Electronic Supplementary Material (ESI) for Energy & Environmental Science.
This journal is © The Royal Society of Chemistry 2020

Supplementary Information for

Cost, performance prediction and optimization on vanadium flow battery by machine-learning

Tianyu Li,[a] Feng Xing,[a] Tao Liu,[a] Jiawei Sun,[a] Dingqin Shi,[a] Huamin Zhang [a] and Xianfeng Li,*[a]

[a] Division of energy storage, Dalian National Laboratory for Clean Energy (DNL), Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Zhongshan Road 457, Dalian 116023, China

* Corresponding author, *E-mail*: lixianfeng@dicp.ac.cn

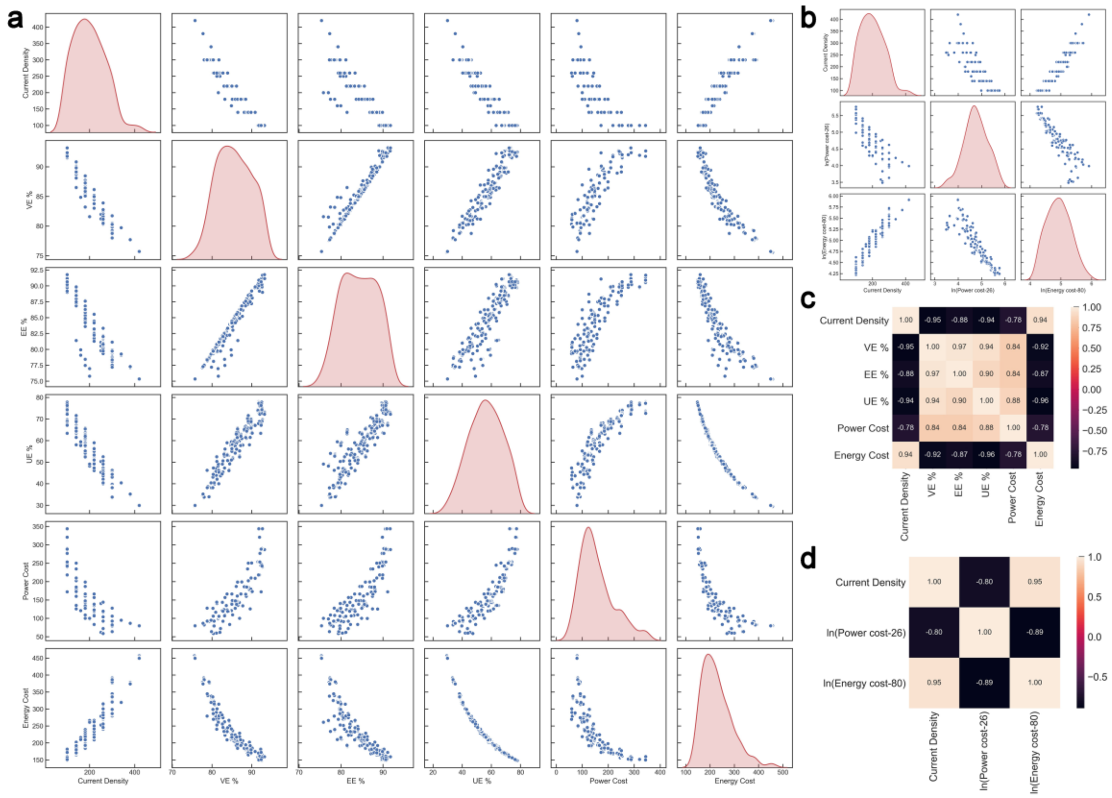


Figure S1 | Pair-plots and Pearson correlation coefficient among Current density, VE, EE, UE, P-cost and E-cost or among Current density, ln(Power cost − 26) and ln(Energy cost − 80). (a) 6×6 pair-plots with 6 variables which include Current density, VE, EE, UE, P-cost and E-cost; (b) 3×3 pair-plots with 3 variables which include Current density, ln(Power cost − 26) and ln(Energy cost − 80) ; Pearson correlation coefficient of (c) 6×6 pair-plots with 6 variables which include Current density, VE, EE, UE, P-cost and E-cost; Pearson correlation coefficient of (d) 3×3 pair-plots with 3 variables which include Current density, ln(Power cost − 26) and ln(Energy cost − 80). The unit of Power cost and Energy cost is ($ (kW h)$^{-1}$ at E/P=4 h).

The definition of Pearson correlation coefficient is as following:

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E\left((X-\mu_X)(Y-\mu_Y)\right)}{\sigma_X \sigma_Y} = \frac{E(XY)-E(X)E(Y)}{\sqrt{E(X^2)-E^2(X)}\sqrt{E(Y^2)-E^2(Y)}} \quad \text{(S1)}$$

For discrete variables, the Pearson correlation coefficient is as following:

$$\rho_{X,Y} = \frac{N\sum XY - \sum X \sum Y}{\sqrt{N\sum X^2 - \left(\sum X\right)^2}\sqrt{N\sum Y^2 - \left(\sum Y\right)^2}} \quad \text{(S2)}$$

where X and Y are two variables. E is the mathematical expectation. $\mathrm{cov}(X,Y)$ is the covariance of X and Y. μ_X and μ_Y are the mean value of X and Y, respectively. σ_X and σ_Y are the standard deviation of X and Y, respectively. N is the number of random variables in X.

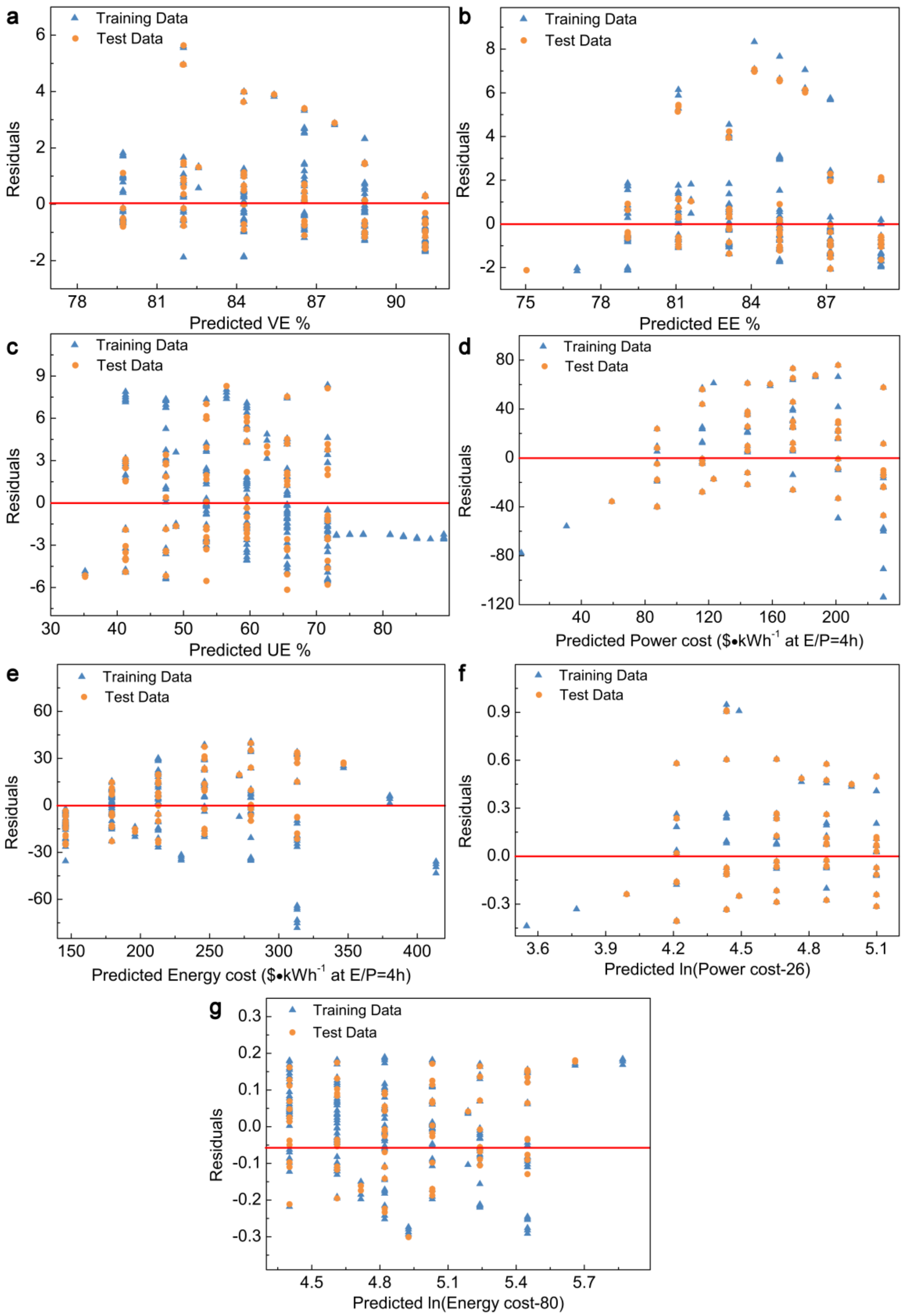
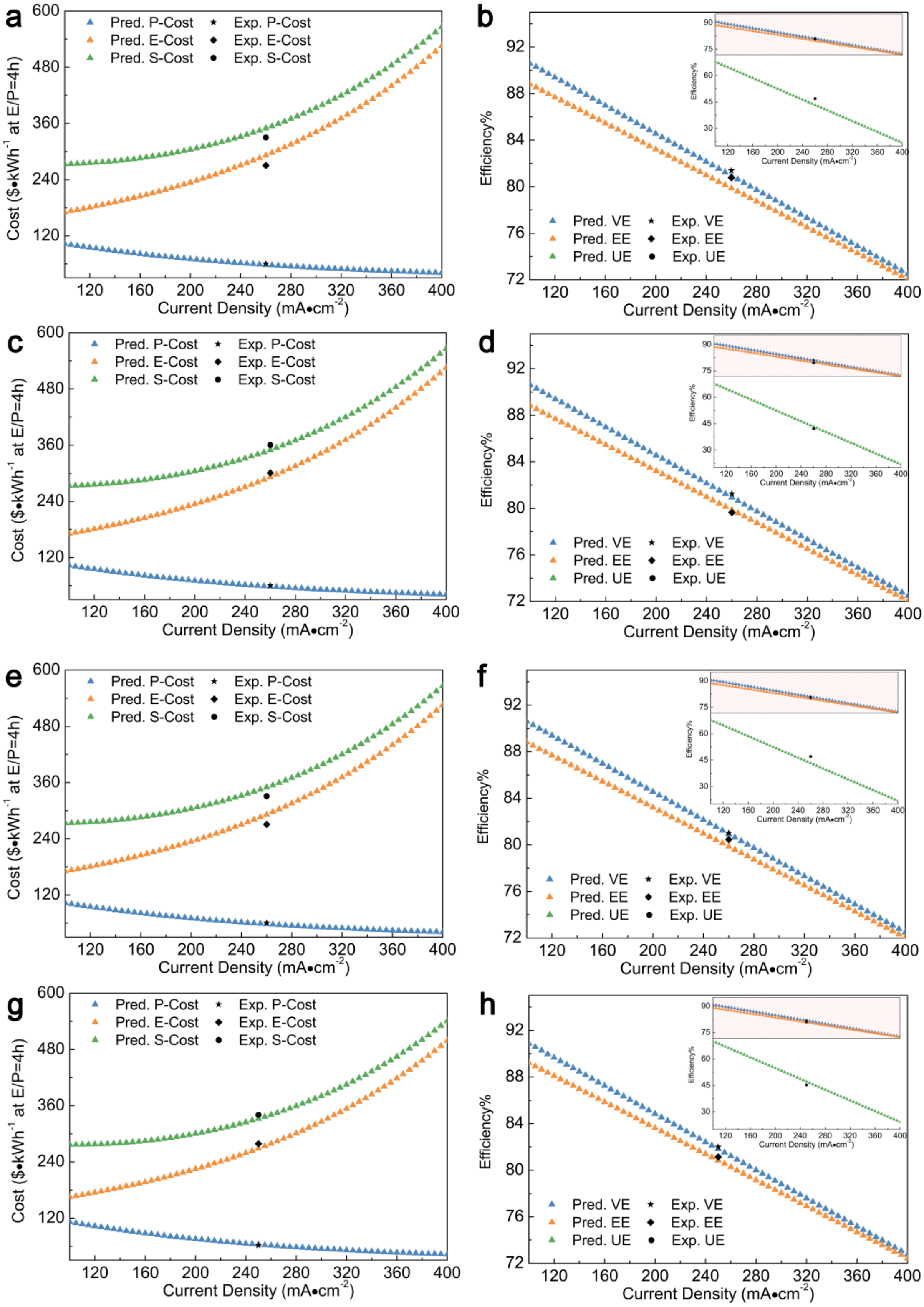
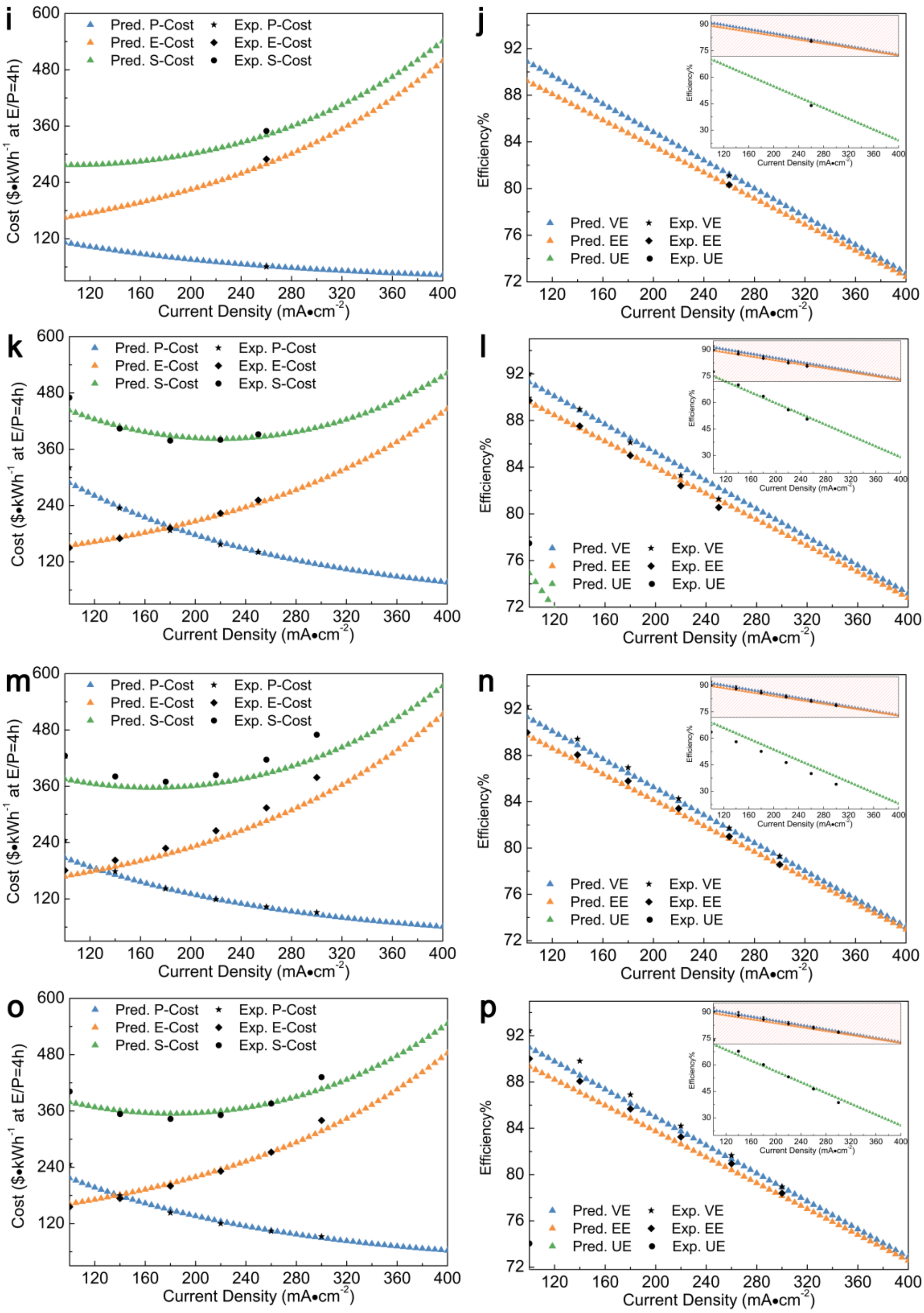
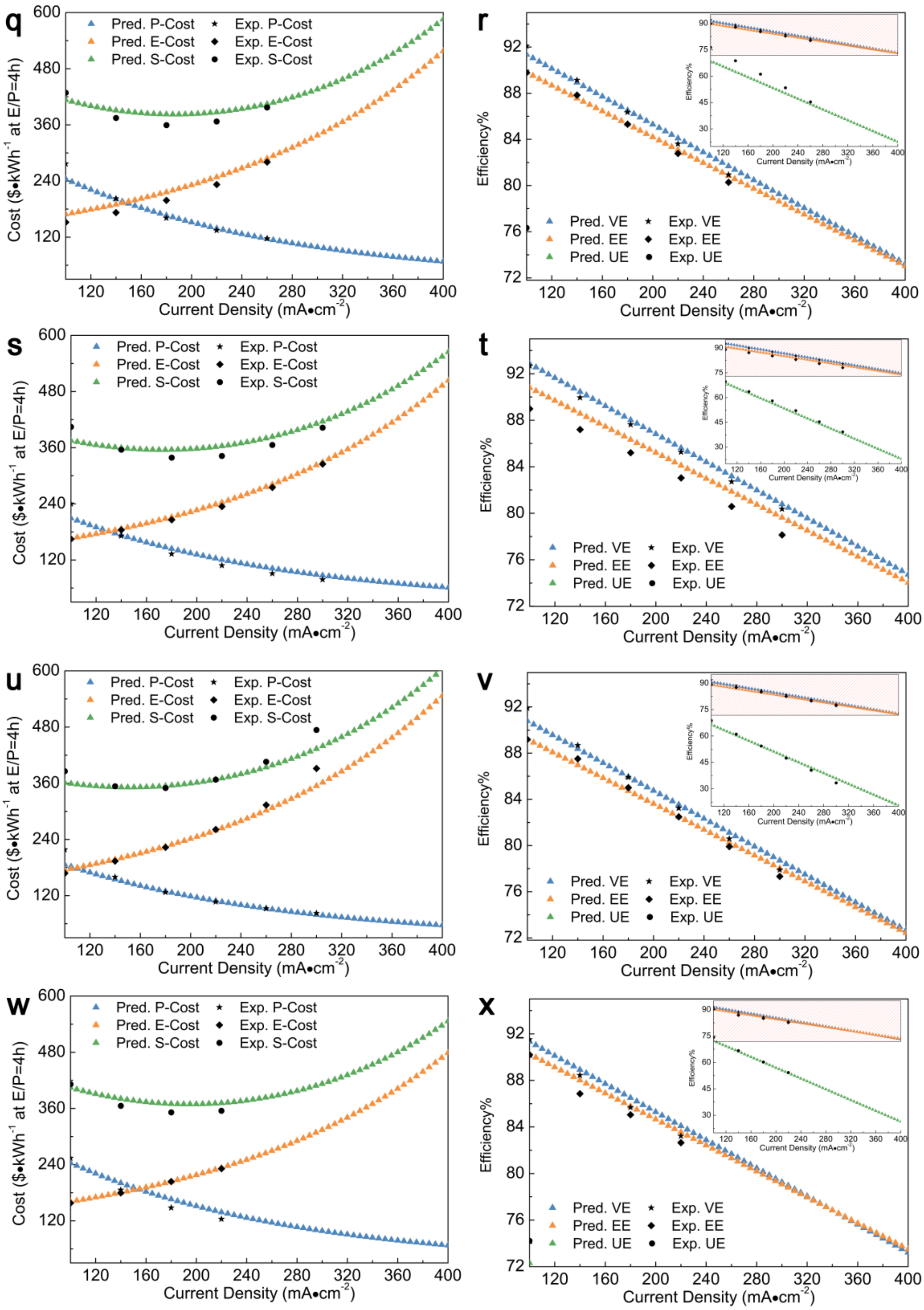


Figure S2 | The residual plots of "only current density" models for Dataset-1. (a) VE, (b) EE, (c) U, (d) ln(Power cost − 26), (e) ln(Energy cost − 80), (f) P-cost and (g) E-cost. The unit of cost is ($ (kW h)$^{-1}$ at E/P=4 h).


a
Pred. P-Cost
Pred. E-Cost
Pred. S-Cost
Exp. P-Cost
Exp. E-Cost
Exp. S-Cost
Cost ($•kWh^{-1} at E/P=4h)
Current Density (mA•cm^{-2})
b
Pred. VE
Pred. EE
Pred. UE
Exp. VE
Exp. EE
Exp. UE
Efficiency%
Current Density (mA•cm^{-2})
c
Pred. P-Cost
Pred. E-Cost
Pred. S-Cost
Exp. P-Cost
Exp. E-Cost
Exp. S-Cost
Cost ($•kWh^{-1} at E/P=4h)
Current Density (mA•cm^{-2})
d
Pred. VE
Pred. EE
Pred. UE
Exp. VE
Exp. EE
Exp. UE
Efficiency%
Current Density (mA•cm^{-2})
e
Pred. P-Cost
Pred. E-Cost
Pred. S-Cost
Exp. P-Cost
Exp. E-Cost
Exp. S-Cost
Cost ($•kWh^{-1} at E/P=4h)
Current Density (mA•cm^{-2})
f
Pred. VE
Pred. EE
Pred. UE
Exp. VE
Exp. EE
Exp. UE
Efficiency%
Current Density (mA•cm^{-2})
g
Pred. P-Cost
Pred. E-Cost
Pred. S-Cost
Exp. P-Cost
Exp. E-Cost
Exp. S-Cost
Cost ($•kWh^{-1} at E/P=4h)
Current Density (mA•cm^{-2})
h
Pred. VE
Pred. EE
Pred. UE
Exp. VE
Exp. EE
Exp. UE
Efficiency%
Current Density (mA•cm^{-2})



i
Pred. P-Cost
Pred. E-Cost
Pred. S-Cost
Exp. P-Cost
Exp. E-Cost
Exp. S-Cost
Cost ($•kWh-1 at E/P=4h)
Current Density (mA•cm-2)
j
Pred. VE
Pred. EE
Pred. UE
Exp. VE
Exp. EE
Exp. UE
Efficiency%
Current Density (mA•cm-2)
k
Pred. P-Cost
Pred. E-Cost
Pred. S-Cost
Exp. P-Cost
Exp. E-Cost
Exp. S-Cost
Cost ($•kWh-1 at E/P=4h)
Current Density (mA•cm-2)
l
Pred. VE
Pred. EE
Pred. UE
Exp. VE
Exp. EE
Exp. UE
Efficiency%
Current Density (mA•cm-2)
m
Pred. P-Cost
Pred. E-Cost
Pred. S-Cost
Exp. P-Cost
Exp. E-Cost
Exp. S-Cost
Cost ($•kWh-1 at E/P=4h)
Current Density (mA•cm-2)
n
Pred. VE
Pred. EE
Pred. UE
Exp. VE
Exp. EE
Exp. UE
Efficiency%
Current Density (mA•cm-2)
o
Pred. P-Cost
Pred. E-Cost
Pred. S-Cost
Exp. P-Cost
Exp. E-Cost
Exp. S-Cost
Cost ($•kWh-1 at E/P=4h)
Current Density (mA•cm-2)
p
Pred. VE
Pred. EE
Pred. UE
Exp. VE
Exp. EE
Exp. UE
Efficiency%
Current Density (mA•cm-2)



q
r
s
t
u
v
w
x
Cost ($•kWh^{-1} at E/P=4h)
Efficiency%
Current Density (mA•cm^{-2})
Pred. P-Cost
Pred. E-Cost
Pred. S-Cost
Exp. P-Cost
Exp. E-Cost
Exp. S-Cost
Pred. VE
Pred. EE
Pred. UE
Exp. VE
Exp. EE
Exp. UE

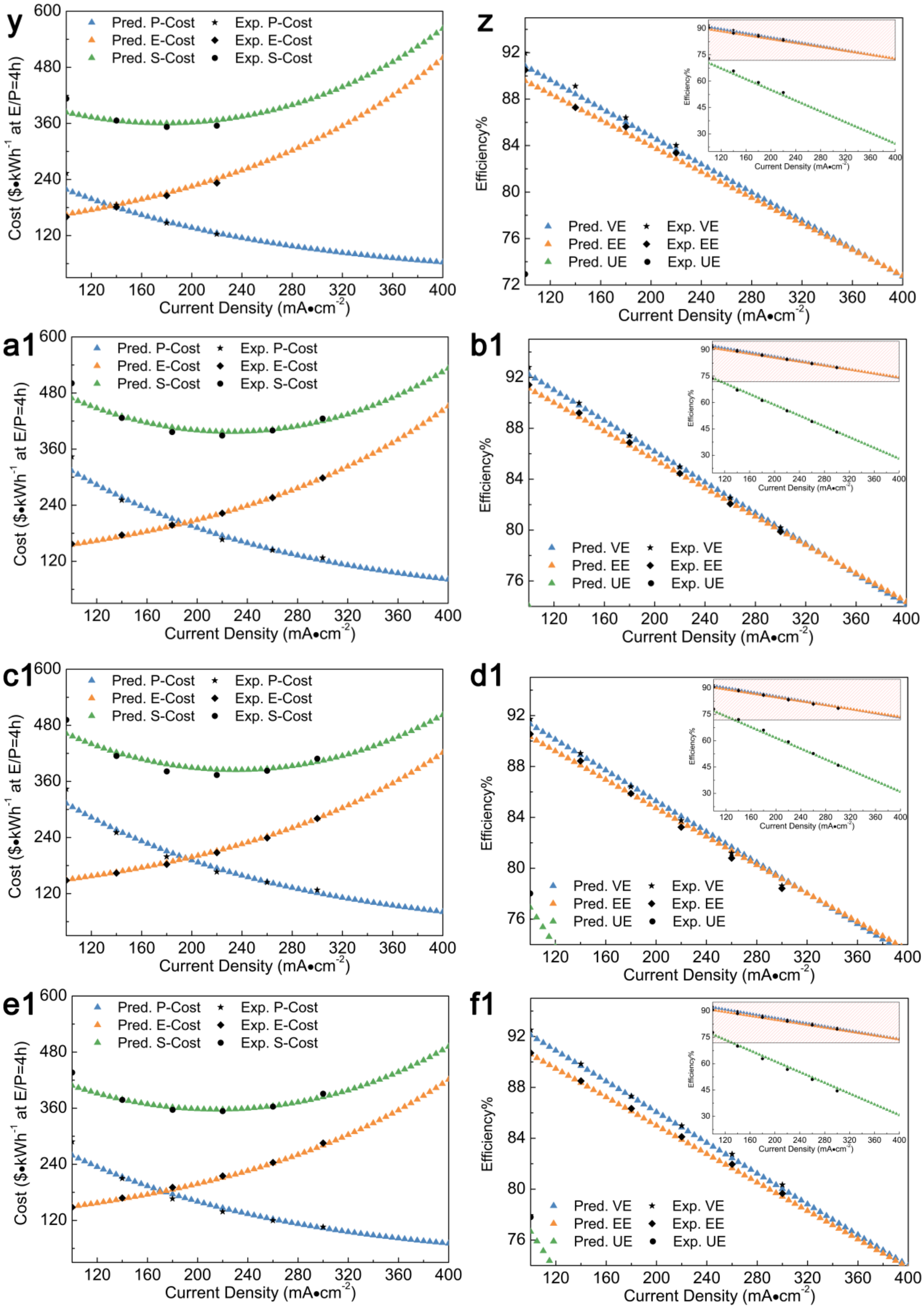


Figure S3 | Comparison of the costs and efficiencies between experiment and "full features" models. (a) costs and (b) efficiencies of "VFB 20190322-1"; (c) costs and (d) efficiencies of "VFB 20190322-2"; (e) costs and (f) efficiencies of "VFB 20190321"; (g) costs and (h) efficiencies of "VFB 20190416"; (i) costs and (j) efficiencies of "VFB

20190415”; (k) costs and (l) efficiencies of “VFB 20181229”; (m) costs and (n) efficiencies of “VFB 20190123”; (o) costs and (p) efficiencies of “VFB 20190129”; (q) costs and (r) efficiencies of “VFB 20190110”; (s) costs and (t) efficiencies of “VFB 20190306”; (u) costs and (v) efficiencies of “VFB 20190126”; (w) costs and (x) efficiencies of “VFB 20190703”; (y) costs and (z) efficiencies of “VFB 20190708”; (a1) costs and (b1) efficiencies of “VFB 20200506”; (c1) costs and (d1) efficiencies of “VFB 20200509”; (e1) costs and (f1) efficiencies of “VFB 20200519”. The unit of cost is (\$ (kW h)$^{-1}$ at E/P=4 h).

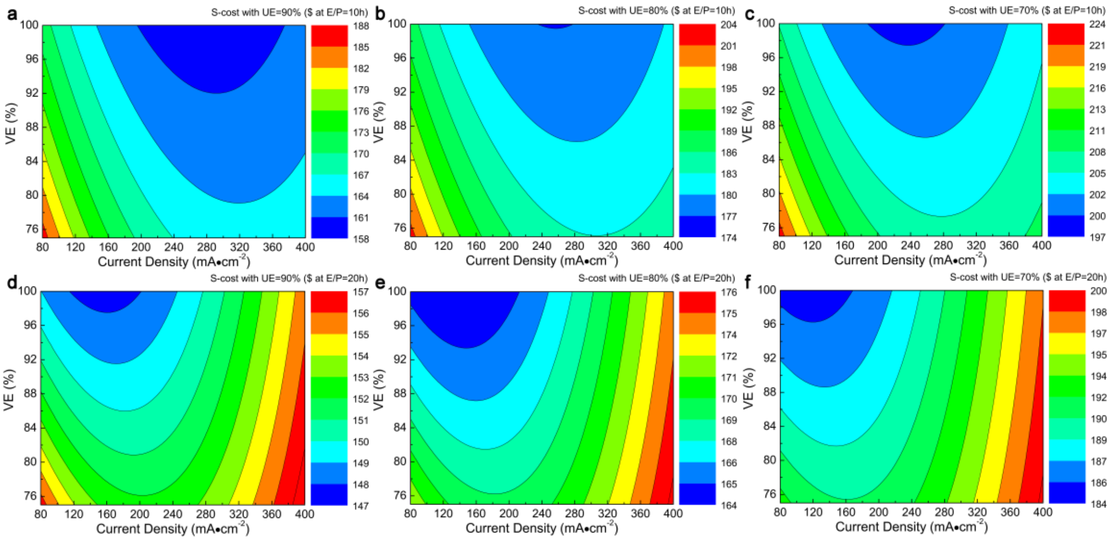


Figure S4 | Prediction S-cost change with operating current density at different VE and UE (E/P=10 h or 20 h). (a) UE=90% (E/P=10 h), (b) UE=80% (E/P=10 h), (c) UE=70% (E/P=10 h); (d) UE=90% (E/P=20 h), (e) UE=80% (E/P=20 h), (f) UE=70% (E/P=20 h).

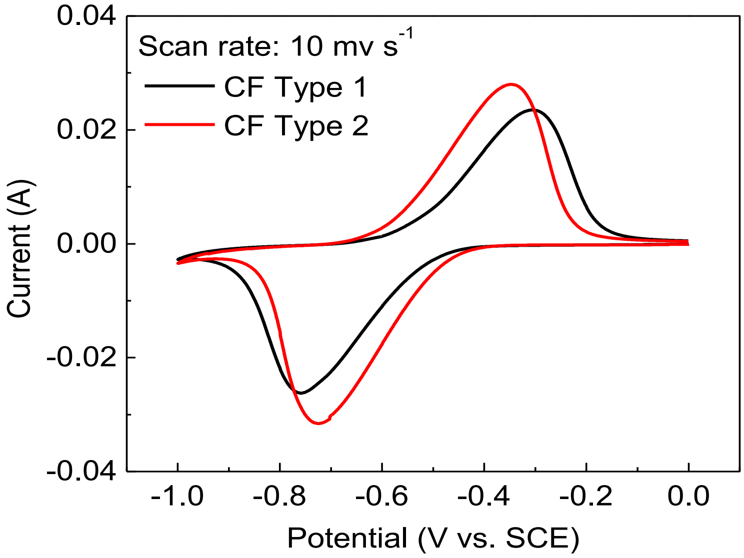


Figure S5 | CV curves of two types of carbon felts.

Table S1 | The maximum absolute relative deviation (MARD) of different target function

	MARD (%)			MARD (%)	
	Training	Test		Training	Test
ln(Power cost)	21.77	15.41	ln(Energy cost)	13.36	11.57
ln(Power cost − 10)	19.57	15.38	ln(Energy cost − 40)	12.94	13.32
ln(Power cost − 26)	16.31	16.29	ln(Energy cost − 80)	11.93	11.29
ln(Power cost − 40)	19.54	19.53	ln(Energy cost − 120)	24.38	12.17

Table S2 | Accuracy evaluations of each "only current density" model for efficiencies and costs. Coefficient of determination (R^2), root mean square error (RMSE) and mean absolute prediction errors (MAPE) are used to evaluate the precision of linear regression models that are used to predict VE %, EE %, UE %, ln(Power cost – 26), ln(Energy cost – 80), P-cost, and E-cost, respectively.

"only current density" models	R^2		RMSE		MAPE (%)	
	Training	Test	Training	Test	Training	Test
VE %	0.9097	0.9071	1.3000	1.2727	1.18	1.14
EE %	0.7731	0.7534	2.0192	1.9798	1.54	1.50
UE %	0.8833	0.8762	4.0062	2.4364	6.50	6.04
ln(Power cost − 26)	0.6479	0.6139	0.8049	0.7835	19.29(P-cost)	20.63(P-cost)
ln(Energy cost − 80)	0.9046	0.8939	0.1235	0.1177	6.38(E-cost)	6.33(E-cost)
Power cost	0.6093	0.6179	41.2443	39.3881	24.18	24.04
Energy cost	0.8853	0.8550	21.8297	19.6793	7.15	7.52

The unit of P-cost and E-cost is ($ (kW h)$^{-1}$ at E/P=4 h).

Table S3 | Properties of materials

Carbon felt type	Electrocatalytic activity	Porosity
CF Type 1	Relatively lower	> 90 %
CF Type 2	Relatively Higher	
Bipolar plate Type	**Electronic conductivities (S/m)**	
BP Type 1	self-made, about 15	
BP Type 2	about 400	
Seal Type		
S Type 1	Face seal	
S Type 2	Line seal	
Membrane type		
M Tpye 1	Self-made membrane	
M Tpye 2	Nafion212	
M Tpye 3	Nafion115	
Flow field type	**Flow field length**	
FF Type 1	Shortest	
FF Type 2	Longest, stack strucure1	
FF Type 3	Longest, stack strucure2	
FF Type 4	Middle	

Table S4 | Comparison of costs and efficiencies between original stack and optimized stack at the optimal operating current density.

	Original Stack	Optimized Stack	Relative percentage %
The optimal operating current density $mA \cdot cm^{-2}$	130	160	23.08
VE %	90.10	89.17	-1.03
EE %	88.62	87.36	-1.42
UE %	63.53	64.95	2.23
P-cost $/(kW h) (at E/P=4 h)	126.08	122.32	-2.98
E-cost $/(kW h) (at E/P=4 h)	184.48	179.89	-2.49
T-cost $/(kW h) (at E/P=4 h)	310.56	302.21	-2.69

Table S5 | Accuracy evaluations for each efficiencies and costs for Dateset-2. Coefficient of determination (R^2), root mean square error (RMSE) and mean absolute prediction error (MAPE) are used to evaluate the precision of linear regression models that are used to predict VE, EE, UE, ln(P-cost), ln(E-cost - 80), P-cost, and E-cost, respectively.

	R^2		RMSE		MAPE (%)	
	Training	Test	Training	Test	Training	Test
VE %	0.9908	0.9839	0.3720	0.4776	0.30	0.40
EE %	0.9765	0.9603	0.4884	0.6113	0.41	1.38
UE %	0.9787	0.9603	2.0381	2.8598	4.03	3.73
ln(Power cost)	0.9977	0.9970	0.0359	0.0371	2.86(P-cost)	3.02(P-cost)
ln(Energy cost - 80)	0.9739	0.9473	0.0709	0.1033	2.89(E-cost)	4.64(E-cost)

The unit of P-cost and E-cost is (\$ (kW h)$^{-1}$ at E/P=4 h).