|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Feature score threshold (%) | Number of features | Train MAE | Test MAE | Train MSE | Test MSE | Train R^2 | Test R^2 | Time |
| 0.00 | 18 | 0.08228 | 0.15022 | 0.01750 | 0.15509 | 0.98409 | 0.94366 | 2706.839 |
| 0.05 | 13 | 0.11509 | 0.18501 | 0.03027 | 0.15672 | 0.97248 | 0.94307 | 2633.780 |
| 0.12 | 12 | 0.26309 | 0.39843 | 0.16507 | 0.44461 | 0.84993 | 0.83849 | 2600.443 |
| 0.80 | 11 | 0.26636 | 0.40548 | 0.16766 | 0.47040 | 0.84757 | 0.82912 | 2573.619 |
| 1.61 | 10 | 0.28637 | 0.46414 | 0.19729 | 0.69194 | 0.82063 | 0.74864 | 2591.059 |
| 1.66 | 9 | 0.28734 | 0.46006 | 0.19638 | 0.67437 | 0.82146 | 0.75503 | 2548.656 |
| 1.86 | 8 | 0.29570 | 0.46535 | 0.20187 | 0.68012 | 0.81647 | 0.75294 | 2540.034 |
| 2.90 | 7 | 0.30291 | 0.48397 | 0.21466 | 0.73709 | 0.80484 | 0.73225 | 2512.508 |
| 3.00 | 6 | 0.30665 | 0.47964 | 0.21598 | 0.71236 | 0.80364 | 0.74123 | 2489.615 |
| 4.50 | 5 | 0.32472 | 0.54059 | 0.24561 | 0.94952 | 0.77670 | 0.65508 | 2455.130 |

Architecture: Number of Features :

Architecture: Number of Features :

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Feature score threshold (%) | Number of features | Train MAE | Test MAE | Train MSE | Test MSE | Train R^2 | Test R^2 | Time |
| 0.00 | 18 | 0.06500 | 0.11512 | 0.01038 | 0.08803 | 0.99056 | 0.96802 | 2921.784 |
| 0.05 | 13 | 0.08184 | 0.14483 | 0.0159 | 0.13244 | 0.98551 | 0.95189 | 2840.543 |
| 0.12 | 12 | 0.24828 | 0.39393 | 0.15508 | 0.40915 | 0.85901 | 0.85137 | 2829.039 |
| 0.80 | 11 | 0.25075 | 0.39059 | 0.15680 | 0.43430 | 0.85744 | 0.84224 | 2784.212 |
| 1.61 | 10 | 0.27865 | 0.45316 | 0.18997 | 0.66173 | 0.82729 | 0.75962 | 2783.221 |
| 1.66 | 9 | 0.28215 | 0.45225 | 0.19082 | 0.66076 | 0.82652 | 0.75997 | 2744.435 |
| 1.86 | 8 | 0.28110 | 0.45287 | 0.19235 | 0.64947 | 0.82513 | 0.76407 | 2718.899 |
| 2.90 | 7 | 0.29484 | 0.47347 | 0.20763 | 0.70978 | 0.81123 | 0.74217 | 2704.924 |
| 3.00 | 6 | 0.29623 | 0.47287 | 0.20884 | 0.70090 | 0.81013 | 0.74539 | 2685.816 |
| 4.50 | 5 | 0.31601 | 0.53125 | 0.23802 | 0.92415 | 0.78360 | 0.66429 | 3057.654 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Feature score threshold (%) | Number of features | Train MAE | Test MAE | Train MSE | Test MSE | Train R^2 | Test R^2 | Time |
| 0.00 | 18 | 0.02529 | 0.04118 | 0.00156 | 0.01384 | 0.99858 | 0.99497 | 4716.297 |
| 0.05 | 13 | 0.03190 | 0.05296 | 0.00239 | 0.02176 | 0.99782 | 0.99209 | 4459.109 |
| 0.12 | 12 | 0.22711 | 0.39452 | 0.13135 | 0.49700 | 0.88058 | 0.81946 | 4416.148 |
| 0.80 | 11 | 0.23055 | 0.38115 | 0.13631 | 0.43788 | 0.87607 | 0.84093 | 4368.005 |
| 1.61 | 10 | 0.26713 | 0.46025 | 0.17285 | 0.69817 | 0.84285 | 0.74638 | 4320.792 |
| 1.66 | 9 | 0.26483 | 0.46508 | 0.17011 | 0.73019 | 0.84534 | 0.73475 | 4354.535 |
| 1.86 | 8 | 0.26534 | 0.45861 | 0.17265 | 0.71426 | 0.84303 | 0.74054 | 4227.313 |
| 2.90 | 7 | 0.28378 | 0.47927 | 0.19322 | 0.75570 | 0.82434 | 0.72549 | 4208.337 |
| 3.00 | 6 | 0.28561 | 0.48334 | 0.19382 | 0.78408 | 0.82379 | 0.71518 | 4185.091 |
| 4.50 | 5 | 0.29979 | 0.52051 | 0.21659 | 0.90289 | 0.80309 | 0.67202 | 4106.297 |

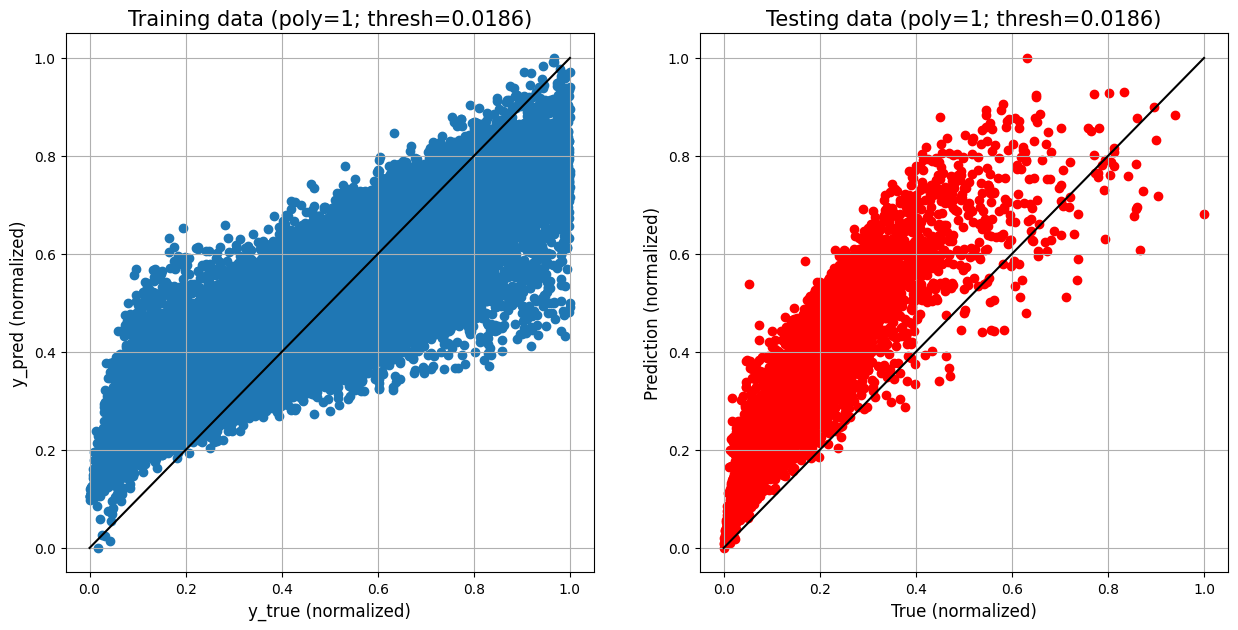
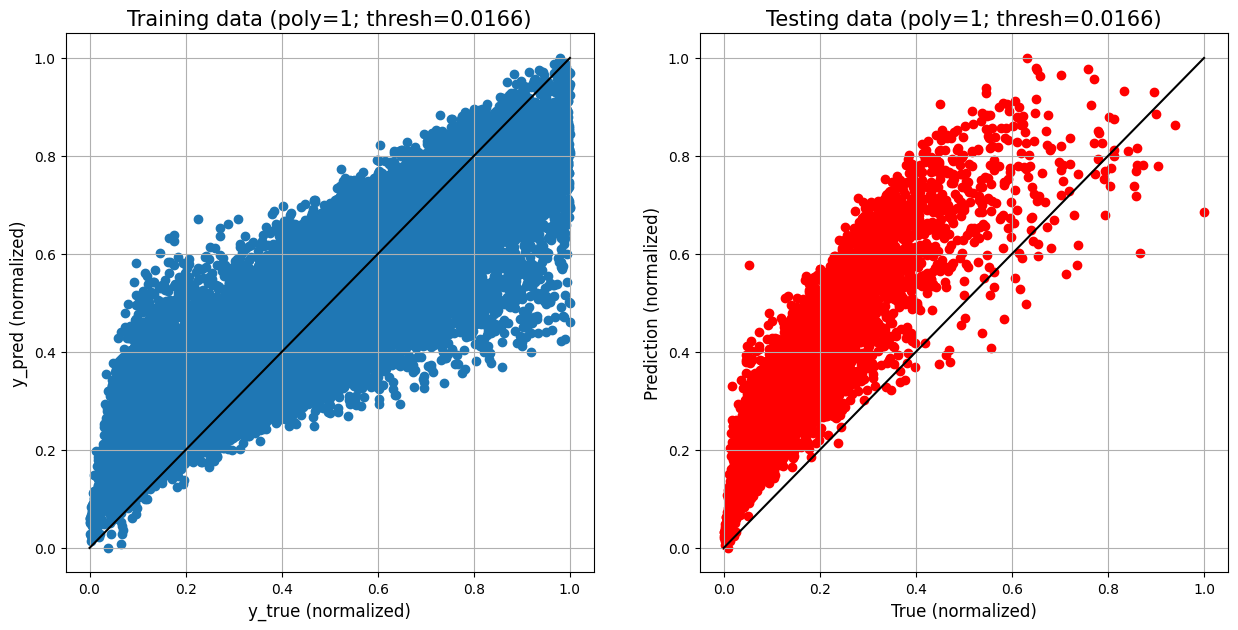
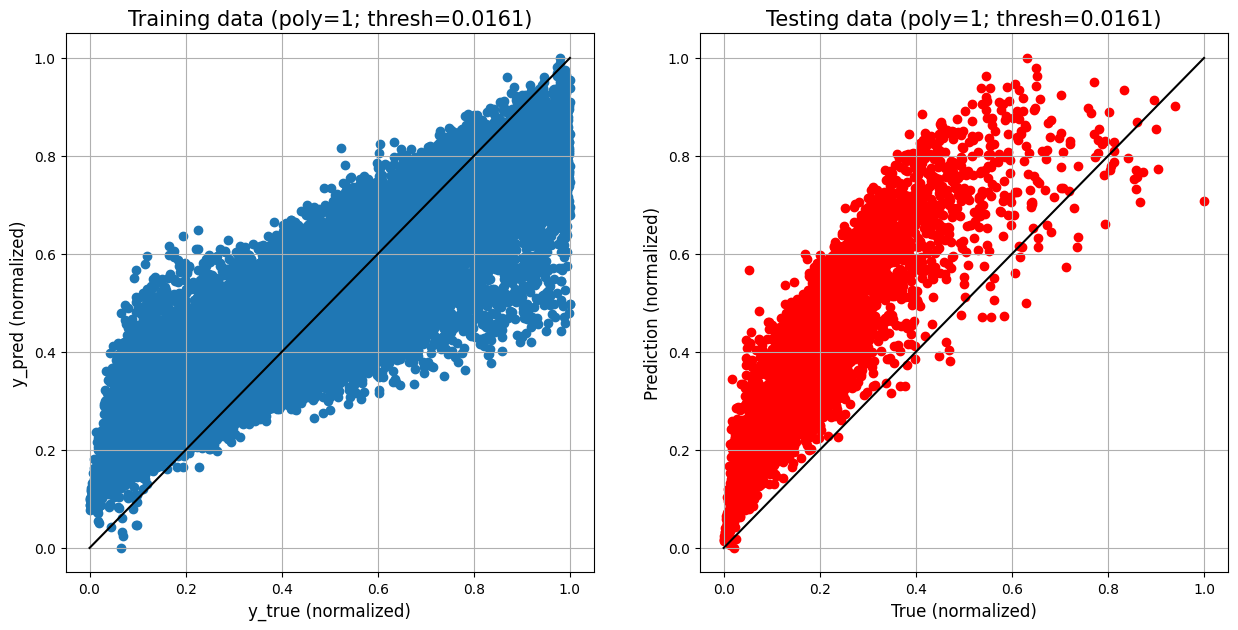
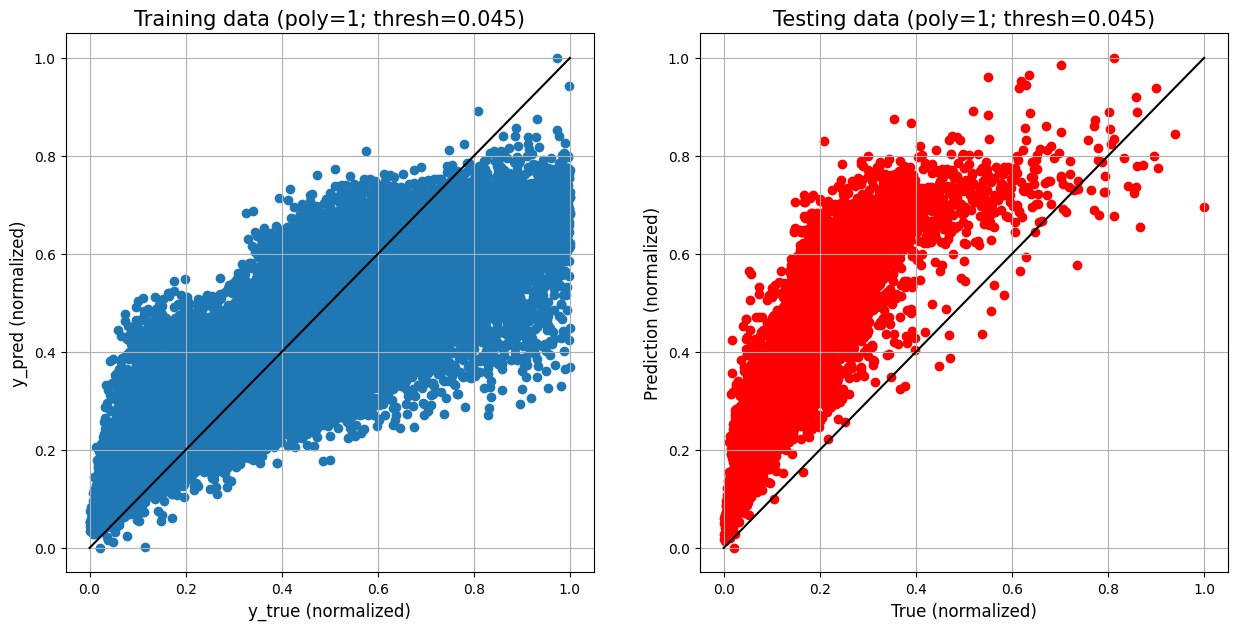
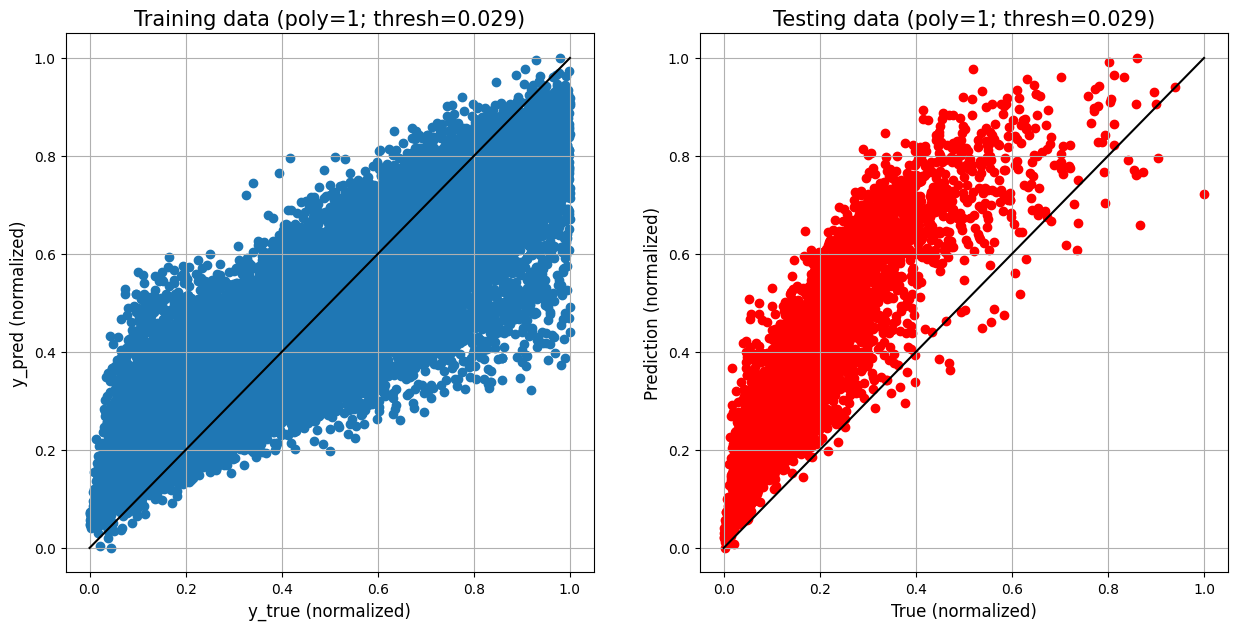
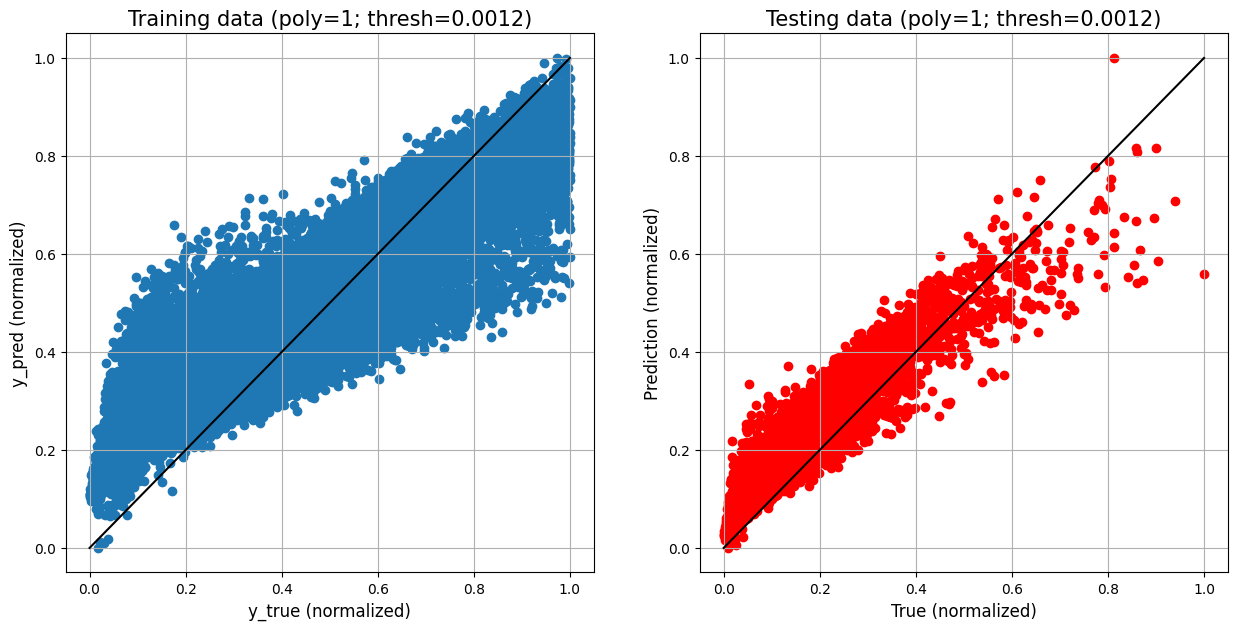
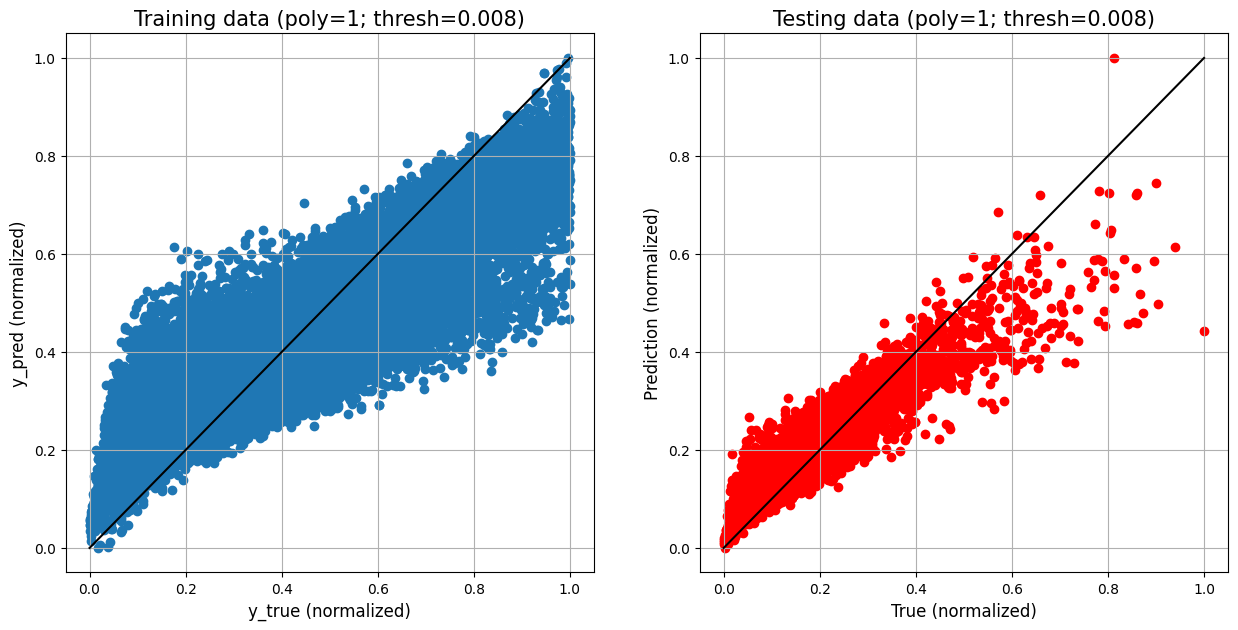
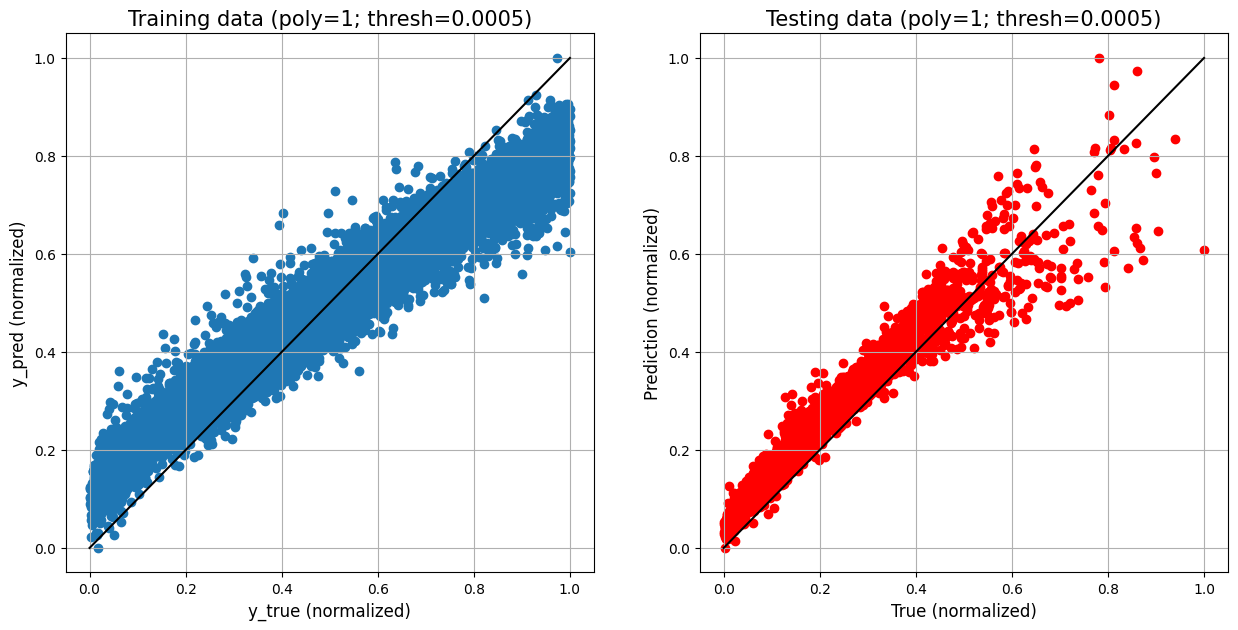
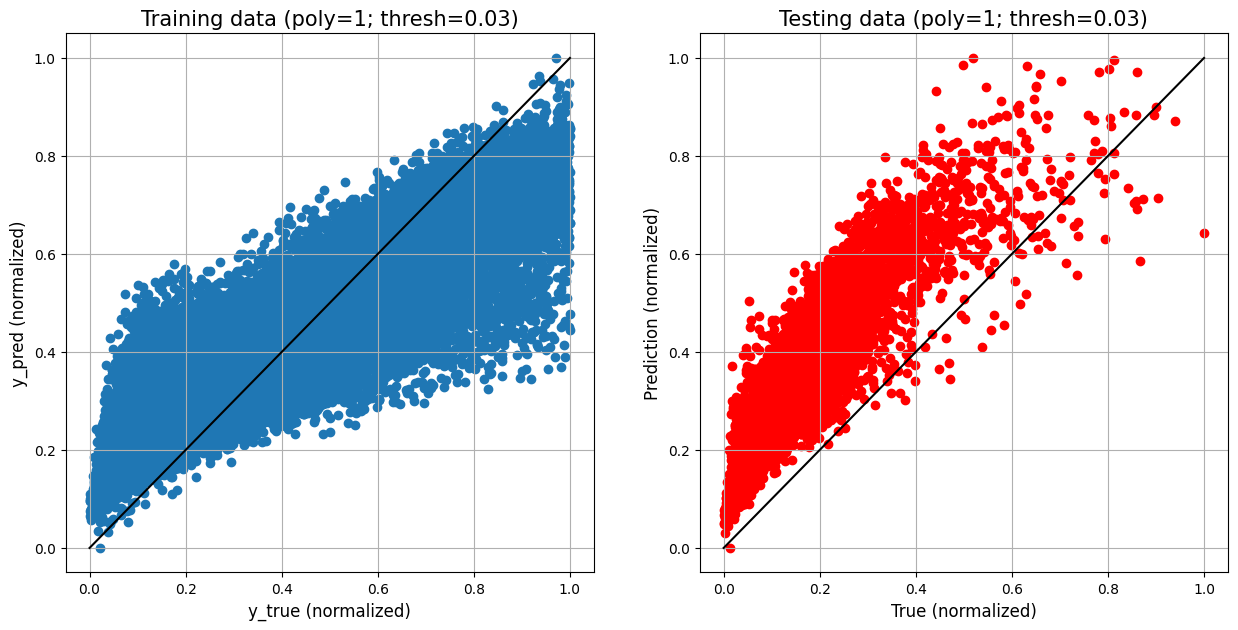
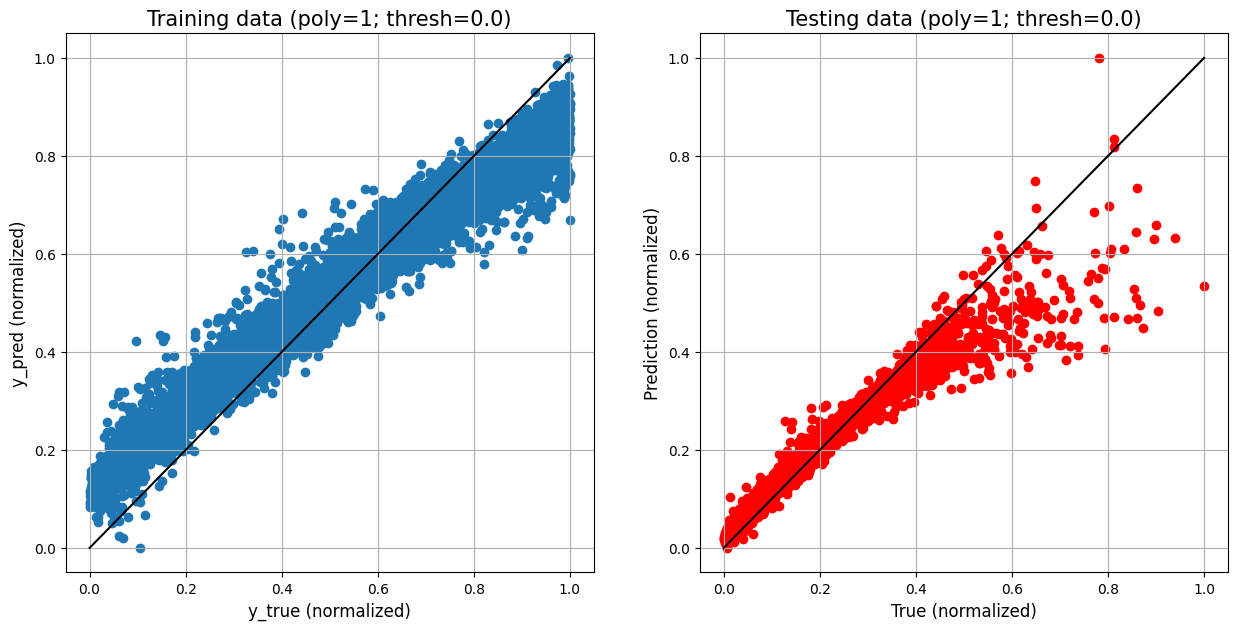
Architecture: Number of Features :

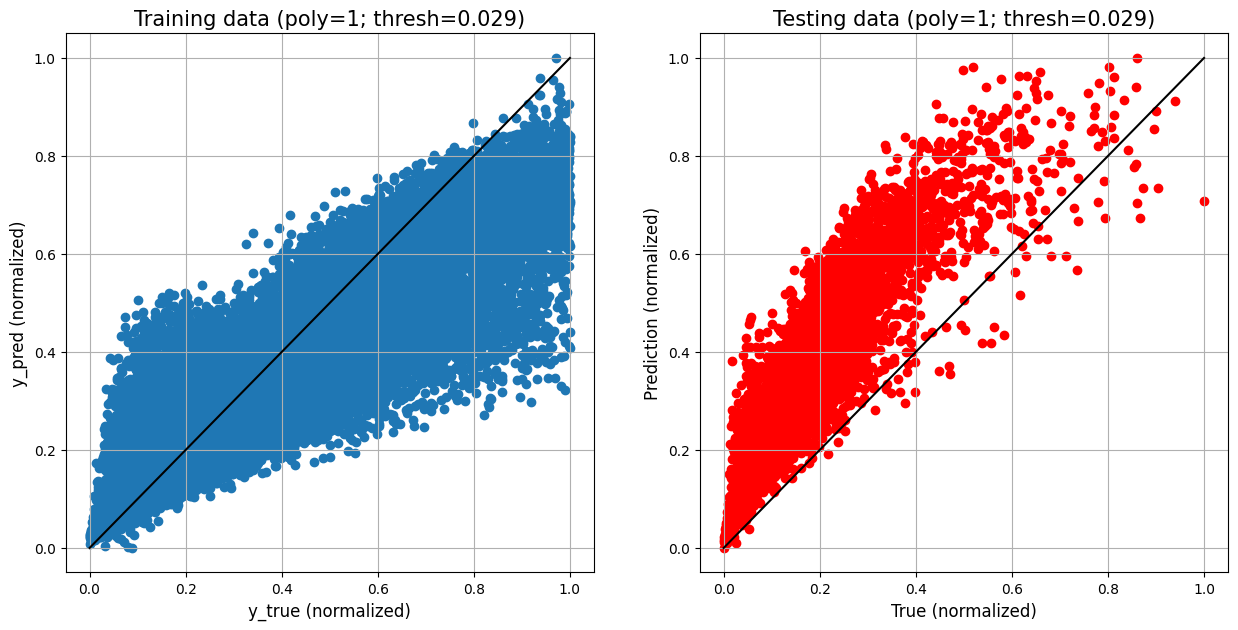
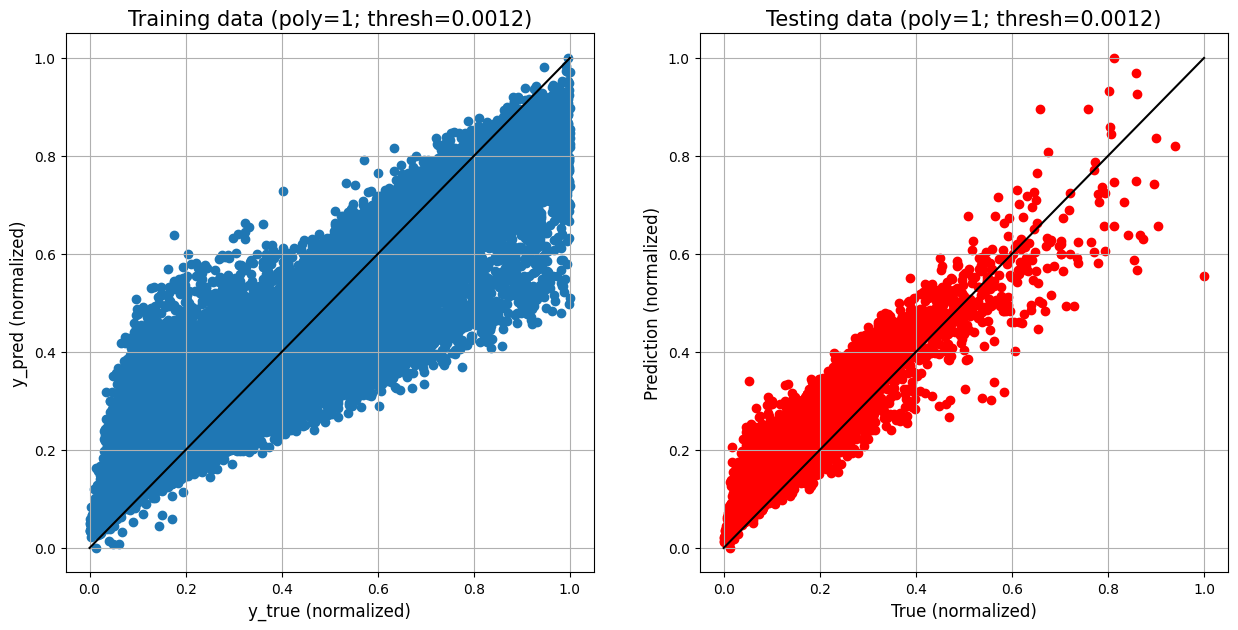
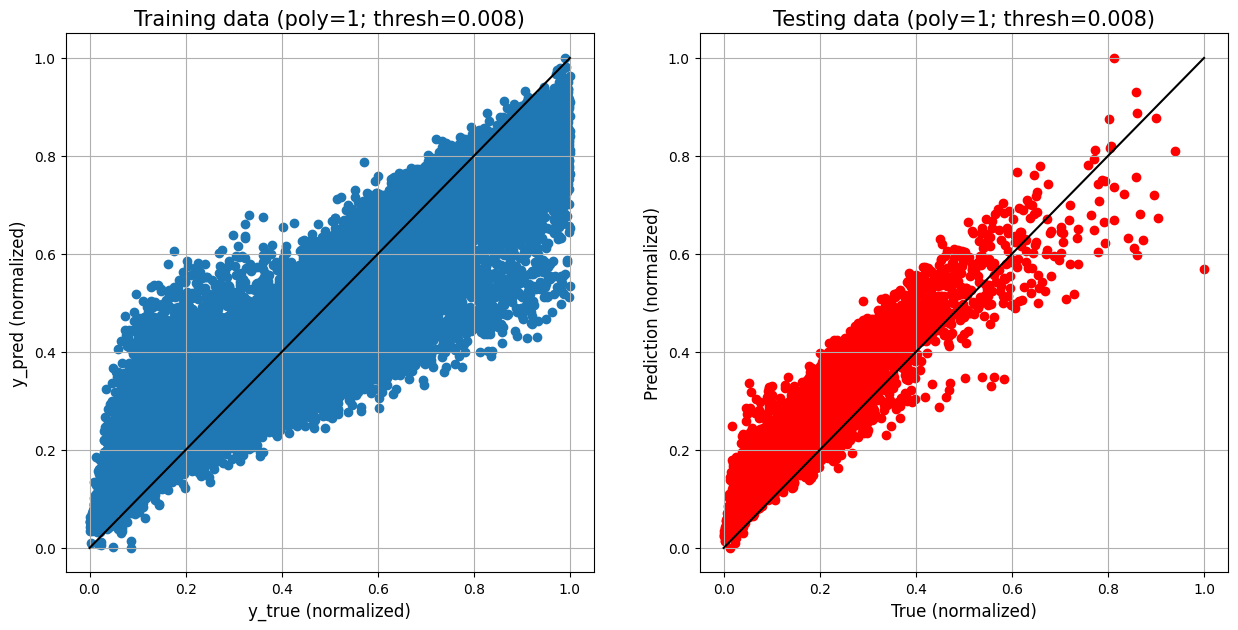
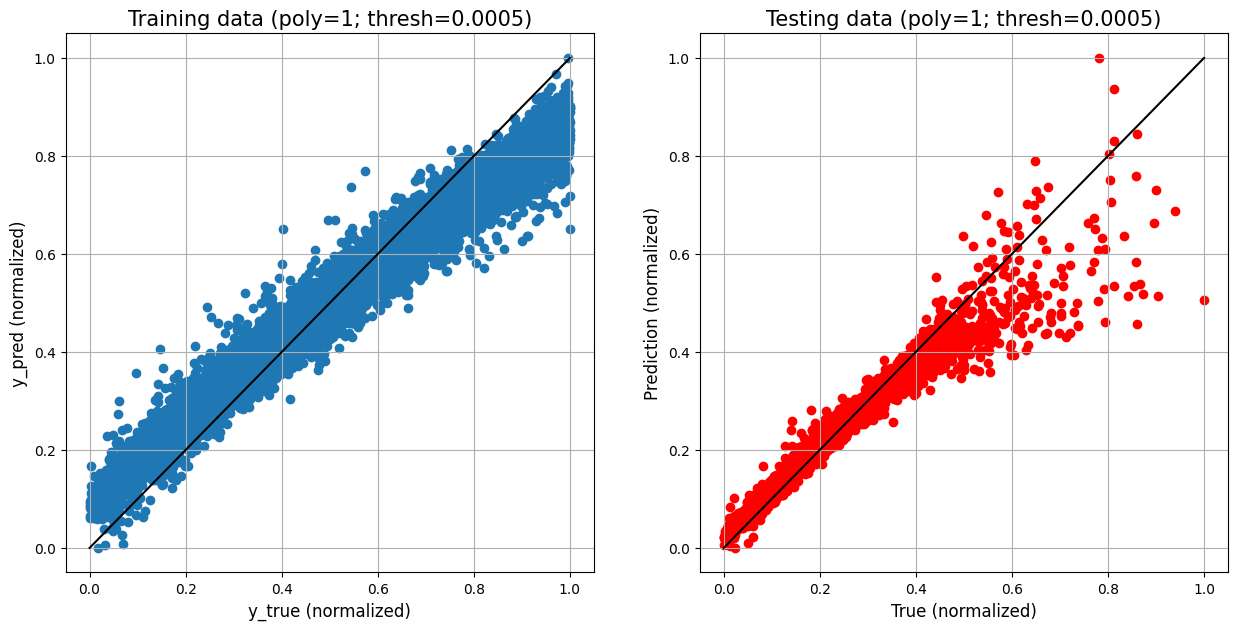
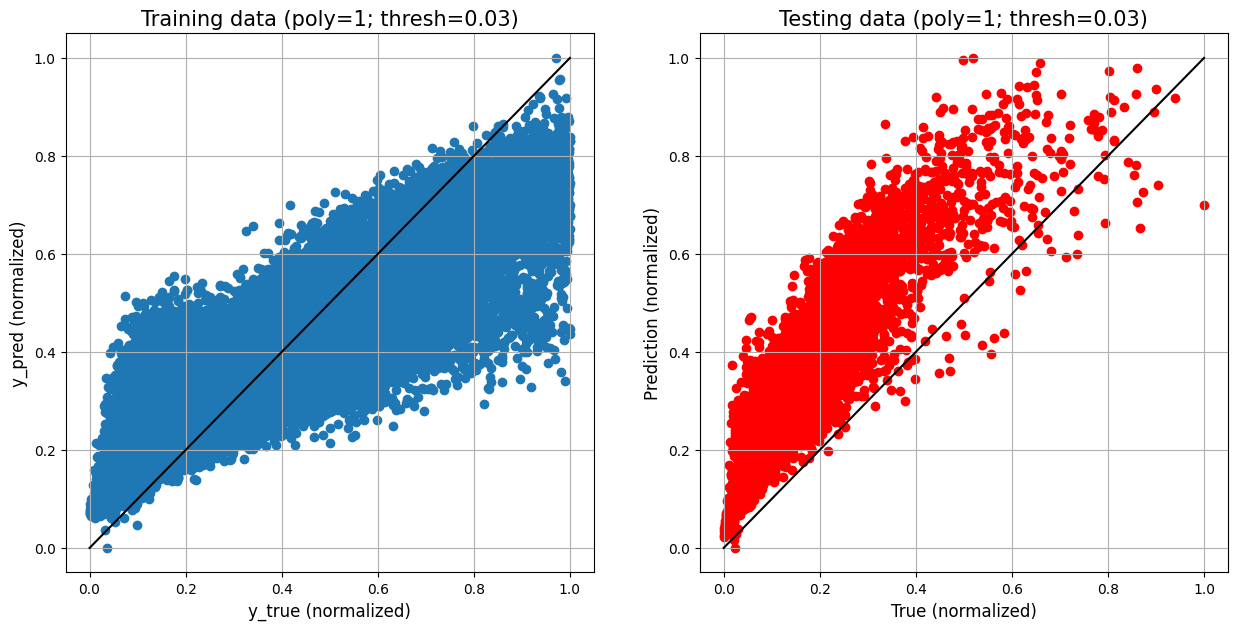
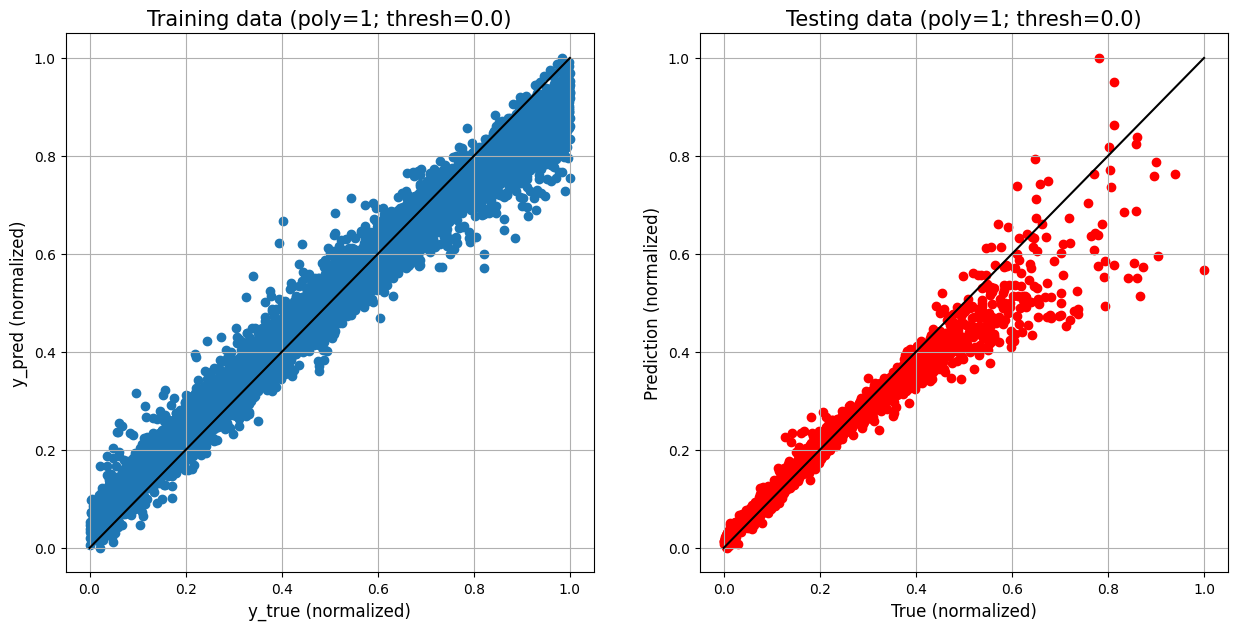
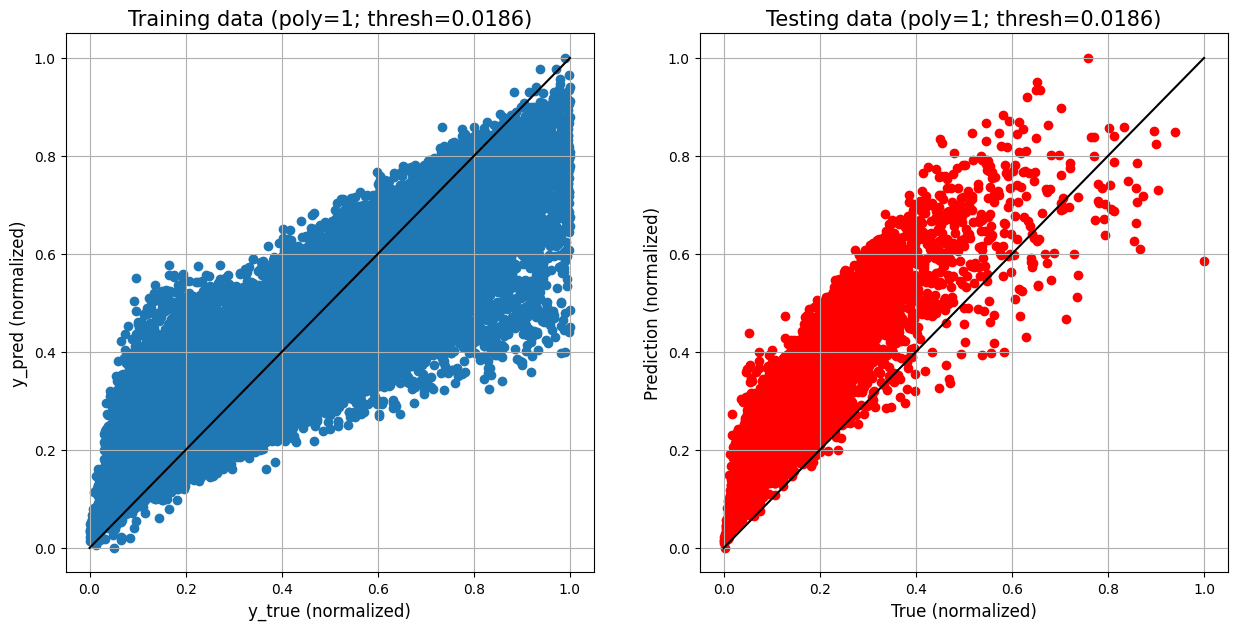
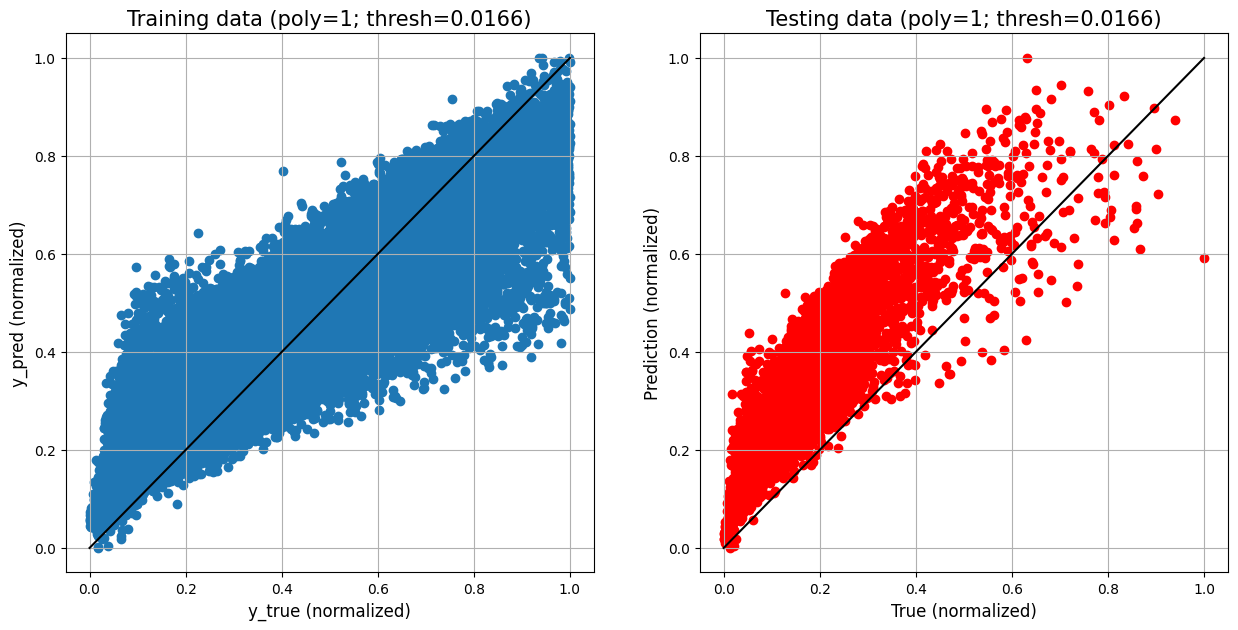
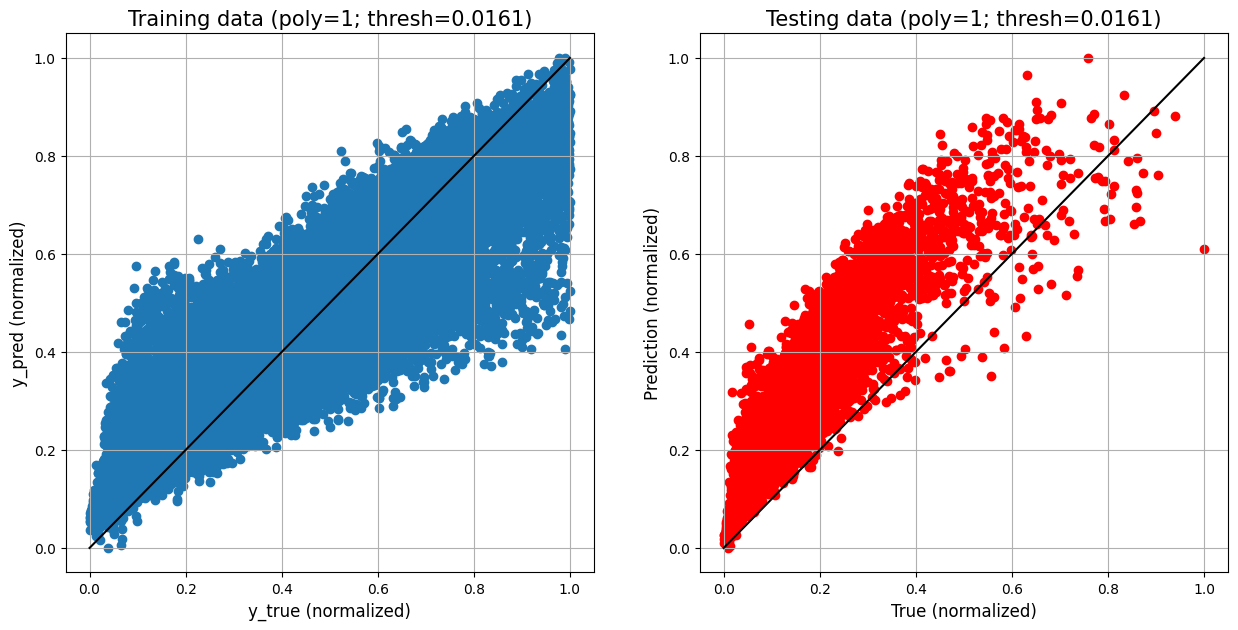
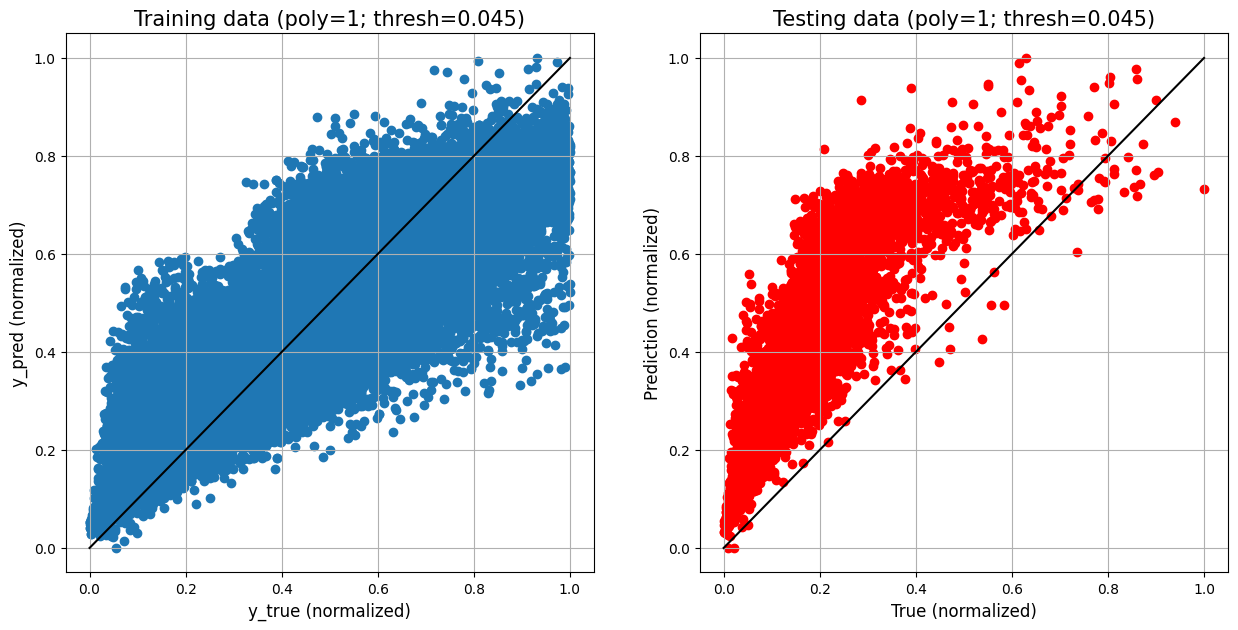
**Pipeline**

1. Creating new compound features in the training dataset.
2. Eliminating the observations that present outliers of the normalized target (y\_norm).
3. Scoring features with Lasso regression (grid search for alpha parameter).
4. Selecting features according to a relative score threshold (percentage of total score).
5. Standardizing the features’ vector.
6. Fitting a neural network with Pyrenn;
7. Applying the same steps to the features from the testing set, except for the elimination of outliers.
8. Predicting the test target feature (y\_norm).
9. Determining the mean absolute error (MAE) and coefficient of determination (R^2) for the training and testing sets.
10. Repeating the process for relative feature score threshold.

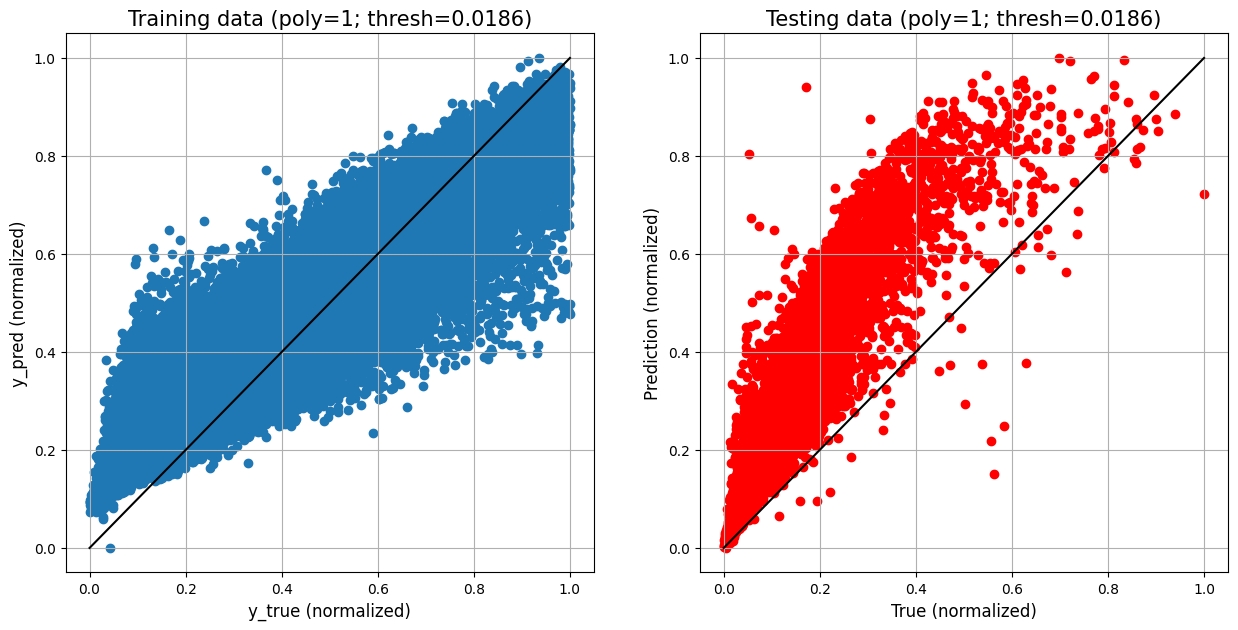
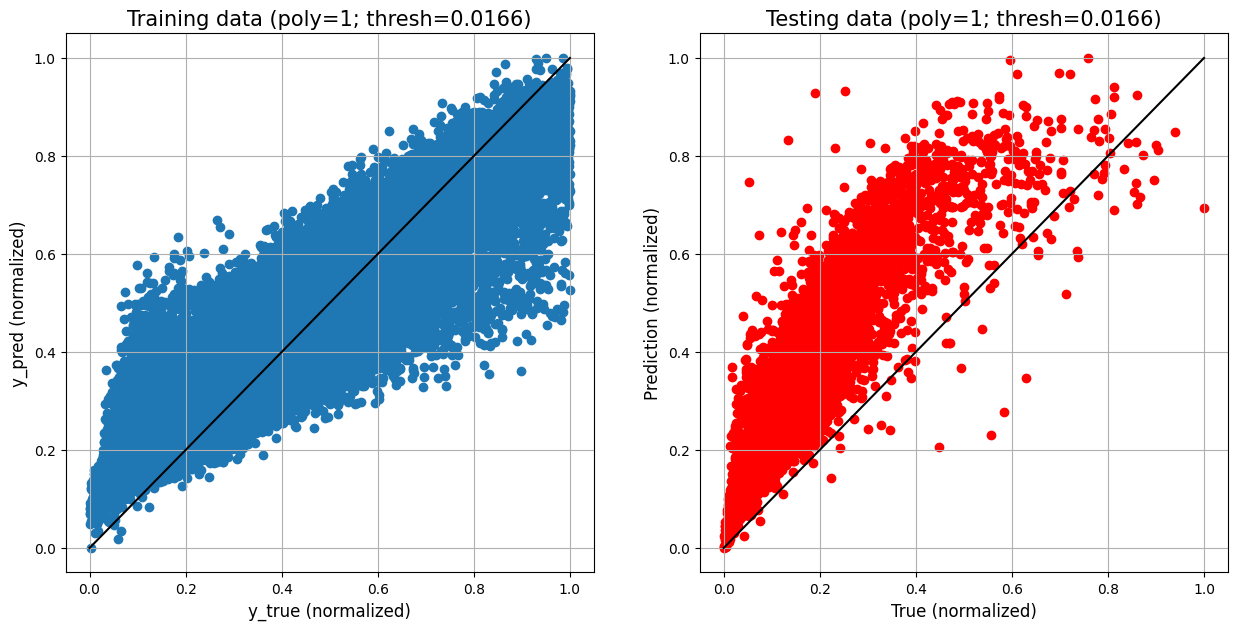
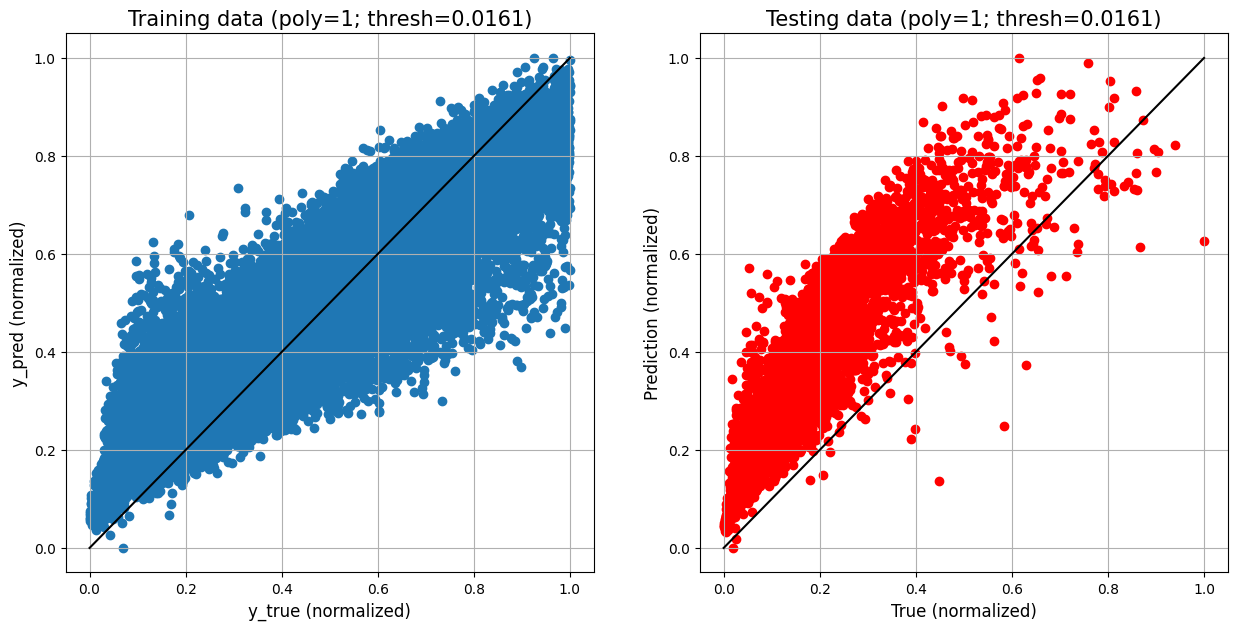
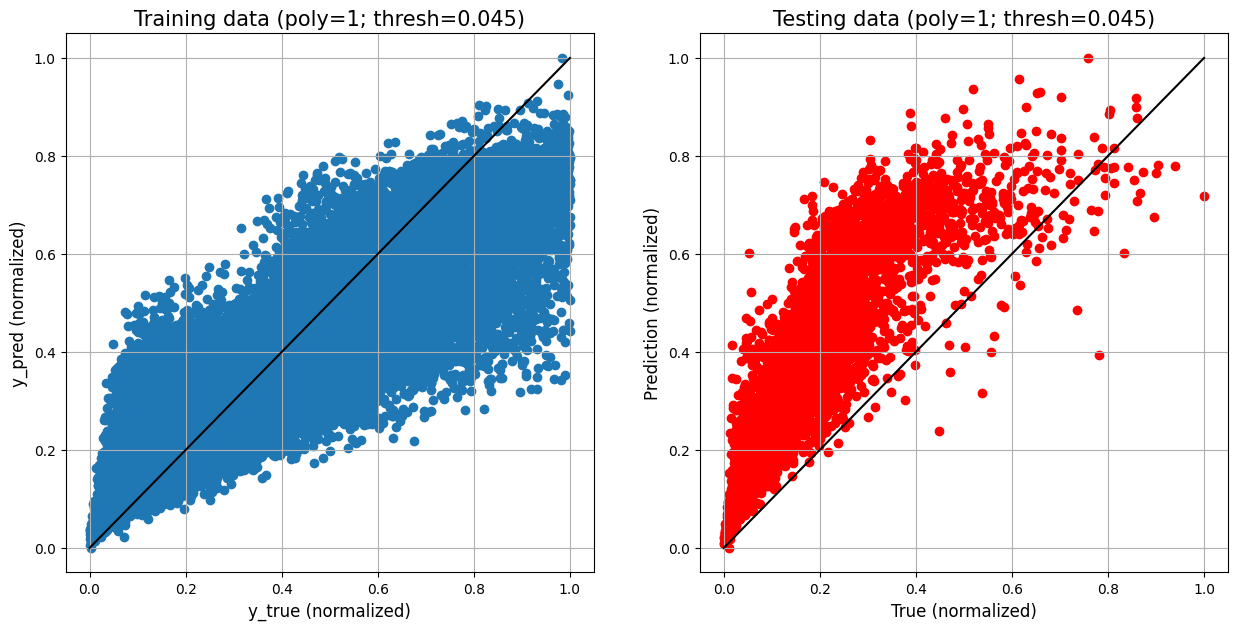
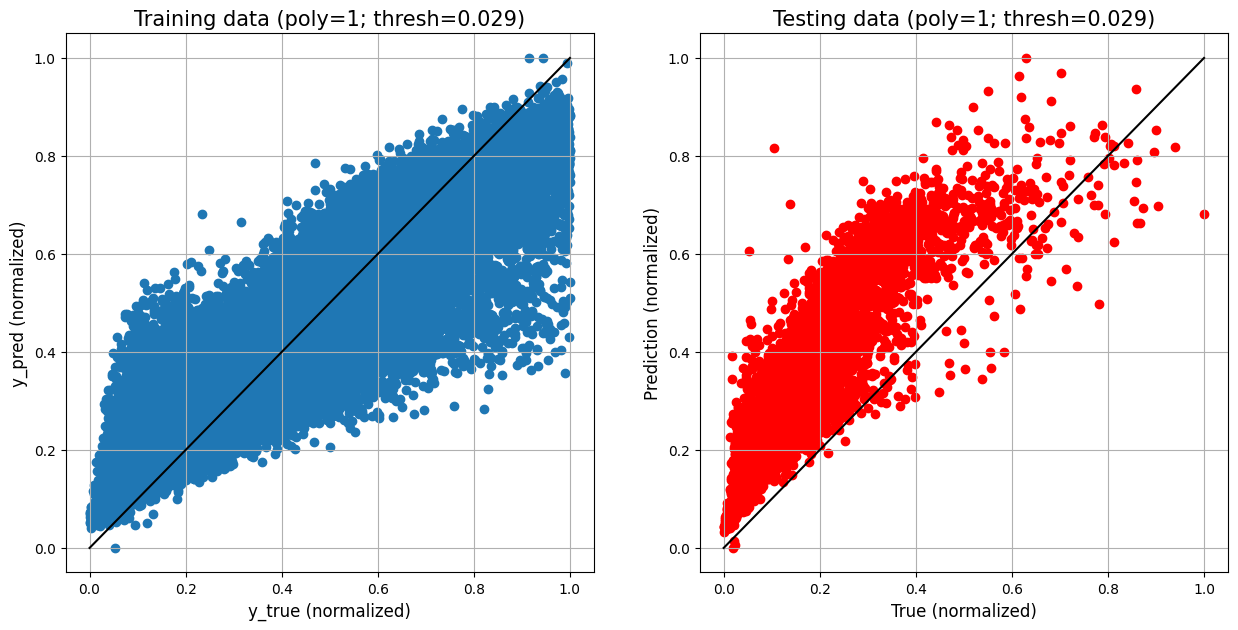
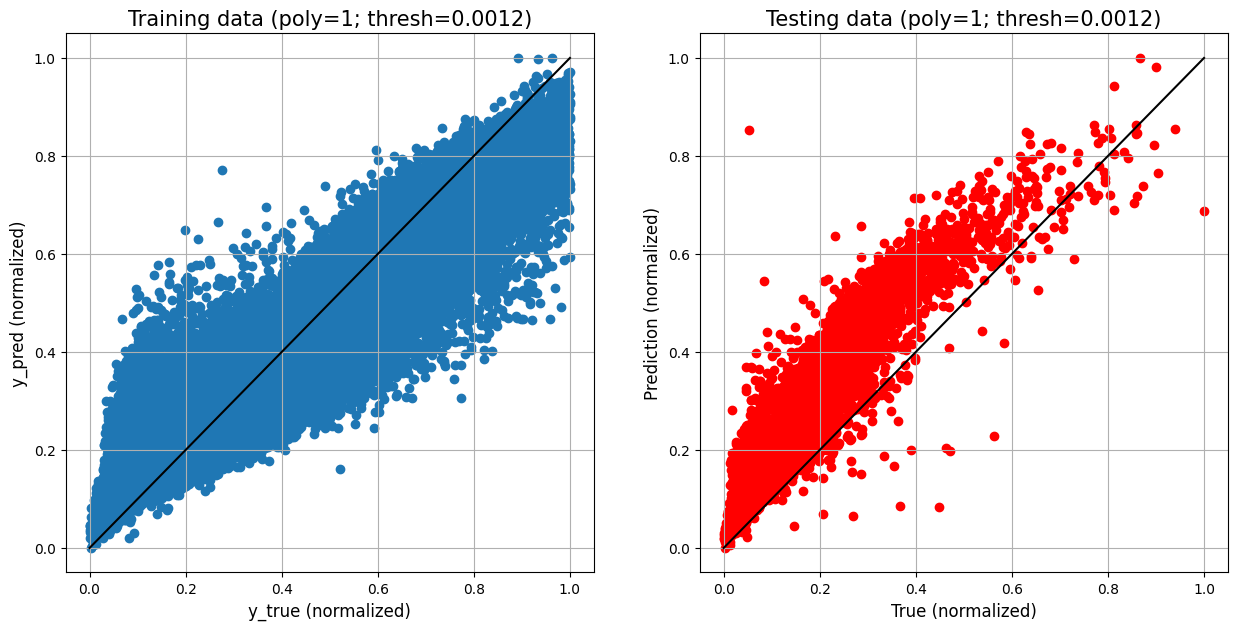
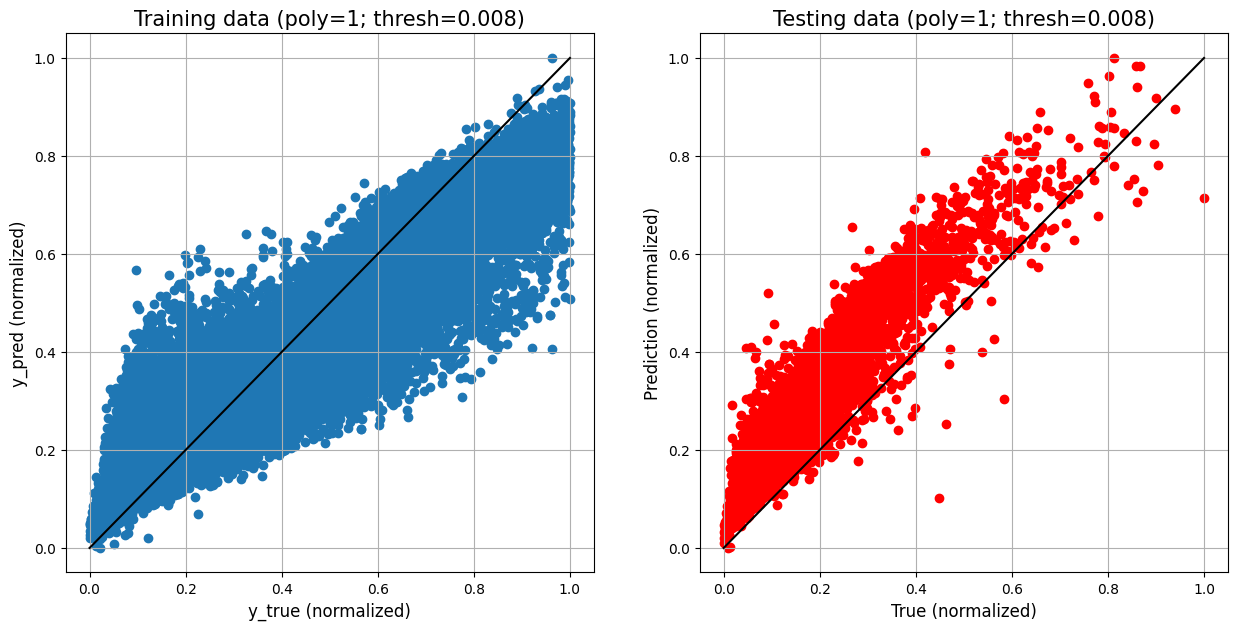
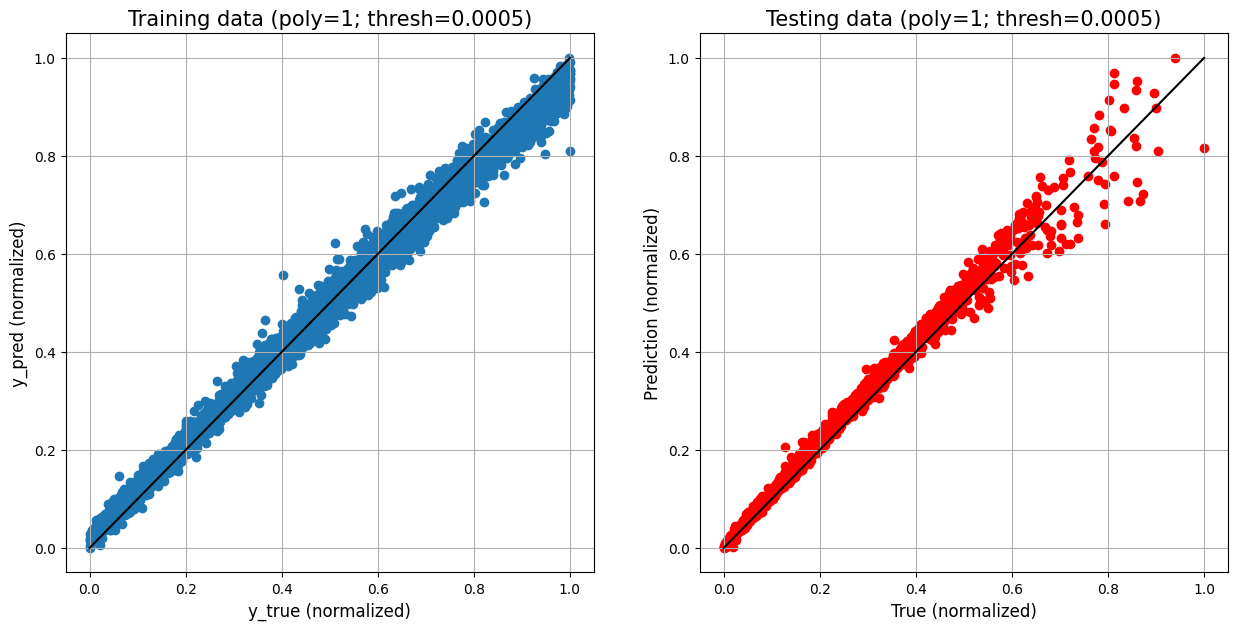
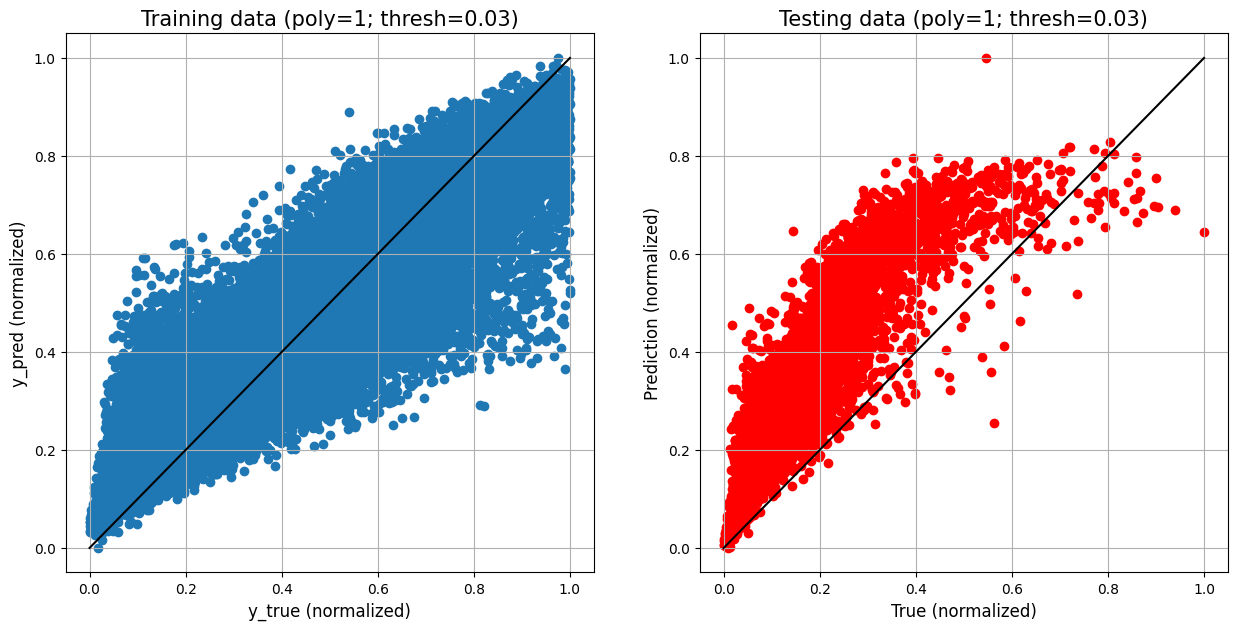
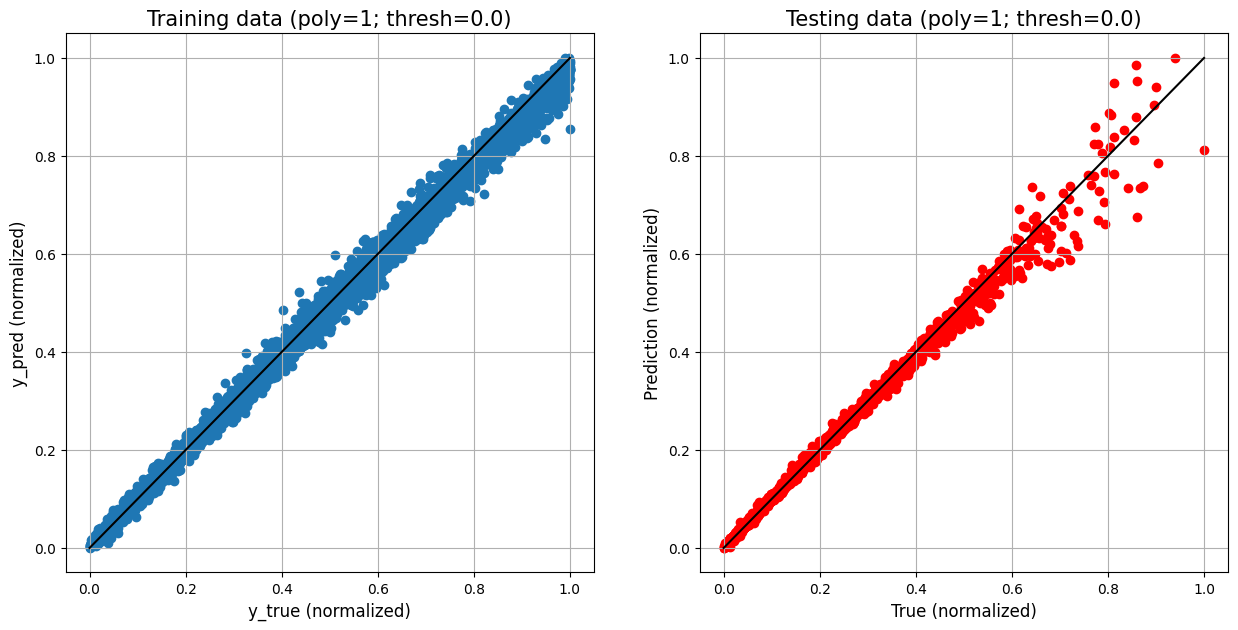
**Performance visualization**

Architecture: Number of Features :

****

Architecture: Number of Features :** **

Architecture: Number of Features :

****