Supplementary Information

What do we talk about, when we talk about single- crystal termination-dependent selectivity of Cu electro- catalysts for CO2 reduction ? A data-driven retrospective

Kevin Rossi[a]

Contents

1 Data and Features

Author	Surface	CN6	CN7	CN8	CN9	CN10	CN11	V
Hahn	100	0.00	0.00	1.00	0.00	0.00	0.00	0.89
Hahn	100	0.00	0.00	1.00	0.00	0.00	0.00	0.97
Hahn	100	0.00	0.00	1.00	0.00	0.00	0.00	1.04
Hahn	100	0.00	0.00	1.00	0.00	0.00	0.00	1.10
Hahn	111	0.00	0.00	0.00	1.00	0.00	0.00	0.89
Hahn	111	0.00	0.00	0.00	1.00	0.00	0.00	0.98
Hahn	111	0.00	0.00	0.00	1.00	0.00	0.00	1.04
Hahn	111	0.00	0.00	0.00	1.00	0.00	0.00	1.10
Hahn	751	0.167	0.167	0.167	0.00	0.167	0.33	0.89
Hahn	751	0.167	0.167	0.167	0.00	0.167	0.33	0.97
Hahn	751	0.167	0.167	0.167	0.00	0.167	0.33	1.04
Hahn	751	0.167	0.167	0.167	0.00	0.167	0.33	1.10

Table 1: table of data points features, listed together with the corresponding surface terminations and data source.

Author	Surface	CN6	CN7	CN8	CN9	CN10	CN11	V
Hori	111	0.00	0.00	0.00	1.00	0.00	0.00	1.15
Hori	100	0.00	0.00	1.00	0.00	0.00	0.00	1.00
Hori	110	0.00	0.5	0.00	0.00	0.00	0.5	1.18
Hori	211	0.00	0.33	0.00	0.33	0.33	0.00	0.98
Hori	311	0.00	0.5	0.00	0.00	0.5	0.00	0.97
Hori	511	0.00	0.33	0.33	0.00	0.33	0.00	0.96
Hori	711	0.00	0.25	0.5	0.00	0.25	0.00	0.94
Hori	911	0.00	0.2	0.6	0.00	0.2	0.00	0.96
Hori	210	0.33	0.00	0.00	0.33	0.00	0.33	1.12
Hori	310	0.25	0.00	0.25	0.25	0.00	0.25	1.02
Hori	510	0.167	0.00	0.5	0.167	0.00	0.167	0.98
Hori	610	0.143	0.00	0.571	0.143	0.00	0.143	0.97
Hori	810	0.111	0.00	0.667	0.111	0.00	0.111	0.98
Hori	331	0.00	0.33	0.00	0.33	0.00	0.33	1.15
Hori	332	0.00	0.167	0.00	0.667	0.00	0.167	1.11
Hori	533	0.00	0.25	0.00	0.5	0.25	0.00	1.02
Hori	650	0.009	0.364	0.00	0.009	0.00	0.455	1.19
Hori	755	0.00	0.167	0.00	0.667	0.167	0.00	1.03
Hori	320	0.2	0.2	0.00	0.2	0.00	0.4	1.12

Table 2: table of data points features, listed together with the corresponding surface terminations and data source.

Author	Surface	CN6	CN7	CN8	CN9	CN10	CN11	V
Huang	100	0.00	0.00	1.00	0.00	0.00	0.00	0.85
Huang	100	0.00	0.00	1.00	0.00	0.00	0.00	0.9
Huang	100	0.00	0.00	1.00	0.00	0.00	0.00	0.95
Huang	100	0.00	0.00	1.00	0.00	0.00	0.00	1.00
Huang	100	0.00	0.00	1.00	0.00	0.00	0.00	1.05
Huang	100	0.00	0.00	1.00	0.00	0.00	0.00	1.10
Huang	100	0.00	0.00	1.00	0.00	0.00	0.00	1.15
Huang	100	0.00	0.00	1.00	0.00	0.00	0.00	1.20
Huang	100	0.00	0.00	1.00	0.00	0.00	0.00	1.25
Huang	110	0.00	0.5	0.00	0.00	0.00	0.5	0.9
Huang	110	0.00	0.5	0.00	0.00	0.00	0.5	1.00
Huang	110	0.00	0.5	0.00	0.00	0.00	0.5	1.05
Huang	110	0.00	0.5	0.00	0.00	0.00	0.5	1.10
Huang	110	0.00	0.5	0.00	0.00	0.00	0.5	1.15
Huang	110	0.00	0.5	0.00	0.00	0.00	0.5	1.20
Huang	110	0.00	0.5	0.00	0.00	0.00	0.5	1.25
Huang	111	0.00	0.00	0.00	1.00	0.00	0.00	0.9
Huang	111	0.00	0.00	0.00	1.00	0.00	0.00	0.95
Huang	111	0.00	0.00	0.00	1.00	0.00	0.00	1.00
Huang	111	0.00	0.00	0.00	1.00	0.00	0.00	1.05
Huang	111	0.00	0.00	0.00	1.00	0.00	0.00	1.10
Huang	111	0.00	0.00	0.00	1.00	0.00	0.00	1.15
Huang	111	0.00	0.00	0.00	1.00	0.00	0.00	1.20
Huang	111	0.00	0.00	0.00	1.00	0.00	0.00	1.25

Table 3: table of data points features, listed together with the corresponding surface terminations and data source.

Aut	Sur	H2	CO	CH4	C2H4	EtOH	PrOH	PrD	MeD	Ac	HCOOH	AlOH
Hahn	100	37.5	5.50	2.50	24.1	2.00	3.40	0.00	0.00	0.00	0.4	15.7
Hahn	100	15.6	0.80	8.40	38.6	14.7	7.20	2.20	0.00	0.00	0.5	2.10
Hahn	100	19.1	0.4	22.8	34.9	11.7	3.50	1.20	0.3	0.00	0.4	0.7
Hahn	100	38.8	0.3	44.3	16.3	5.80	0.8	0.00	0.00	0.00	0.2	0.4
Hahn	111	47.1	15.3	0.9	3.00	0.00	0.00	0.00	0.00	0.00	0.00	29.3
Hahn	111	27.5	7.60	9.40	24.2	5.30	5.20	1.80	0.8	0.00	0.7	7.30
Hahn	111	15.6	1.50	34.3	21.0	9.40	3.50	1.40	0.5	0.00	0.5	3.40
Hahn	111	29.4	0.7	42.3	15.6	7.70	1.50	0.4	0.2	0.00	0.3	0.8
Hahn	751	36.2	10.1	0.8	10.1	2.80	4.00	0.00	0.00	0.00	1.30	19.6
Hahn	751	24.7	3.90	13.2	28.5	8.90	6.60	2.30	0.7	0.00	0.5	6.50
Hahn	751	23.0	0.9	23.2	27.5	12.7	3.30	1.00	0.2	0.00	0.3	1.20
Hahn	751	38.0	0.6	29.2	22.6	11.0	2.00	0.9	0.2	0.00	0.5	0.5

Table 4: table of data points selectivities, together with the corresponding surface terminations, applied potential, and data source.

Aut	Surf	H2	CO	CH4	C2H4	EtOH	PrOH	PrD	MeD	Ac	HCOOH	AlOH
Hori	100	6.80	0.9	30.4	40.4	9.70	1.50	0.8	1.60	2.80	1.00	3.00
Hori	111	16.3	6.40	46.3	8.30	2.60	0.00	0.7	2.10	0.6	1.50	11.5
Hori	110	3.10	13.9	6.90	13.5	10.5	0.04	0.00	19.9	1.30	20.8	10.1
Hori	211	11.2	2.10	45.6	17.8	3.40	1.30	0.7	0.3	2.00	0.5	13.6
Hori	311	13.3	2.60	36.0	23.8	3.30	1.50	0.4	1.10	2.30	0.6	14.0
Hori	511	18.1	1.90	11.4	39.0	12.2	3.30	1.60	1.40	3.00	0.8	8.80
Hori	711	15.6	1.10	5.00	50.0	7.40	4.60	2.20	1.20	5.20	0.9	4.60
Hori	911	12.7	0.00	5.70	50.9	16.9	4.50	2.50	1.20	2.40	1.10	3.50
Hori	210	7.00	2.20	64.0	13.4	6.60	0.5	0.2	0.9	0.6	0.7	5.50
Hori	310	12.8	0.00	17.7	34.6	29.9	1.90	0.9	1.70	2.60	1.60	2.70
Hori	510	10.5	2.10	8.10	42.3	26.1	1.70	1.70	3.00	2.60	2.10	2.90
Hori	610	9.00	0.9	7.60	44.71	26.0	2.00	1.30	1.50	1.20	1.60	1.40
Hori	810	8.70	1.40	6.40	45.1	26.0	1.90	0.9	0.9	1.10	1.60	1.50
Hori	320	5.30	5.40	52.4	13.7	6.50	0.4	0.3	3.20	0.6	4.80	5.80
Hori	331	5.70	7.70	13.8	16.6	15.6	0.4	0.00	7.10	0.5	7.50	9.10
Hori	332	10.3	6.10	39.6	9.90	7.10	0.2	0.3	5.10	0.2	3.40	9.40
Hori	533	4.70	3.00	62.9	13.0	1.90	0.4	0.5	0.6	0.7	0.5	9.70
Hori	650	2.50	14.5	10.5	16.2	10.9	0.00	0.00	16.2	0.8	20.6	6.10
Hori	755	6.90	4.40	62.9	11.5	5.30	0.6	0.7	0.6	1.20	0.5	12.3

Table 5: table of data points selectivities, together with the corresponding surface terminations, applied potential, and data source.

Ref	Surf	H2	CO	CH4	C2H4	EtOH	PrOH	PrD	MeD	Ac	HCOOH	AlOH
Huang	100	66.6	4.87	1.87	15.8	0.00	0.00	0.00	0.00	0.00	0.00	9.02
Huang	100	62.2	2.73	2.38	18.8	0.84	0.00	0.55	0.00	0.00	0.00	8.21
Huang	100	54.7	2.24	5.76	26.0	4.09	0.00	0.24	0.00	0.00	0.00	4.93
Huang	100	35.7	1.41	17.6	30.6	6.49	0.00	0.77	0.00	0.00	0.00	4.06
Huang	100	37.3	1.24	27.	21.9	2.72	0.00	0.68	0.00	0.00	0.00	4.23
Huang	100	59.5	0.11	30.4	6.82	1.73	0.00	0.07	0.00	0.00	0.00	1.33
Huang	100	68.6	0.04	27.6	1.01	0.00	0.00	0.00	0.00	0.00	0.00	1.22
Huang	100	80.4	0.04	22.1	0.47	0.00	0.00	0.00	0.00	0.00	0.00	0.48
Huang	100	77.2	0.02	18.4	0.16	0.00	0.00	0.00	0.00	0.00	0.00	0.28
Huang	110	63.9	5.95	0.91	3.19	0.00	0.00	0.00	0.00	0.00	0.00	22.7
Huang	110	57.5	2.54	6.91	17.1	6.40	0.00	0.00	0.00	0.00	0.00	9.27
Huang	110	33.7	2.63	21.6	25.2	7.41	0.00	1.15	0.00	0.00	0.00	4.98
Huang	110	29.0	2.23	31.4	23.8	6.76	0.00	0.65	0.00	0.00	0.00	2.29
Huang	110	61.0	1.08	25.4	6.63	3.26	0.00	0.55	0.00	0.00	0.00	1.54
Huang	110	71.7	0.34	22.2	3.43	1.28	0.00	0.25	0.00	0.00	0.00	0.44
Huang	110	76.9	0.21	15.8	1.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Huang	111	64.1	10.3	0.24	0.54	0.00	0.00	0.00	0.00	0.00	0.00	19.3
Huang	111	58.7	9.76	2.87	3.81	0.00	0.00	0.00	0.00	0.00	0.00	20.8
Huang	111	55.6	4.86	17.5	10.7	0.00	0.00	0.00	0.00	0.00	0.00	11.8
Huang	111	36.2	4.89	28.1	16.8	2.60	0.00	0.81	0.00	0.00	0.00	6.80
Huang	111	33.2	0.57	42.1	16.3	4.08	0.00	0.3	0.00	0.00	0.00	3.02
Huang	111	52.0	0.16	43.5	2.79	0.49	0.00	0.23	0.00	0.00	0.00	2.42
Huang	111	76.0	0.15	28.4	0.43	0.00	0.00	0.00	0.00	0.00	0.00	1.71
Huang	111	77.5	0.08	22.9	0.1	0.00	0.00	0.00	0.00	0.00	0.00	0.94

Table 6: table of data points selectivities, together with the corresponding surface terminations, applied potential, and data source.

2 First Attempt - Out of the box ML

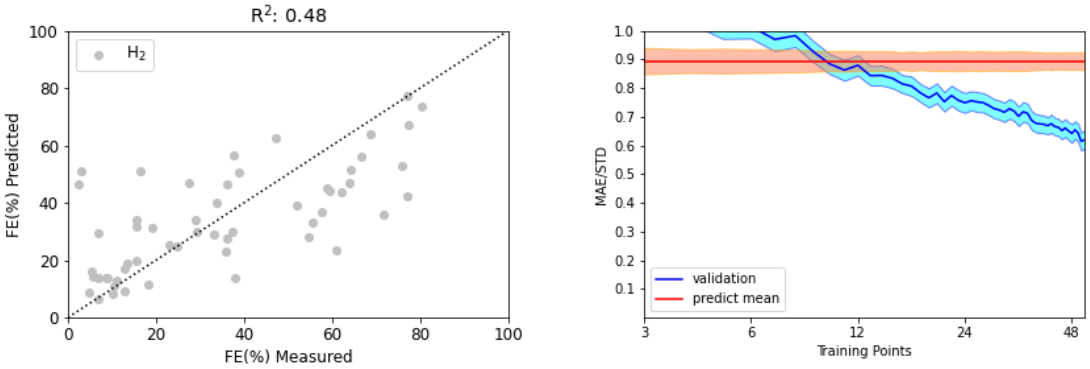

Figure S1: Left Panel: Parity plot between the predicted and measured FEs for H_2 production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction H_2 production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

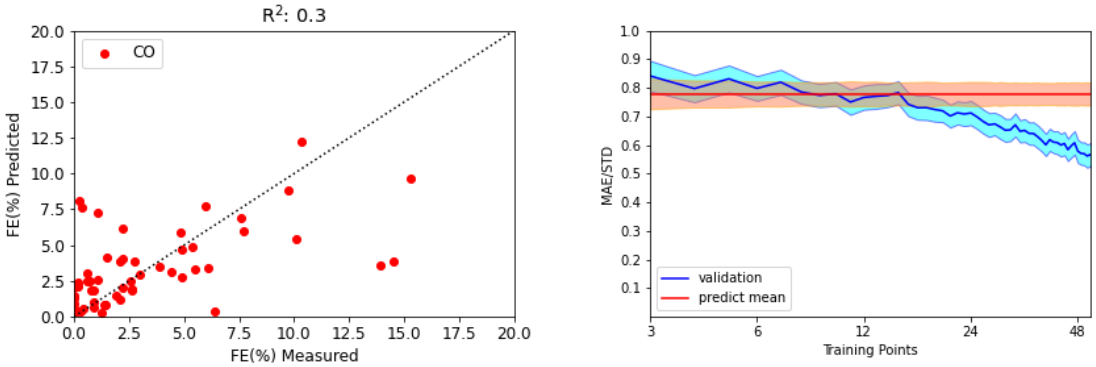

Figure S2: Left Panel: Parity plot between the predicted and measured FEs for CO production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction CO production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

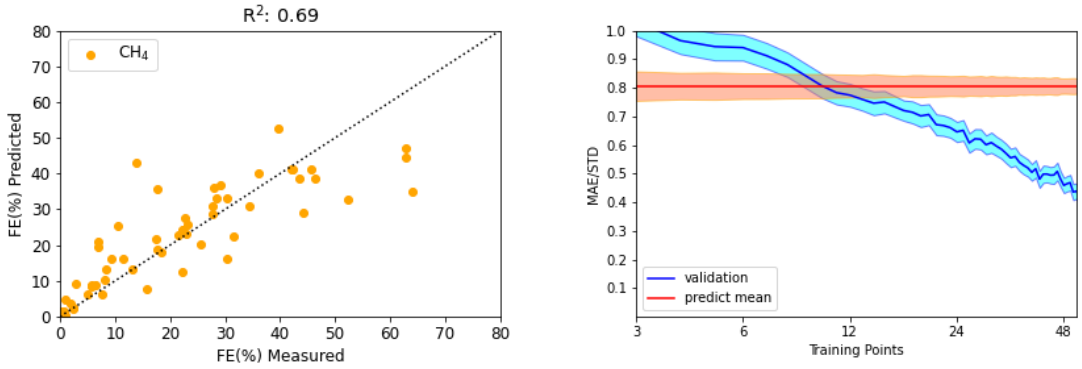

Figure S3: Left Panel: Parity plot between the predicted and measured FEs for CH_4 production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction CH_4 production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

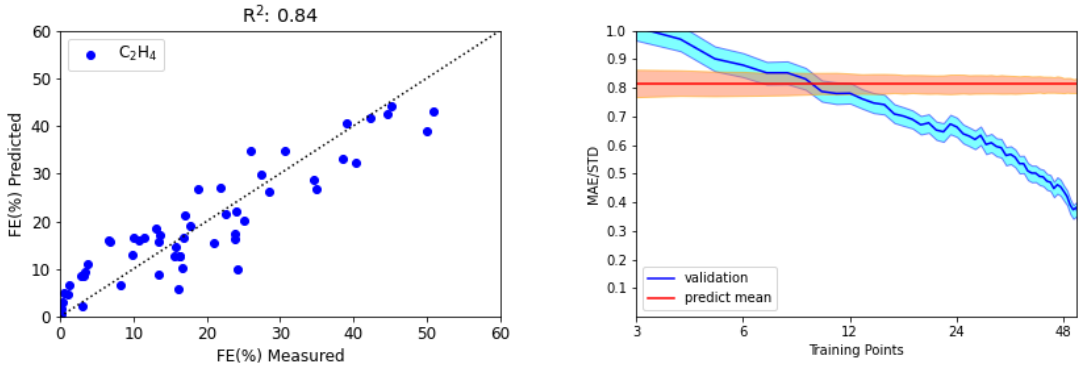

Figure S4: Left Panel: Parity plot between the predicted and measured FEs for C_2H_4 production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction C_2H_4 production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

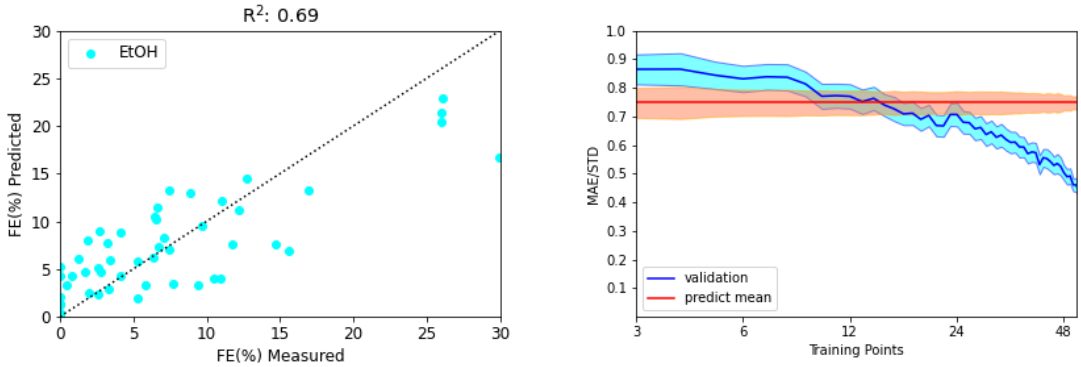

Figure S5: Left Panel: Parity plot between the predicted and measured FEs for EtOH production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction EtOH production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

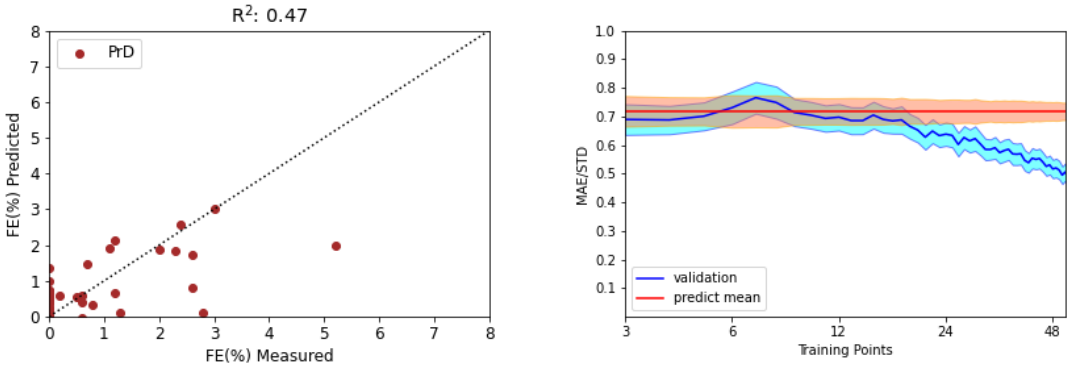

Figure S6: Left Panel: Parity plot between the predicted and measured FEs for PrD production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction PrD production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

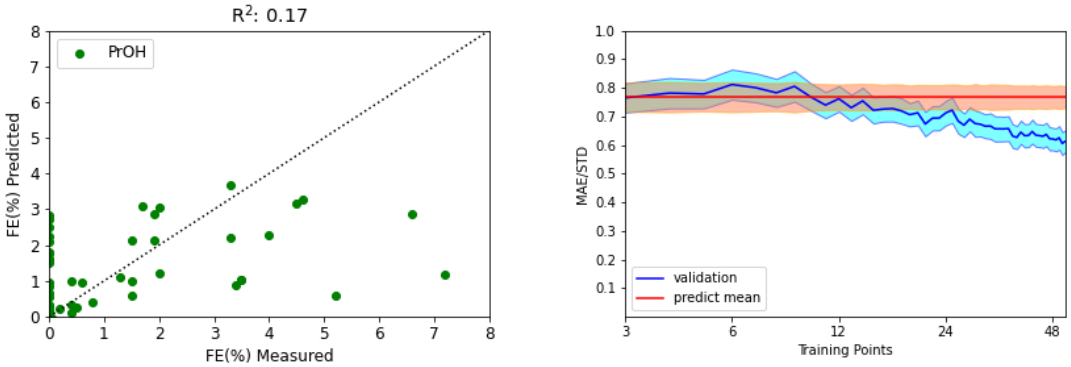

Figure S7: Left Panel: Parity plot between the predicted and measured FEs for PrOH production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction PrOH production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

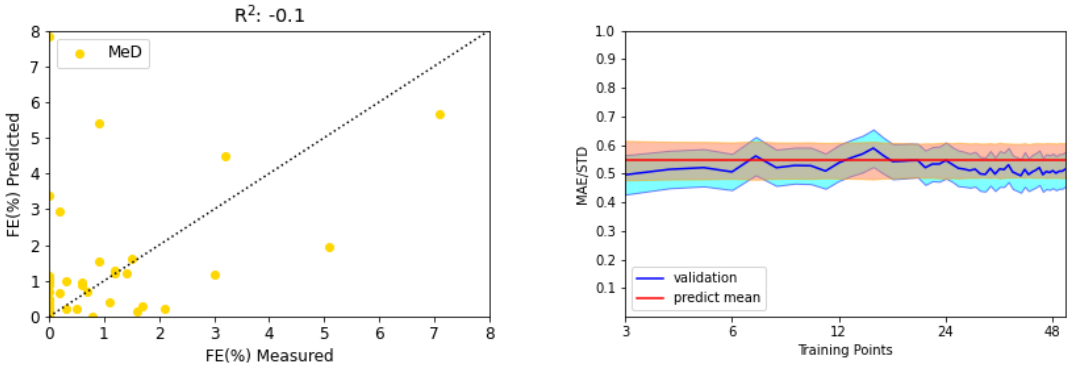

Figure S8: Left Panel: Parity plot between the predicted and measured FEs for MeD production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction MeD production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

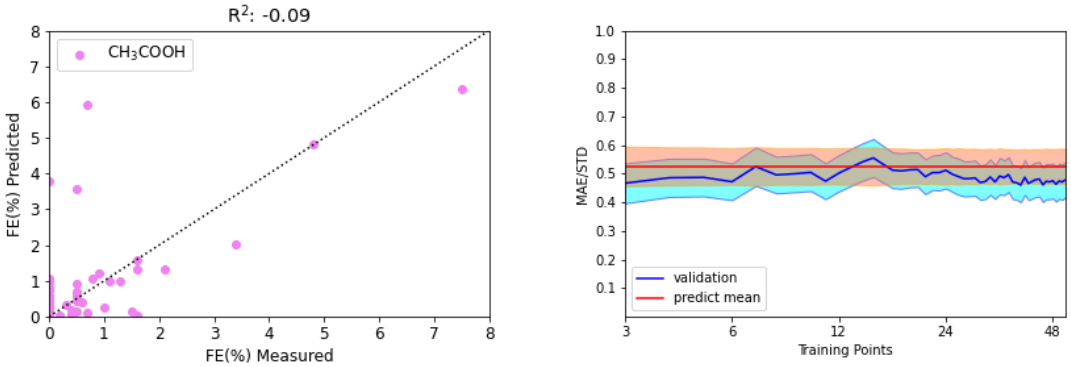

Figure S9: Left Panel: Parity plot between the predicted and measured FEs for CH_3COOH production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction CH_3COOH production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

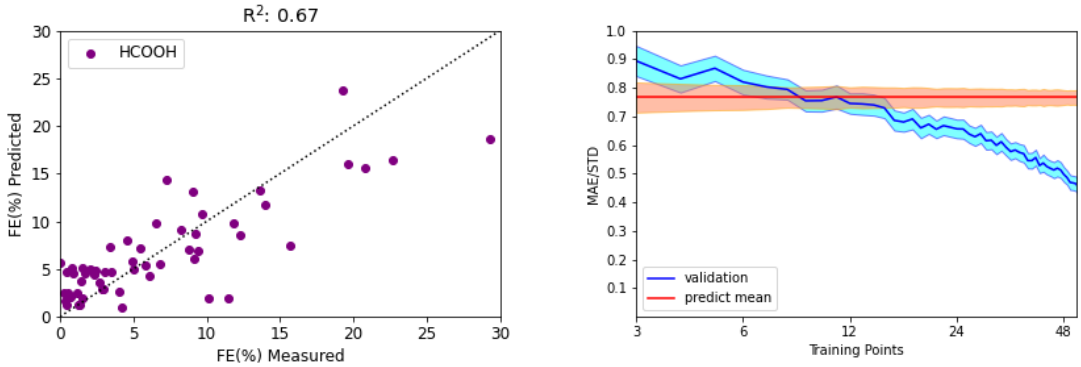

Figure S10: Left Panel: Parity plot between the predicted and measured FEs for HCOOH production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction HCOOH production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

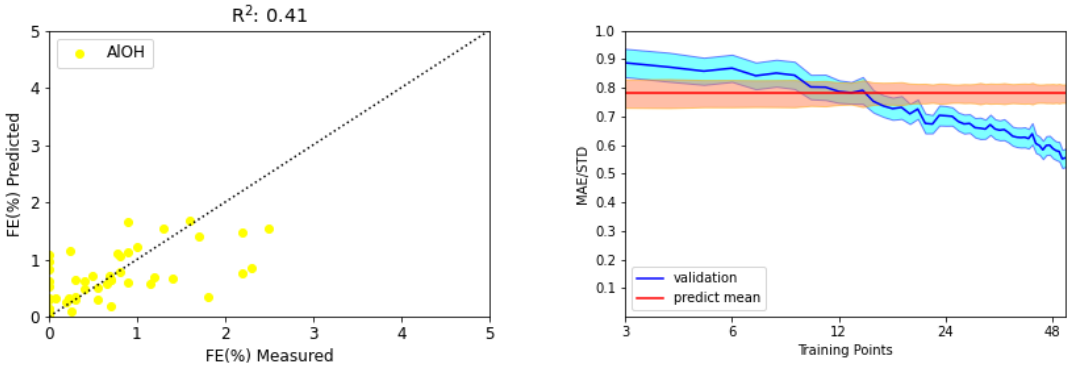

Figure S11: Left Panel: Parity plot between the predicted and measured FEs for AlOH production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction AlOH production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

3 Synthetic Data

Synthetic data are generated according to the following criterion: i) all products display non-trivial, non-monotonic trends as a function of the overpotential and coordination distributions; ii) hydrogen is the majority product at high and low overpotentials; iii) FEs for CH_4 or C_2H_4 are highest at moderate overpotentials; iii) surfaces abundant in 9-coordinated atoms (e.g., (111)-terminated) are selective for CH_4, while surfaces rich in 8-coordinated atoms (e.g., (100)-terminated) are selective for C_2 products; iv) the synthetic data FEs standard deviations are comparable to the one of the experimental data (albeit the FEs distributions also show relatively different skewness and kurtosis). By design, the analytical equations for generating synthetic data are written down as a function of the exact same features that are adopted by the ML model. This benchmark is therefore built solely to verify the capability of the model in learning non-trivial equations, in a low-data regime, in the scenario where the model input feature are exactly related to the generated synthetic data. By the same reason, the 55 datapoints are obtained from the generating equations (Equation S1-S11) for the exact same voltages and coordination distributions reported by Hori2003, Huang2017, or Hahn2017.

The following analytical equations are used as the generators of the FE distributions, as a function of the applied potential and coordination distributions:

$$fe_{H2} = exp(0.1 * fe^0_{H2})/fe_{tot} \tag{1}$$

$$fe_{co} = exp(0.1 * fe^0_{co})/fe_{tot} \tag{2}$$

$$fe_{CH4} = exp(0.1 * fe^0_{CH4})/fe_{tot} \tag{3}$$

$$fe_{C2H4} = exp(0.1 * fe^0_{C2H4})/fe_{tot} \tag{4}$$

$$fe_{eth} = exp(0.1 * fe^0_{eth})/fe_{tot} \tag{5}$$

$$fe_{prd} = exp(0.1 * fe^0_{prd})/fe_{tot} \tag{6}$$

$$fe_{proh} = exp(0.1 * fe^0_{proh})/fe_{tot} \tag{7}$$

$$fe_{aloh} = exp(0.1 * fe^0_{aloh})/fe_{tot} \tag{8}$$

$$fe_{hcooh} = exp(0.1 * fe^0_{hcooh})/fe_{tot} \tag{9}$$

$$fe_{ch3cooh} = exp(0.1 * fe^0_{ch3cooh})/fe_{tot} \tag{10}$$

$$fe_{md} = exp(0.1 * fe^0_{md})/fe_{tot} \tag{11}$$

where

$$\begin{aligned} fe_{tot} =& exp(0.1 * fe^0_{H2}) + exp(0.1 * fe^0_{co}) + exp(0.1 * fe^0_{CH4}) + exp(0.1 * fe^0_{C2H4}) + \\ & exp(0.1 * fe^0_{eth}) + exp(0.1 * fe^0_{prd}) + exp(0.1 * fe^0_{proh}) + exp(0.1 * fe^0_{aloh}) + \\ & exp(0.1 * fe^0_{ch3cooh}) + exp(0.1 * fe^0_{md}) \end{aligned} \tag{12}$$

and

$$fe^0_{H2} = ((voltage + 1.1) * *2) * (300) * (6 * cn6 + 5 * cn7 + 4 * cn8 + 3 * cn9 + 2 * cn10 + 1 * cn11) + 10 * (6 * cn6 + 5 * cn7 + 4 * cn8 + 3 * cn9 + 2 * cn10 + 1 * cn11 \tag{13}$$

$$fe^0_{co} = ((-1.1 - voltage) * *2) * (-250) * (6 * cn6 + 3 * cn7 + 2 * cn8 + 1 * cn9 + 1 * cn10 + 1 * cn11) + 10 * (4 * cn6 + 2 * cn7 + 2 * cn8 + 1 * cn9 + 1 * cn10 + 1 * cn11) \tag{14}$$

$$fe^0_{CH4} = ((-1.2 - voltage) * *2) * (-250) * (1 * cn6 + 2 * cn7 + 3 * cn8 + 6 * cn9 + 2 * cn10 + 1 * cn11) + 10 * (1 * cn6 + 2 * cn7 + 3 * cn8 + 6 * cn9 + 2 * cn10 + 1 * cn11) \tag{15}$$

$$fe^0_{C2H4} = ((-1.0 - voltage) * *2) * (-250) * (2 * cn6 + 6 * cn7 + 6 * cn8 + 3 * cn9 + 2 * cn10 + 1 * cn11) + 10 * (2 * cn6 + 2 * cn7 + 6 * cn8 + 2 * cn9 + 1 * cn10 + 1 * cn11) \tag{16}$$

$$fe^0_{eth} = ((-1.1 - voltage) * *2) * (-250) * (3 * cn6 + 4 * cn7 + 5 * cn8 + 3 * cn9 + 2 * cn10 + 1 * cn11) + 10 * (1 * cn6 + 2 * cn7 + 5 * cn8 + 2 * cn9 + 1 * cn10 + 1 * cn11) \tag{17}$$

$$fe^0_{prd} = ((-1.1 - voltage) * *2) * (-250) * (2 * cn6 + 1 * cn7 + 1 * cn8 + 1 * cn9 + 2 * cn10 + 1 * cn11) + 10 * (2 * cn6 + 1 * cn7 + 1 * cn8 + 1 * cn9 + 2 * cn10 + 1 * cn11) \tag{18}$$

$$fe^0_{proh} = ((-1.1 - voltage) * *2) * (-250) * (1 * cn6 + 2 * cn7 + 1 * cn8 + 1 * cn9 + 1 * cn10 + 2 * cn11) + 10 * (1 * cn6 + 2 * cn7 + 1 * cn8 + 1 * cn9 + 1 * cn10 + 2 * cn11) \tag{19}$$

$$fe^0_{aloh} = ((-1.1 - voltage) * *2) * (-250) * (1 * cn6 + 1 * cn7 + 2 * cn8 + 1 * cn9 + 2 * cn10 + 1 * cn11) + 10 * (1 * cn6 + 1 * cn7 + 2 * cn8 + 1 * cn9 + 2 * cn10 + 1 * cn11) \tag{20}$$

$$fe^0_{hcooh} = ((-0.8 - voltage) * *2) * (-250) * (2 * cn6 + 3 * cn7 + 1 * cn8 + 2 * cn9 + 2 * cn10 + 1 * cn11) + \quad (21)$$

$$fe^0_{ch3cooh} = ((-1.1 - voltage) * *2) * (-250) * (1 * cn6 + 2 * cn7 + 2 * cn8 + 1 * cn9 + 1 * cn10 + 1 * cn11) + 10 * (1 * cn6 + 2 * cn7 + 2 * cn8 + 1 * cn9 + 1 * cn10 + 1 * cn11) \quad (22)$$

$$fe^0_{md} = ((-1.1 - voltage) * *2) * (-250) * (1 * cn6 + 1 * cn7 + 1 * cn8 + 1 * cn9 + 1 * cn10 + 1 * cn11) + 10 * (1 * cn6 + 1 * cn7 + 1 * cn8 + 1 * cn9 + 1 * cn10 + 1 * cn11) \quad (23)$$

As shown in Figures S23-S33, a model trained on these data displays steep learning rates as well as almost perfect accuracy metrics for leave-one-out cross-validation across the whole range of products.

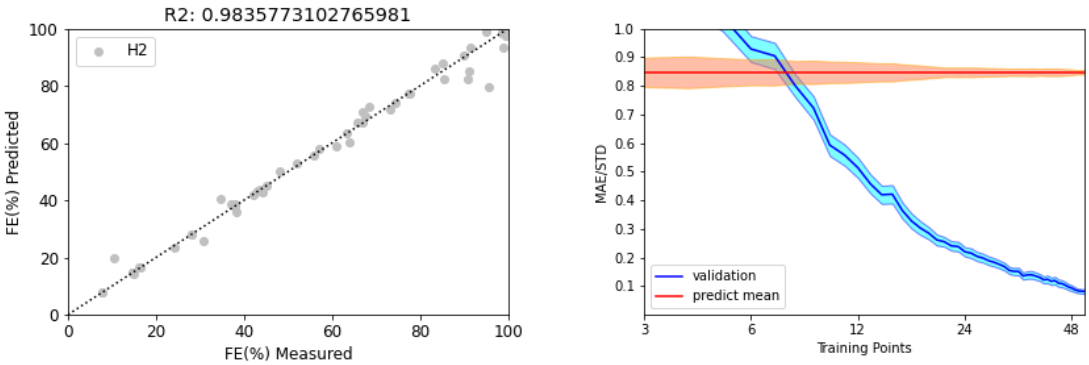

Figure S12: Left Panel: Parity plot between the predicted and measured FEs for H_2 production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction H_2 production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

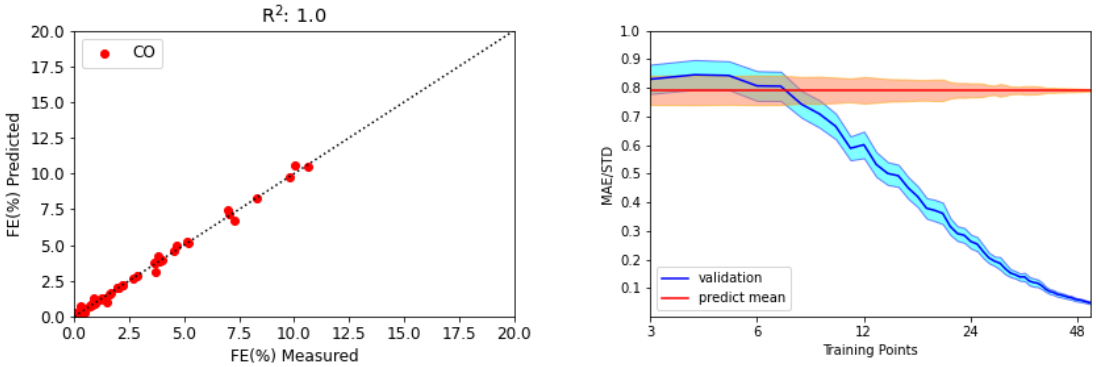

Figure S13: Left Panel: Parity plot between the predicted and measured FEs for CO2 production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction CO2 production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

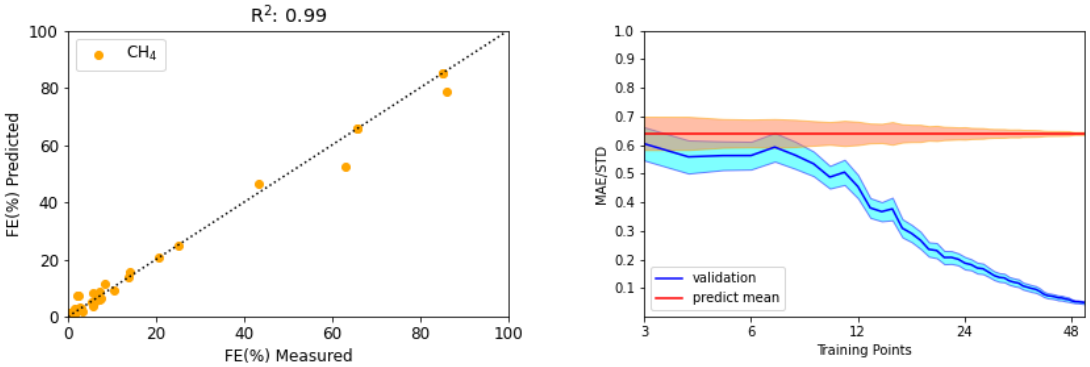

Figure S14: Left Panel: Parity plot between the predicted and measured FEs for CH_4 production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction CH_4 production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

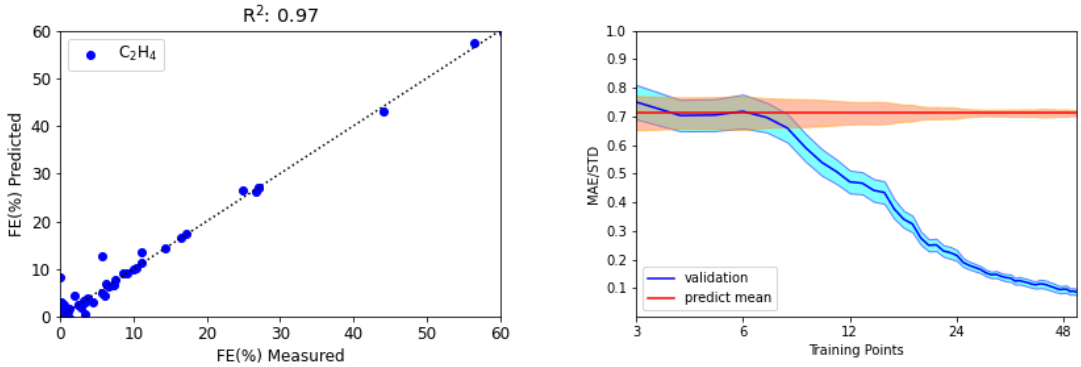

Figure S15: Left Panel: Parity plot between the predicted and measured FEs for C_2H_4 production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction C_2H_4 production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

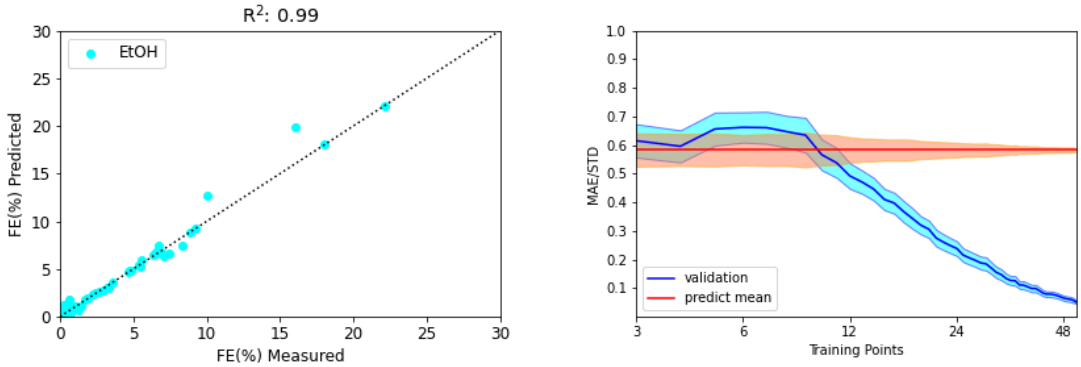

Figure S16: Left Panel: Parity plot between the predicted and measured FEs for EtOH production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction EtOH production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

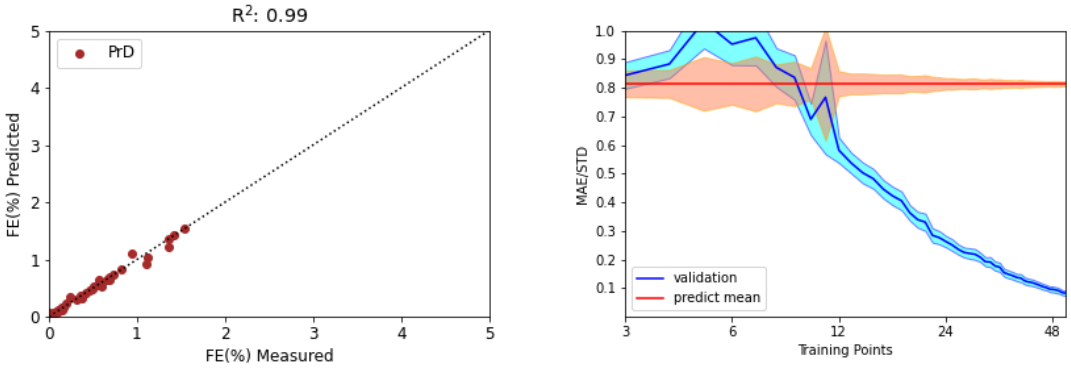

Figure S17: Left Panel: Parity plot between the predicted and measured FEs for PrD production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction PrD production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

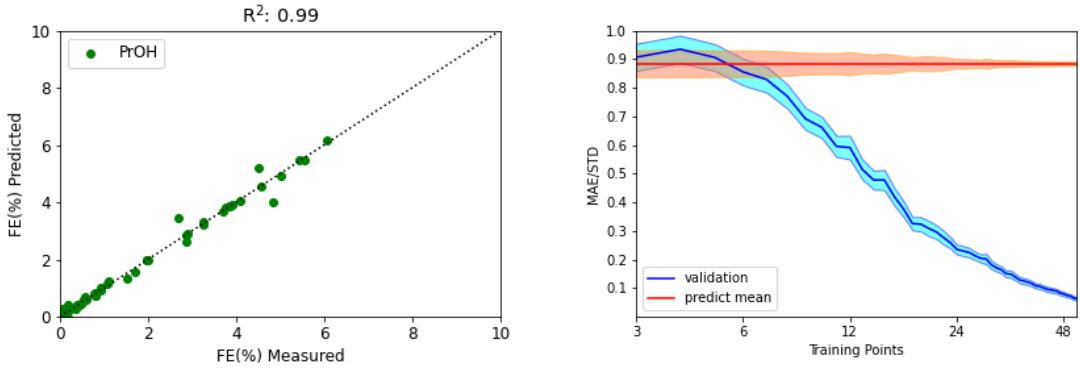

Figure S18: Left Panel: Parity plot between the predicted and measured FEs for PrOH production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction PrOH production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

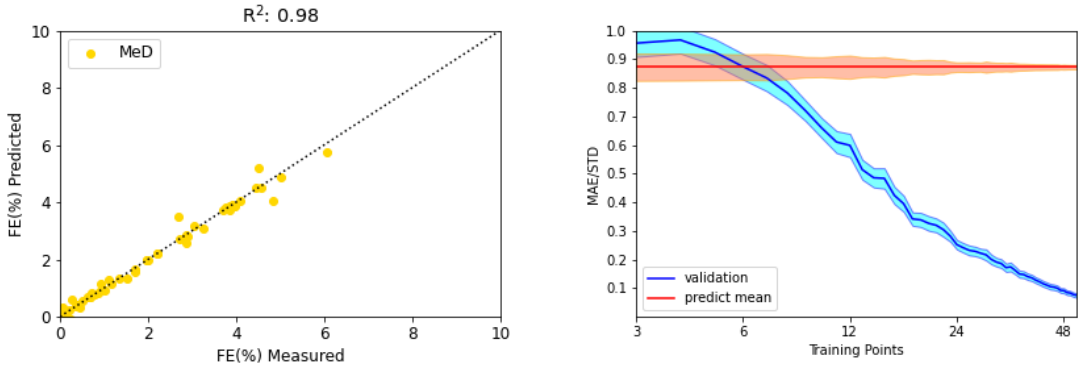

Figure S19: Left Panel: Parity plot between the predicted and measured FEs for MeD production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction MeD production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

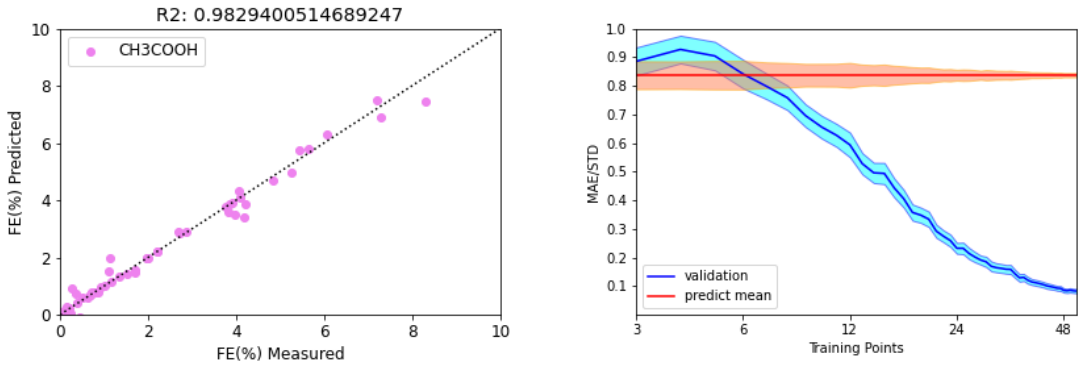

Figure S20: Left Panel: Parity plot between the predicted and measured FEs for CH_3COOH production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction CH_3COOH production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

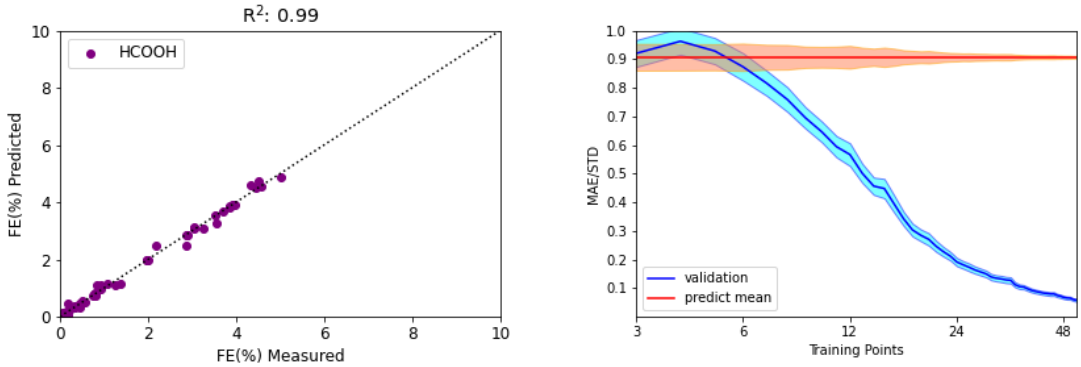

Figure S21: Left Panel: Parity plot between the predicted and measured FEs for HCOOH production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction HCOOH production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

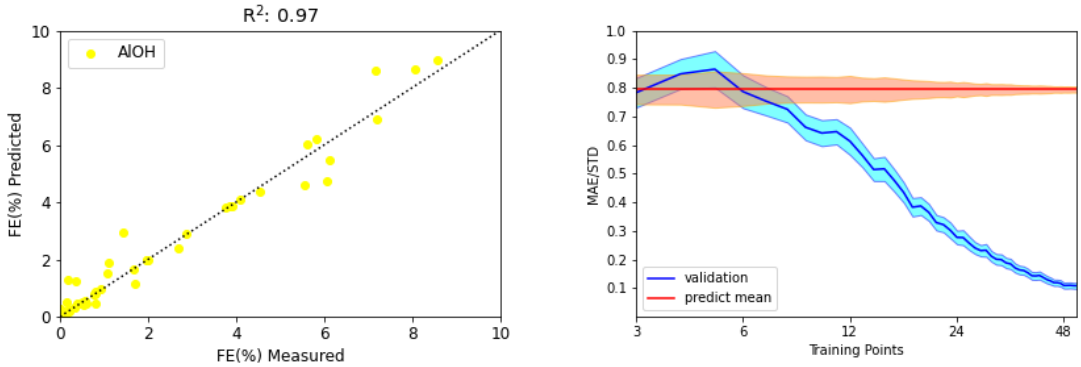

Figure S22: Left Panel: Parity plot between the predicted and measured FEs for AlOH production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction AlOH production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

4 Outlier Detection

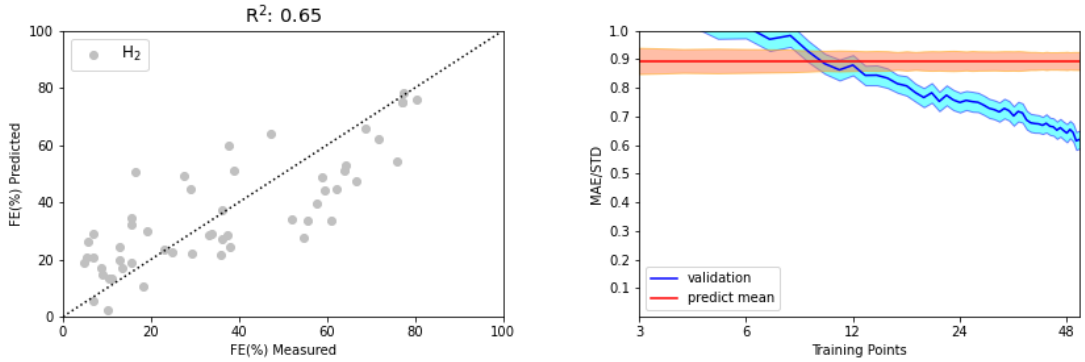

Figure S23: Left Panel: Parity plot between the predicted and measured FEs for H_2 production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction H_2 production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

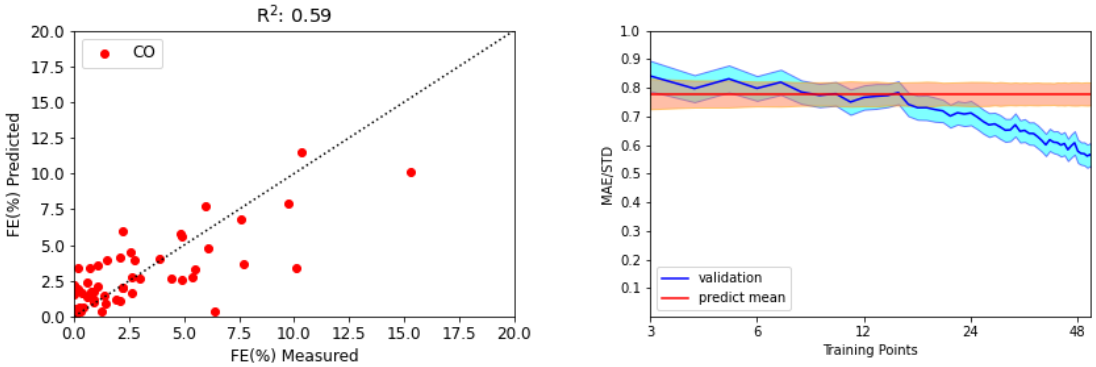

Figure S24: Left Panel: Parity plot between the predicted and measured FEs for CO2 production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction CO2 production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

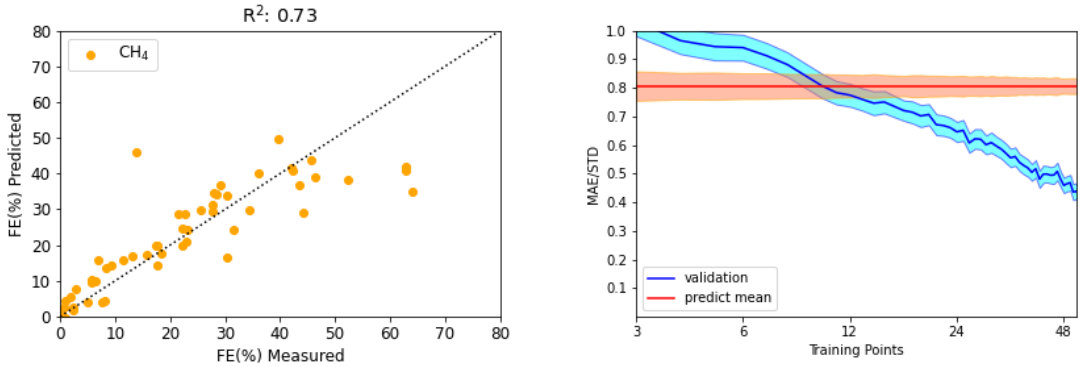

Figure S25: Left Panel: Parity plot between the predicted and measured FEs for CH_4 production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction CH_4 production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

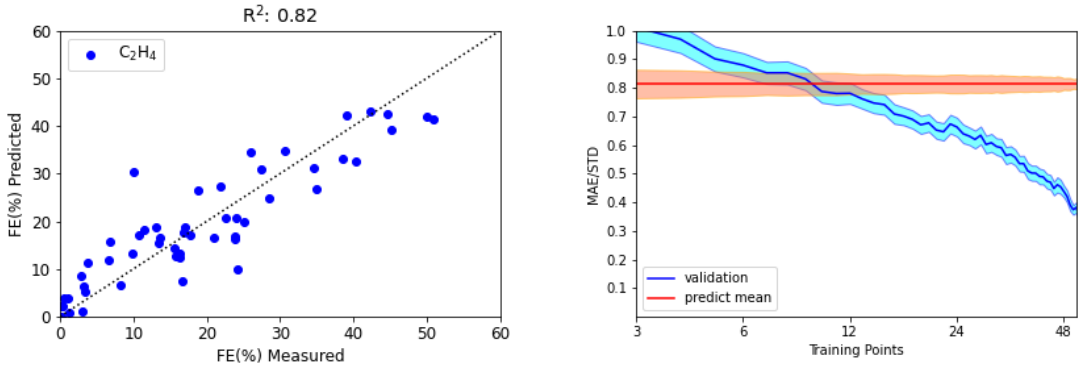

Figure S26: Left Panel: Parity plot between the predicted and measured FEs for C_2H_4 production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction C_2H_4 production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

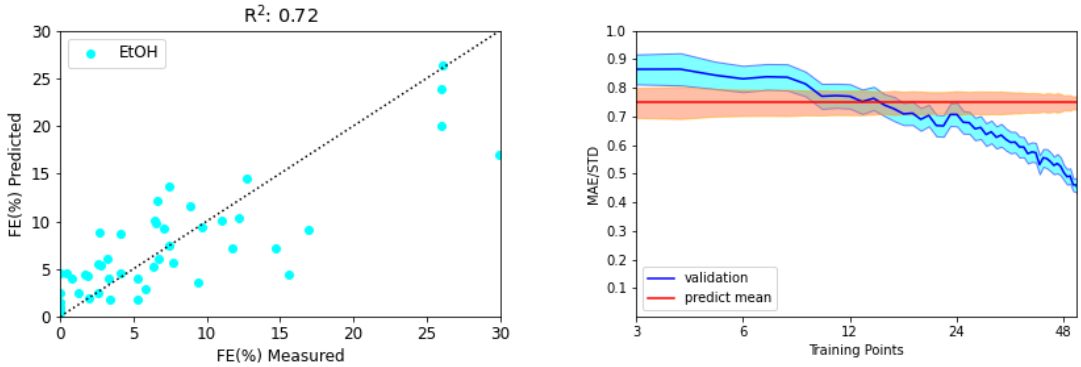

Figure S27: Left Panel: Parity plot between the predicted and measured FEs for EtOH production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction EtOH production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

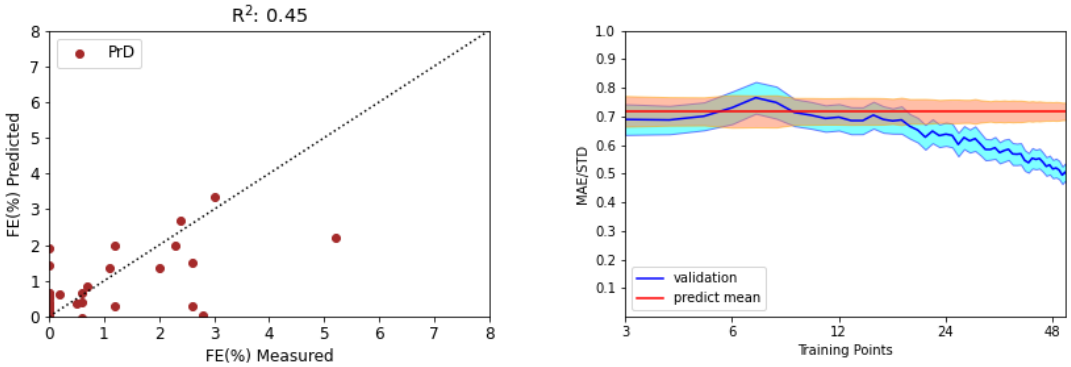

Figure S28: Left Panel: Parity plot between the predicted and measured FEs for PrD production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction PrD production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

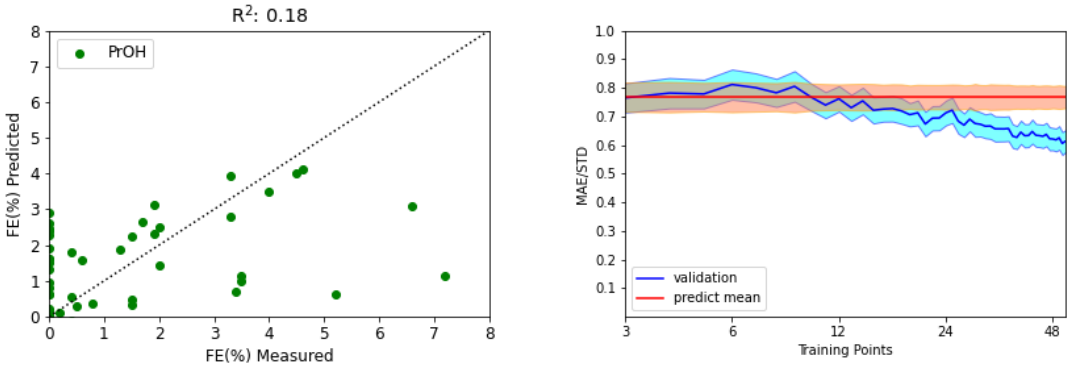

Figure S29: Left Panel: Parity plot between the predicted and measured FEs for PrOH production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction PrOH production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

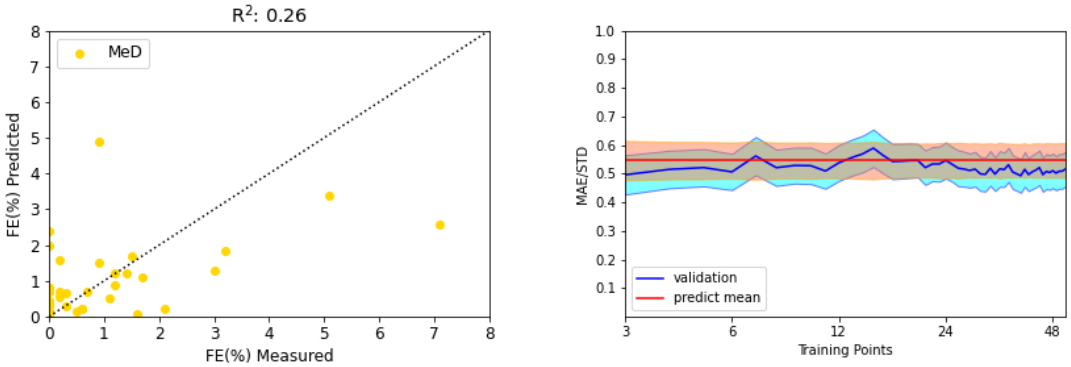

Figure S30: Left Panel: Parity plot between the predicted and measured FEs for MeD production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction MeD production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

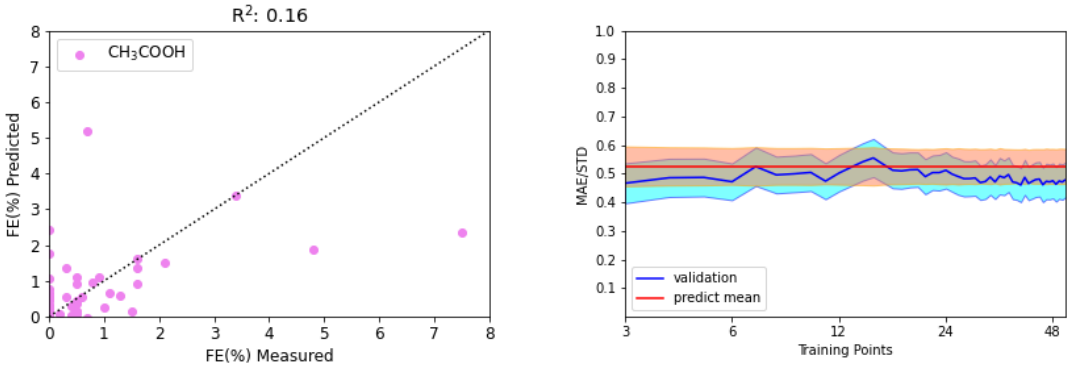

Figure S31: Left Panel: Parity plot between the predicted and measured FEs for CH_3COOH production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction CH_3COOH production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

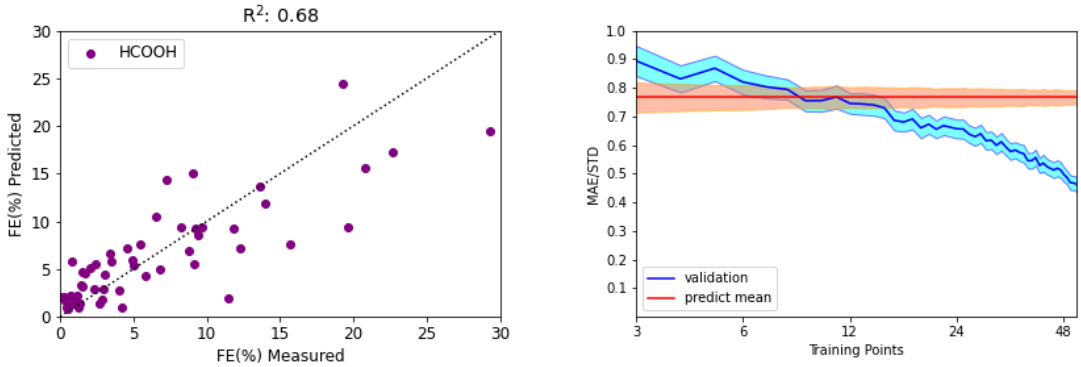

Figure S32: Left Panel: Parity plot between the predicted and measured FEs for HCOOH production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction HCOOH production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

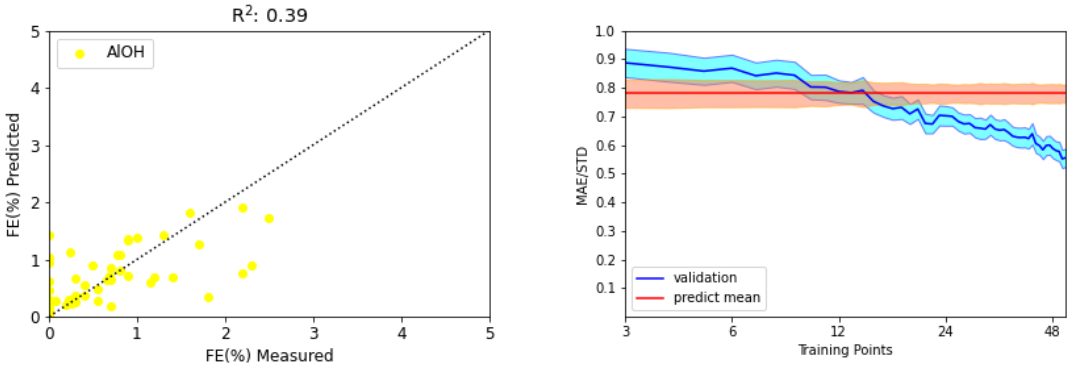

Figure S33: Left Panel: Parity plot between the predicted and measured FEs for AlOH production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction AlOH production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

5 Data recalibration

Learning curves for this case study are built under the constrain that at least 2 training data from each of the Hori2003,Huang2018, and Hahn2017 report are always included in the training set.

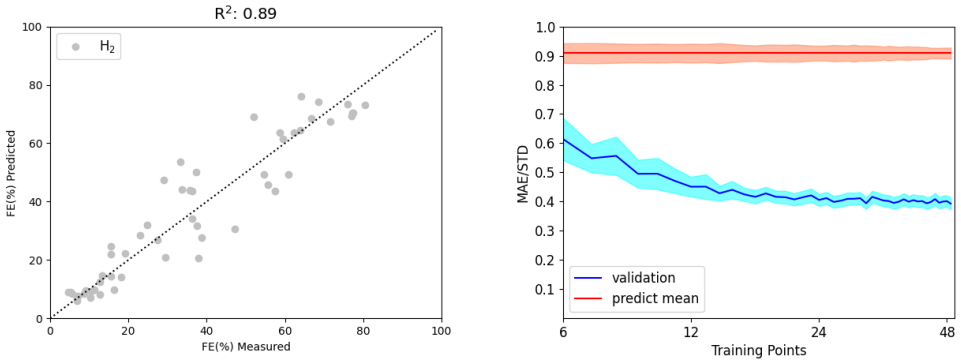

Figure S34: Left Panel: Parity plot between the predicted and measured FEs for H_2 production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction H_2 production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

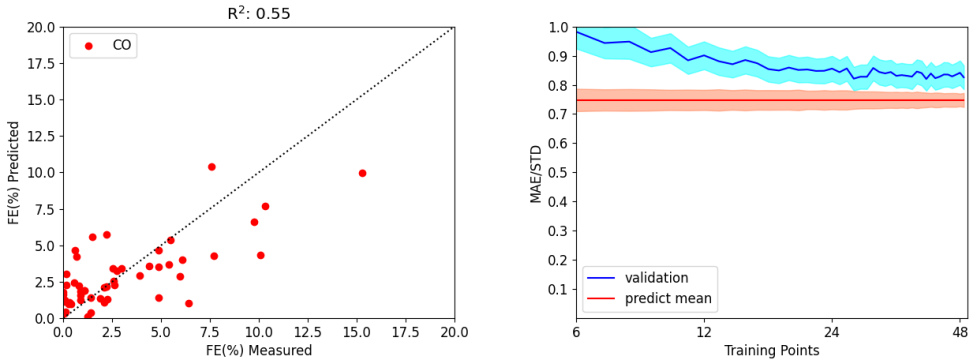

Figure S35: Left Panel: Parity plot between the predicted and measured FEs for CO production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction CO production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

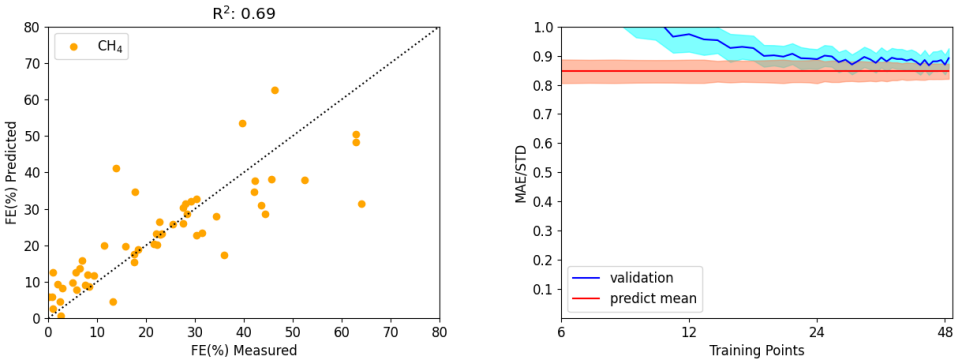

Figure S36: Left Panel: Parity plot between the predicted and measured FEs for CH_4 production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction CH_4 production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

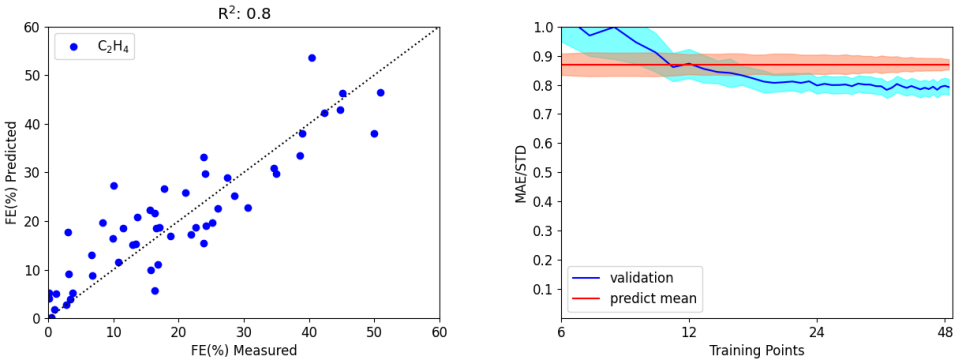

Figure S37: Left Panel: Parity plot between the predicted and measured FEs for C_2H_4 production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction C_2H_4 production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

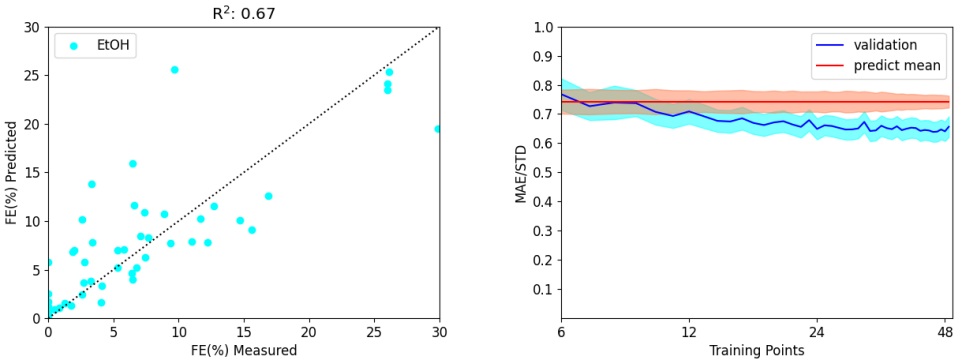

Figure S38: Left Panel: Parity plot between the predicted and measured FEs for EtOH production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction EtOH production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

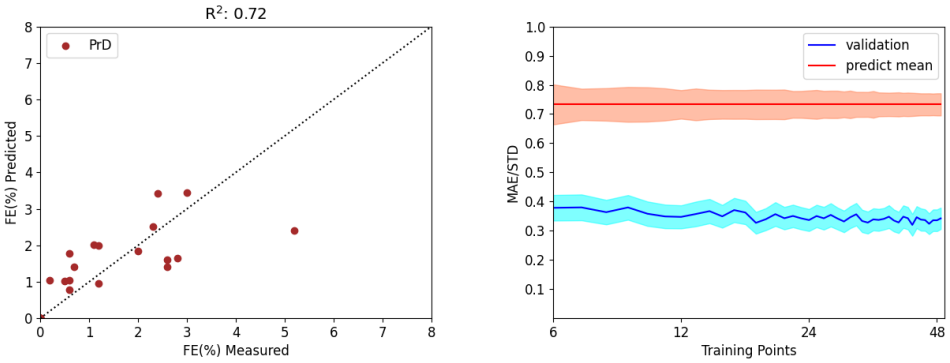

Figure S39: Left Panel: Parity plot between the predicted and measured FEs for PrD production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction PrD production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

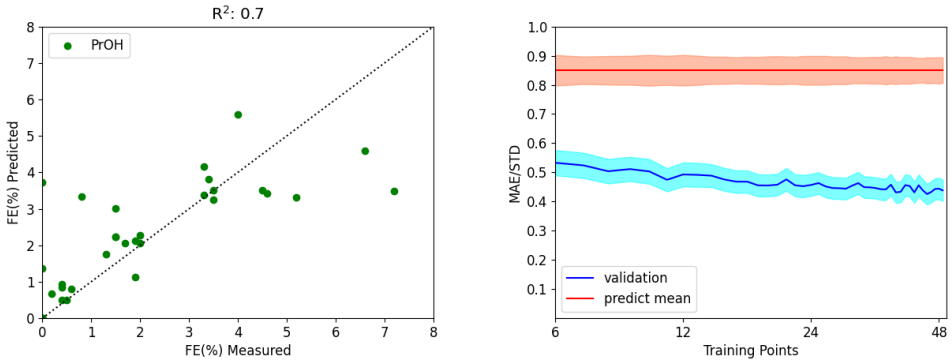

Figure S40: Left Panel: Parity plot between the predicted and measured FEs for PrOH production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction PrOH production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

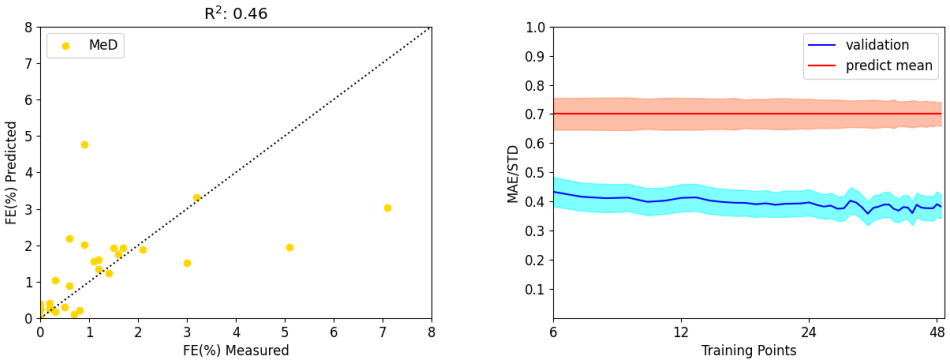

Figure S41: Left Panel: Parity plot between the predicted and measured FEs for MeD production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction MeD production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

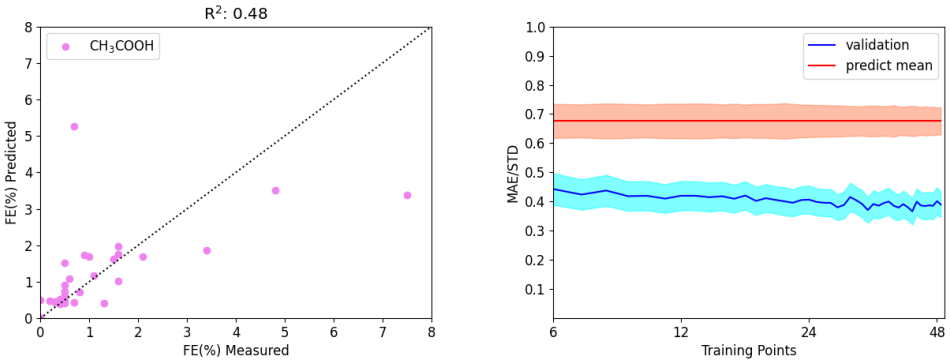

Figure S42: Left Panel: Parity plot between the predicted and measured FEs for CH_3COOH production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction CH_3COOH production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

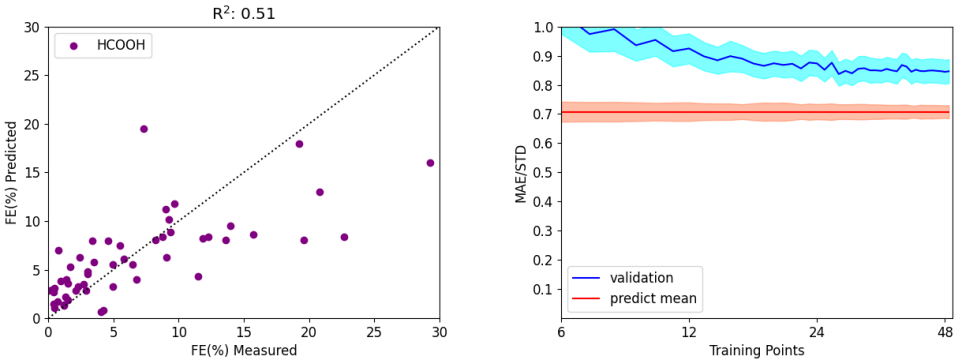

Figure S43: Left Panel: Parity plot between the predicted and measured FEs for HCOOH production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction HCOOH production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

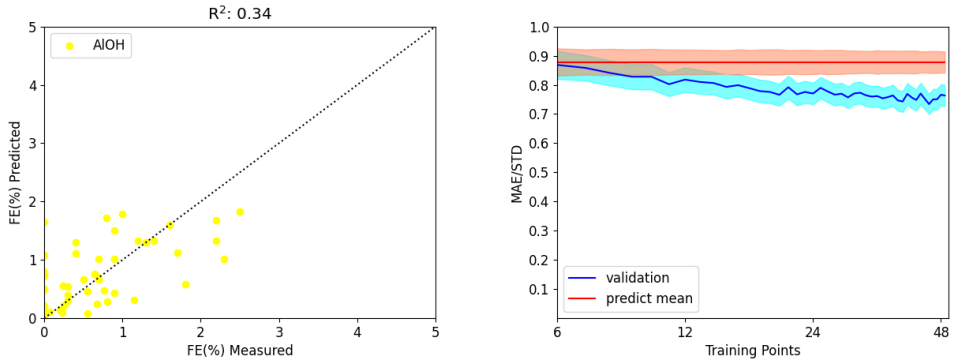

Figure S44: Left Panel: Parity plot between the predicted and measured FEs for AlOH production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction AlOH production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

6 One-Hot Encoding

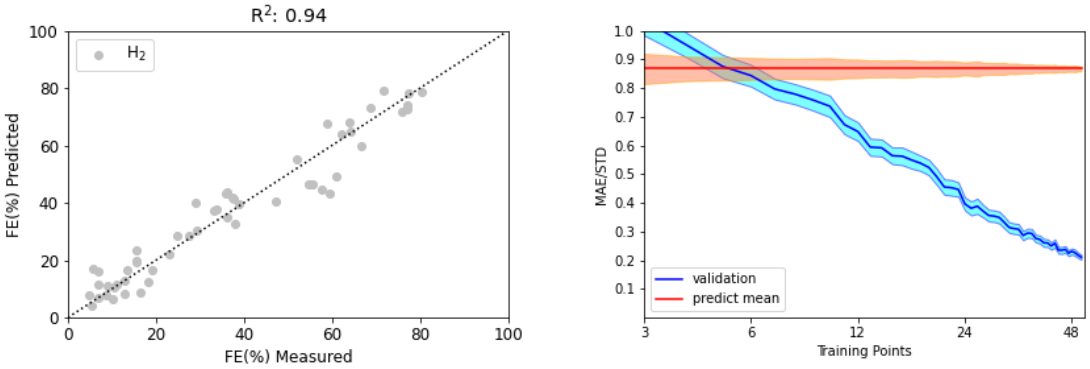

Figure S45: Left Panel: Parity plot between the predicted and measured FEs for H_2 production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction H_2 production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

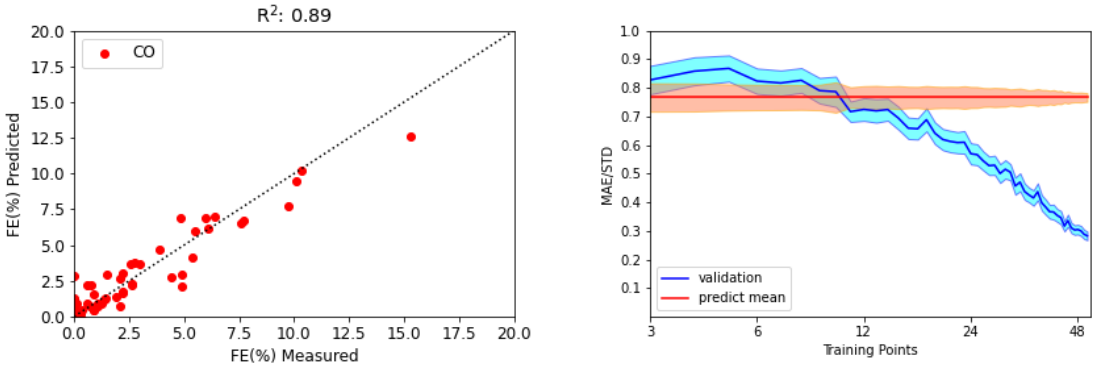

Figure S46: Left Panel: Parity plot between the predicted and measured FEs for CO production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction CO production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

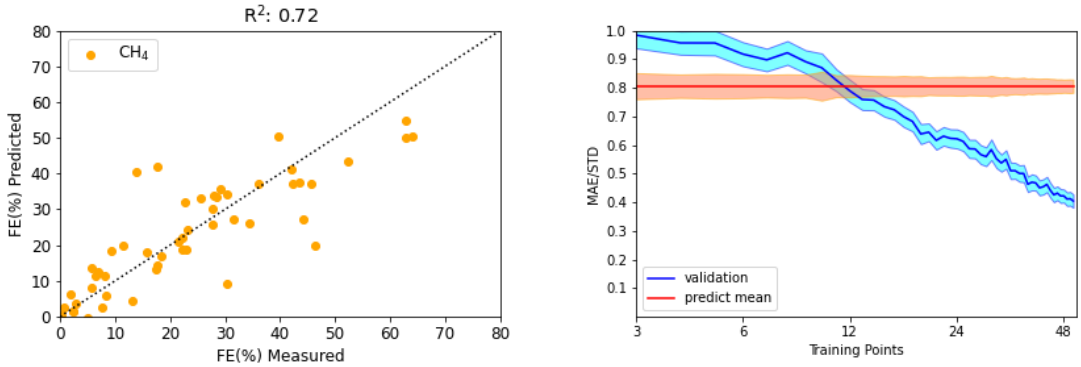

Figure S47: Left Panel: Parity plot between the predicted and measured FEs for CH_4 production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction CH_4 production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

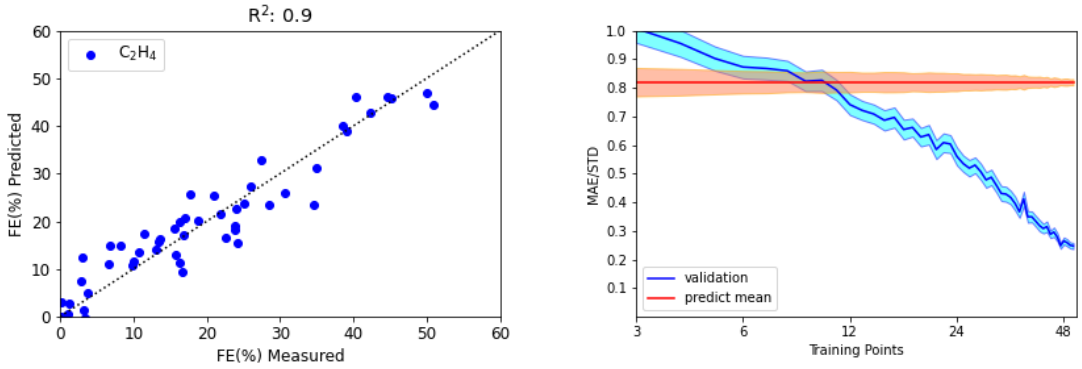

Figure S48: Left Panel: Parity plot between the predicted and measured FEs for C_2H_4 production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction C_2H_4 production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

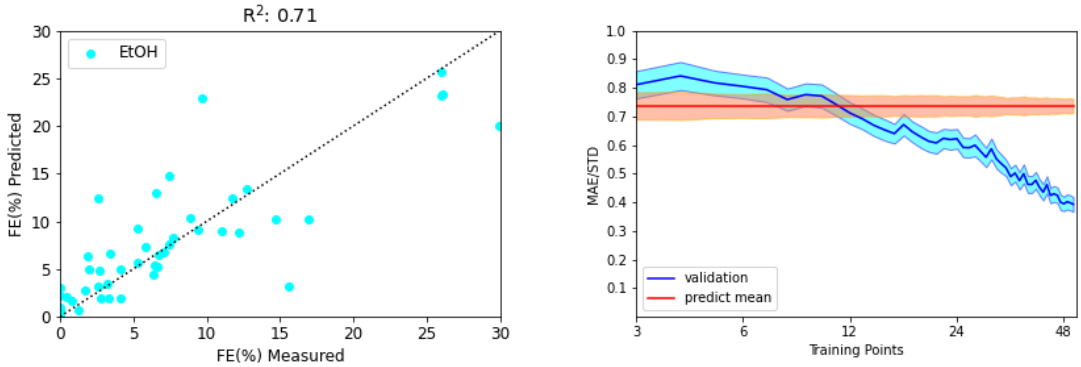

Figure S49: Left Panel: Parity plot between the predicted and measured FEs for EtOH production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction EtOH production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

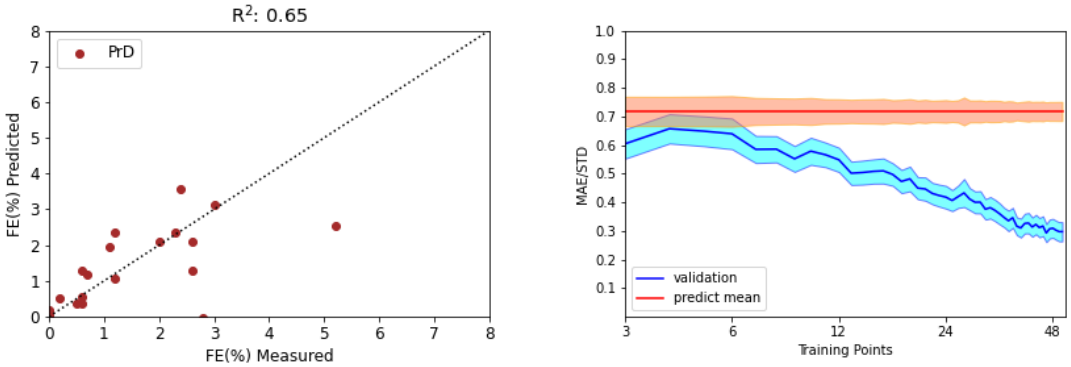

Figure S50: Left Panel: Parity plot between the predicted and measured FEs for PrD production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction PrD production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

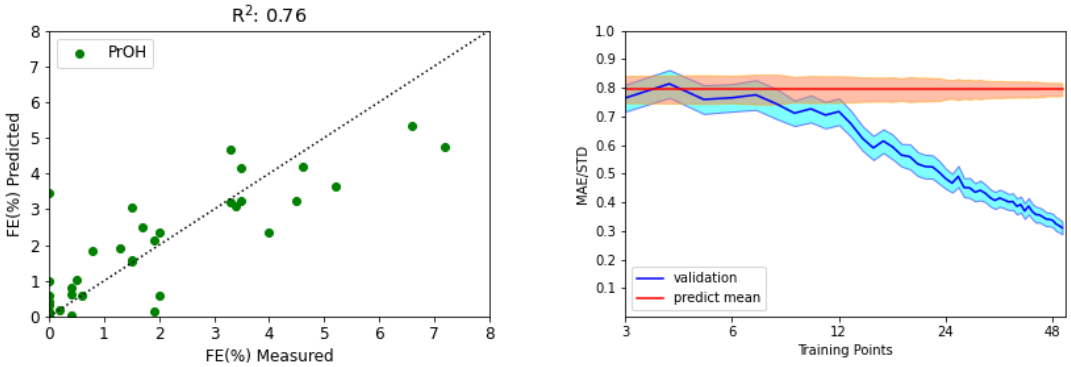

Figure S51: Left Panel: Parity plot between the predicted and measured FEs for PrOH production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction PrOH production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

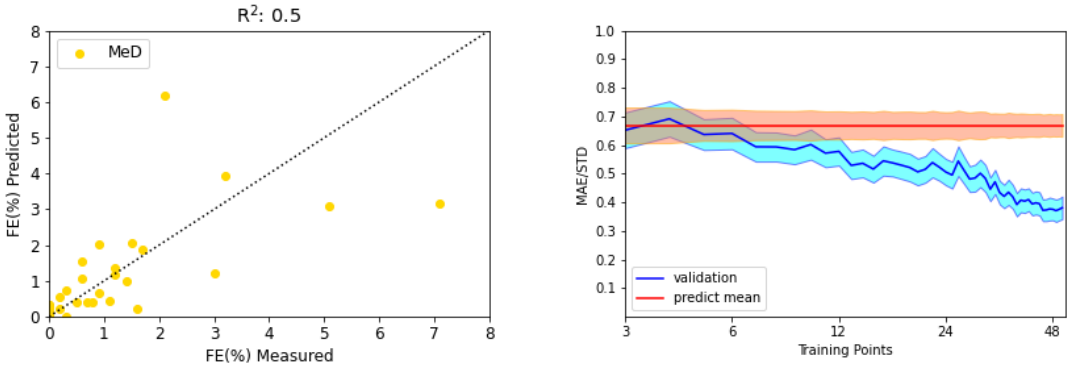

Figure S52: Left Panel: Parity plot between the predicted and measured FEs for MeD production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction MeD production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

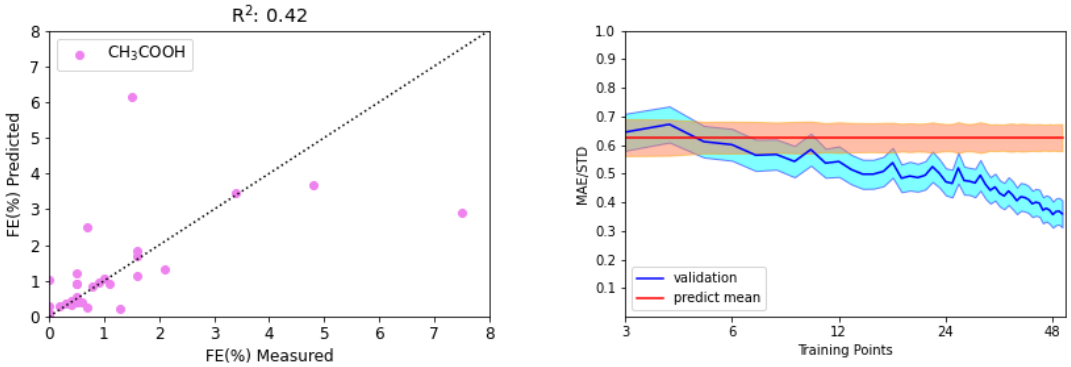

Figure S53: Left Panel: Parity plot between the predicted and measured FEs for CH_3COOH production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction CH_3COOH production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

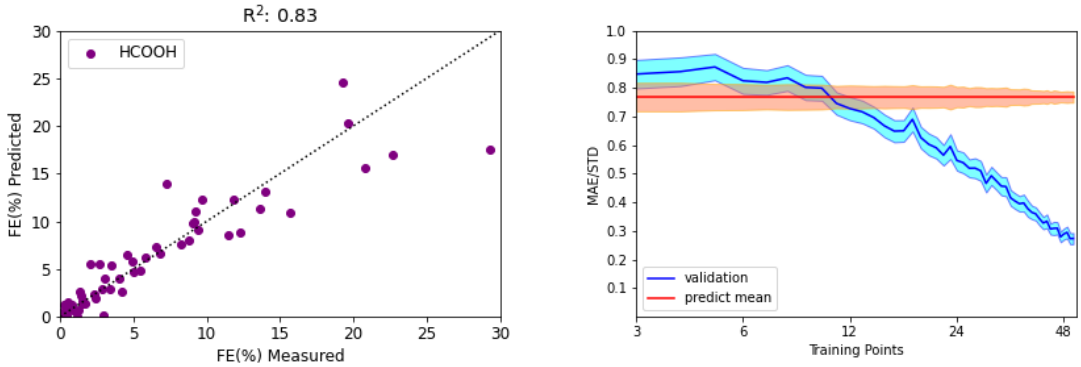

Figure S54: Left Panel: Parity plot between the predicted and measured FEs for HCOOH production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction HCOOH production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

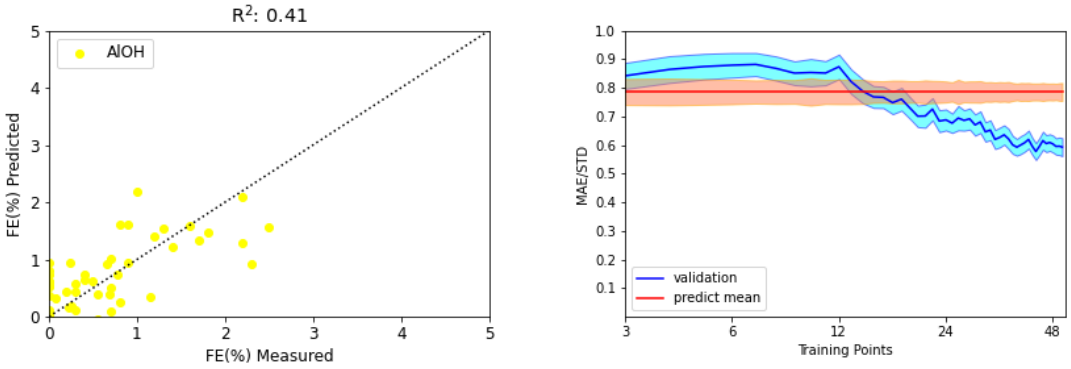

Figure S55: Left Panel: Parity plot between the predicted and measured FEs for AlOH production during CO_2 electroreduction on Cu single crystals with diverse terminations and at a number of applied voltages. Right Panel: MAE/STD (blue line) found for ML models prediction AlOH production during CO_2RR as a function of the number of their training points. For reference, the orange line shows the MAE/STD incurred by a model which predicts the mean FE among the data in the full database regardless of the surface topology and the applied potential.

References