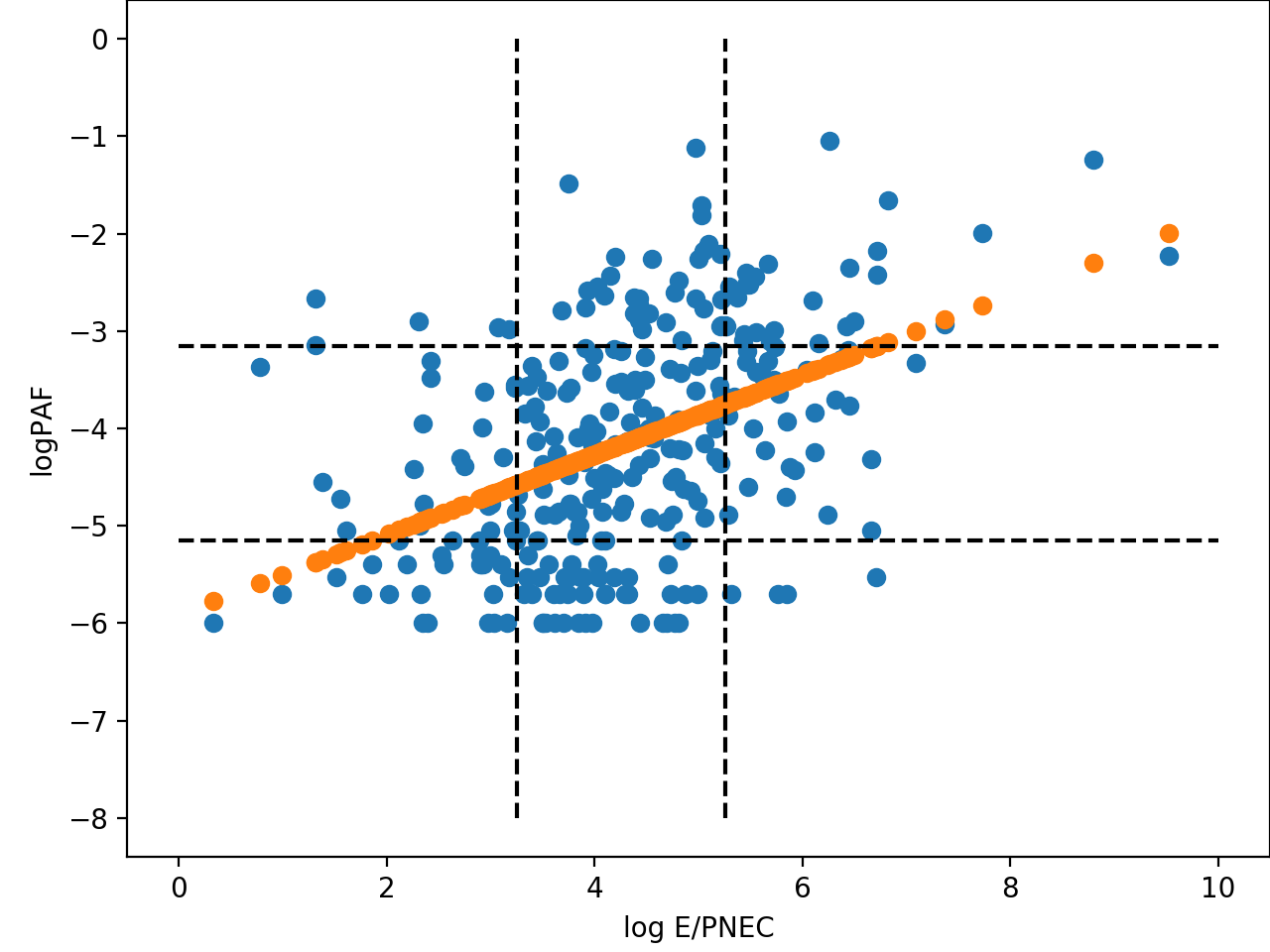
**Data investigation – SOLUTIONS**

In this exercise we try to determine if there are explanatory variables able describe the average basin toxic pressure of a chemical (PAF)

Figure 1 shows the extent to which the toxic pressure can be linearly predicted by the ratio of the mass emitted over the unit weight toxicity.



*Figure 1 – A regression of log PAF by E/PNEC. Dashed lines indicate 1 log unit deviations from the median of the population*

OLS Regression Results

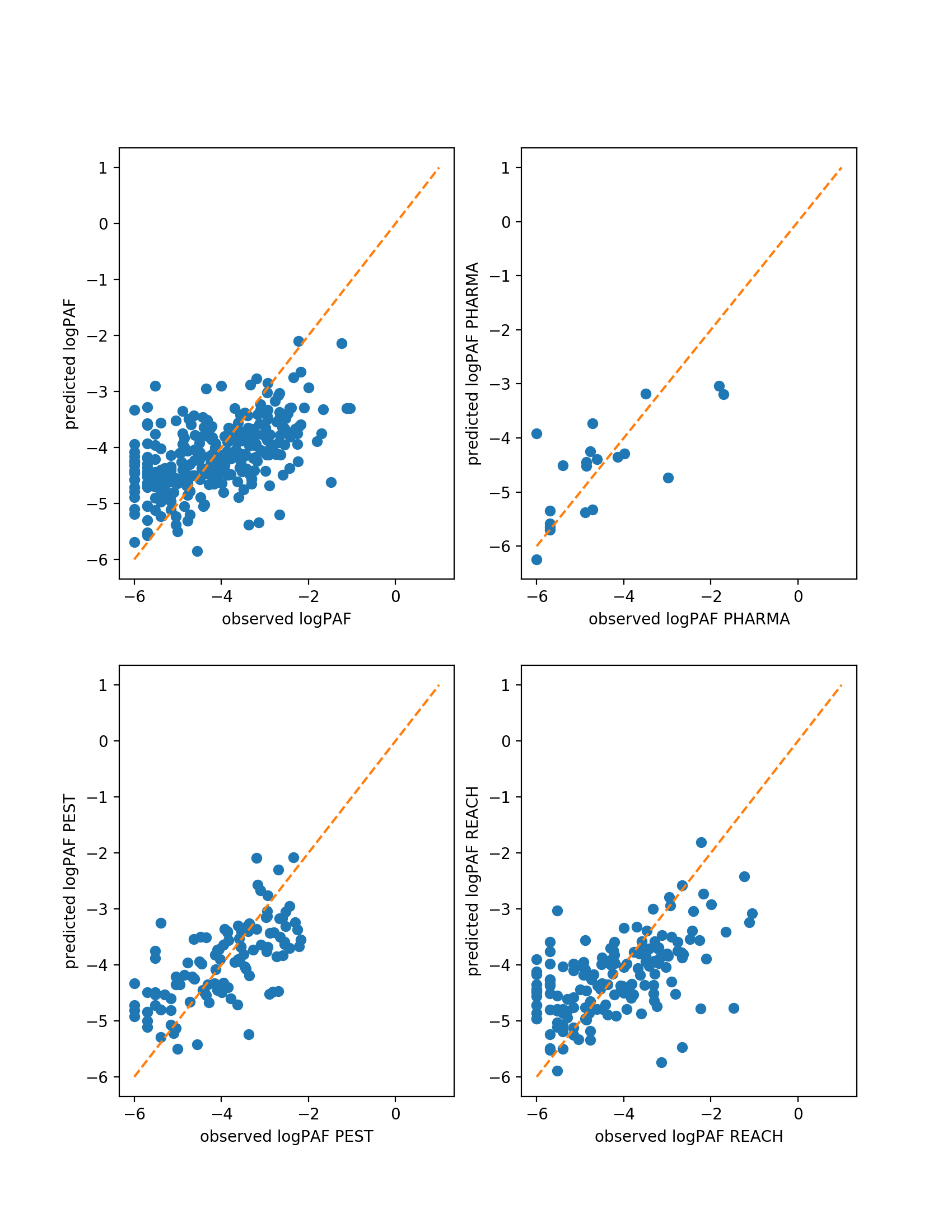
coef std err t P>|t| [0.025 0.975]

------------------------------------------------------------------------------

const -5.9112 0.213 -27.772 0.000 -6.330 -5.492

E/PNEC 0.4106 0.048 8.643 0.000 0.317 0.504

Figure 2 shows the extent to which the toxic pressure can be predicted by the linear combination of 12 independent variables. The first image is for a model created for all substances for which PAF > 0. Following that, the substances are broken down according to their type and the linear model is re-created. Following the figure are the tables describing the coefficients for each variable included in the multivariate linear regression.



*Figure 2 - A multi-linear regression of log PAF by 12 regressors. Dashed lines indicate 1 log unit deviations from the median of the population*

**REGMODEL (R = 0.27)**

coef std err t P>|t| [0.025 0.975]

----------------------------------------------------------------------------------

const 2522.3479 1936.961 1.302 0.194 -1290.507 6335.203

E2SNO (kmol/y) -3374.7880 2582.522 -1.307 0.192 -8458.412 1708.836

E2S1 (kmol/y) -3371.5092 2582.553 -1.305 0.193 -8455.194 1712.175

E2RIV (kmol/y) -3371.0774 2582.571 -1.305 0.193 -8454.797 1712.642

E/PNEC 0.4152 0.053 7.816 0.000 0.311 0.520

%Stor 2.4554 0.978 2.512 0.013 0.531 4.380

%Export 1.1887 0.955 1.245 0.214 -0.691 3.068

%Decay 1.9357 0.882 2.196 0.029 0.200 3.671

%Ion 841.2175 645.651 1.303 0.194 -429.728 2112.163

%Prec 839.9113 645.643 1.301 0.194 -431.018 2110.841

%Fug 841.2191 645.668 1.303 0.194 -429.761 2112.199

logKOW0 0.0241 0.036 0.666 0.506 -0.047 0.095

kdw -0.1888 0.091 -2.077 0.039 -0.368 -0.010

**PHARMAMODEL (R = 0.52)**

coef std err t P>|t| [0.025 0.975]

----------------------------------------------------------------------------------

const -5357.3904 1.7e+04 -0.315 0.761 -4.45e+04 3.38e+04

E2SNO (kmol/y) 8426.2457 2.3e+04 0.367 0.723 -4.46e+04 6.14e+04

E2S1 (kmol/y) 7959.0802 2.24e+04 0.355 0.732 -4.38e+04 5.97e+04

E2RIV (kmol/y) 7977.7028 2.25e+04 0.355 0.732 -4.39e+04 5.98e+04

E/PNEC 0.5064 0.281 1.801 0.109 -0.142 1.155

%Stor -342.0634 2251.062 -0.152 0.883 -5533.021 4848.894

%Export -272.4093 2273.013 -0.120 0.908 -5513.986 4969.167

%Decay -287.5408 2268.213 -0.127 0.902 -5518.050 4942.968

%Ion -2351.9353 6469.081 -0.364 0.726 -1.73e+04 1.26e+04

%Prec -651.3536 8459.484 -0.077 0.941 -2.02e+04 1.89e+04

%Fug -2354.1015 6469.319 -0.364 0.725 -1.73e+04 1.26e+04

logKOW0 0.0341 0.317 0.108 0.917 -0.697 0.765

kdw -0.0286 1.375 -0.021 0.984 -3.200 3.143

**PESTMODEL (R = 0.44)**

coef std err t P>|t| [0.025 0.975]

----------------------------------------------------------------------------------

const 119.4591 2317.174 0.052 0.959 -4477.748 4716.666

E2SNO (kmol/y) -187.4447 3089.745 -0.061 0.952 -6317.411 5942.521

E2S1 (kmol/y) -179.5922 3089.639 -0.058 0.954 -6309.348 5950.163

E2RIV (kmol/y) -158.1911 3089.941 -0.051 0.959 -6288.546 5972.164

E/PNEC 0.4676 0.079 5.898 0.000 0.310 0.625

%Stor 0.7548 3.710 0.203 0.839 -6.606 8.115

%Export 0.4495 3.594 0.125 0.901 -6.682 7.581

%Decay 0.5559 3.537 0.157 0.875 -6.461 7.573

%Ion 52.9741 772.510 0.069 0.945 -1479.664 1585.612

%Prec 13.8389 772.234 0.018 0.986 -1518.251 1545.928

%Fug 52.6461 772.597 0.068 0.946 -1480.165 1585.457

logKOW0 0.1186 0.068 1.753 0.083 -0.016 0.253

kdw -0.2257 0.150 -1.507 0.135 -0.523 0.071

**REACHMODEL (R = 0.31)**

coef std err t P>|t| [0.025 0.975]

----------------------------------------------------------------------------------

const 4538.6750 2942.897 1.542 0.125 -1276.850 1.04e+04

E2SNO (kmol/y) -6043.2251 3923.997 -1.540 0.126 -1.38e+04 1711.073

E2S1 (kmol/y) -6061.4618 3923.762 -1.545 0.125 -1.38e+04 1692.372

E2RIV (kmol/y) -6057.0192 3923.801 -1.544 0.125 -1.38e+04 1696.892

E/PNEC 0.5287 0.078 6.769 0.000 0.374 0.683

%Stor 1.0861 1.270 0.855 0.394 -1.423 3.595

%Export -1.1757 1.267 -0.928 0.355 -3.679 1.328

%Decay 0.3089 1.055 0.293 0.770 -1.777 2.395

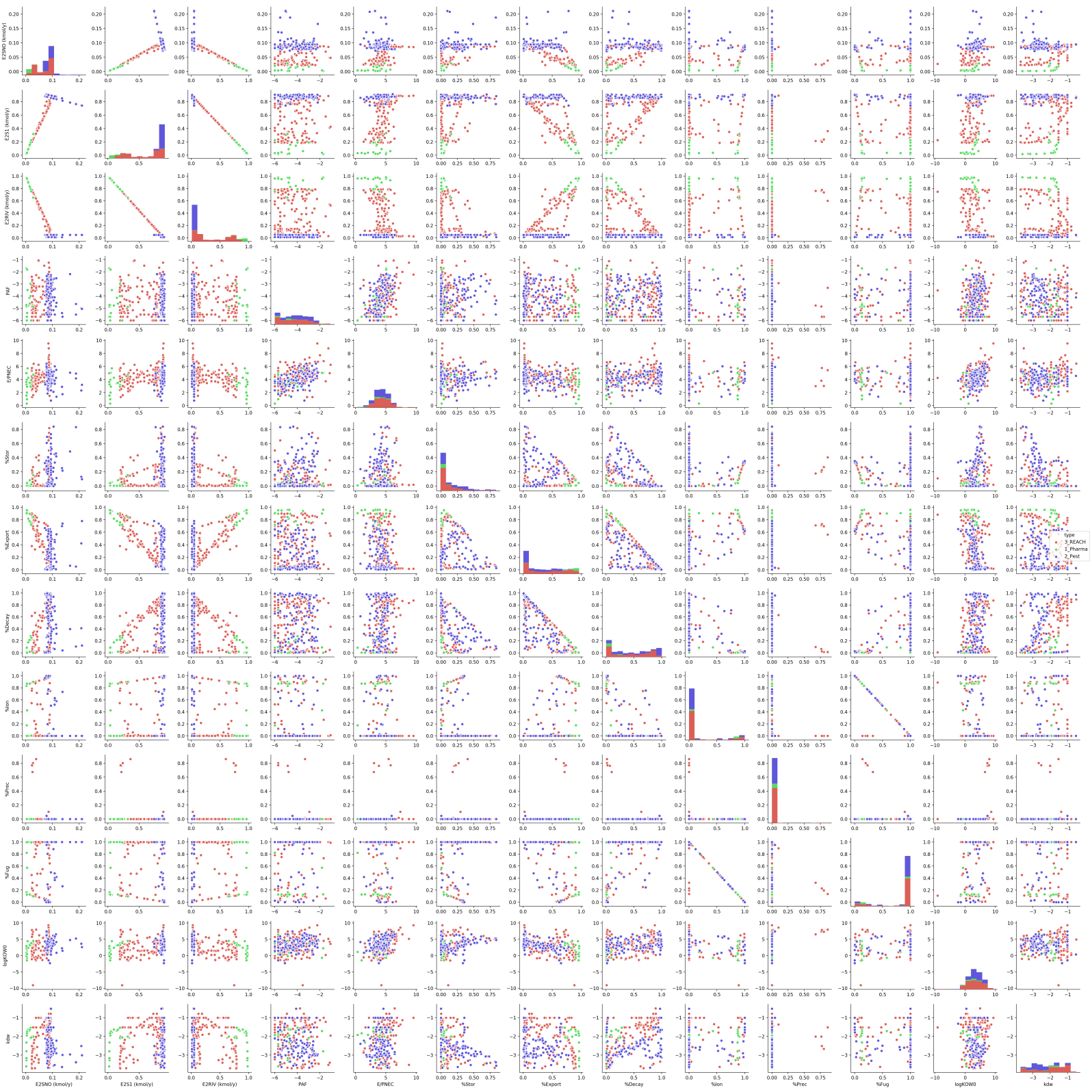
%Ion 1513.6222 980.981 1.543 0.125 -424.916 3452.160

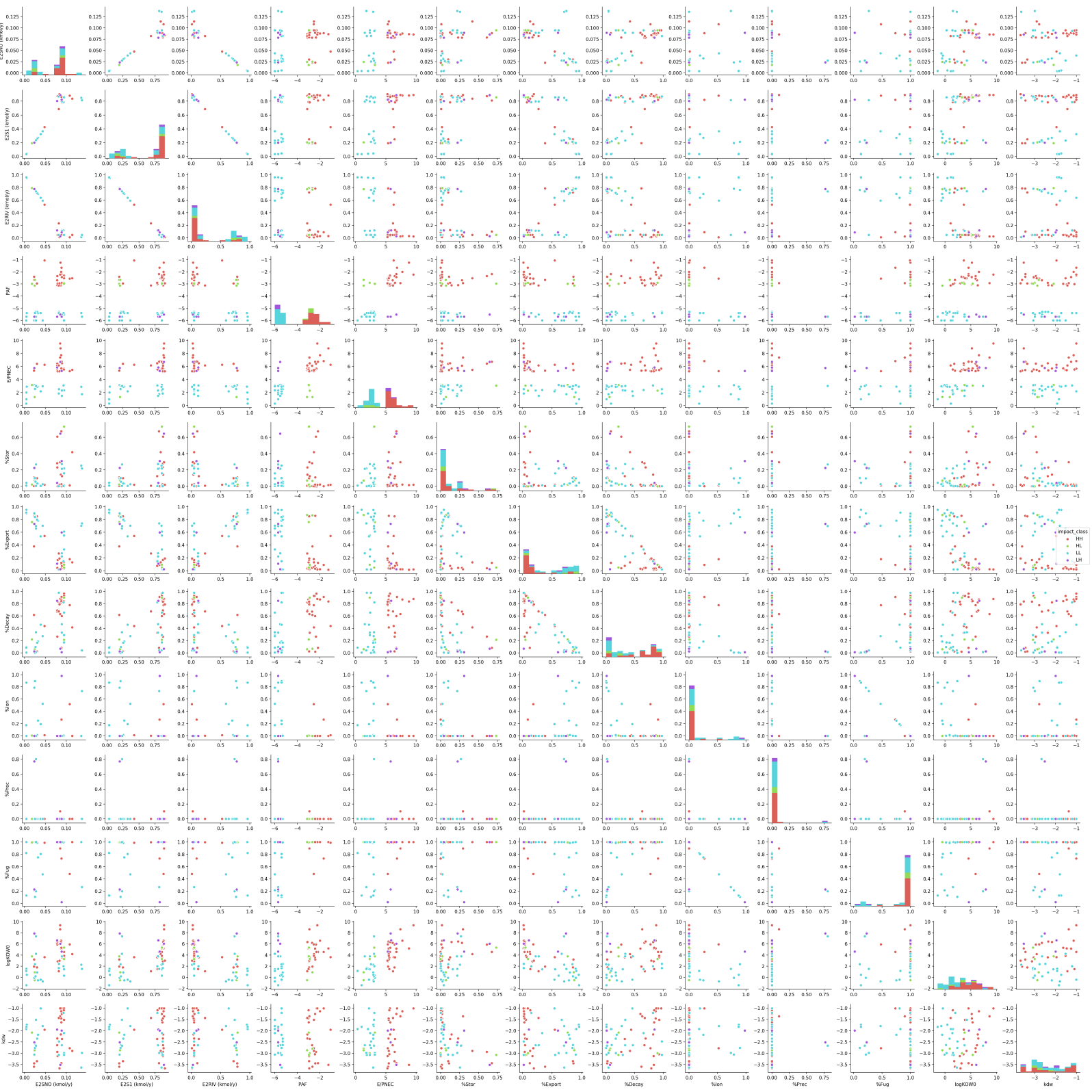
%Prec 1511.5709 980.949 1.541 0.125 -426.904 3450.046

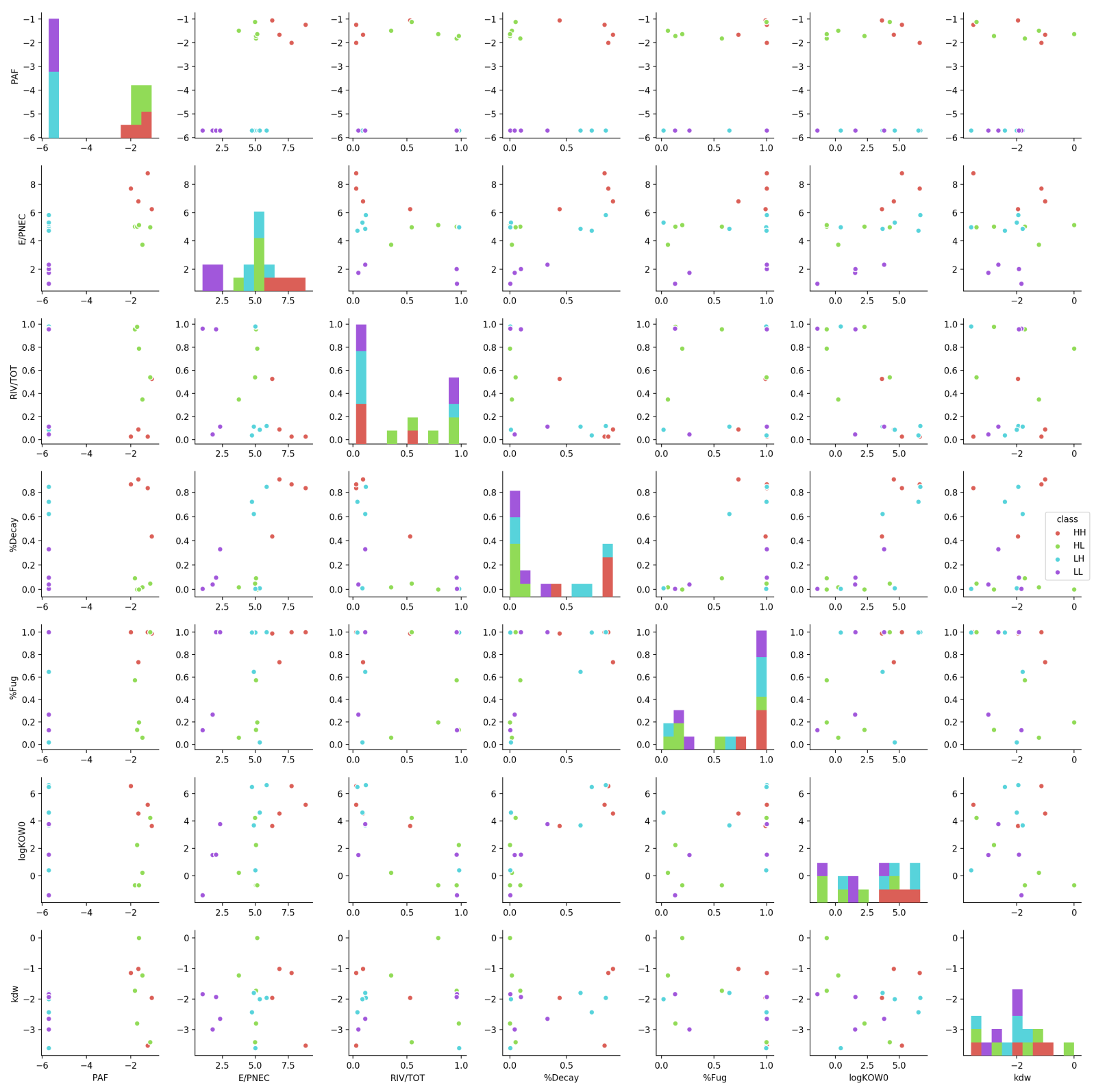
%Fug 1513.4820 980.968 1.543 0.125 -425.030 3451.994

logKOW0 0.0597 0.055 1.077 0.283 -0.050 0.169

kdw -0.0109 0.134 -0.081 0.935 -0.275 0.253

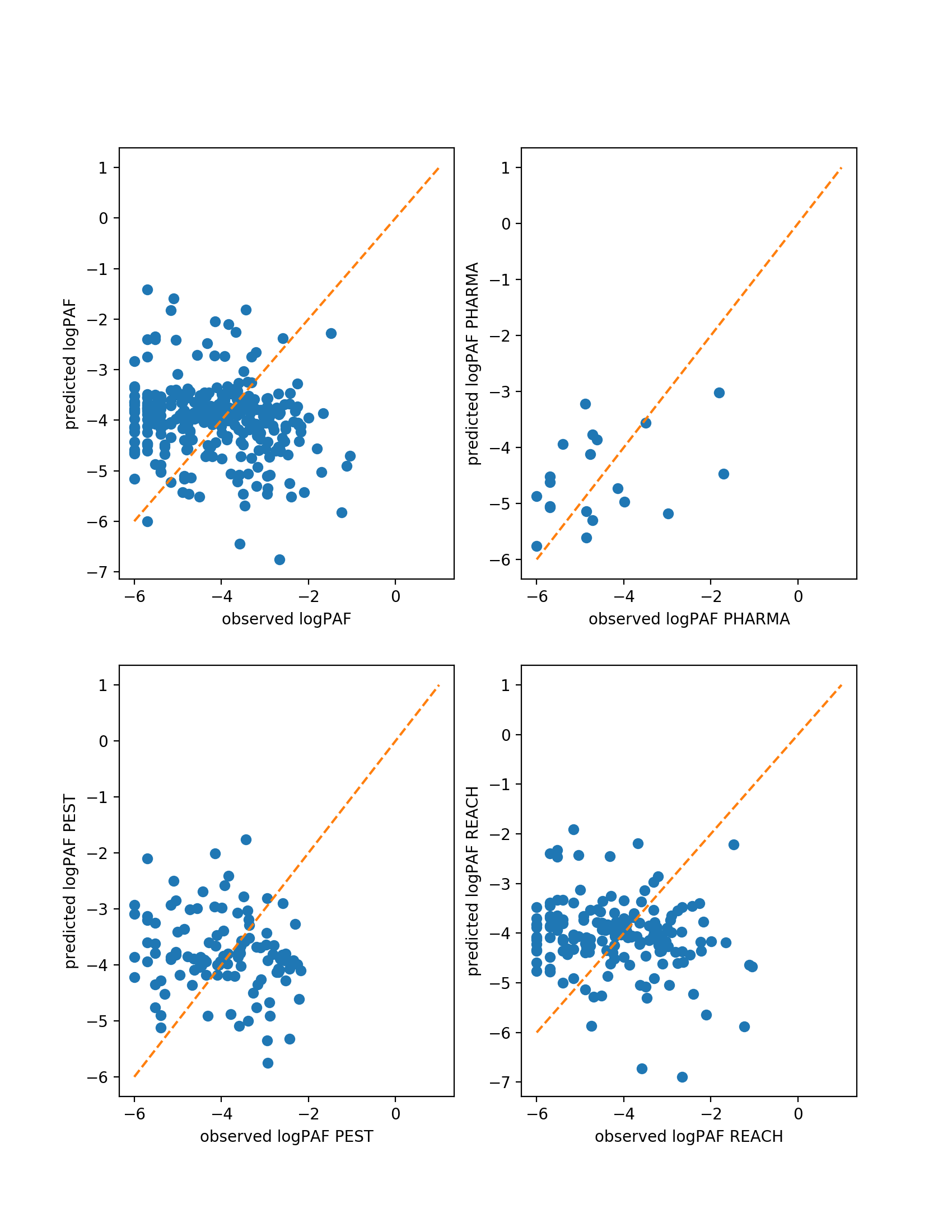




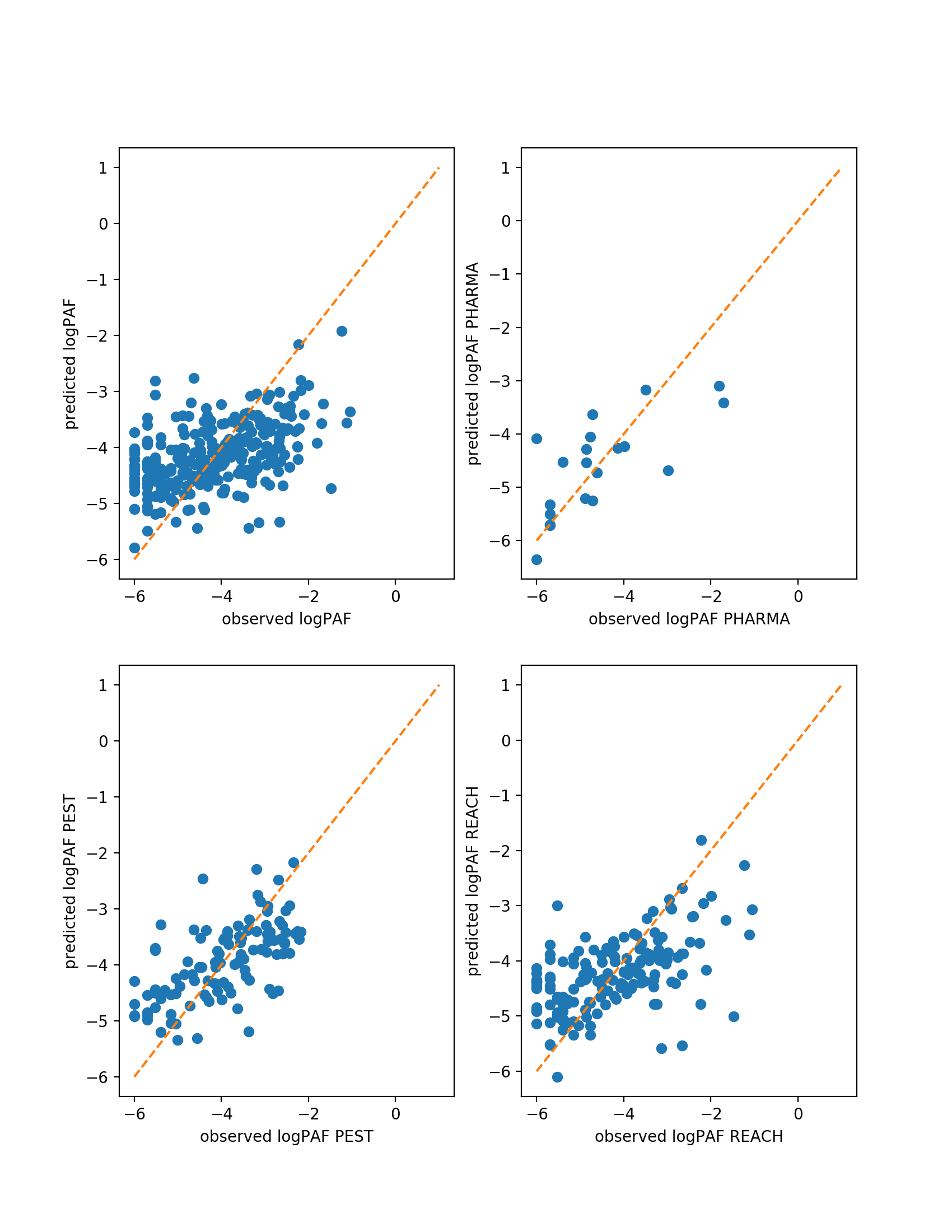


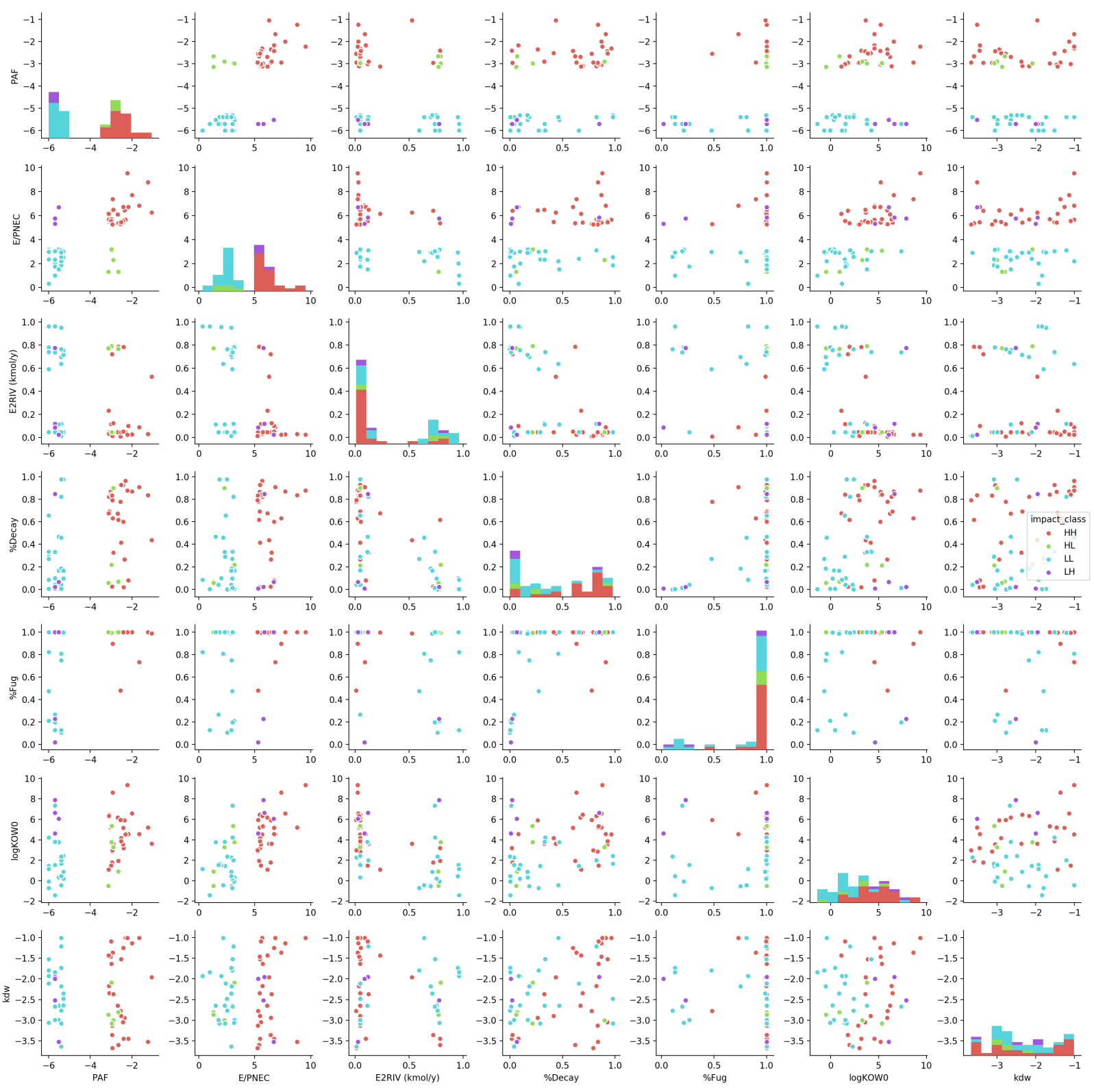
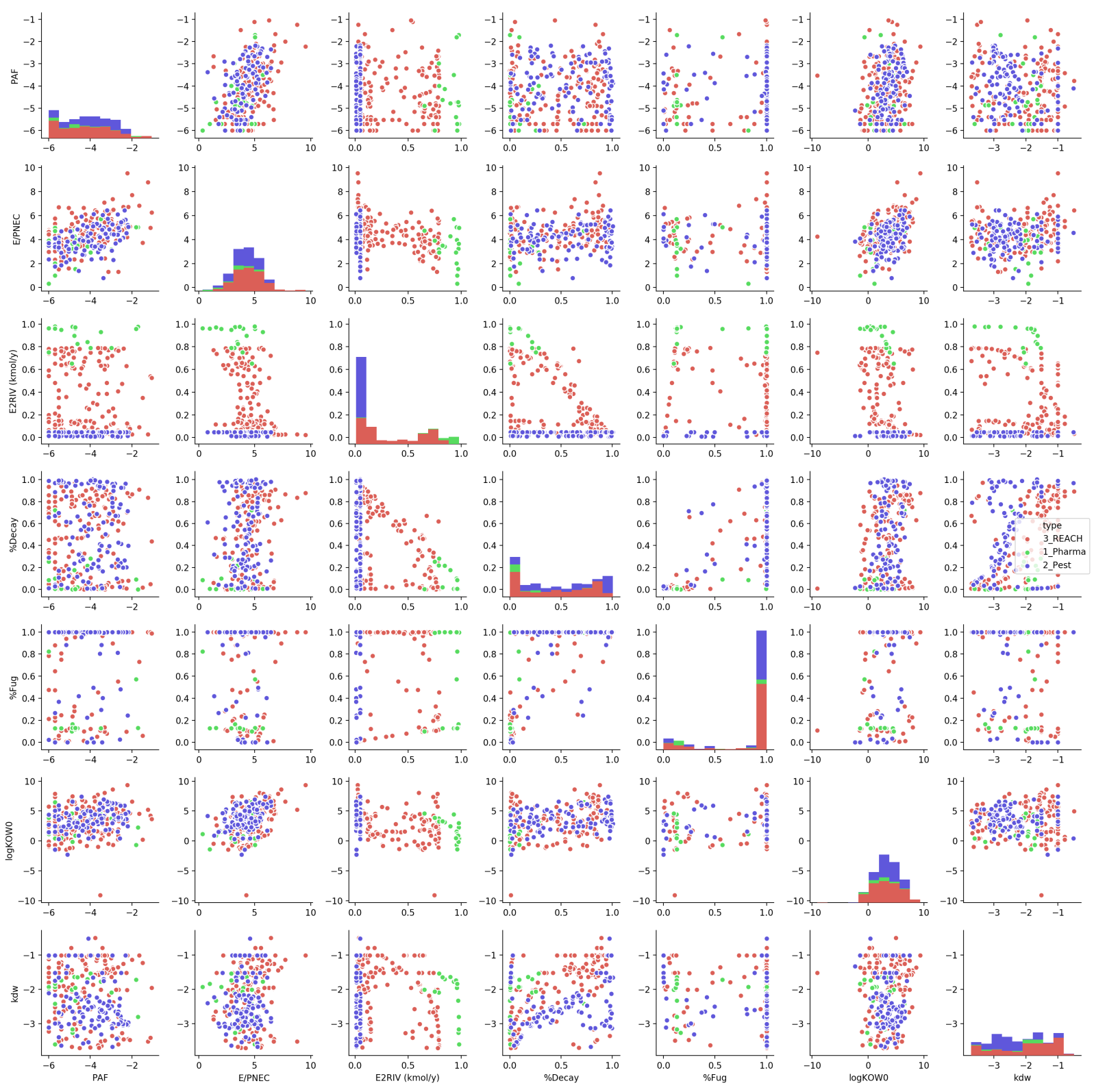
To try and reduce collinearity in the multi-linear regression, fewer variables were used. Collinearity was still present.

colinreg = regression[['type','impact\_class','PAF', 'E/PNEC', 'E2RIV (kmol/y)' ,'%Decay' ,'%Fug','logKOW0' ,'kdw']].copy()



colinreg = regression[['type','impact\_class','PAF', 'E/PNEC', 'E2RIV (kmol/y)' ,'E2S1 (kmol/y)', '%Export','%Decay' ,'%Fug', '%Ion','logKOW0','kdw']].copy()





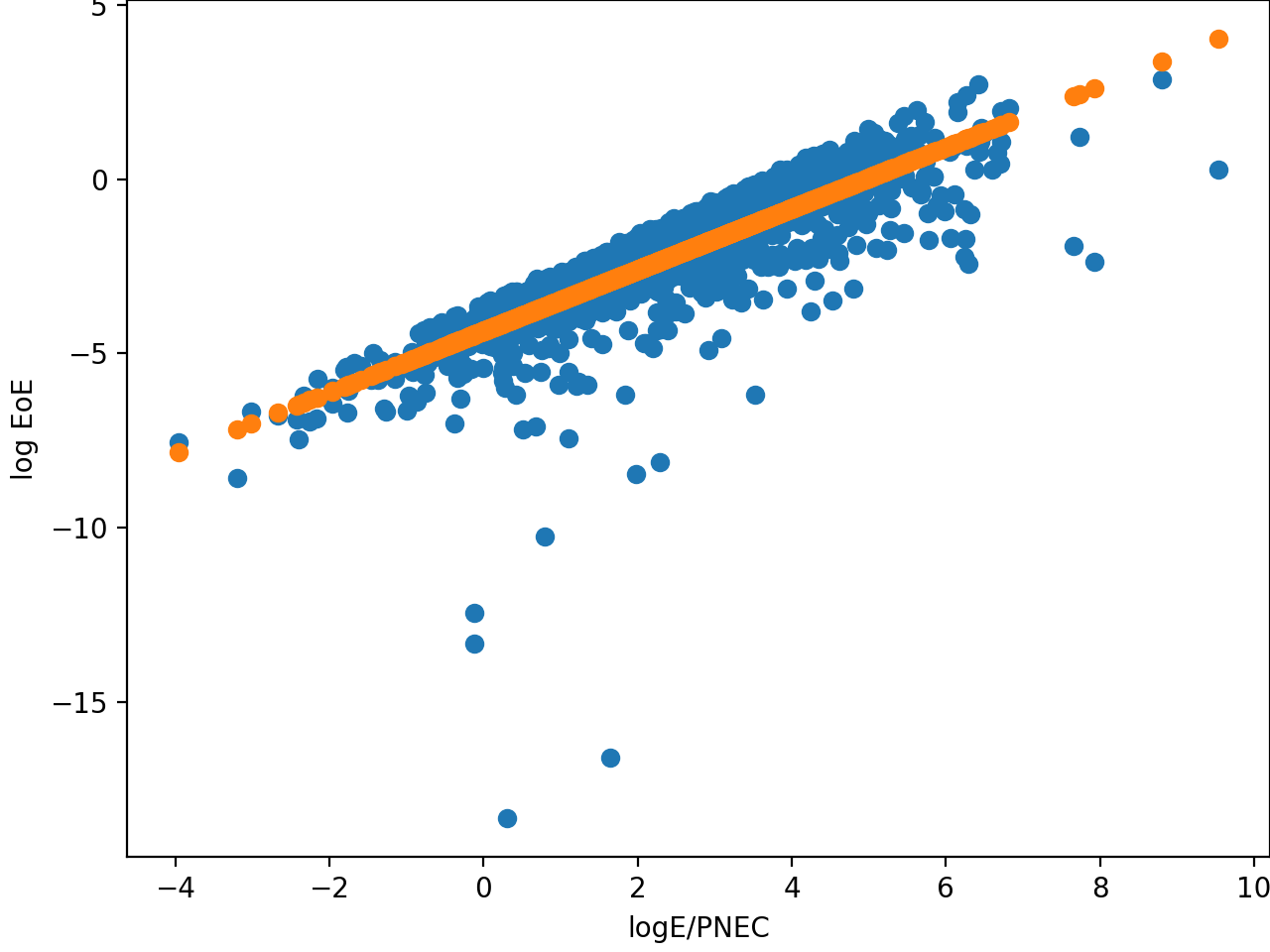
No significant trends have been identified

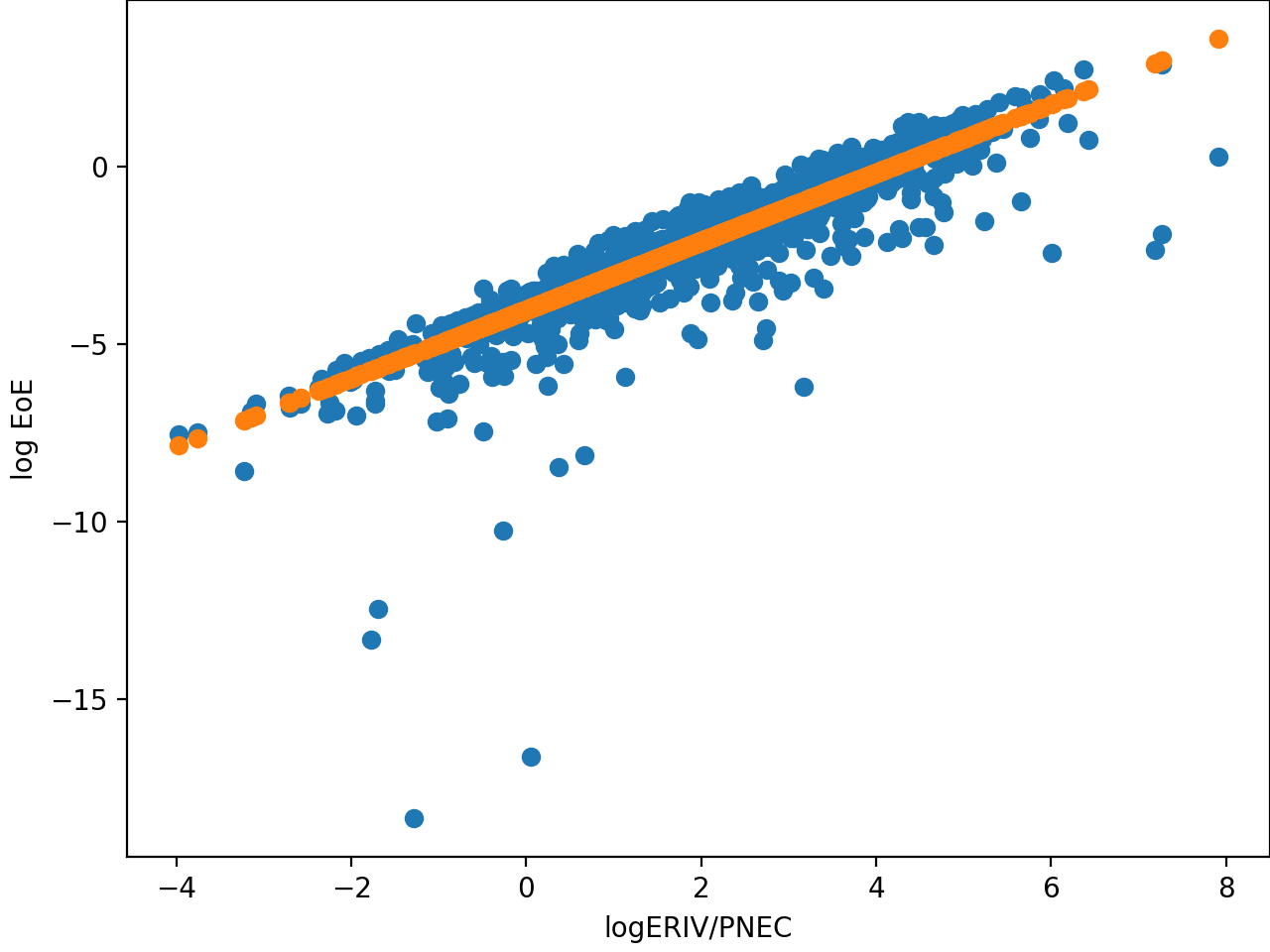
The following qualitative observations are made:

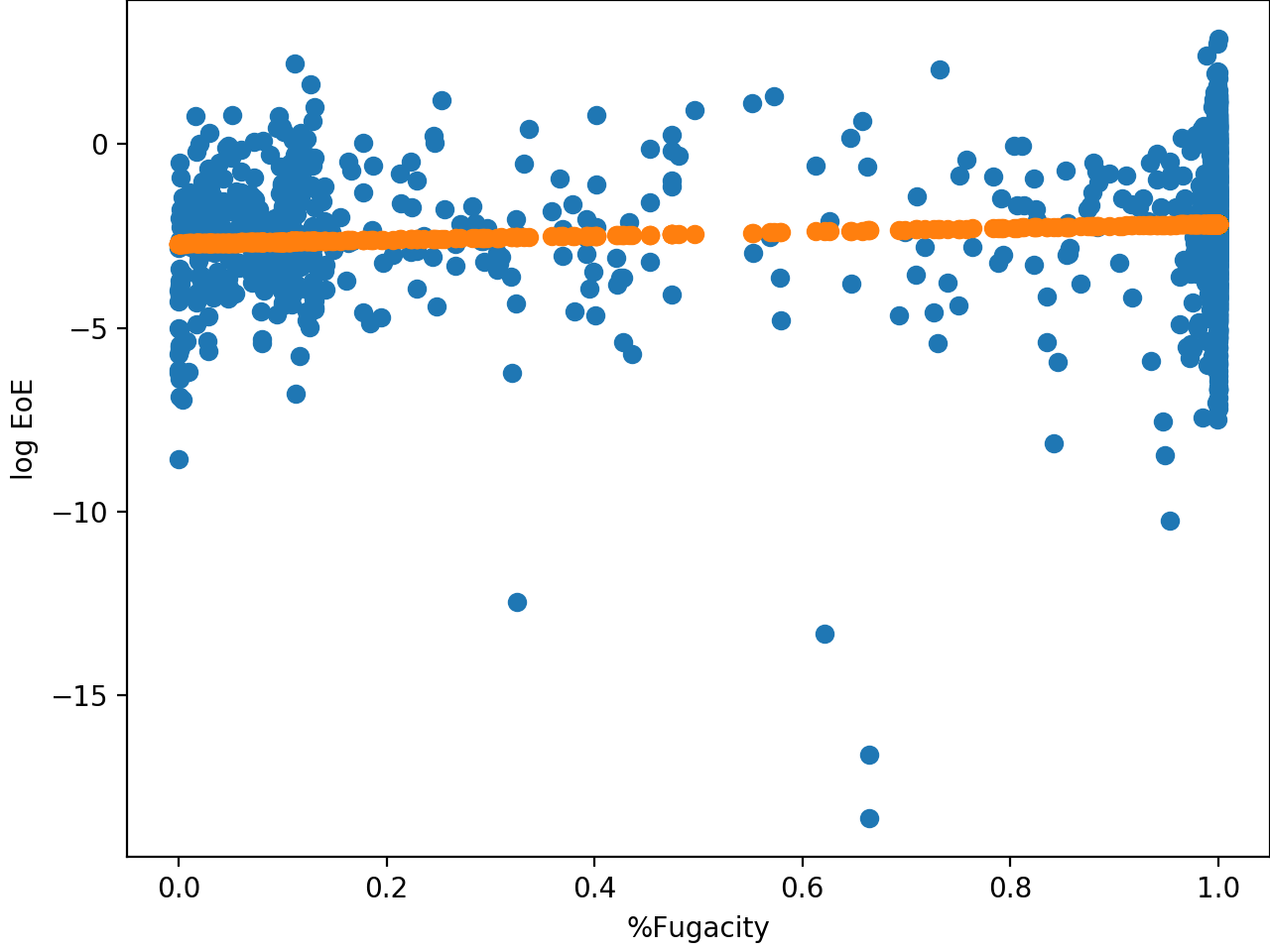
1. HH chemicals tend to stay in the system (not exported) and decay. They also tend to favour the fugacity phase
2. They have a higher KoW and are emitted to soil
3. HL chemicals are often released directly to River mostly
4. All HH chemicals are reach
5. Pharmaceuticals tend to be exported

**EoE and FoE examination**

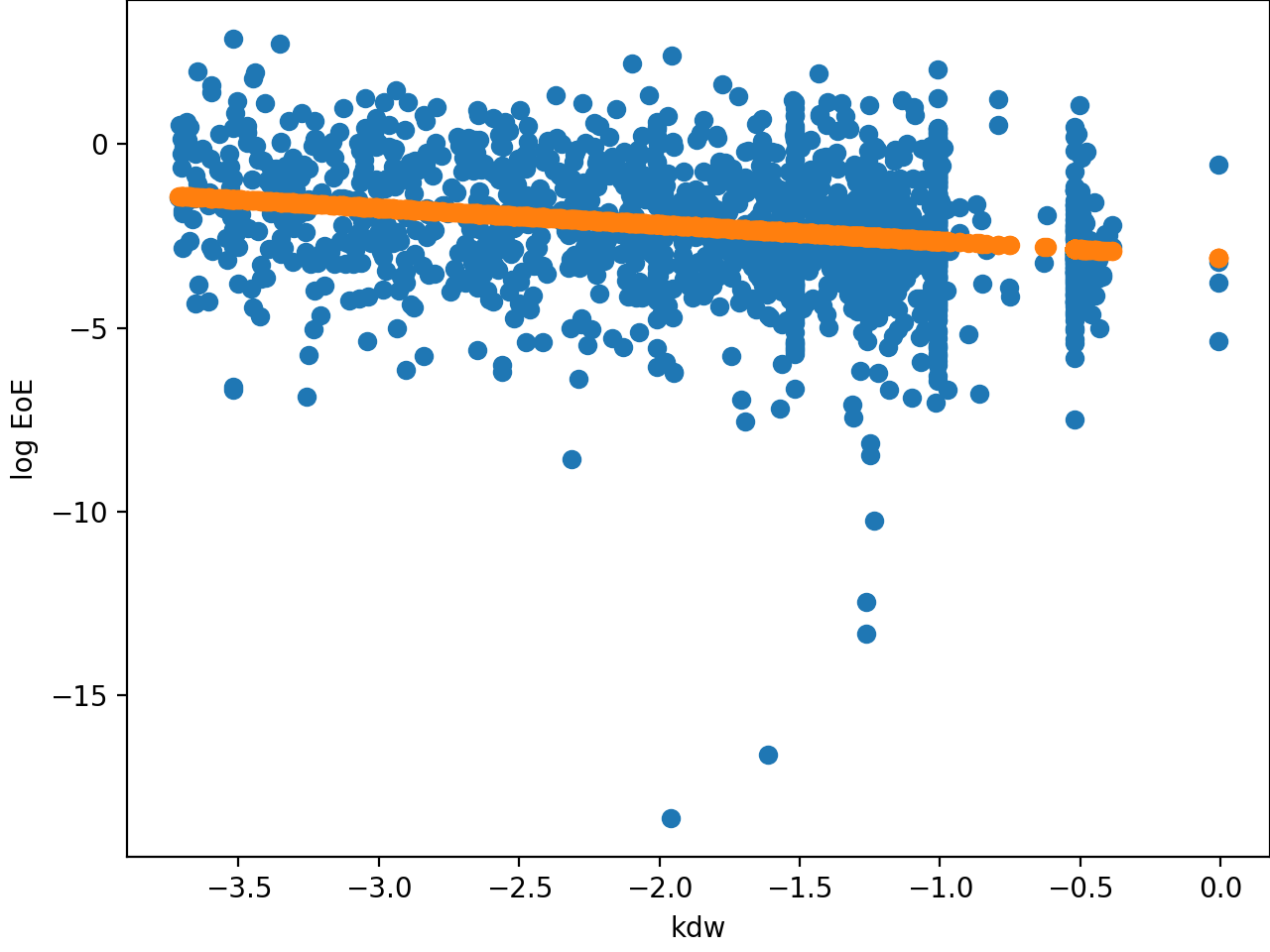
We expect now that using EoE instead of PAF will provide better relationships, as there is less inherent randomness/uncertainty in EoE compared to PAF

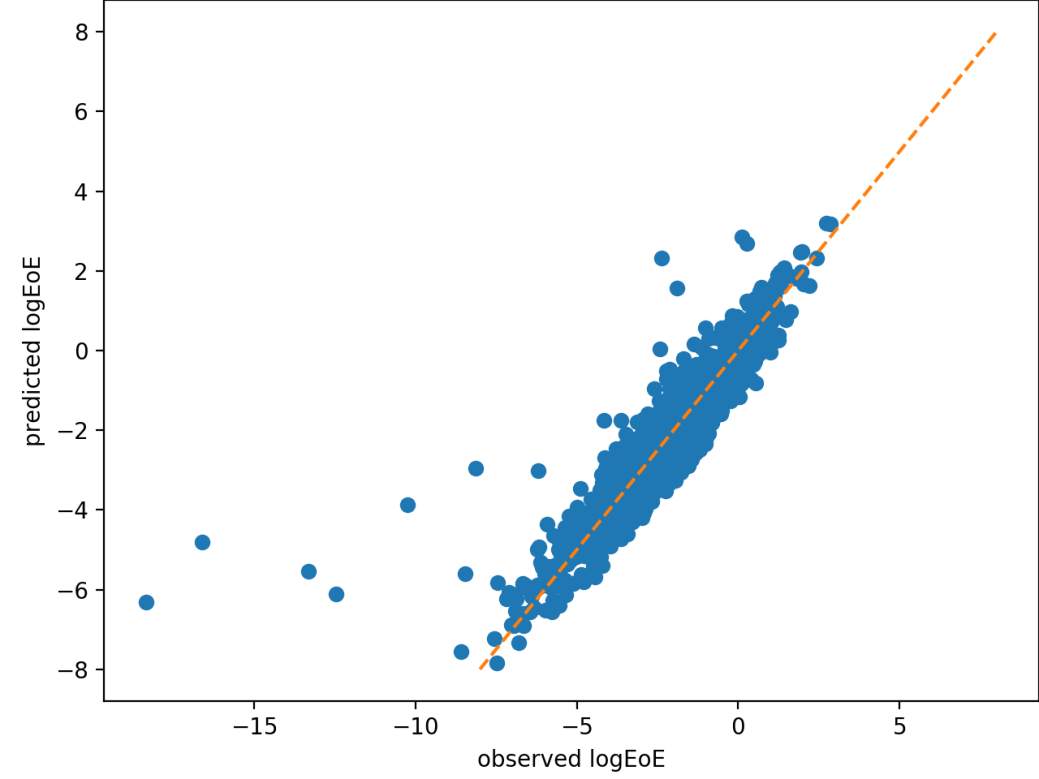


****

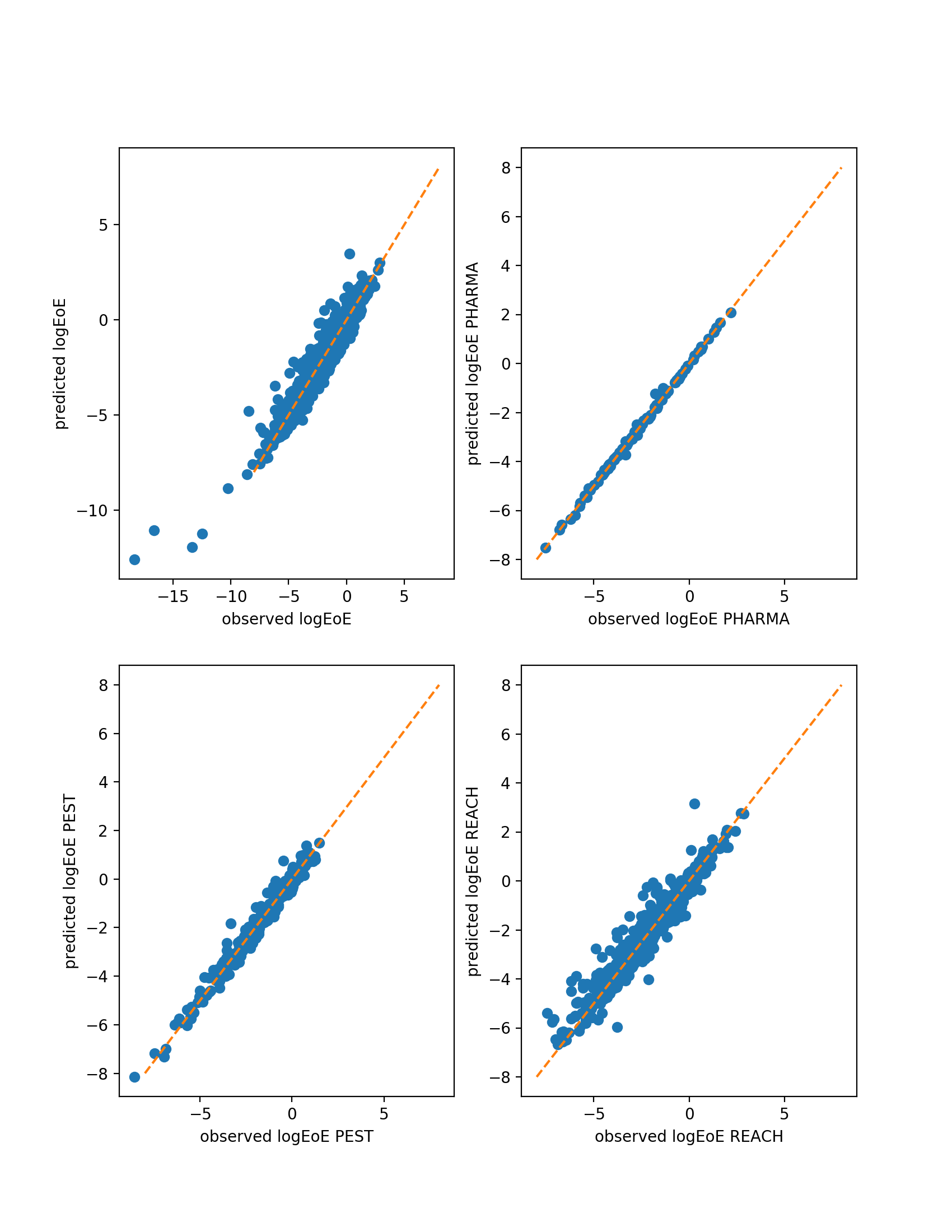


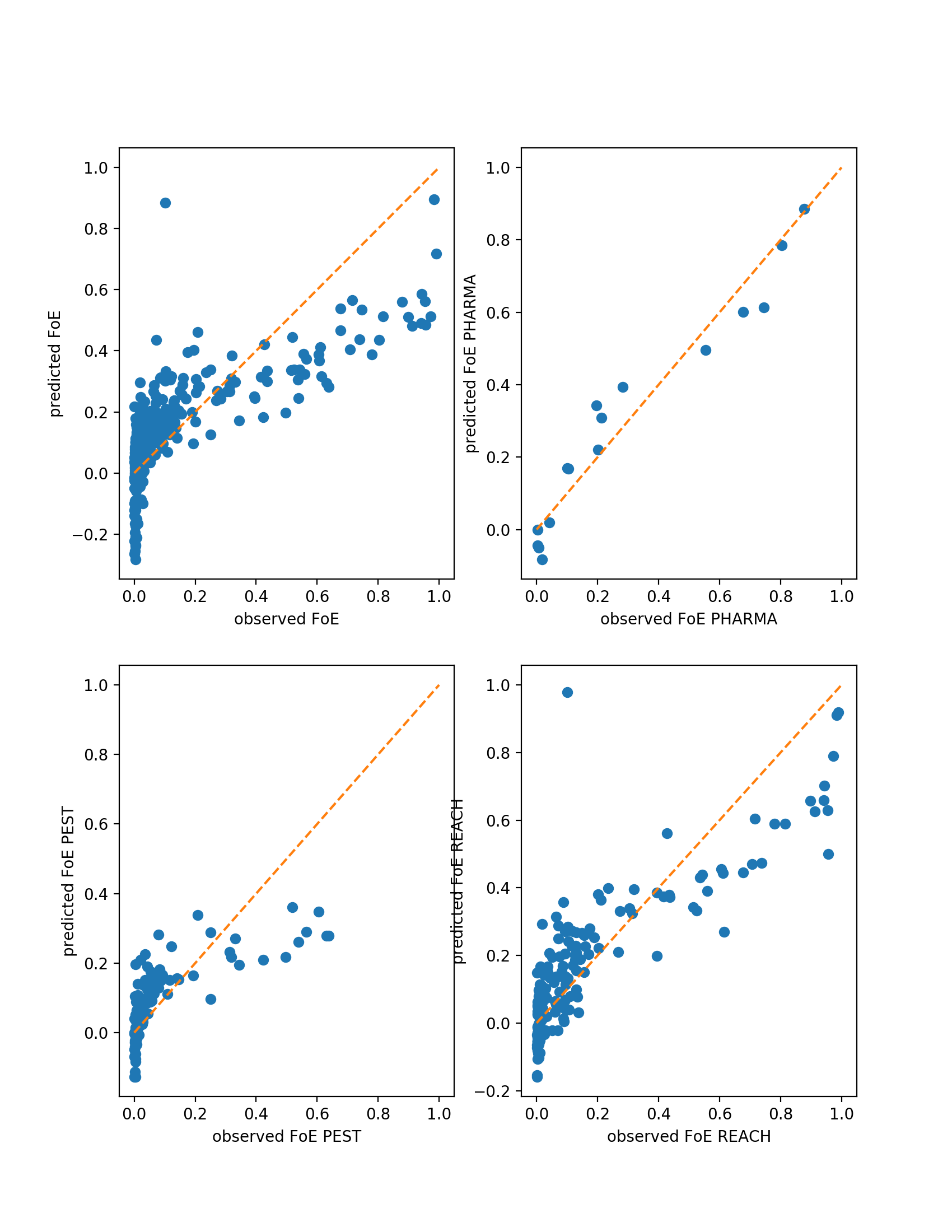
It is odd that there is no relationship whatsoever here. Chemicals with < 1% fugacity can have similar effects as those with 100%. Groups are not separated by chemical type. No indication of additional separable regressor.

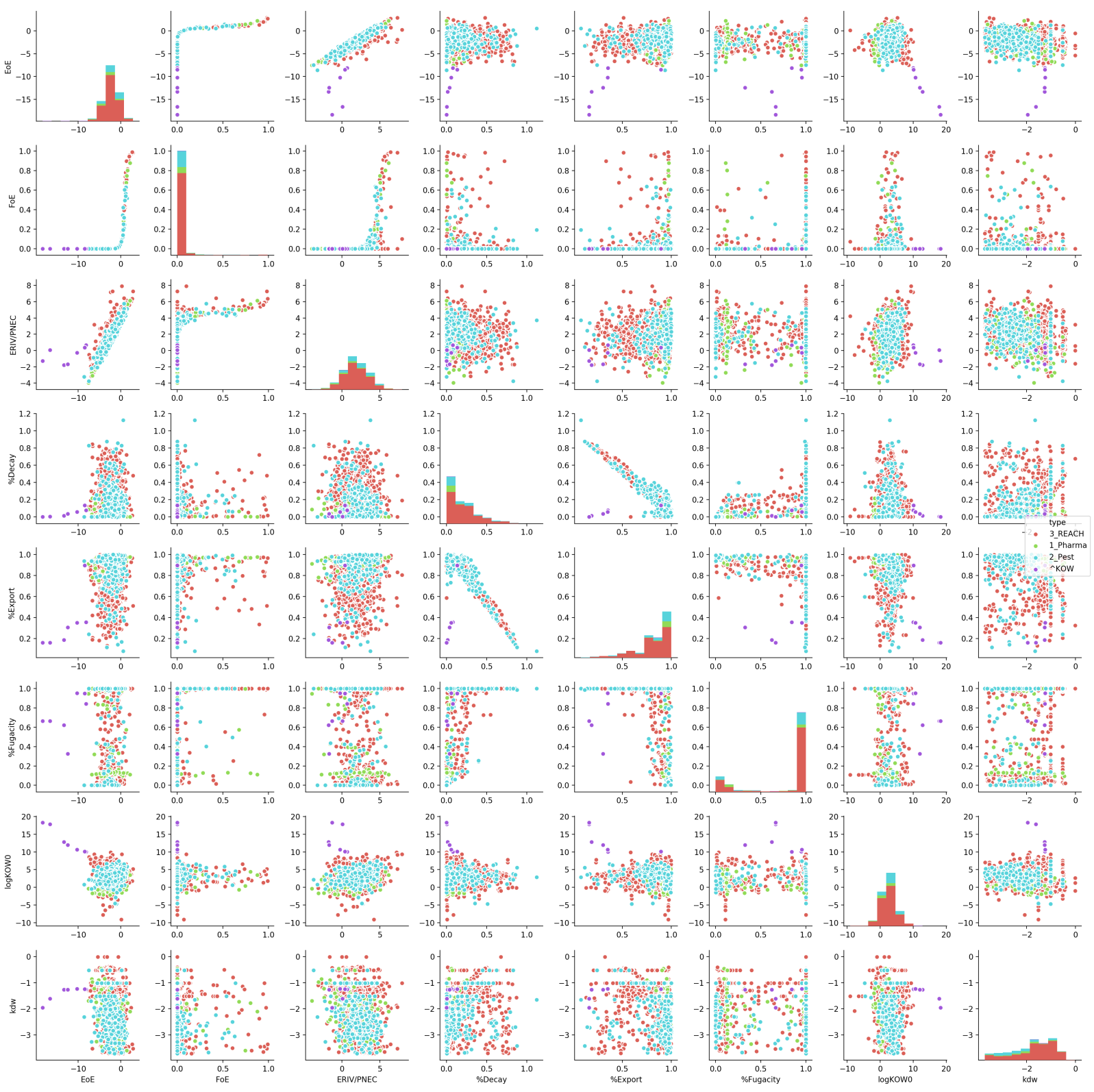




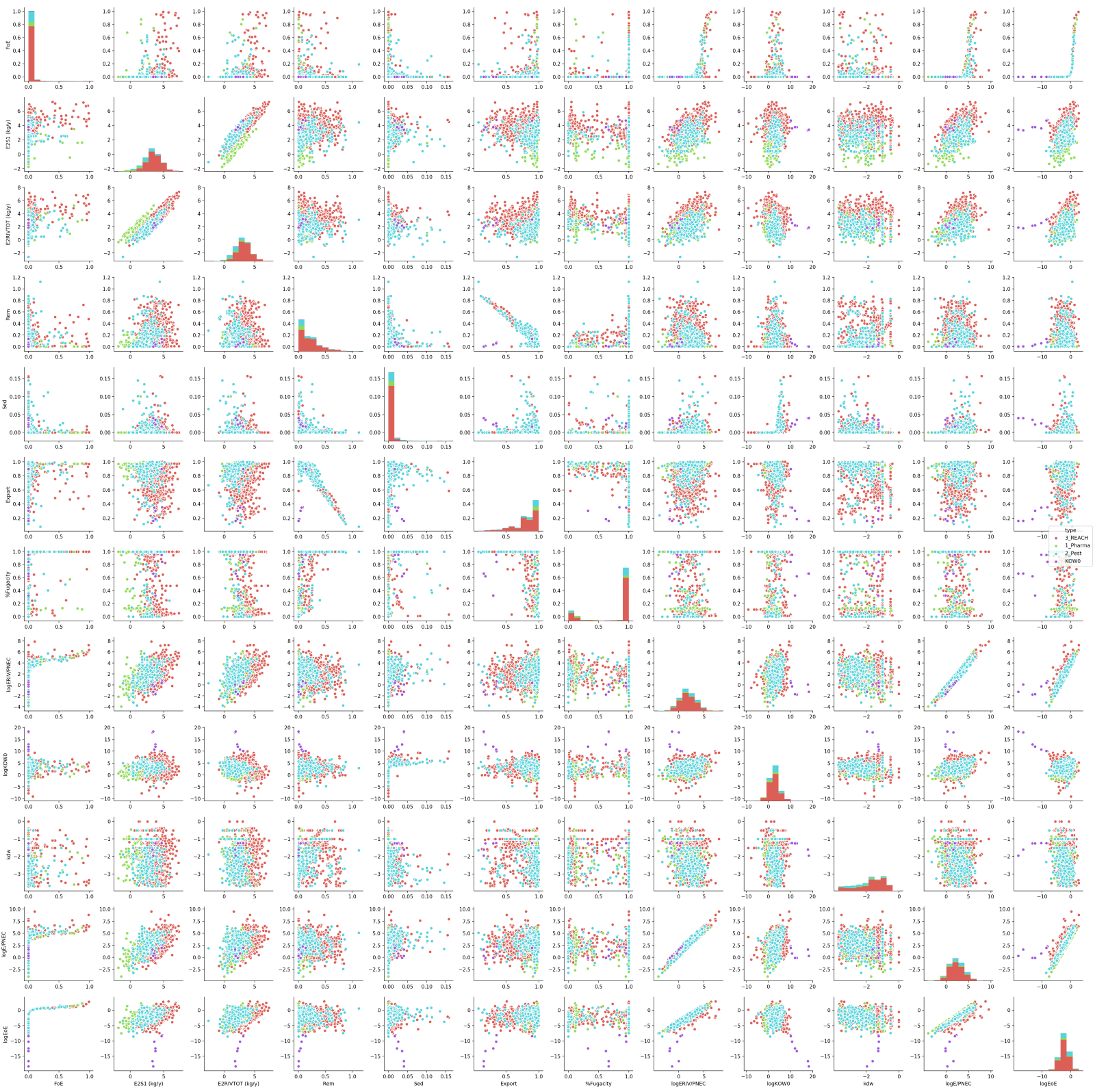
R2 = 0.931







Notice here how the extreme outliers are those Reach chemicals with high Kow values



Full set