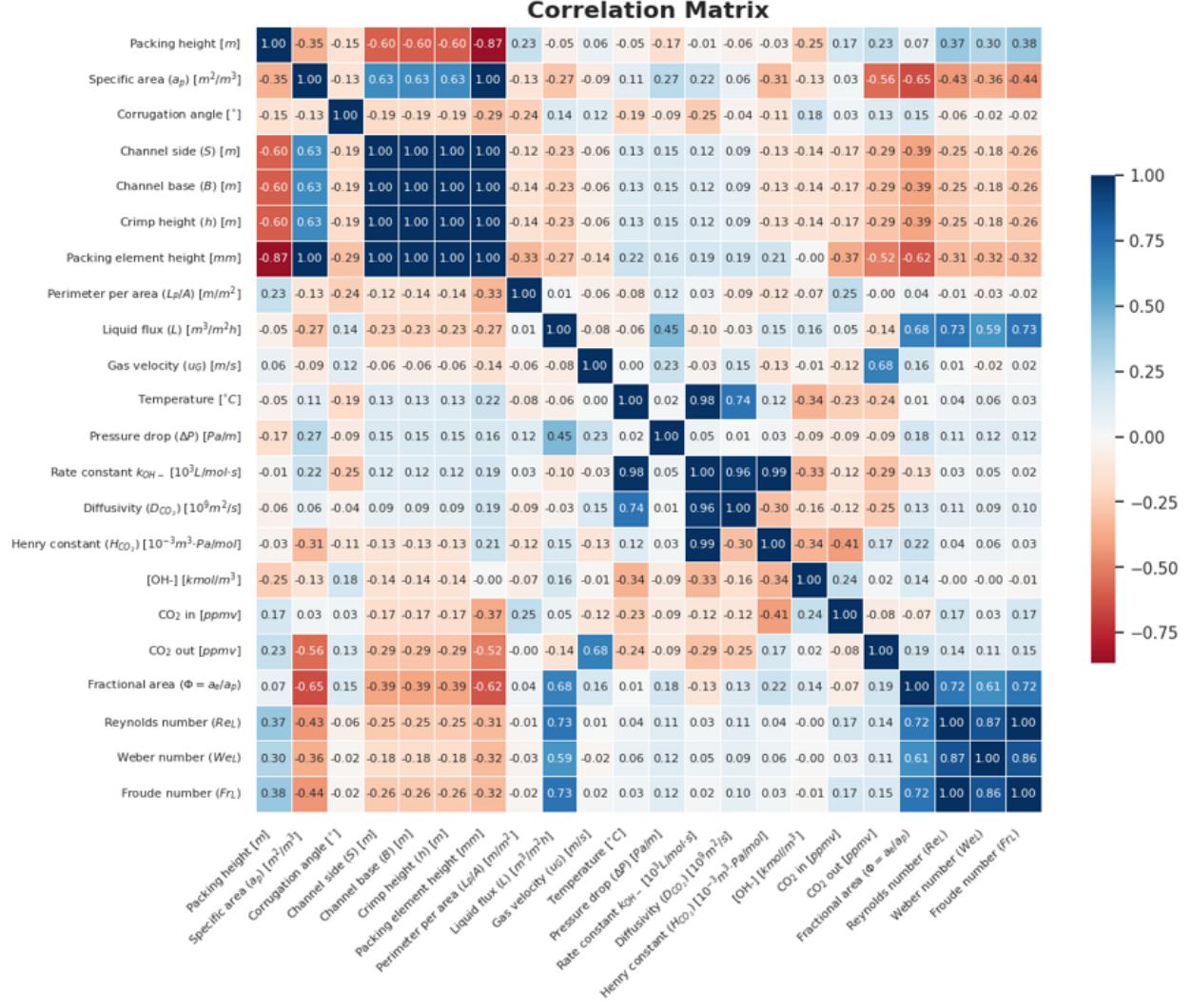


Figure 4: Correlation matrix heat map for the complete dataset.

4 Modeling & Validation

4.1 Artificial Neural Network (ANN)

After data pre-processing and scaling via `MinMaxScaler`, three different models were developed: the base model, the reduced model, and the “best” model. The base model used as many inputs variables as possible for the dataset to be looked at, the reduced model only used L , u_G , and T as input variables for predicting a_e , while the “best” model is the base model with the hyperparameters optimized via `GridSearchCV`. The overall dataset was stratified into three different versions: one for UT-SRP random packing tests, one for UT-

SRP structured packing tests, and one with all data (UT-SRP and others). For the random packing experiments, void fraction is the only packing-specific variable passed to the model compared to the corrugation angle, channel base width and height of the structured packing. For the combined dataset, the lack of variables reported by the additional papers meant that the only variables able to be passed to the ANN models were packing height, packing specific area, L , and u_G . Thus, there is no reduced model developed for the combined dataset.

The default parameters used for the base and reduced ANN models are shown in Table 1, while the parameter grid used for grid search optimization is shown in Table 2.

Solver	Activation Function	Hidden Layer	α
LBFGS	Logistic	[700,700]	$1 \times e^{-5}$
Random State	Maximum Iterations	Learning Rate	
1	10,000	$1.8 \times e^{-4}$	

Table 1: Default Parameters used for base and reduced ANN models

Hidden Layer Sizes	(100),(100,100),(100,100,100)
Activation Function	'Identity', 'Logistic', 'Tanh', 'ReLU'
α	$1 \times e^{-5}, 1 \times e^{-4}, 1 \times e^{-3}$
Learning Rate	0.1, 0.01, 0.001

Table 2: Grid used for ANN hyperparameter optimization

For each of the datasets, each of the three models were trained on a random 80/20 train/test split of the datasets for 50 iterations, with the R^2 , MAE, RMSE, MSE, and explained variance scores saved for each iteration. The best model’s hyperparameter optimization was conducted prior to the 50 iterations in the interest of time. Mean values for each of these statistical metrics were computed. The importance of each feature fed to the various models was ascertained via the permutation importance function within `sklearn`, which is defined as the decrease in the score of a machine learning model when a singular feature value is randomly altered. For visualization purposes, each set of predicted test and train outputs were visualized on a parity plot using error margin lines for a 95% confidence interval, or dashed lines signifying a static 95% confidence interval.

Additional Work, Structured Packing Dataset

For the structured packing dataset, the best model was trained on a smaller subset of the structured UT-SRP dataset variables, corresponding to those reported by the non-UT-SRP sources. The non-UT-SRP dataset was then fed to the trained model and scored based on how well it predicted the reported a_e values using R^2 and a parity plot for visualization purposes.

Partial dependence plots, which allow for visualization of the interaction between the output variable and either a single or two input features, and a SHAP plot, which allows for visualization of each feature towards the overall model prediction, were also generated for the

outputs of the best model trained on the original UT-SRP structured packing dataset. The variables visualized are L , u_G , T , CO₂ diffusivity, specific area of packing, and corrugation angle. The combination of variables visualized in the PDPs are corrugation angle – specific area, L – specific area, u_G – specific area, and $u_G - L$.

4.2 Random Forest

The dataset used for the Random Forest Regressor underwent data scaling using the transformers in the `sklearn` library, similar to the pre-processing of datasets in the ANN model. This was done to prevent the model from giving more weight to larger values and less weight to smaller values due to the varying magnitude of the predictors. The `MinMaxScaler` was specifically used to rescale the dataset, bringing it into the 0 to 1 range, which reduces the impact of the actual values. One advantage of using `MinMaxScaler` is that it does not affect the relationship between the data and the target or predictors. In addition, missing values were filled using the mean value of each feature. The study involved the development of three distinct models: the base model, the reduced model, and the optimized model. The base model used the majority of input variables in each specific dataset, while the reduced model used only L , u_G , and T as input variables as we did in the case of ANN model training. To identify the "best" model, hyperparameter tuning was conducted using `GridSearchCV`. The base models including the majority of input variables were then trained on three different versions of the dataset: UT-SRP random packing tests, UT-SRP structured packing tests, and the combined data from all UT-SRP tests. For each dataset, the three models were trained on an 80/20 train-test split for 50 iterations, and the R^2 , MAE, and MSE were calculated and recorded accordingly in each iteration. The mean values and standard deviations of these statistical metrics were calculated. Additionally, the permutation importance function within `sklearn` was used to determine the importance of each feature for the various models. To visualize the results, each set of predicted test and train outputs was plotted on a parity plot with the actual value and the predicted value on the x axis and y axis, respectively, including the 0.2 error margin lines. The default parameters used for the base and reduced Random Forest models are shown in Table 3, while the parameter grid used for grid search optimization is shown in Table 4.

Max Depth	Number of Estimators	Criterion
20	75	friedman mse

Table 3: Default Parameters used for base and reduced Random Forest models.

Number of Estimators	50,100,150
Max Depth	10, 20, 30
Max Features	'auto', 'sqrt', 'log2'
Min Samples Leaf	1, 2, 4

Table 4: Grid used for Random Forest hyperparameter optimization.

4.3 Support Vector Regression (SVR)

Three support vector regression models were trained using only structured packing experiments in the UT-SRP portion of the dataset, a total of 902 observations. All numeric input features are scaled to the range [0,1] using `MinMaxScaler` and then missing values are imputed using the mean value of each feature. Features which are missing an excessive amount of data ($> 20\%$) are removed, leaving 11 numeric features. The “base” SVR model is trained using all of these features, using a randomly sampled 80%/20% training/testing split. Goodness-of-fit on both the training set and testing set are measured using R^2 , MSE, and MAE. The random train/test split and model fitting is repeated 100 times to calculate mean and standard deviation of the goodness-of-fit criteria.

Next, a “best” model is identified by feature selection followed by hyperparameter tuning. Out of the 11 available features, it is assumed based on domain knowledge that a_p , L , and u_G are the most important predictors of a_e . These three features are fixed and must be included in the best model. All possible combinations of the three fixed features plus up to three additional features are then generated, giving 93 combinations ($\binom{8}{3} + \binom{8}{2} + \binom{8}{1} + \binom{8}{0}$). The feature selection algorithm then re-trains the SVR model over 5 random 80%/20% train/test splits to select the feature combination giving the lowest MSE as the “best” combination. Then the optimal hyperparameters are selected by grid search using `GridSearchCV` over the range [0.1,30] for C (regularization parameter), [0.001,10] for ϵ , and over three choices of kernel (RBF, linear, or polynomial).

The final model is a “reduced” model which takes only a_p , L , u_G , and T as inputs. These features are the most complete in the dataset both for UT-SRP experiments and non-UT experiments, which allows predictions from this model to be compared to the non-UT data. This is also a more practical set of inputs which would be expected to be easily available if this regression model were applied in a process modeling application. This model is trained using only UT-SRP data similar to the other models. C, ϵ , and kernel are optimized by grid search similarly to the best model.

5 Results

5.1 ANN

Figures 5, 6, and 7 show parity plots for the reduced, base, and best models for the UT-SRP structured packing, UT-SRP random packing, and combined packing datasets, respectively. Feature importance through permutation analysis is displayed in Figures 8, 9, and 10 for the structured, random and combined datasets.

Statistical metrics for the structured, random, and combined datasets are shown in Tables 5, 6, and 7. The optimized hyperparameters as determined by the grid search method are show in Table 8 for the structured, random, and combined datasets.

A ln-ln plot of theoretical structured packing data overlaid with predictions from Song, Wang, and the best ANN model are shown in Figure 11. Prediction results for the non-UT-SRP packing data on the best UT-SRP structured packing model are shown in Figure 12 as a parity plot, with the R^2 for prediction of the a_e overlaid on the graph.

Single and double variable PDP, and SHAP plots are displayed for the best structured packing model in Figures 13, 14, and 15.

Model	R^2	MAE	RMSE	MSE
Base, Train				
Mean	0.74	0.065	0.085	0.007
Standard Deviation	0.008	0.0008	0.001	0.0002
Base, Test				
Mean	0.72	0.065	0.085	0.007
Standard Deviation	0.03	0.003	0.004	0.0007
Best, Train				
Mean	0.97	0.019	0.025	0.0006
Standard Deviation	0.006	0.002	0.003	0.0002
Best, Test				
Mean	0.94	0.03	0.039	0.0016
Standard Deviation	0.01	0.002	0.004	0.0003
Reduced, Train				
Mean	0.51	0.088	0.11	0.013
Standard Deviation	0.016	0.0013	0.0017	0.0004
Reduced, Test				
Mean	0.51	0.089	0.11	0.013
Standard Deviation	0.06	0.0051	0.0066	0.0015

Table 5: Statistical Metrics for ANN models trained on structured packing data.

Model	R^2	MAE	RMSE	MSE
Base, Train				
Mean	0.96	0.044	0.054	0.0029
Standard Deviation	0.0042	0.0022	0.0023	0.00025
Base, Test				
Mean	0.94	0.049	0.06	0.0038
Standard Deviation	0.022	0.0079	0.011	0.0014
Best, Train				
Mean	0.999	0.0042	0.0053	$3 \times e^{-5}$
Standard Deviation	0.00026	0.0011	0.0014	$1.7 \times e^{-5}$
Best, Test				
Mean	0.986	0.021	0.027	0.00085
Standard Deviation	0.0096	0.0067	0.0098	0.00058
Reduced, Train				
Mean	0.657	0.12	0.149	0.022
Standard Deviation	0.034	0.0071	0.0063	0.0019
Reduced, Test				
Mean	0.614	0.12	0.15	0.024
Standard Deviation	0.14	0.021	0.025	0.0075

Table 6: Statistical Metrics for ANN models trained on random packing data.

Model	R^2	MAE	RMSE	MSE
Base, Train				
Mean	0.598	0.074	0.097	0.0094
Standard Deviation	0.0128	0.0009	0.0015	0.0003
Base, Test				
Mean	0.593	0.075	0.097	0.0095
Standard Deviation	0.037	0.0031	0.0042	0.0008
Best, Train				
Mean	0.805	0.051	0.0068	0.0046
Standard Deviation	0.0157	0.0021	0.0026	0.0003
Best, Test				
Mean	0.757	0.057	0.075	0.0057
Standard Deviation	0.0277	0.0027	0.0038	0.0006

Table 7: Statistical Metrics for ANN models trained on combined packing data.

Parameter	Structured	Random	Combined
Activation Function	ReLU	ReLU	ReLU
α	0.001	0.001	0.0001
Hidden Layer Sizes	(100,100)	(100,100,100)	(100)
Learning Rate	0.001	0.001	0.001

Table 8: Optimal Hyperparameters for structured, random and combined datasets as determined by grid search method.

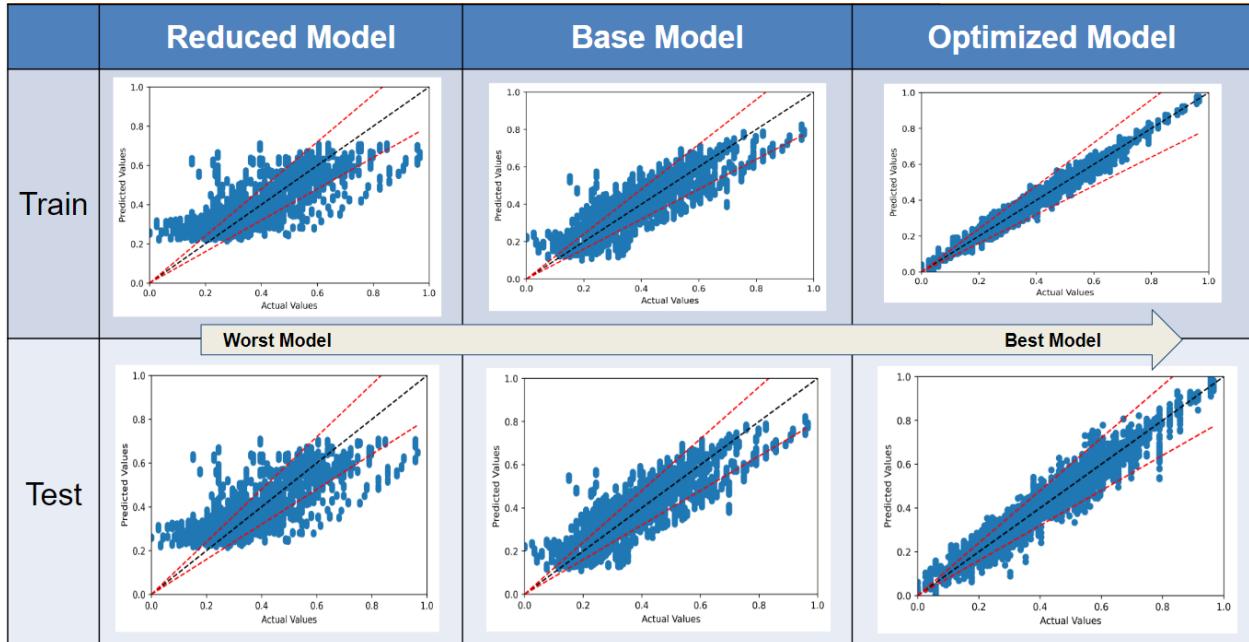


Figure 5: Parity plots for UT-SRP structured packing dataset trained on ANN models.

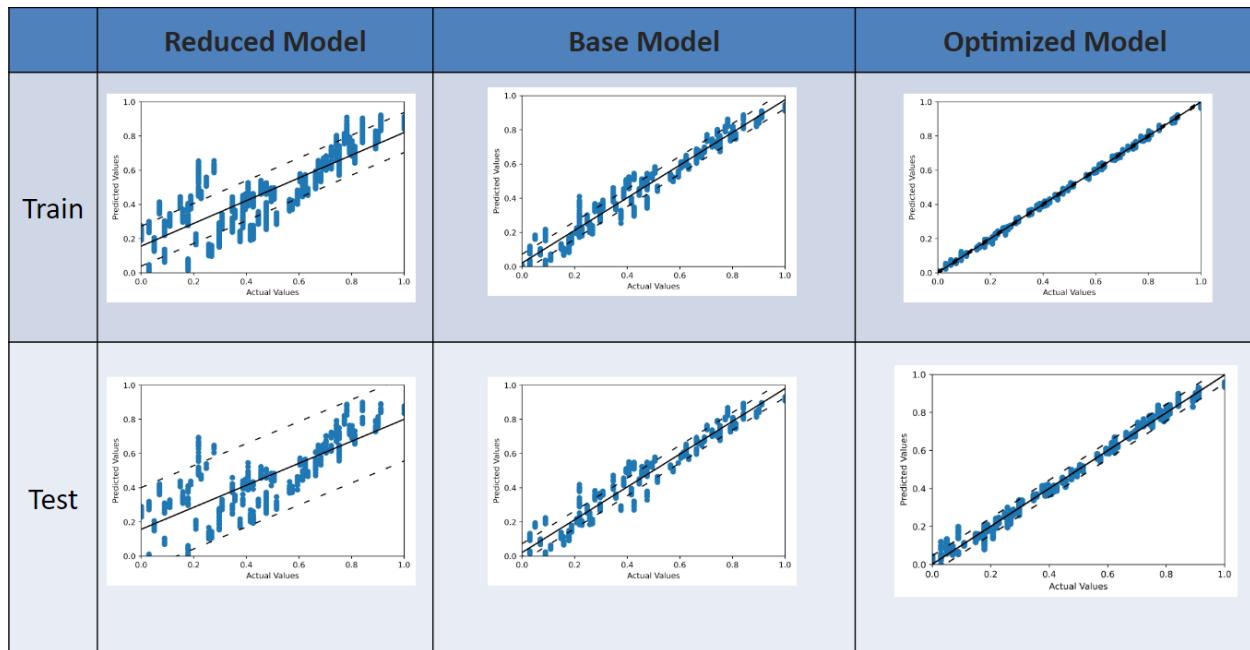


Figure 6: Parity plots for UT-SRP random packing dataset trained on ANN models.

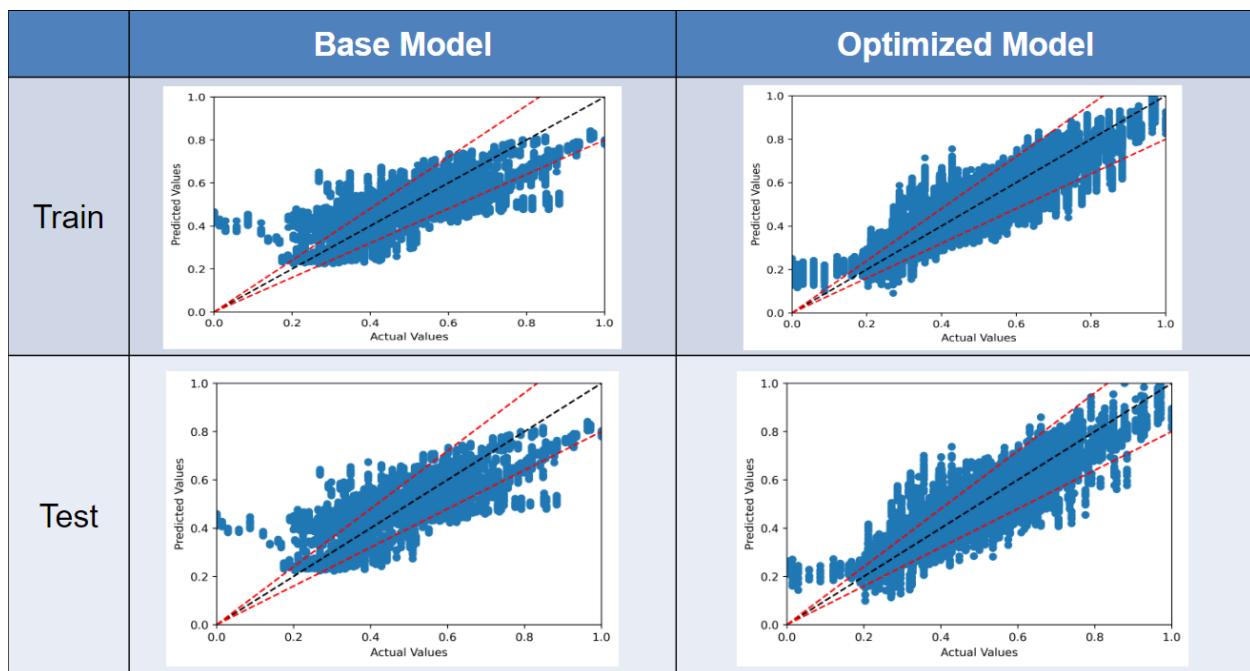


Figure 7: Parity plots for combined packing dataset trained on ANN models.

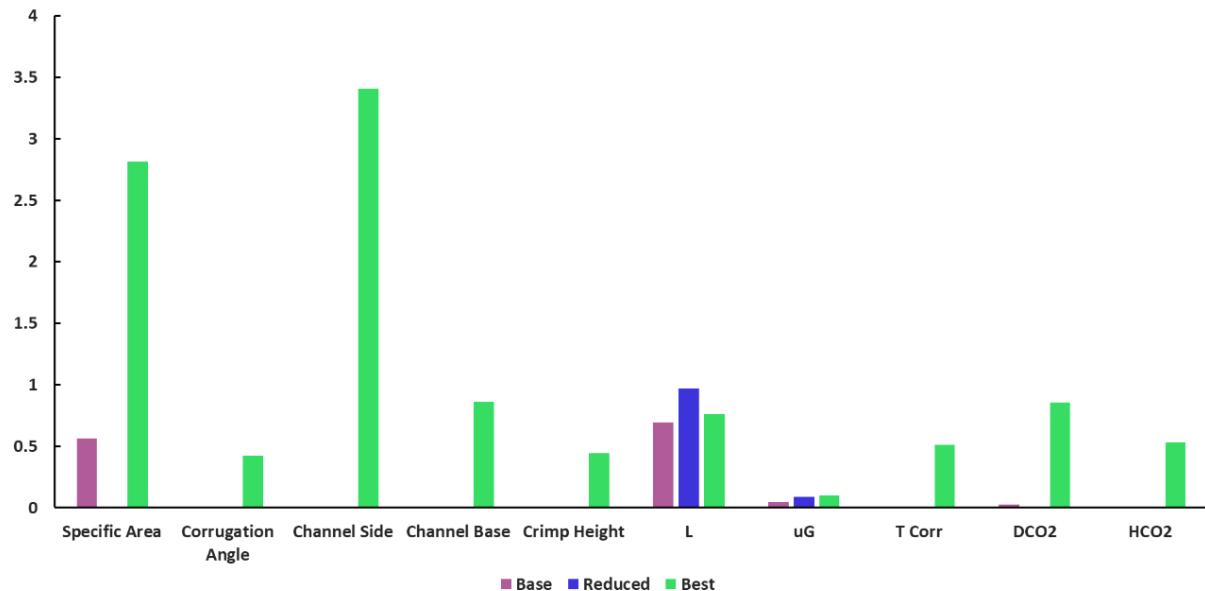


Figure 8: Feature importance via permutation importance for variables fed to ANN models trained on UT-SRP structured packing dataset.

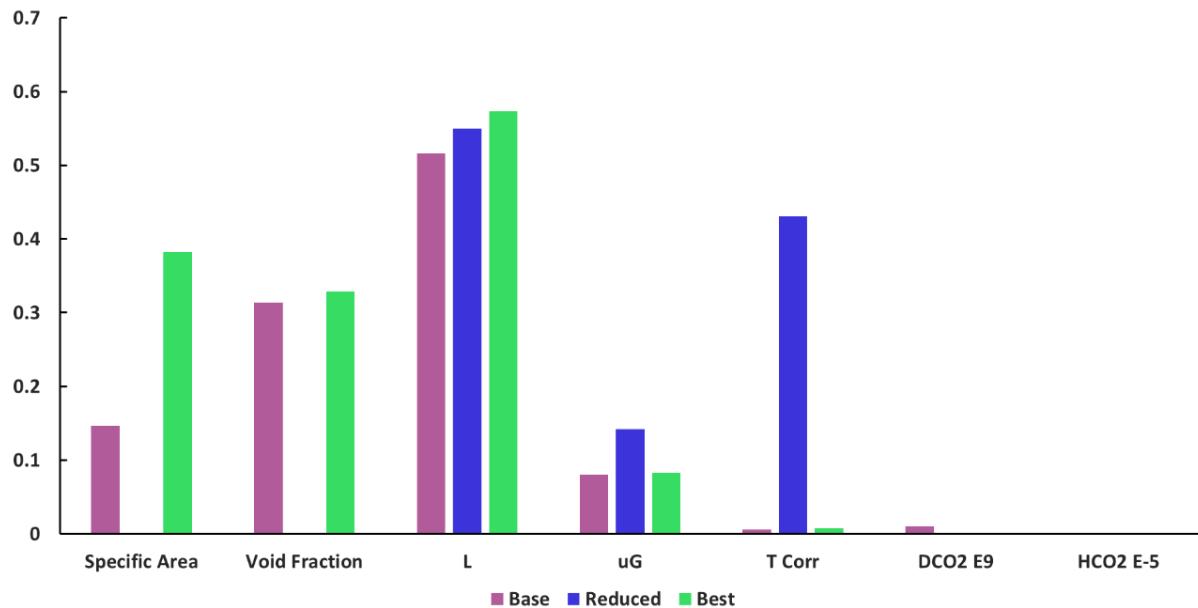


Figure 9: Feature importance via permutation importance for variables fed to ANN models trained on UT-SRP random packing dataset.

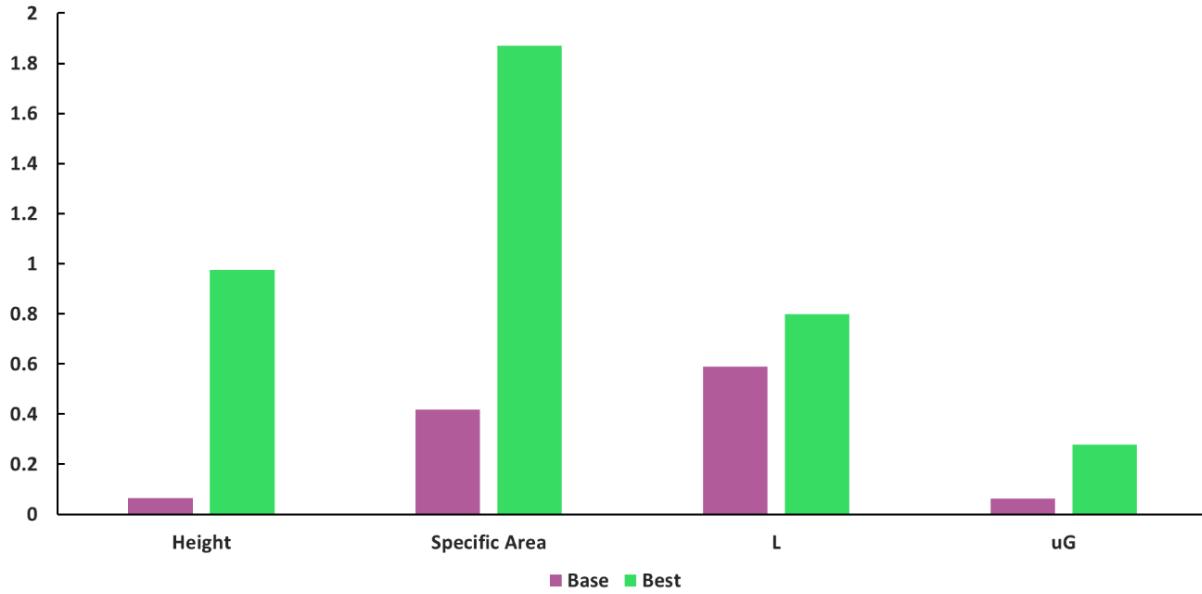


Figure 10: Feature importance via permutation importance for variables fed to ANN models trained on combined packing dataset.

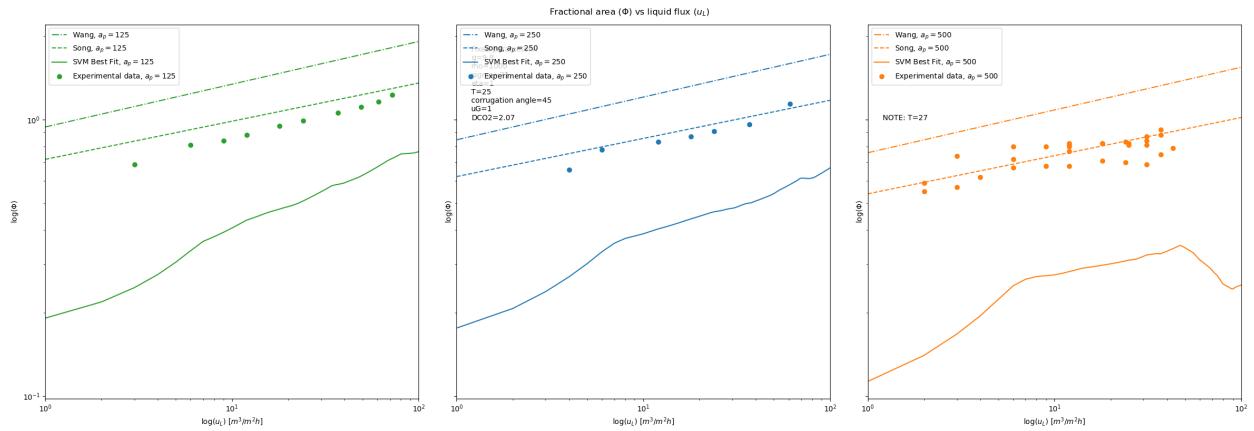


Figure 11: Visual comparison of Song, Wang, and ANN model for predicting theoretical structured packing data.

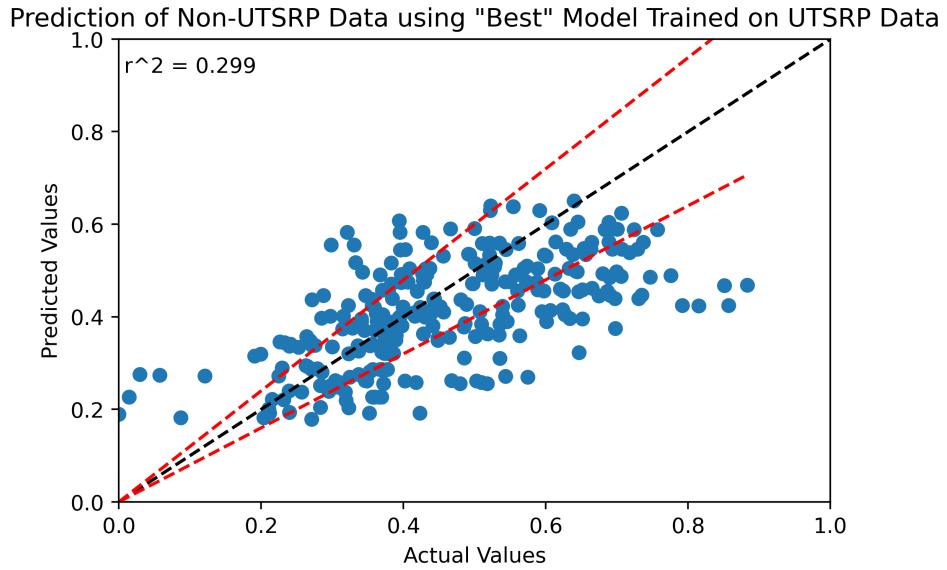


Figure 12: Parity Plot of predicted versus actual a_e values for non-UTSRP packing data trained on best structured UT-SRP model.

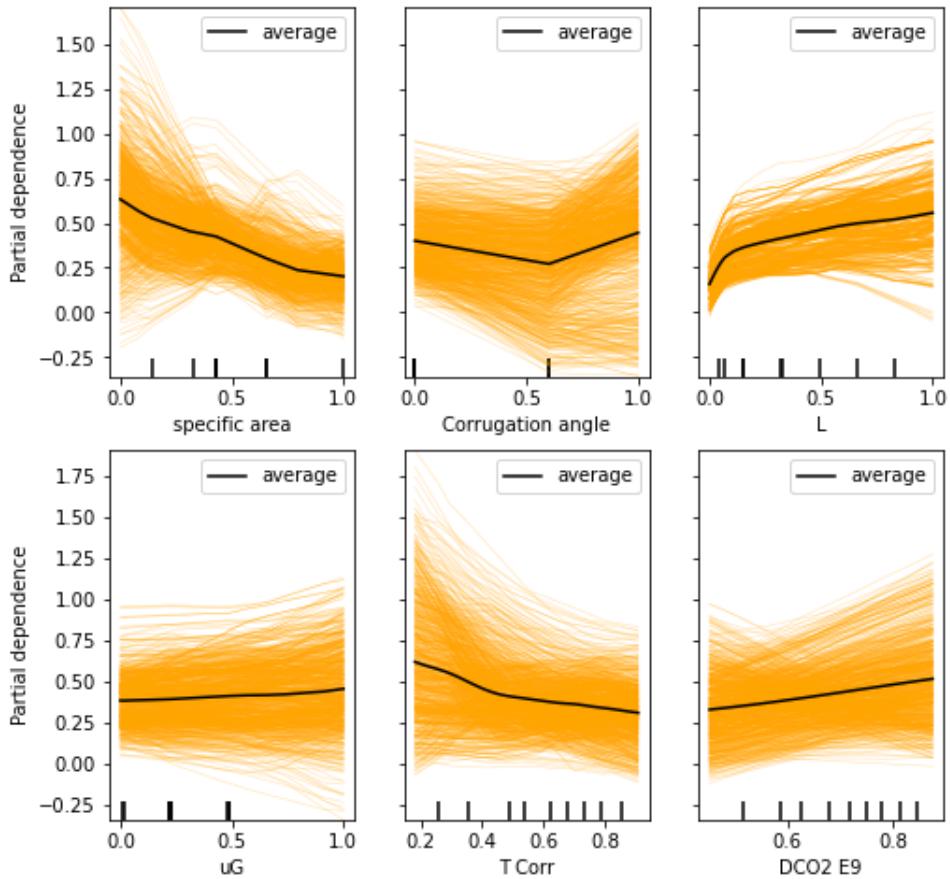


Figure 13: Single Variable PDP for Structured Packing dataset, Best Model.

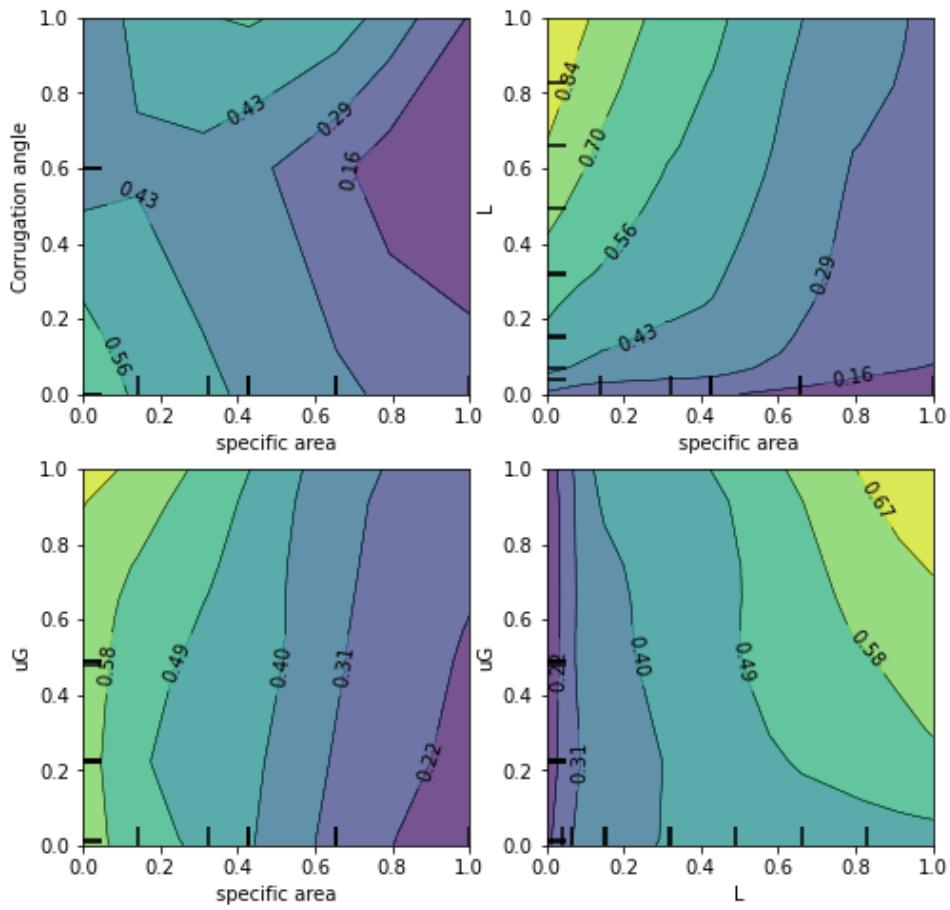


Figure 14: Double Variable PDP for Structured Packing dataset, Best Model.

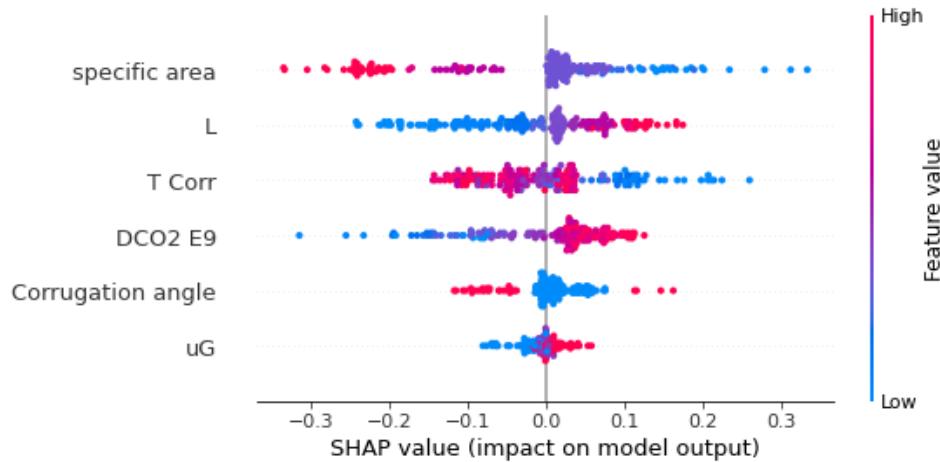


Figure 15: SHAP plot for Structured Packing dataset, Best Model.

5.2 Random Forest

Figure 16 shows the parity plots of the reduced, base, and optimized models for the UT-SRP structured packing dataset. The base model adopted the hyperparameters from (Hazare et al. 2022) as a starting point, which shows a good agreement between the actual values and the predicted values. While, the best model shows better performance after hyperparameter tuning by Grid Search method. The reduced model suffers from having very limited input variables, showing a more scattered pattern. The hyperparameters that have been optimized using the Grid Search approach for the structured, random, and combined datasets are shown in Table 11. Figure 17 displayed the parity plots of the base and optimized models for combined packing datasets. The patterns exhibited for combined dataset appear to be more scattered compared to those observed in the structured packing dataset. This is likely due to differences in the data collection methodology or difference in origin of the data source. Statistical metrics for the structured packing are shown in Table 9. Table 10 displays the statistical metrics of the optimized Random Forest models trained on random packing and combined datasets, respectively. Feature importance was evaluated by permutation importance, SHAP values, and partial dependence plots (PDP), respectively. Figure 18 shows the permutation feature importance of the reduced, base and optimized models for the structured packing datasets. It turned out that the two most important features in each model are a_p and L , which is in accordance with what's reported by Wang and Song. Single and double variable PDP plots for the best structured packing model by Random Forest Regression are displayed in Figures 19, 20. Figure 21 shows the SHAP value of input features by optimized Random Forest model.

Model	R^2	MAE	MSE
Best, Test			
Mean	0.95	0.0276	0.0013
Standard Deviation	0.007	0.0015	0.0002
Reduced, Test			
Mean	0.59	0.080	0.01
Standard Deviation	0.038	0.0037	0.001

Table 9: Statistical Metrics for Random Forest models trained on structured packing data.

5.3 SVR

Table 12 lists the features and goodness-of-fit statistics for each of the three SVR models. The base model is the most flexible, but without any hyperparameter tuning it performs relatively poorly. The best model, after feature selection and tuning, gives the highest accuracy on both training and testing data. Reducing the feature count further for the reduced model decreases accuracy, but the performance on testing data is comparable to the base model.

Figure 22 shows parity plots of prediction of Φ by each SVR model for both training and testing data. Figure 23 shows parity plots of predicted vs actual Φ for the reduced model, with the data points color-coded by the author conducting the experiment or by the nominal specific area (a_p) of the packing in that experiment. There is no obvious difference between data collected by different authors. Finer packings (larger a_p) tend to have lower

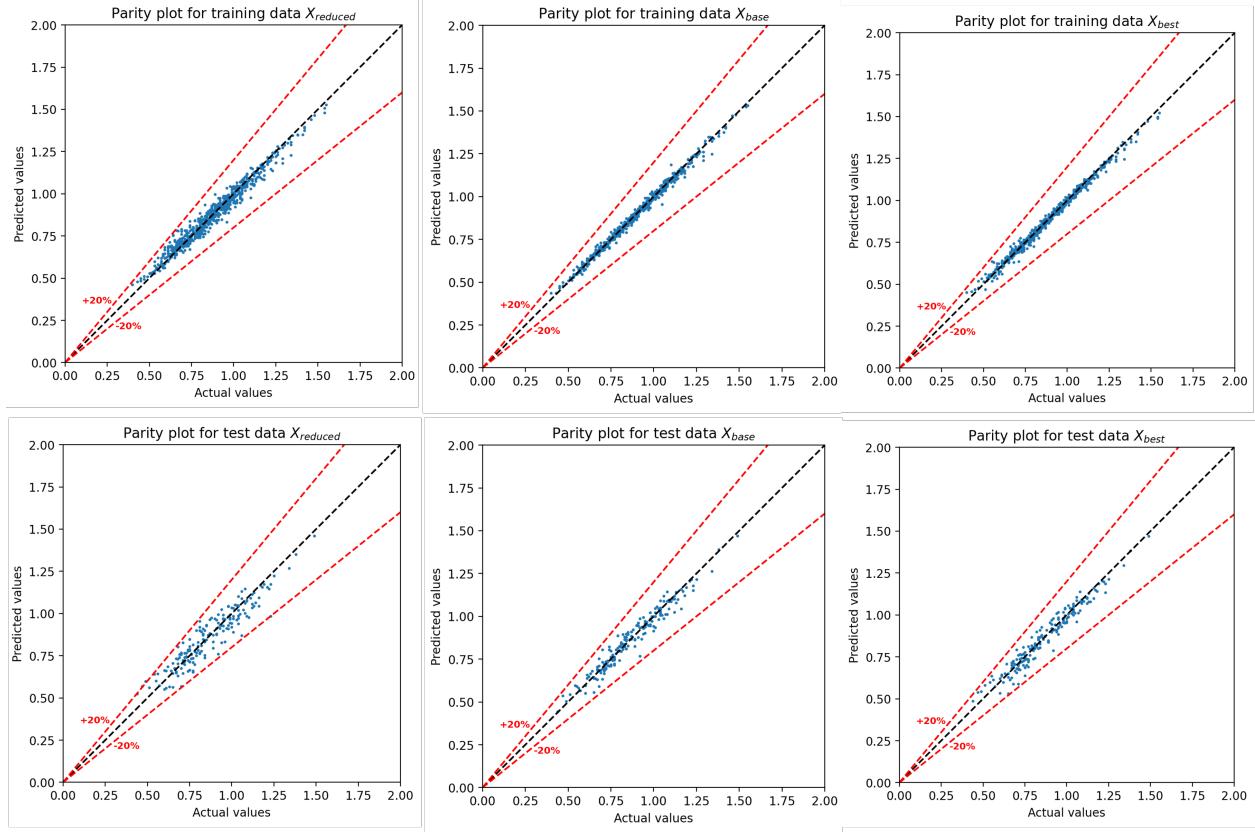


Figure 16: Parity plots for UT-SRP structured packing dataset trained on Random Forest.

Model	R^2	MAE	MSE
Best Test Random			
Mean	0.9647	0.037	0.0022
Standard Deviation	0.0116	0.0067	0.0007
Best Test Combine			
Mean	0.8214	0.046	0.0042
Standard Deviation	0.031	0.0032	0.0007

Table 10: Statistical Metrics for optimized Random Forest models trained on the random packing and combined datasets, respectively.

Parameter	Structured	Random	Combined
Max Depth	30	20	10
Max Features	log2	auto	sqrt
Min Samples Leaf	1	1	1
Number of Estimators	100	100	100

Table 11: Optimal Hyperparameters of Random Forest Regressor for structured, random and combined datasets as determined by grid search method.

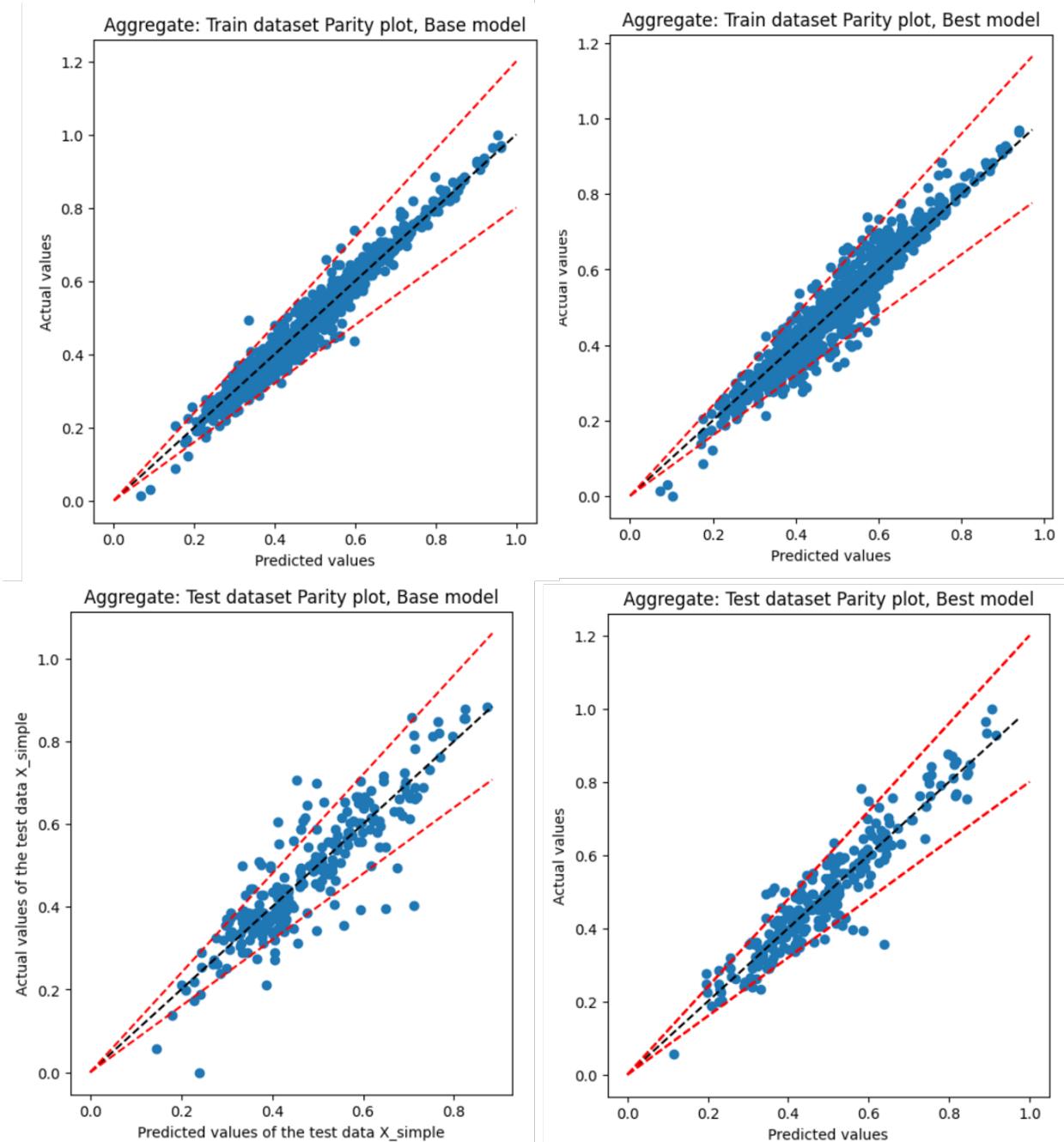


Figure 17: Parity plots for UT-SRP combined dataset trained on Random Forest.

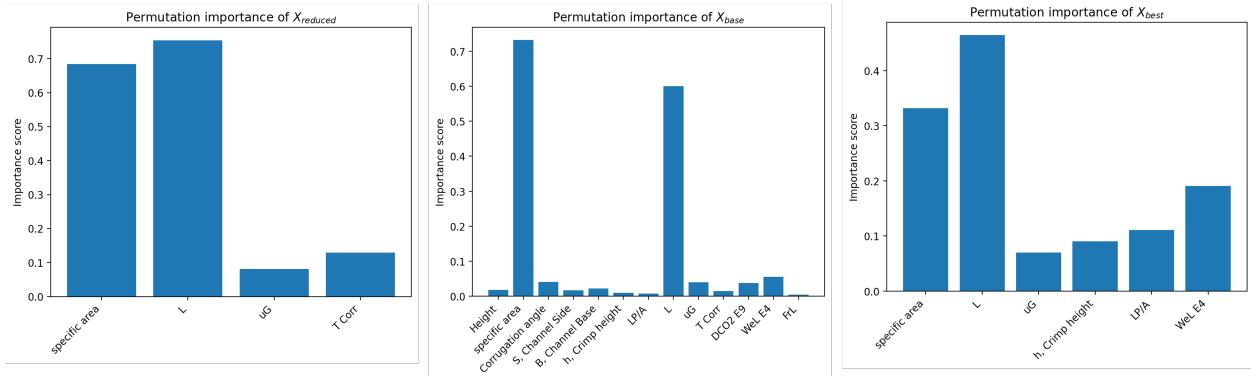


Figure 18: Feature importance for variables fed to Random Forest trained on UT-SRP structured packing dataset.

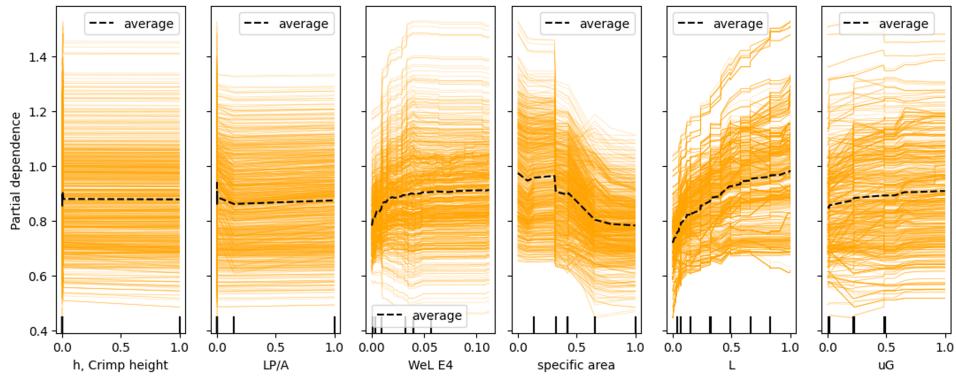


Figure 19: Single Variable PDP for Structured Packing dataset by optimized Random Forest model.

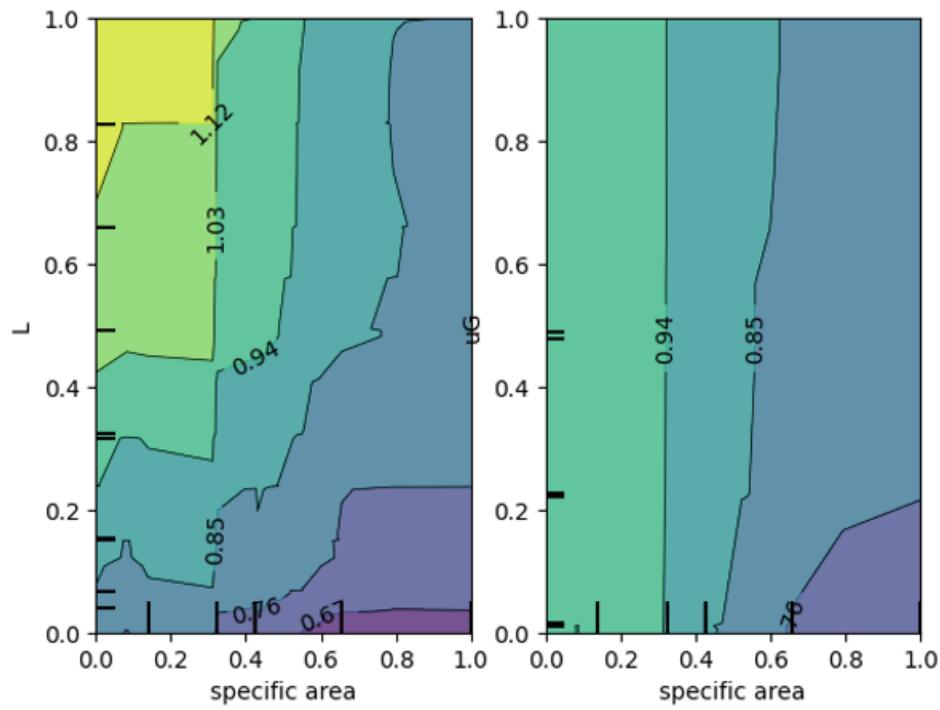


Figure 20: Double Variable PDP Plot for Structured Packing dataset by optimized Random Forest model.

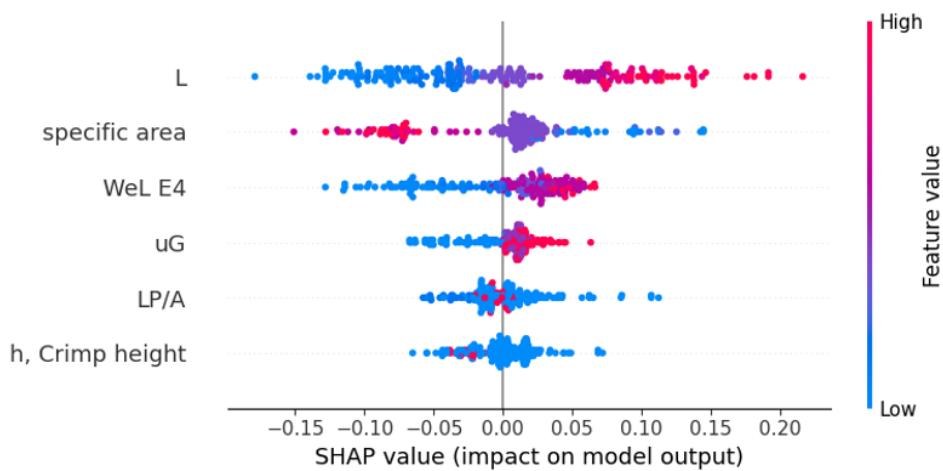


Figure 21: SHAP Plot for Structured Packing dataset by optimized Random Forest model.

Table 12: SVR model features and goodness-of-fit statistics.

Features	Base model	Best model	Reduced model
Packing height	Specific area (a_p)	Specific area (a_p)	Specific area (a_p)
Specific area (a_p)	Liquid flux (L)	Liquid flux (L)	Liquid flux (L)
Corrugation angle	Gas velocity (u_G)	Gas velocity (u_G)	Gas velocity (u_G)
Channel side (S)	Corrugation angle	Corrugation angle	Temperature (T)
Channel base (B)	Temperature (T)	Temperature (T)	Temperature (T)
Crimp height (h)	Diffusivity (D_{CO_2})	Diffusivity (D_{CO_2})	Diffusivity (D_{CO_2})
Perimeter/area (L_P/A)			
Liquid flux (L)			
Gas velocity (u_G)			
Temperature (T)			
Diffusivity (D_{CO_2})			
Training R^2	$88.1\% \pm 0.1\%$	$95.5\% \pm 0.0\%$	$90.7\% \pm 0.1\%$
Training MSE	4.55×10^{-3}	1.70×10^{-3}	3.55×10^{-3}
Training MAE	5.59×10^{-2}	2.73×10^{-2}	4.04×10^{-2}
Testing R^2	$86.3\% \pm 0.4\%$	$91.9\% \pm 0.3\%$	$84.9\% \pm 0.5\%$
Testing MSE	5.14×10^{-3}	1.70×10^{-3}	5.68×10^{-3}
Testing MAE	5.80×10^{-2}	4.09×10^{-2}	5.54×10^{-2}

values of Φ than coarse packings, which is expected. Although there is no apparent bias at higher or lower a_p , there is more scatter when predicting Φ for finer packings. Further work is needed to understand this issue, and it would be helpful to have more experiments with finer packings operated at higher Φ and coarser packings operated at lower Φ to better cover the sample space.

Feature importance was evaluated by three methods: permutation importance, Shapley values (using the `SHAP` package), and partial dependence plots (PDP). Figure 24 shows the permutation feature importance for all three SVR models. The two most important features in every model are a_p and L , which supports their selection as “fixed” features for the best model. u_G , the final fixed feature, is less important than was assumed initially. Permutation importance quantifies the importance of a feature to the given model rather than its true importance in predicting the model output. This is reflected in how features like T and D_{CO_2} have higher permutation importance in the better-performing models, but have lower importance in the base model than some of the other features.

Shapley values, shown in Figure 25, are another way of explaining how features affect the SVR model predictions. These plots again show that L and a_p have a strong impact on Φ .

Figure 26 shows partial dependence plots for only the reduced model. The top four panels show one-way partial dependence (PDP) and individual conditional expectation (ICE) plots for all four features of the model. The bottom four panels show two-way partial dependence for selected combinations of two features. The two-way plot of Φ vs a_p and L in particular shows an expected trend. A real mass transfer column would be designed with a diameter and packing type selected such that the column operates near $\Phi = 1.0$. There is a trade-off involved because finer packing offers more mass transfer area but requires a lower liquid flux and therefore larger column diameter.

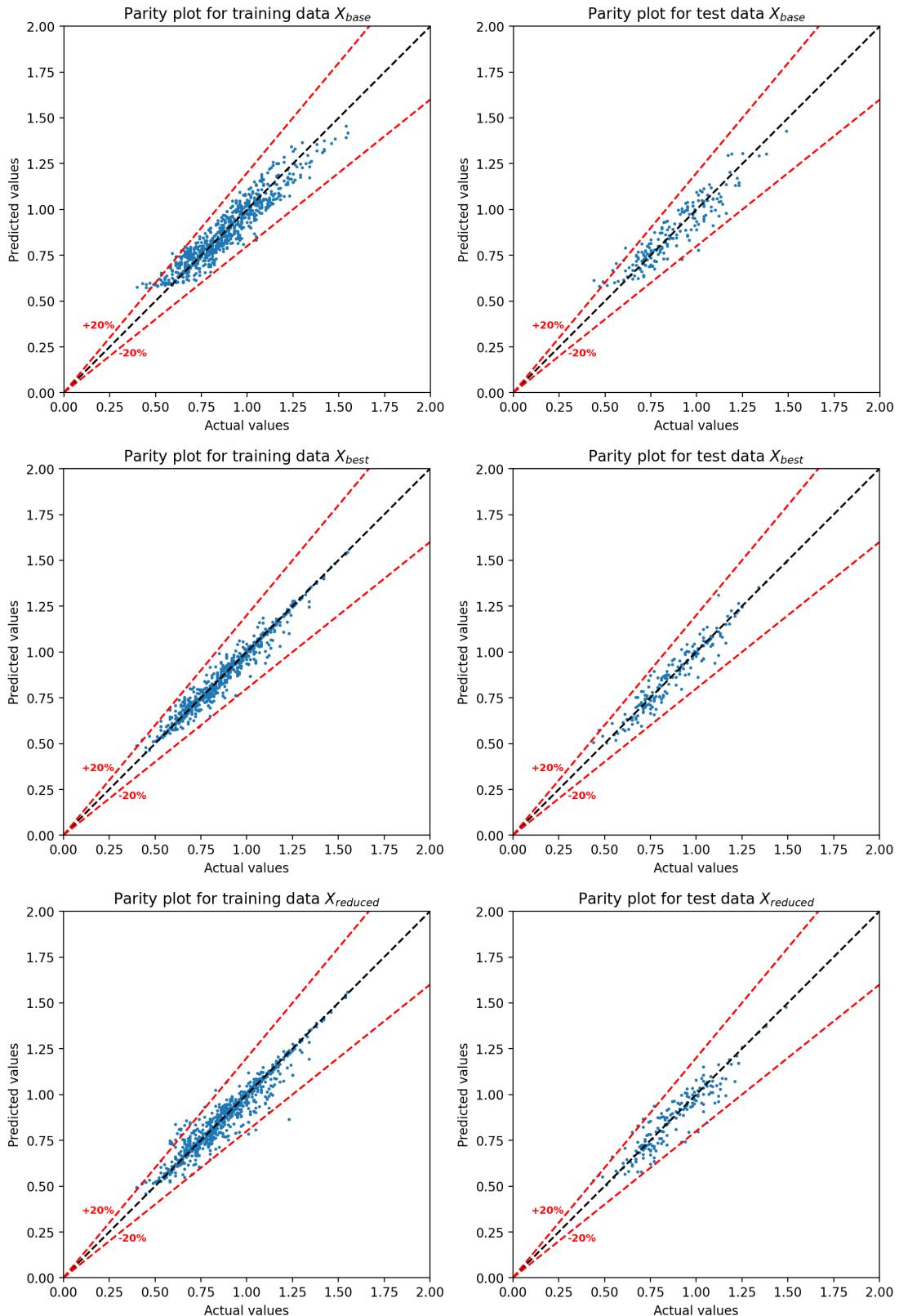


Figure 22: Parity plots showing accuracy of Φ predictions on training and testing data for all three SVR models.

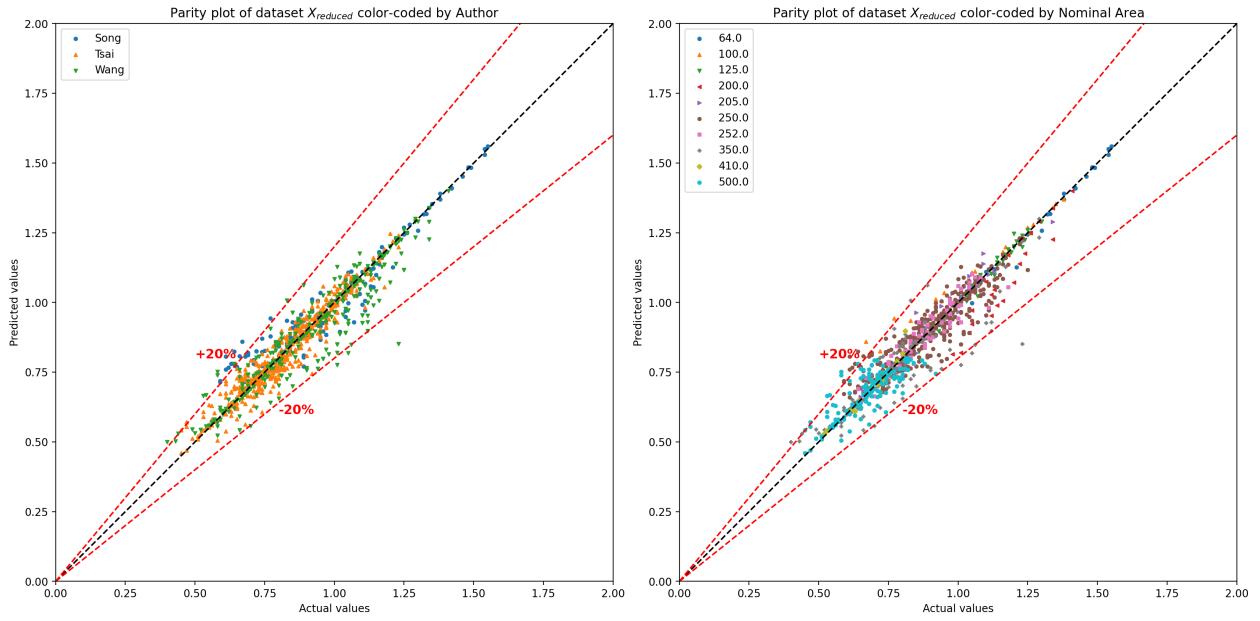


Figure 23: Parity plots of predicted vs actual Φ using the SVR reduced model on the complete UT-SRP dataset, color-coded by author (left) and packing specific area (right).

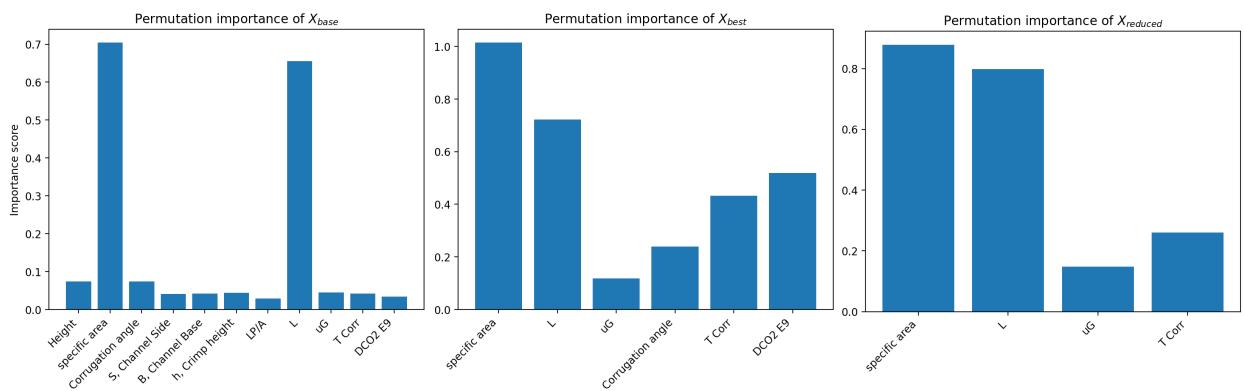


Figure 24: Permutation importance of SVR model features. Base model (left), best model (middle), and reduced model (right).

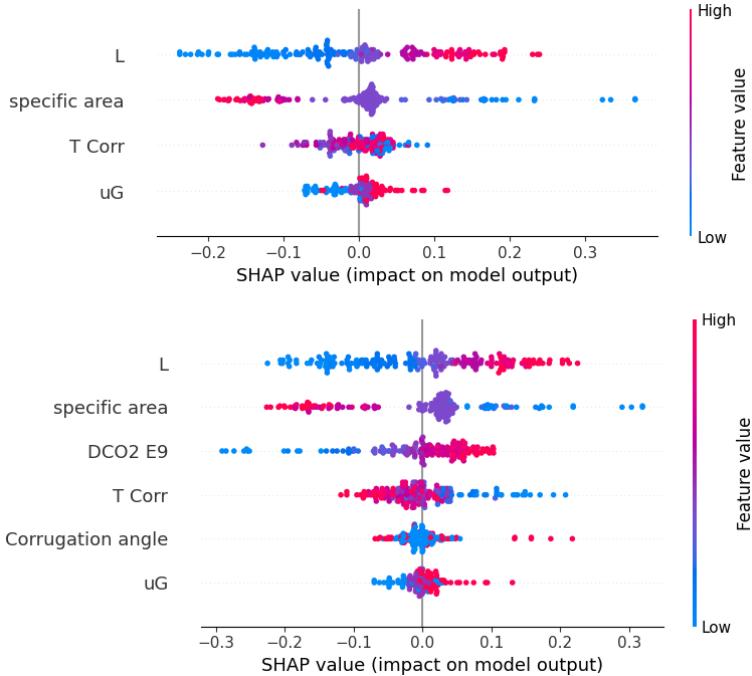


Figure 25: SHAP values of SVR reduced model (top) and best model (bottom).

6 Discussion

6.1 ANN

The grid search - optimized ANN models learned the relation between target and feature variables, with the caveat that there is reduced accuracy in the learning of the predictor-target relation for the combined dataset. This is surmised to be due to the non - UT-SRP data containing fewer features than the UT-SRP data; since the ANN models cannot be trained on missing or invalid data, variables reported by UT-SRP such as D_{CO_2} , packing dimensions like corrugation angle, and temperature were unable to be utilized.

As a general trend, the reduced models tended to train well for median a_e values, while underpredicting and overpredicting low and high a_e values, respectively. As the model improved, these under and overpredictions were improved in the base model until mostly fixed in the best model.

The most important features as selected by permutation importance were Channel Side - Specific Area - Channel Base - D_{CO_2} - physical fluid properties for the structured packing best model, with contrasts with the base and reduced models only considering L and specific area of importance. For random packing, the best model considered L , specific area, and void fraction to be the most importance variables, while for the combined dataset the specific area was the most important variable. For the ANN models in their current form, the optimized models consider the physical properties of the packing to be the most important for predicted effective wetted area.

From the ln-ln plot, the best ANN model for structured packing seems to underpredict effective wetted area as compared to Song and Wang, which runs counter to previous

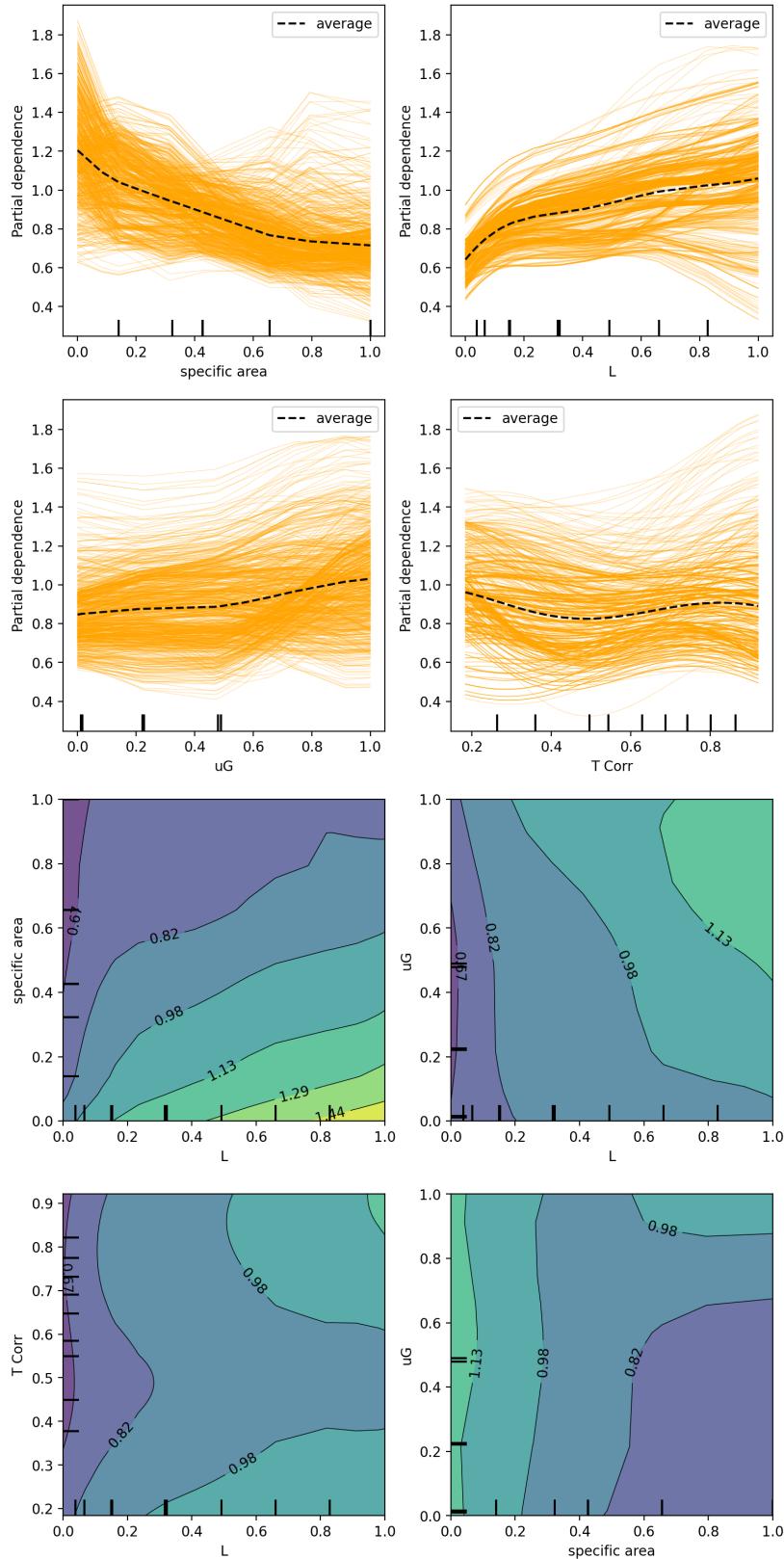


Figure 26: PDP plots for SVR reduced model. Top four plots show one-way partial dependence for all four features of the reduced model. Bottom four plots show two-way dependence for four combinations of features.

assertions. This could be an error in the code when computing the a_e for this step.

The non-UT-SRP data was not accurately predicted by the best UT-SRP trained model, with outliers on the parity plot at both high and low a_e likely skewing the R^2 . Nonetheless, the median wetted area values were mostly within the error margins, suggesting that the prediction of datasets not included in the training of the model should hold true for median predicted values of a_e .

PDP plots show that of the UT-SRP variables considered for predicting structured packing a_e , wetted area is not sensitive to changes in the value of u_G and D_{CO_2} . The dependence on specific area and temperature decreases at larger area and T values, has a v-shaped dependence on corrugation angle, and a quasi-logarithmic dependence on L . The largest dependencies come from combinations of high L - low specific area, high u_G - low specific area, and high L - high u_G . The SHAP plots show similar impacts on the model output, with high specific area, low D_{CO_2} , low L , and high T having negative impacts on predicted model output. Conversely, low specific area, high L , low T , and high D_{CO_2} have greater impact on the predicted model output values.

6.2 Random Forest

The Random Forest models shared similar trends to the ANN models, indicating that the structured packing dataset produced the highest level of agreement between the predicted and actual values. However, the fitting of the combined dataset, which includes random packing data as well as data collected outside UT-SRP, resulted in less optimal outcomes. This was evident from the more scattered parity plots produced by the base and optimized Random Forest models.

When it comes to the structured packing dataset, both the Random Forest and ANN models concurred that the reduced model underperformed due to a lack of input variables. However, the optimized model with a R^2 value of 0.95 exhibited superior performance. Surprisingly, the Random Forest model exhibits an outstanding fit to the random packing dataset, yielding an R^2 value of 0.96. This result is counterintuitive because random packing is typically less predictable than structured packing. The relatively small number of data points, in comparison to the large number of estimators, raises concerns about overfitting. Therefore, one need to be cautious to adapt to this model as other factors may need further consideration.

The permutation importance score, SHAP value as well as the PDP plots indicate that the most important features are a_p and L . PDP plots also indicated that in the optimized Random Forest model for structure packing dataset, the target wetted area is not sensitive to changes in the value of u_G , h and LP/A . The dependence on specific area decreases at larger area and a gradually increasing dependence on L is observed. The double variable partial dependence plots show two-way partial dependence for selected combinations of two features, where the double variable plot of Φ vs a_p and L shows an expected relationship.

6.3 SVR

The right panel of Figure 23 shows greater scatter in predictions of fine packings ($a_p = 500[m^2/m^3]$). It is possible that this issue is related to secondary area effects (i.e., mass transfer that occurs on “secondary” surfaces like the walls of the column or liquid droplets as opposed to on the packing surface). However, it is expected that secondary effects would

be more significant for coarser packings (Song 2017).

Another way of visualizing model performance is shown in Figure 27. For this analysis, pure water properties and a temperature of 25[°C] are assumed. Nine different combinations of a_p and u_G are selected, and $\log(\Phi)$ vs $\log(L)$ is plotted for each combination. With the other variables fixed, the Wang and Song models give straight lines on log-log axes. The best and reduced SVR models are used to predict $\log(\Phi)$, giving the orange and green lines respectively. Finally, experimental data is filtered for points collected at similar operating conditions which are plotted in red for UT-SRP data (used for training and testing) and purple for non-UT data (not used in training/testing). These plots give a visual comparison of the various models and show that the SVR models are closer to the Song model, which is the current best model for this application. The SVR models also mostly trend with the relevant UT-SRP data, as expected. Unfortunately, much of the non-UT data was collected at different conditions so only a small amount of it was relevant to this analysis. The SVR models appear to have relatively good performance for coarser packings (left column), but at lower u_G and for finer packings (towards top right) there is unusual behavior where the SVR models predict a decrease in Φ . There does not appear to be an obvious explanation for this effect. Further work is needed to refine the model, and more external validation data (at similar conditions) would be useful for evaluating the model.

7 Conclusions

SVM, Random Forest, and ANN models available within scikit-learn were fitted to random and structured packing data corresponding to experiments conducted by UT-SRP and other sources for the purpose of determining the effective wetted area for carbon capture applications. The target-predictor response was able to be learned by all three flavours of machine learning model, with Random Forest, SVM and ANN in order of most to least accurate for the grid search-optimized models trained on the UT-SRP structured packing data. All optimized models, when compared like-for-like to the Song model, were able to more accurately predict a_e .

The reduced form of the SVM model had a similar accuracy as compared to the base structured packing model, suggesting that optimized SVM models with fewer input variables compared to the base model should be able to predict effective wetted area well. However, the Random Forest and ANN reduced models produced markedly worse models in terms of accuracy, statistical metrics when compared to the base and best models.

Prediction of non-UT-SRP data was attempted via a variation of the best structured ANN model, which was unable to satisfactorily learn the target-predictor response. This was, however, skewed by significant over and underpredictions of low and high a_e values, respectively.

Permutation feature importance was conducted to identify the most important variables for predicting a_e . The most important variables for the ML models were any of the variables associated with the physical dimensions of the packing, such as corrugation angle, channel width and height, and specific area of packing, follow by system variables such as the liquid flowrate L, superficial vapor velocity u_G , temperature and CO₂ diffusivity. Similarly, SHAP and PDP plots suggest the models are typically most sensitive to variations in packing specific area, liquid flowrate, and temperature.

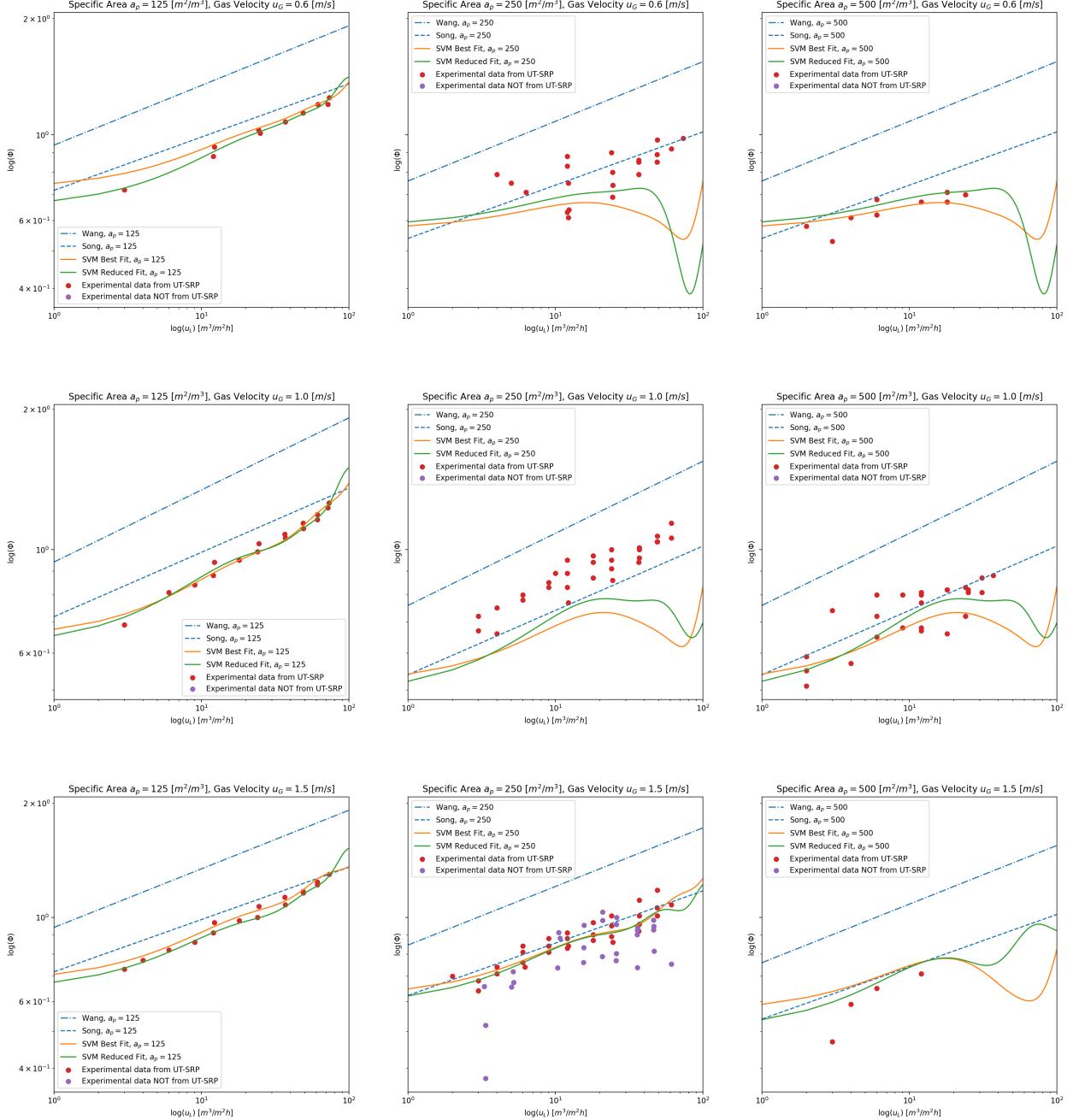


Figure 27: Plots of $\log(\Phi)$ vs $\log(L)$ at different combinations of a_p and u_G . Lines indicate predictions by Wang model (blue dash-dot), Song model (blue dashed), SVR best model (orange) and SVR reduced model (green). Experimental data collected at similar operating conditions is also shown as points. UT-SRP (red) data is from the training/testing sets and non-UT-SRP (purple) data is shown for validation only.

8 Acknowledgements

The authors thank Miguel Abreu for providing the compiled and collated data which was the basis of the dataset studied in this work. The authors also thank Prof. Arya Farahi for providing feedback and guidance on this project.

Percentage contribution of each team member:

Benjamin Drewry	100%
Jorge Martorell	100%
Yuying Wang	100%

References

- Abreu, Miguel and Gary Rochelle (Nov. 2022). “CO₂ absorber intensification by high liquid flux operation”. In: Proceedings of the 16th Greenhouse Gas Control Technologies Conference, GHGT-16. DOI: <https://dx.doi.org/10.2139/ssrn.4282100>.
- Aferka, Saïd, Pierre Marchot, Michel Crine, and Dominique Toye (2010). “Interfacial area measurement in a catalytic distillation packing using high energy X-ray CT”. In: *Chemical Engineering Science* 65.1, pp. 511–516.
- Billet, Reinhard and Michael Schultes (1999). “Prediction of mass transfer columns with dumped and arranged packings: updated summary of the calculation method of Billet and Schultes”. In: *Chemical Engineering Research and Design* 77.6, pp. 498–504.
- Bolenz, Lukas, Florian Fischer, Dominique Toye, and Eugeny Y Kenig (2019). “Tomographische Untersuchung der Flüssigkeitsdynamik viskoser Systeme in Packungskolonnen”. In: *Chemie-Ingenieur-Technik* 91.12, pp. 1892–1896.
- Bolles, W L and James R Fair (1979). In: *Inst. Chem. Eng. Symp. Ser* 3.35.
- Hazare, Sumit R, Shivam V Vala, Chinmay S Patil, Aniruddha J Joshi, Jyeshtsharaj B Joshi, Vivek S Vitankar, and Ashwin W Patwardhan (2022). “Correlating Interfacial Area and Volumetric Mass Transfer Coefficient in Bubble Column with the Help of Machine Learning Methods”. In: *Ind. Eng. Chem. Res.* DOI: <https://doi.org/10.1021/acs.iecr.2c02820>.
- IPCC (2014). *Climate Change 2014: Mitigation of Climate Change: Working Group III Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Ed. by O. Edenhofer, Pichs-Madruga R., Sokona Y., Minx J.C., Farahani E., Kadner S., Seyboth K., Adler A., Baum I., Brunner S., Eickemeier P., Kriemann B., Savolainen J., Schlömer S., von Stechow C., and Zwickel T. Cambridge University Press.
- Janzen, Anna, Julia Steube, Saïd Aferka, EY Kenig, Michel Crine, Pierre Marchot, and Dominique Toye (2013). “Investigation of liquid flow morphology inside a structured packing using X-ray tomography”. In: *Chemical Engineering Science* 102, pp. 451–460.
- Onda, Kakusaburo, Hiroshi Takeuchi, and Yoshio Okumoto (1968). “Mass transfer coefficients between gas and liquid phases in packed columns”. In: *Journal of chemical engineering of Japan* 1.1, pp. 56–62.
- Rocha, J Antonio, Jose L Bravo, and James R Fair (1993). “Distillation columns containing structured packings: a comprehensive model for their performance. 1. Hydraulic models”. In: *Ind. Eng. Chem. Res.* 32.4, pp. 641–651.
- Song, Di (2017). “Effect of liquid viscosity on liquid film mass transfer for packings”. PhD thesis. University of Texas at Austin.
- Tsai, Robert E (2010). “Mass Transfer Area of Structured Packing”. PhD thesis. University of Texas at Austin.
- Tsay, Calvin, Richard C Pattison, Yue Zhang, Gary T Rochelle, and Michael Baldea (2019). “Rate-based modeling and economic optimization of next-generation amine-based carbon capture plants”. In: *Applied Energy* 252, p. 113379. DOI: <https://doi.org/10.1016/j.apenergy.2019.113379>.
- Vega, F, FM Baena-Moreno, Luz M Gallego Fernández, E Portillo, B Navarrete, and Zhien Zhang (2020). “Current status of CO₂ chemical absorption research applied to CCS”.

- Towards full deployment at industrial scale". In: *Applied Energy* 260, p. 114313. DOI: [10.1016/j.apenergy.2019.114313](https://doi.org/10.1016/j.apenergy.2019.114313).
- Wang, Chao (2015). "Mass transfer coefficients and effective area of packing". PhD thesis. University of Texas at Austin.
- Zakeri, Ali, Aslak Einbu, and Hallvard F Svendsen (2011). "A Comparison of Pressure drop, Liquid hold-up and Effective mass transfer area in three Different Structured packings". In: *6th Conference on CO₂ Capture Transport and Storage (TCCS6)*, pp. 06–14.