Support Information

Physics-informed Gaussian process regression of *in-operando* capacitance for carbon supercapacitors

Runtong Pan[1], Mengyang Gu[2] and Jianzhong Wu[1]

[1]Department of Chemical and Environmental Engineering, University of California, Riverside, CA 92521, USA

[2]Department of Statistics and Applied Probability, University of California, Santa Barbara, CA 93106, USA

In this support information, we provide the formulas for calculating capacitance, the input data used for training the machine learning models and the methodology of Gaussian Process Regression (GPR). All the data are calculated from the CV curves collected from the literature. [1-7]

Capacitive behavior

The specific integral capacitance is given from the CV curves by

$$C_{sp} = \frac{\int_{V_{start}}^{V_{end}} i(V)dV}{2\upsilon m\Delta V} = \frac{\bar{I}\frac{\Delta V}{\upsilon}}{m\Delta V} = \frac{\bar{I}\Delta t}{m\Delta V} \tag{S1}$$

where υ is the scan rate (V/s), i is the electrical current, m is the electrode mass, ΔV is the potential window, $\bar{I}$ is the average current, Δt=ΔV/υ is charging/discharging time, and C_{sp} stands for specific integral capacitance of the electrode.

The energy density is defined as

$$E = \frac{C_{cell}\Delta V^2}{2} = \frac{C_{sp}\Delta V^2}{8} = \frac{\bar{I}\Delta t\Delta V}{8m} \tag{S2}$$

where C_{cell} is the specific capacitance of a two-electrode symmetrical supercapacitor. The power density is calculated from

$$P = \frac{E}{\Delta t} = \frac{\bar{I}\Delta V}{8m} \quad . \tag{S3}$$

Gaussian Process Regression (GPR) models

GPR is a non-parametric Bayesian method for solving regression problems.[8, 9] The supervised ML can capture different kinds of relationships by using an appropriate kernel to capture the unknown relations between the independent and dependent variables.[10] By introducing a theoretically infinite number of parameters, kernels are widely used in supervised ML methods including not only GPR but also support vector machine (SVM), principal components analysis (PCA), canonical correlation, and ridge regression. The kernel functions empower the ML methods to operate in a high-dimensional, implicit feature space by computing the inner products between the images of all pairs of data in the feature space.

In this work, the predictors used in regression are all standardized, so they are unitless values in regression. $X = [z_1, z_2, \ldots, z_n]$

$$z_i = \frac{x_i - \mu_i}{\sigma_i} \tag{S4}$$

where μ_i, σ_i are the mean and standard deviation of original input predictor x_i), so they are unitless values in the model.

Specifically, GPR provides the mapping from a predictor matrix $X = [x_1, x_2, \ldots, x_n]$ to a response vector **y**.[9, 11, 12]. Consider one input observation $x = [z_1, z_2, \ldots, z_n]$ (in this work, $x = \left[z_v,\ z_{S_{micro}},\ z_{S_{meso}}\right]$, the standardized input observation of $[v, S_{micro}, S_{meso}]$), and its corresponding response value y, the mapping is assumed to be an unknown function, $y = f(x) + \varepsilon$, where $\varepsilon \sim N(0, \sigma^2)$ is an independent zero-mean Gaussian noise with a standard deviation of σ. GPR defines a probability distribution with function

$$f(x) \sim MN\left(m(x), \kappa\left(x, x^{'}\right)\right) \tag{S5}$$

where $m(x)$ and $\kappa(x,x')$ are the mean and covariance functions:

$$\{m(x) = E[f(x)]\ \kappa(x,x') = E\left[(f(x) - m(x))(f(x') - m(x'))\right] \quad \text{(S6)}$$

Usually, the mean function is assumed to be a basis function in the form as:

$$m(x) = H(x)\beta \quad \text{(S7)}$$

where $H(x)$ is the basis matrix, β is a vector of basis coefficients.

Common selections of the basis matrix include the constant basis ($H = 1$), the linear basis ($H = [1,X]$), and the 'pure Quadratic' basis:

$$H = [1,X,X_2] \quad \text{(S8)}$$

where X_2 is half-vectorization of the quadratic form of the predictors.

The prediction mean $\dot{y}$ and variance y^{sd} of the response value at a given point x^* are:

$$\dot{y}(x^*) = m(x^*) + K(x^*,X)^T\left[K(X.X) + \sigma^2 I_n\right]^{-1}(y - u) \quad \text{(S9)}$$

$$y^{sd} = K(x^*,x^*) - K(x^*,X)^T\left[K(X.X) + \sigma^2 I_n\right]^{-1}K(X,x^*) \quad \text{(S10)}$$

where

$$K(x^*,X) = \left[\kappa(x^*,x_1),\kappa(x^*,x_2),\ldots,\kappa(x^*,x_n)\right] \quad \text{(S11)}$$

The covariance function (or kernel function) is the major component of a GP model. Under the stationary condition, $\kappa(x,x') = \sigma^2 f_\kappa(x - x')$, where σ^2 being a variance parameter, which is the signal standard deviation and f_κ is a correlation function, with $f_\kappa(0) = 1$. The covariance function is assumed to be isotropic, i.e., $\kappa(x,x') = \sigma^2 f_\kappa(d)$, where $d = ||x - x'|| = \sqrt{\sum_{i=1}^{n} (x_i - x_i')^2}$ being the Euclidean distance between x and x'. The frequently used covariance functions include power exponential correlation, Matérn correlation and Rational Quadratic correlation.[9] We can see that the correlation function can be normalized by a length scale γ. so

$$\kappa(x,x') = \sigma^2 f_\kappa(x - x') = \sigma^2 f_\kappa(d) = \sigma^2 f_{\kappa 0}(r) \quad \text{(S12)}$$

where $r = \frac{d}{\gamma} = \frac{|x - x'|}{\gamma}$ being the related radius.

It's also possible to use a separate length scale γ_m for each predictor m, called automatic relevance determination (ARD). This can be done by replacing all the related distance $\frac{d}{\gamma}$ by related radius r with separate length scale for each predictor:

$$r = \sqrt{\sum_{m=1}^{D} \frac{(x_{im} - x_{jm})^2}{\gamma_m^2}} \tag{S13}$$

The power exponential correlation kernel is given by:

$$\kappa(x,x') = \sigma^2 expexp\{ - (r)^{\alpha}\} \tag{S14}$$

Where σ is the signal standard deviation, $\alpha \in (0,2]$ is the roughness parameter of the kernel function. When $\alpha = 2$, the kernel function is called Squared Exponential Kernel or Gaussian kernel, which is infinitely differentiable.

The Matérn correlation kernel is

$$\kappa(x,x') = \sigma^2 \frac{1}{2^{\alpha - 1}\Gamma(\alpha)}(r)^{\alpha}\mathrm{K}_{\alpha}(r) \tag{S15}$$

where $\mathrm{K}_{\alpha}, \alpha$ are the modified Bessel function of the second kind and the roughness parameter. This kernel is $\lceil \alpha \rceil - 1$ differentiable, where $\lceil \alpha \rceil$ means the ceiling integer of α. $\alpha = \frac{3}{2} or \frac{5}{2}$ are most frequently used Matérn kernels. When $\alpha = \frac{1}{2}$, it becomes the exponential kernel function. When $\alpha \to \infty$, it converges to the Gaussian kernel.

The Rational Quadratic correlation kernel has the following form

$$\kappa(x,x') = \sigma^2 \left(1 + \frac{r^2}{2\alpha}\right)^{-\alpha} \tag{S16}$$

where α is a positive-valued scale-mixture parameter. This kernel is infinitely differentiable as the Gaussian kernel. It can be interpreted as an infinite sum of different Gaussian kernels with different characteristic length scales. Here, α means the weighting between different length scales. When $\alpha \to \infty$, it converges to the Gaussian kernel.

In this work, we have tested all these covariance functions except exponential correlation, which is not smooth, and using the ARD kernels since it's clear that the scan rate needs a different scaling length comparing to the surface area.

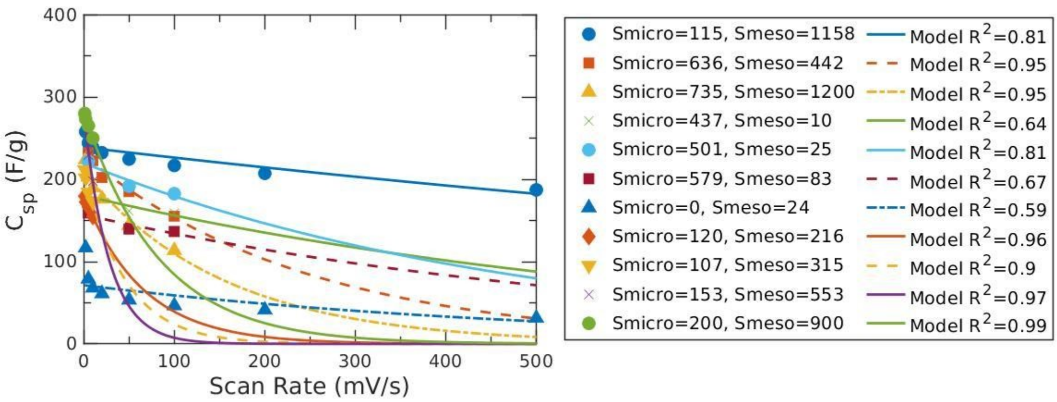

Fig S1 Correlation of the specific capacitance-scan rate relationship based on a semi-empirical physical model (Eqn (1)) for the carbon materials with enough data points.

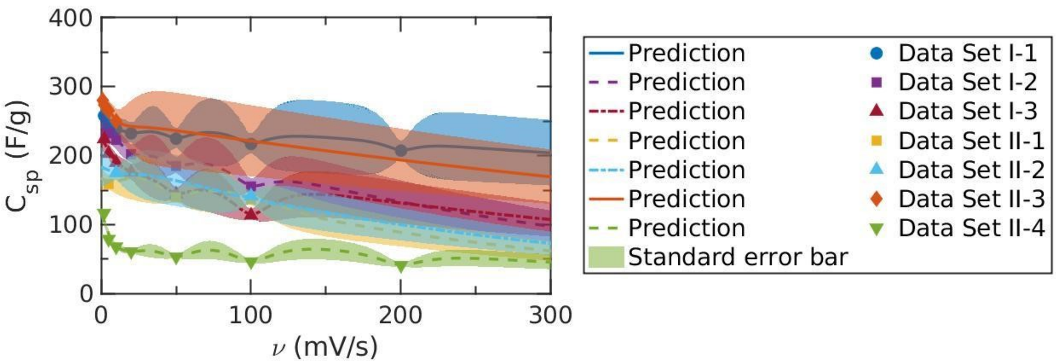

Fig S2 The specific capacitance versus the scan rate predicted by PhysGPR with non-ARD Matérn 3/2 kernel, shows strong overfitting. The specific surface areas of electrode materials are: Data Set I-1: $S_{micro} = 115\ m^2/g$, $S_{meso} = 1158\ m^2/g$; Data Set I-2: $S_{micro} = 636\ m^2/g$, $S_{meso} = 442\ m^2/g$; and Data Set I-3: $S_{micro} = 735\ m^2/g$, $S_{meso} = 1200\ m^2/g$. Data Set II-1: $S_{micro} = 579\ m^2/g$, $S_{meso} = 83\ m^2/g$, Data Set II-2: $S_{micro} = 481\ m^2/g$, $S_{meso} = 193\ m^2/g$, Data Set II-3: $S_{micro} = 200m^2/g$, $S_{meso} = 900\ m^2/g$, Data Set II-4: $S_{micro} = 0\ m^2/g$, $S_{meso} = 24\ m^2/g$..

Table S1 Dataset for the capacitive performance of carbon electrodes. Columns: C_{sp}: specific capacitance. E: Energy density. P: Power density. SA_{micro}: Specific micropore surface area (d<2nm) SA_{meso}: specific mesopore surface area (2nm<d<50nm) ν: cyclic voltammetry scan rate

#	C_{sp}/(F/g)	*E*/(Wh/kg)	*P*/(kW/kg)	SA_{micro}/(m^2/g)	SA_{meso}/(m^2/g)	υ/(mV/s)
1	0	0	0	0	0	0
2	0	0	0	0	0	5
3	0	0	0	0	0	10
4	188.58	6.548	0.118	1990	879	5
5	232.27	8.065	0.145	636	442	5
6	222.77	7.735	0.278	636	442	10
7	202.29	7.024	0.506	636	442	20
8	185.15	6.429	1.157	636	442	50
9	155.41	5.396	1.943	636	442	100
10	185.11	6.428	0.116	713	290	5
11	170.51	5.921	0.213	457	126	10
12	101.47	3.523	1.268	457	126	100
13	160.84	5.585	0.201	429	188	10
14	115.52	4.011	1.444	429	188	100
15	175.29	6.086	0.219	481	193	10
16	141.55	4.915	1.769	481	193	100
17	253.90	8.816	0.317	1118	504	10
18	203.05	7.050	2.538	1118	504	100
19	224.15	7.783	0.056	735	1200	2
20	202.99	7.048	0.127	735	1200	5
21	189.89	6.593	0.237	735	1200	10
22	176.24	6.119	0.441	735	1200	20
23	144.14	5.005	0.901	735	1200	50
24	113.60	0.394	0.142	735	1200	100

25	241.54	8.387	0.151	1506	269	5
26	212.44	7.376	0.266	1506	269	10
27	207.12	7.192	0.518	1506	269	20
28	197.94	6.873	1.237	1506	269	50
29	198.00	6.875	2.475	1506	269	100
30	182.58	6.340	0.114	437	10	5
31	161.70	5.615	1.011	437	10	50
32	158.97	5.520	1.987	437	10	100
33	221.86	7.703	0.139	501	25	5
34	191.76	6.658	1.198	501	25	50
35	182.59	6.340	2.282	501	25	100
36	159.09	5.524	0.099	579	83	5
37	139.68	4.850	0.873	579	83	50
38	136.66	4.745	1.708	579	83	100
39	116.85	4.057	0.029	0	24	2
40	79.11	2.747	0.049	0	24	5
41	68.03	2.362	0.085	0	24	10
42	61.20	2.125	0.153	0	24	20
43	53.48	1.857	0.334	0	24	50
44	46.58	1.617	0.582	0	24	100
45	41.30	1.434	1.033	0	24	200
46	31.42	1.091	1.964	0	24	500
47	257.94	8.956	0.064	115	1158	2
48	244.31	8.483	0.153	115	1158	5
49	238.34	8.276	0.298	115	1158	10
50	232.37	8.068	0.581	115	1158	20
51	224.65	7.800	1.404	115	1158	50

52	216.90	7.531	2.711	115	1158	100
53	207.36	7.200	5.184	115	1158	200
54	187.26	6.502	11.704	115	1158	500
55	179.60	6.236	0.022	120	216	1
56	172.40	5.986	0.043	120	216	2
57	166.30	5.774	0.104	120	216	5
58	155.00	5.382	0.194	120	216	10
59	211.60	7.347	0.026	107	315	1
60	201.60	7.000	0.050	107	315	2
61	184.20	6.396	0.115	107	315	5
62	172.60	5.993	0.216	107	315	10
63	277.00	9.618	0.035	153	553	1
64	259.60	9.014	0.065	153	553	2
65	229.50	7.969	0.143	153	553	5
66	198.10	6.878	0.248	153	553	10
67	280.10	9.726	0.035	200	900	1
68	273.5	9.497	0.068	200	900	2
69	265.2	9.208	0.166	200	900	5
70	250.1	8.684	0.313	200	900	10

Note:1-3: Artificial zero surface area points. 4-10[1]; 11-18[2]; 19-24[3]; 25-29[4]; 30-38[5]; 39-54[6]; 55-70[7];

References

1. Li, Y.T., et al., Hierarchical porous active carbon from fallen leaves by synergy of K2CO3 and their supercapacitor performance. Journal of Power Sources, 2015. **299**: p. 519-528.
2. Zhang, D.D., et al., Scalable synthesis of hierarchical macropore-rich activated carbon microspheres assembled by carbon nanoparticles for high rate performance supercapacitors. Journal of Power Sources, 2017. **342**: p. 363-370.
3. Zhang, J.T., et al., Preparation of activated carbon from waste Camellia oleifera shell for supercapacitor application. Journal of Solid State Electrochemistry, 2012. **16**(6): p. 2179-2186.

4. Jiang, L., et al., High rate performance activated carbons prepared from ginkgo shells for electrochemical supercapacitors. Carbon, 2013. **56**: p. 146-154.

5. Yang, W., et al., Template-free synthesis of ultrathin porous carbon shell with excellent conductivity for high-rate supercapacitors. Carbon, 2017. **111**: p. 419-427.

6. Jiang, L.L., et al., Construction of nitrogen-doped porous carbon buildings using interconnected ultra-small carbon nanosheets for ultra-high rate supercapacitors. Journal of Materials Chemistry A, 2016. **4**(29): p. 11388-11396.

7. Wu, H., et al., The effect of activation technology on the electrochemical performance of calcium carbide skeleton carbon. Journal of Solid State Electrochemistry, 2012. **16**(9): p. 2941-2947.

8. Kennedy, M.C. and A. O'Hagan, *Bayesian calibration of computer models.* Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2001. **63**(3): p. 425-464.

9. Rasmussen, C.E. and C.K.I. Williams, *Gaussian Processes for Machine Learning*. 2005, Cambridge, Massachusetts: The MIT Press.

10. Hofmann, T., B. Schölkopf, and A.J. Smola, *Kernel methods in machine learning.* The Annals of Statistics, 2008. **36**(3).

11. Pedregosa, F., et al., *Scikit-learn: Machine Learning in Python.* Journal of Machine Learning Research, 2011. **12**: p. 2825-2830.

12. Wu, J. and M. Gu, Emulating the first principles of matter: A probabilistic roadmap. arXiv preprint arXiv:2010.05942, 2020.