D11.2 – Identification of interphase descriptor dynamics for test system

VERSION

VERSION	DATE
1	29/4-2022

PROJECT INFORMATION

GRANT AGREEMENT NUMBER	957189
PROJECT FULL TITLE	Battery Interface Genome - Materials Acceleration Platform
PROJECT ACRONYM	BIG-MAP
START DATE OF THE PROJECT	1/9-2020
DURATION	3 years
CALL IDENTIFIER	H2020-LC-BAT-2020-3
PROJECT WEBSITE	big-map.eu

DELIVERABLE INFORMATION

WP NO.	11
WP LEADER	Arghya Bhowmik (DTU)
CONTRIBUTING PARTNERS	DTU, FZJ
NATURE	Report
AUTHORS	Arghya Bhowmik, Eibar Flores, Jonas Busk, Williams Agyei Appiah
CONTRIBUTORS	Peter Bjørn Jørgensen, Isidora Cekic-Laskovic, Christian Wolke
CONTRACTUAL DEADLINE	M20
DELIVERY DATE TO EC	29/4-2022
DISSEMINATION LEVEL (PU/CO)	PU

ACKNOWLEDGMENT

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957189. The project is part of BATTERY 2030+, the large-scale European research initiative for inventing the sustainable batteries of the future.

ABSTRACT

Identification of the multitude of descriptors encoding for the spatio-temporal evolution of batteries is a prerequisite for the inverse design process. This deliverable uses machine learning (ML) to make the descriptor discovery process semi-autonomous and semi-supervised. Black box machine learning modelling can provide a high accuracy or a quick optimization but does little to progress a widely applicable scientific understanding. The holy grail of ML in science is in the possibility of discovering descriptors, mechanistic insights, and governing equations. High dimensional functions that represent complex systems beyond human analytics can be learned from noisy experimental data or a large number of simulations from such a synergism. With data from WP3 and WP6, the demonstration is done towards autonomous descriptor discovery – both local and global. The infusion of domain knowledge through predefined chemical and physical variables makes the ML framework more data-efficient.

Furthermore, it helps create and disseminate battery design principles at different scales accelerated by keeping the ML framework explainable. We have used two types of ML frameworks for these demonstrators – symbolic regression and Gaussian processes. The demonstrators have helped us identify (a) physics-abiding governing equations for lithium-ion transport that are universal to composition, salt concentration, and temperature, and (b) key physical parameters that control battery capacity fade and SEI growth at the cell level and variability in their contribution across the parameter space.

Contents

1 DESCRIPTORS AND DATA-DRIVEN AUTOMATED DISCOVERY 3

1.1 WHY DESCRIPTORS 3
1.2 DATA DRIVEN AUTOMATED DESCRIPTOR DISCOVERY WITH MACHINE LEARNING 3

2 LEARNING THE GOVERNING EQUATIONS OF ELECTROLYTE TRANSPORT USING SYMBOLIC REGRESSION 5

2.1 BACKGROUND 5
2.2 DATA AND MODEL SETUP 6
2.3 DESCRIPTOR AND SCIENTIFIC KNOWLEDGE FROM SR 11
2.4 CODE AND MODEL AVAILABILITY 13

3 LOCAL AND GLOBAL DESCRIPTORS FOR BATTERY DEGRADATION WITH GAUSSIAN PROCESS 13

3.1 BACKGROUND 14
3.2 DATA AND METHOD SETUP 14
3.3 GLOBAL DESCRIPTORS 17
3.4 LOCAL DESCRIPTORS 18
3.5 CODE AND MODEL AVAILABILITY 21

4 REFERENCES 22

1 Descriptors and data-driven automated discovery

1.1 Why descriptors?

Descriptor search for key properties has been central to chemical and materials sciences as they guide exploration of design space efficiently for target materials properties with limited experimental/computational cost[1]. Simple descriptors can also be seen as input to human-understandable design principles, which are a simple, actionable form of the underlying theoretical paradigm. Even approximate property estimation through descriptors is valuable as a multi-step screening-based[2] design protocol. Both empirical observation-based and physically grounded patterns in structure-property correlation help us design better functional materials by narrowing down the phase space with a high probability of desired activity. By describing observed properties that are typically emergent from a range of different phenomena as a function of one or more easy-to-obtain (experimentally or computationally) parameters of the material called descriptors, we simplify the process of fast low-fidelity screening of materials. For example, researchers choose a few reactivity descriptors in catalysts to correlate surface reaction energies and activation barriers to the catalytic process. In this way, a high-dimensional kinetic model based on the activation and reaction energies of all elementary reaction steps is reduced to a few dimensions described by a series of reactivity descriptors, which can predict trends in the catalytic performance. Such simplifications come from scaling relationships between different reaction intermediate adsorbents[3,4]. These catalyst correlations have a physical basis in the d-band theory. While catalyst design has long benefited from a descriptor-based design paradigm, it is a relatively new concept in the battery domain but with success both with atomistic and macro-scale descriptors. For example, the electronegativity of neighbouring atoms might be a simple descriptor to optimize the ion intercalation potential in battery cathode materials[5], although the electronic structure effects causing the correlation are highly complex and nonlinear. Mobility of ions is a key property for electrode and electrolyte materials, for which descriptors have been discovered[6]. Unlike catalysts, a combination of descriptors is often needed for useful accuracy for diffusive properties of battery materials[7–10]. Electrolyte stability against oxidation can also be estimated via density function simulation derived descriptor-based approach[11]. Experimentally available descriptors have been used for identifying the ion storage process[12]. Simple materials properties like crystal volume change can also be a descriptor to macroscale behavior like battery cycle that arises from multi-time/length scale phenomena[13]. Battery science provides ample opportunity to expedite the design processes with a descriptor-based approach from atomistic to macro scale, with both experiment- and simulation-based descriptors.

1.2 Data driven automated descriptor discovery with machine learning

Identifying the multiple descriptors (or genes) for the spatio-temporal evolution of battery interfaces and interphases is a prerequisite for the inverse design process and exceeds the capabilities of conventional approaches to understand complicated phenomena as the interrelations are complex. Although descriptor discovery primarily occurred through empirical observations and theoretical deductions, descriptors can also be discovered from

statistical/symbolic models or machine learning (ML) models with limited complexity[14,15]. These approaches have become popular in the last few years in the materials science community, specifically for batteries and for data-driven descriptor search[16,17].

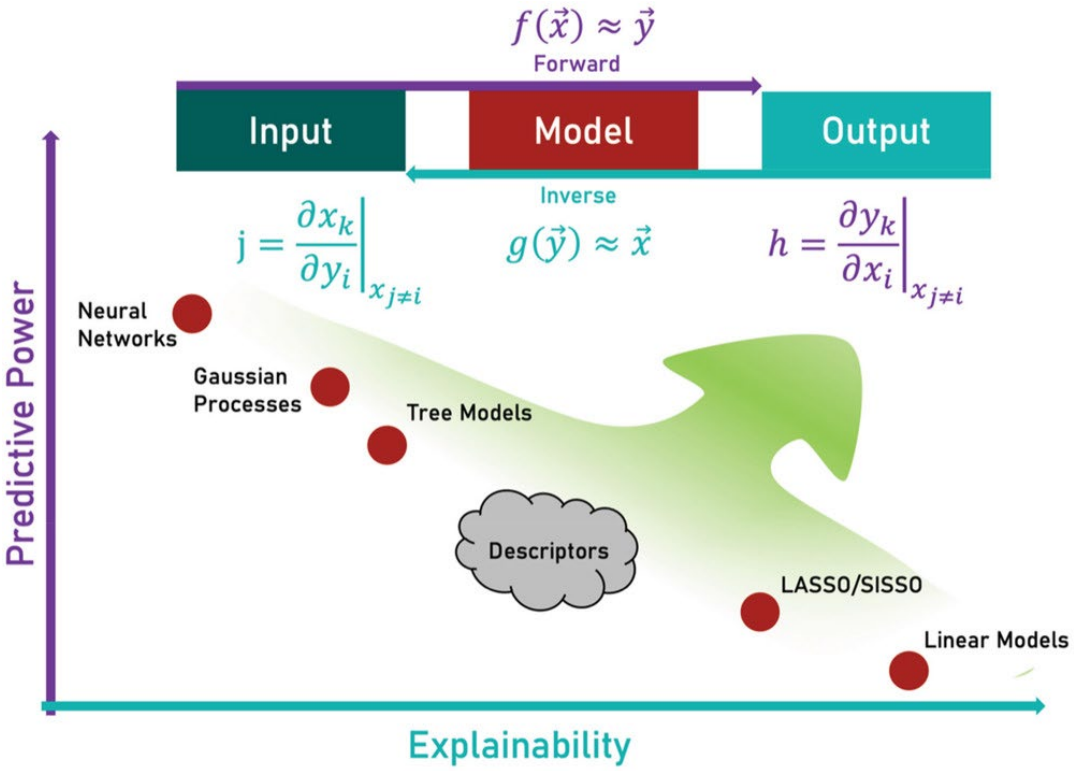

Figure 1. The model complexity and its explainability are typically inversely correlated. AI methods are used to first make a prediction, and then those trained models can be used to inversely predict input parameters determining the descriptors that are most crucial in the target property. Explainable AI methods effectively allow studying the partial (high dimensional and sparse) derivative of the output by the input.

ML models are especially good when the underlying correlations are complex and high-dimensional. We can use deep learning to understand some of the most complex systems like battery interphases, where descriptors might be in high-dimensional space beyond the capacity of manual construction. But the black-box nature of complex ML models stops us from obtaining scientific insights from them. To make ML models physics-driven, we also need to ensure that the predictions from deep models come from the combinations of descriptors appropriate in the context of the underlying physical phenomena. Artificial intelligence (AI) methods for interpreting deep learning models are generally known as 'explainable AI (XAI)'[18]. Using XAI to understand how complex ML models work can provide new scientific insight and design principles. Machine learning models can also autonomously generate scientific hypotheses and complex descriptors without human intervention[19]. With XAI, we can discover complex descriptors[20] going beyond simple linear descriptors. In ML model development, we compromisn accuracy and explainability (Figure 1). The most accurate models (e.g., deep neural networks) are the least explainable models, and the most explainable models (e.g., linear regression) have low accuracy. As the understanding of complex multi-scale inter-correlated phenomena at the battery interface and other such systems is limited,

comprehensible models with reasonable accuracy have been adopted[21,22]. However, there is a clear trend in deep learning models that are quickly gaining popularity in this application space, opening up the possibility of applying XAI methods to make those models interpretable. The goal of the deep learning XAI models is to make deep models intelligible to humans by explaining how the output is formed from what kind of salient features are in the input variable. Explainable AI is also "reliable AI," as the predictions come from the correct reasoning. An alternative way to interpret and explain black-box ML models is to design inherently interpretable models without significant loss of accuracy. As ML models are increasingly utilized as a key tool in data-driven materials modelling in battery materials and interfaces, XAI should be used for the data-driven realization of chemical laws and design principles. Structure-property trends and descriptors derived from data-driven methods need to have physical reasoning and be verified.

The type of XAI models used for Task 11.2 in WP11 ranges from (a) symbolic regression, (b) Gaussian processes, (c) graph message passing neural network, (d) variational auto-encoders to (e) long-short term memory recurrent neural networks. The models were deployed with both experimental and computational data at short and long time/length scales. As most of the work is recent, unpublished, and confidential while this report is public, we have chosen to demonstrate it here

1. Model (a) deployment with high throughput experimental dataset from WP6 and
2. Model (b) deployed with simulated battery degradation data from the WP3 dataset.

2 Learning the governing equations of electrolyte transport using symbolic regression

For better-performing Li-ion batteries, electrolyte conductivity is critical. A widely applicable model describing ionic transport on practical electrolyte formulations is highly desired. We apply symbolic regression to find a suitable surrogate model of the conductivity of a LiPF6-based electrolyte, using a large experimental dataset from high-throughput conductivity measurements (WP6). We demonstrate the emergence of an expression outstanding for being straightforward, accurate, consistent, and generalizable. Although discovered from a purely statistical approach, the expression inherits functional aspects from established thermodynamic limiting laws, indicating it to be grounded on the physical mechanisms underpinning electrolyte transport.

2.1 Background

Non-aqueous solutions are state-of-the-art electrolytes for Li-ion batteries (LIBs) as they better comply with the strict requirements to operate safely, long-lasting, and highly performing Li-ion cells. These electrolytes consist of a Li salt dissolved in a mixture of organic solvents and complemented with performance-enhancing additives. Electrolyte formulations balance multiple and often competing properties, among which the ionic conductivity is arguably the most important.[23–26] The choice of solvent, salts, and their proportion usually aims at achieving fluid and dielectric electrolytes capable of keeping ions dissociated.[23,27,28] However, the conductivity is not the only property to tailor: the electrochemical stability window, chemical compatibility with electrodes, thermal stability, costs, and toxicity, are all additional factors to consider.[29,30] In this multi-objective optimization scenario, researchers in the field would greatly benefit from a

predictive, thermodynamic model for electrolyte conduction, enabling them to quickly explore how a promising formulation would affect the ionic conductivity of electrolytes without needing additional experiments. Such a model would be ideally denoted as a universal and straightforward closed-form expression, i.e., an equation with a few algebraic terms, relating easily measurable variables with fundamental physical constants, and without fitting parameters.

Despite significant progress in thermodynamic descriptions of electrolyte transport,[31] such a utopic model only exists for highly dilute electrolytes. At infinite dilution, an ion's conductivity is directly proportional to its concentration in solution *c.*[32,33] However; this model fails at the dilute domain ($0 < c < 10^{-3}$ mol/l) since the conductivity depends additionally on a squared root term of the salt concentration.[34] Kohlraush formulated these findings into an empirical law with an adjustable parameter,[34,35] later addressed by Onsager by considering that ions are dragged not only by hydrodynamic effects but also by electrophoretic and relaxation phenomena as in the Debye-Huckel theory. The Debye-Huckel-Onsager (DHO) theory effectively upgrades Kohlraush's law into a fully theoretical law without adjustable parameters:[36]

$$\kappa = \kappa^0 - \left(\frac{z^2 eF}{3\pi}\right)\left(\frac{2}{\varepsilon RT}\right)^{1/2}\left(\frac{F}{\eta} + \frac{qz}{8eRT}\right)c^{3/2} \qquad \text{Eqn. 1}$$

Despite the success of the DHO law on strong electrolytes, it fails at describing the concentrated ($c > 1$ mol/l) and weak electrolyte formulations used in Li-ion batteries. In its place, researchers formulate expressions following two main approaches. Semi-empirical approaches extend non-electrolyte thermodynamic theories by including long-ranged ion-ion interactions from the DHO theory.[37,38] Instead, phenomenological approaches assume the conductivity to depend on salt concentration and temperature via an arbitrarily-chosen functional expression (e.g., polynomial, exponential), expanded to enough terms to reach a good fit of experimental data;[39–41] While these models might fit the data well, they are ill-posed to generalize and provide physical insight, given the arbitrary choice of functional expression and all the parameters that need to be adjusted for every new system.

In this work, we propose an alternative approach – Symbolic Regression[15] (SR) – to gain insights into the laws of electrolyte transport on non-aqueous electrolytes. In essence, SR simultaneously learns both adjustable parameters and the functional form relating to electrolyte conductivity. While the usage of different ML models is gaining popularity in battery research[42–44], SR has not been explored. In the broader energy materials area, SR has been proven to be capable of identifying suitable descriptor catalyst activities from experimental data, promoting accelerated discovery[45].

2.2 Data and model setup

We make use of a high-throughput setup (Figure 2) to collect thousands of conductivity measurements of a $LiPF_6$-based electrolyte system at different temperatures. More details about the data acquisition are present in deliverable D6.3. With over 3000 experimental data points collected this way, and using a simple SR approach (Figure 3), we train multiple candidate expressions and show that a particular expression emerges as a clear candidate for complying with multiple strict and often competing criteria: i) prediction accuracy, ii) expression simplicity and iii) model consistency.

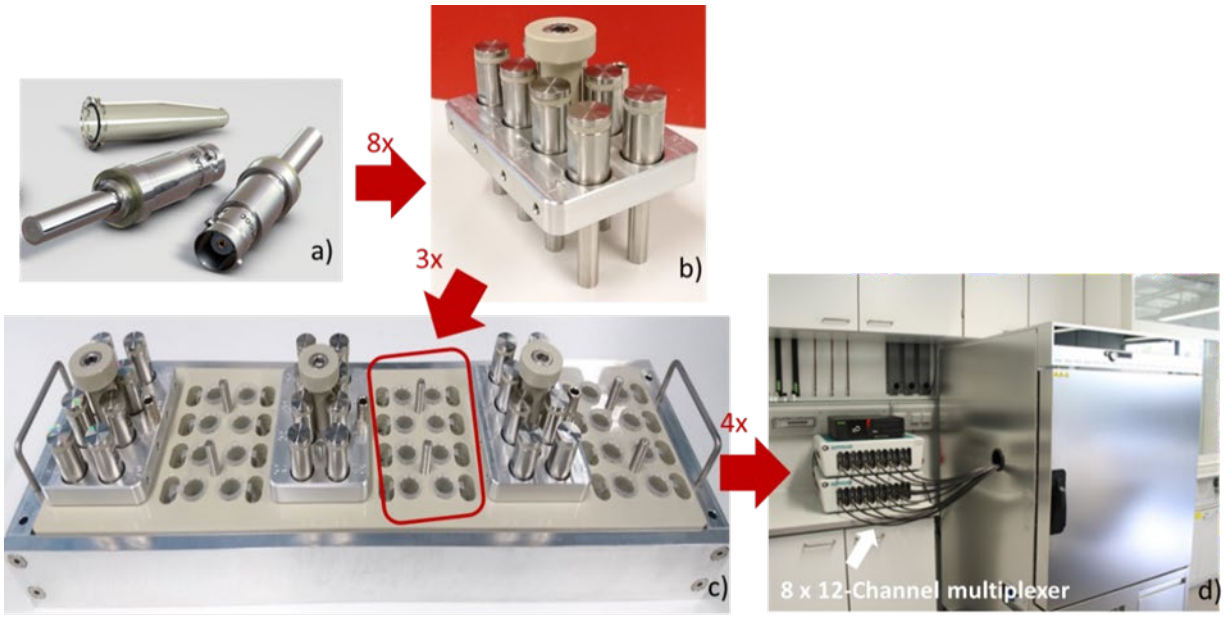

Figure 2. High-throughput conductivity module comprising: a) in-house developed impedance electrode and Eppendorf tube, b) small rack containing 8 electrodes, c) big rack with 24 conductivity cells, and d) potentiostat/galvanostat with an 8x12-channel multiplexer and temperature chamber.

Our SR approach approximates the conductivity with a generalized linear model, not on the original predictors, but on a non-linear transformation of these (i.e., candidate features). Formally:

$$\kappa \approx \sum_k \beta_k \boldsymbol{\Theta_k}(T, c, r) \qquad \text{Eqn.2}$$

Where β_k is the k-th coefficient and $\boldsymbol{\Theta_k}$ the k-th transformation on the predictors: temperature T, salt concentration c, and PC:EC molar ratio r. The conductivity is assumed to depend not on all possible candidate features but a much-reduced set of these; i.e., the solution of Eqn. 2 is sparse. Figure 3 illustrates the methodology, split into feature generation and selection steps. The training process involves defining a transformation set (e.g., inverse, logarithms, exponentials), then applying it to the initial predictors to generate a library of candidate features, a few of which are selected to form a candidate expression.

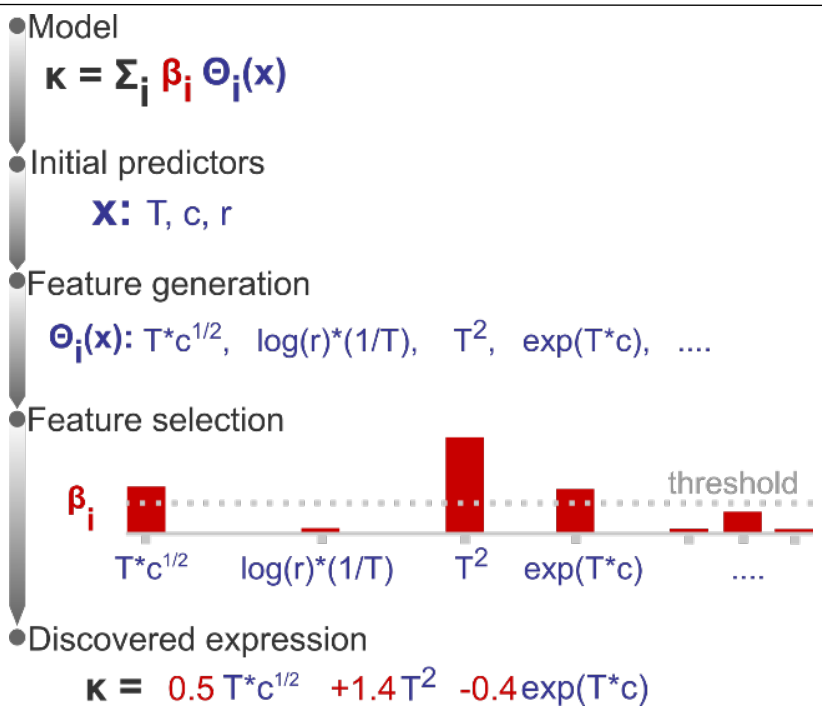

Figure 3. Representation of the symbolic regression method

The discovered expressions are not unique: candidate features might combine in multiple ways to result in similarly accurate expressions. Consequently, instead of using all 515 samples on the training set, we train instead of on numerous, random subsamples of 50, 100, 250, and 400 data points to evaluate whether a discovered expression is consistent across 20 independent training sessions. We use the validation set to evaluate the performance of the discovered expressions. The first benchmark uses the three initial predictors as the simplest approximation. The second uses 3rd-order polynomial expansions of the initial predictors, as in phenomenological models.[46] The third assumes exponential transformations as in Arrhenius-based models. The fourth uses exponential transformations on the 3rd-order polynomial expansion of the initial predictors, such as in the extended Castel-Amis model.[28]
During the evaluation, we search for an expression being not only i) accurate, i.e., yielding a low mean squared error (MSE), but also ii) parsimonious, quantified as the number of terms of the expression, and iii) stable, represented by the number of times the expression repeats across training sessions.

Figure 4a presents the accuracy vs. complexity trade-off from the expressions found. Each data point represents an expression whose color references its parent transformation set. As expected, more significant expressions fit the data better but at the expense of increased complexity; this is the case of the expressions originating from exponential and logarithmic transformations (MSE < 2 but 10+ terms). Interestingly, the expressions populating the Pareto-frontier of the figure originate from transformation sets, including square-root operations; i.e., they offer the best compromise between MSE and the number of terms.
Note that most expressions only appear once, highlighting these to be highly sensitive to the training subsample and that there is no unique solution. Figure 4b shows the most frequent expressions

across the training sessions, where expressions with square-root operations are highlighted in green. Unlike most expressions, the model:

$$\kappa = \beta_1 c + \beta_2 T + \beta_3 c^{5/2} r^{1/2} + \beta_4 c^{1/4} T^{5/2} \quad \text{Eqn. 3}$$

is by far the most frequent and was discovered 15 times out of 20 training sessions. We, therefore, select this expression as it stands out from the other competing models for being not only consistent but also parsimonious (four terms), comparatively accurate in the training set (MSE < 0.75), and generalizable, as evidenced by a good accuracy in the validation set. Table 1 summarizes the coefficients and performance metrics of our selected expression Eqn. 3.

Table 1. Coefficients of Eqn. 5 and associated performance metrics after training on the full training set of 515 samples.

Attribute	Value
β_1	- 5.11
β_2	- 0.040
β_3	-0.35
β_4	2.73×10^{-4}
Training MSE	1.08
Training R^2	0.92
Validation MSE	1.22
Validation R^2	0.90

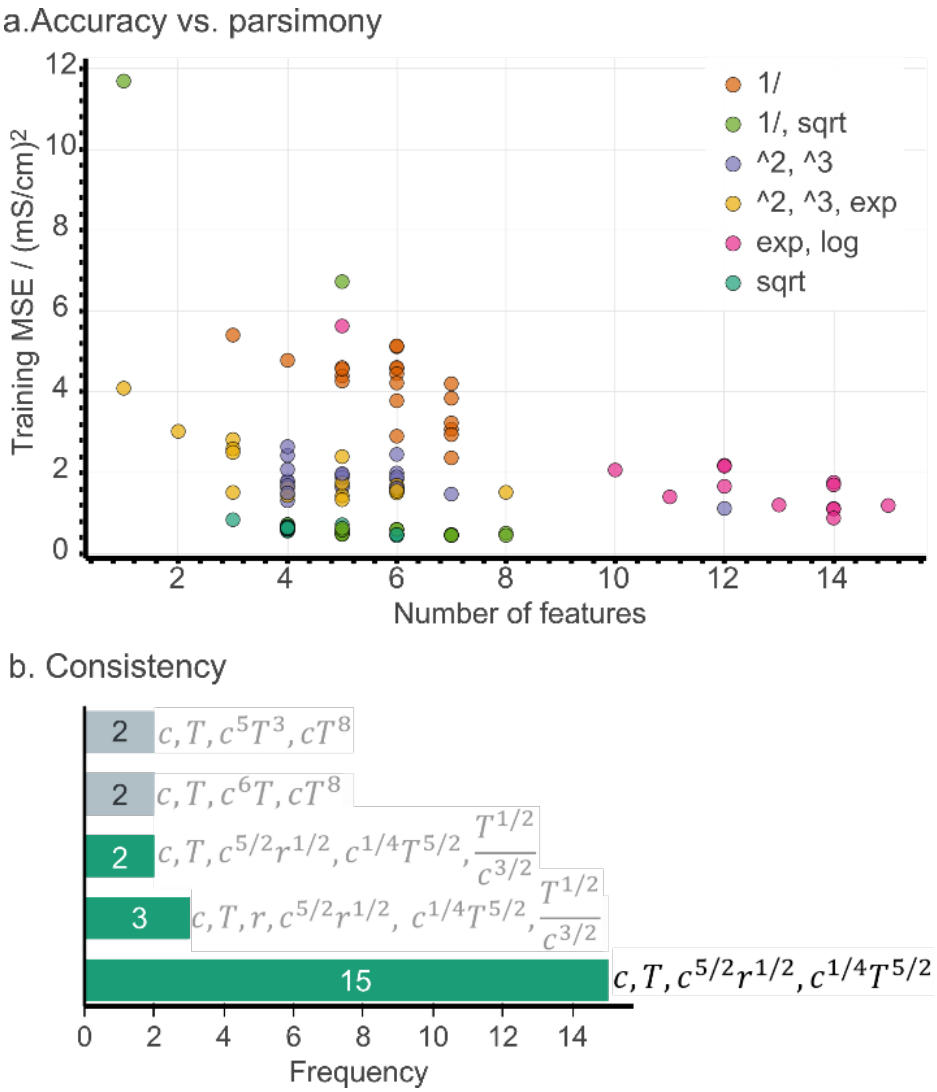

Figure 4. Accuracy, parsimony, and consistency of discovered expressions throughout multiple training sessions. All expressions were trained with the constrain $k_0 = 0$.

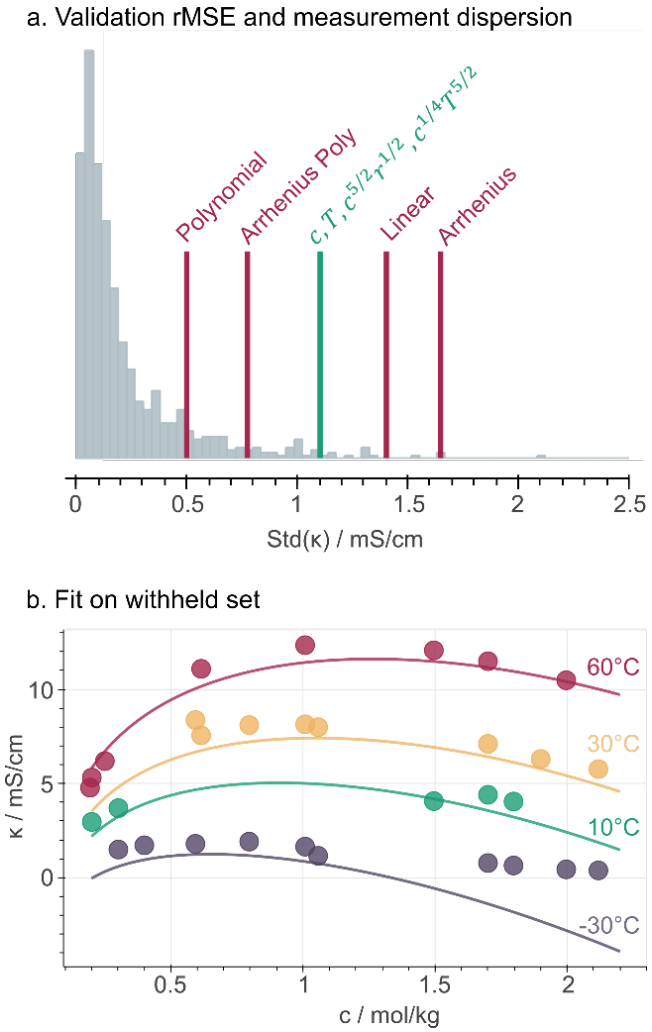

Figure 5.a) Root mean square error of selected model (green) and benchmarks (red) on the validation set, compared to measurement dispersion (grey). b) Fit of the selected model on the withheld (validation and test) set at r=1.0.

2.3 Descriptor and scientific knowledge from SR

Constraining models might be beneficial when selecting promising surrogate models. To illustrate why, we repeated the 20 training sessions with the same transformation sets, this time allowing the intercept to vary freely. Expectedly, removing the intercept constrain results in slightly improved accuracy but significantly worsens model stability. Implementing domain-knowledge constraints is an effective filter for discovering consistent and physically sound expressions. However, constrained models have substantially higher data requirements. The non-constrained expression converges to optimal accuracy already with 100 samples; in contrast, the constrained model fails at almost all sample sizes and only approaches optimal accuracy when using all 515 training samples. Such data requirements need to be weighed when modelling the often-small datasets available from experiments.

Figure 5a compares the accuracy of the selected constrained expressions on the validation set relative to the measurement dispersion and benchmark models. We use the root mean squared error (rMSE) to describe the prediction accuracy in the same mS/cm units as the conductivity measurements. As expected, the simpler benchmarks such as linear and straightforward Arrhenius models are less accurate. Instead, the more complex models (polynomial and Arrhenius polynomial)

are prone to overfitting, as their prediction errors are smaller than a non-negligible fraction of measurement dispersion values. Notably, the selected model stands in the middle with a validation-set rMSE of 1.1 mS/cm, indicating that it is accurate up to the measurement noise and so it does not overfit the dataset. At first glance, the selected model seems to yield only a minor improvement (0.3 mS/cm) compared to the basic linear model; however, the square-root dependence in the selected model reproduces the curvature and maxima in the data, and ii) by having no intercept, it complies with the physical constraint of no conductivity at c, T, r = 0.

Figure 5b illustrates that the selected model generally fits the data not used in training (i.e., validation and testing sets). However, the fit generally underestimates the measurements. The same expression trained with an intercept fits the withheld data without such bias, indicating that the underestimation in Figure 5b is a result of imposing the physically-motivated $y_0 = 0$ constrain. In addition, the model is not flexible enough to describe the conductivities measured at -10 °C and concentrations above 1 mol/kg; in these extreme regimes, the high salt concentration and low temperature result in a highly viscous medium that might promote a non-vehicular type of ionic transport.[47] However, we highlight that in most of the experimental range, the fit reproduces the concentration- and temperature-dependent conductivity maxima observed in the data and in previous studies, which is a key attribute for implementing the discovered model as part of multi-target optimization and/or active learning frameworks.[48]

Assigning a physical meaning to the discovered expression is not straightforward. For one, any comparison to the thermodynamically-derived DHO law would require explicitly account for the solution's viscosity and electric permittivity, measurements that are not available in the dataset. Second, there are no constraints to avoid unphysical values, like the negative conductivities at sub-zero temperatures and high salt concentrations (Figure 5b). Third, the solution to our symbolic regression approach is generally not unique, i.e., there are multiple expressions equally accurate to fit the dataset. Despite these limitations, we observe that the best compromise between simplicity and accuracy is achieved by expressions sharing square-root transformations. We, therefore, believe that our method is learning square-root trends inherent to the data manifold, which indicates that some functional aspects of the DHO law, i.e., its square-root trends on temperature and concentration, are still valid to describe electrolyte conductivity in concentrated solutions.

Physical insights can be drawn not only from the expression itself but also from its predictions. Higher temperatures increase the conductivity and shift the conductivity maxima towards higher salt concentrations (0.74 mol/kg at -30 °C to 1.70 mol/kg at 60 °C), as expected (see Figure 5b). However, the role of the cyclic carbonate is more subtle. Note first that all conductivities peak when the electrolyte formulation is EC-pure (PC:EC ratio = 0). Second, the tails along the salt concentration axis elongate at higher concentrations as the formulations become increasingly EC-pure. From a fundamental standpoint, conductivity depends on a compromise between the ionic mobility, mainly influenced by viscosity, and the number of charge carriers available for migration, mainly controlled by the electrolyte's dielectric constant (c.f. see Bjerrums criterion[1] for ionic association).[27,28] As EC has a higher dielectric constant compared to PC,[49] EC-pure solutions are more effective at preventing ion association and so improve electrolyte conductivity. This effect is especially pronounced at high salt concentrations, where ionic association becomes a critical limiting factor for electrolyte transport.[50,51] Such EC-driven improvement of conductivity, observed experimentally,[52] is not evident neither in the pair-plots nor in the correlation maps. Despite this, our SR approach manages to capture these subtle effects that align with our current understanding of the interplay between the solvent's dielectric properties and ionic migration.

At this point, we emphasize we have only implemented two domain-knowledge decisions – including square-root transformations and constraining the intercept to zero – on an otherwise purely statistical approach. Even in these circumstances, we observe the emergence of an expression outstanding from competing models for being accurate without overfitting, simple, consistent, with a square-root functional structure resembling the DHO law, and agreeing with our understanding of ionic migration. In other words, our expression is not only an appropriate surrogate model from a machine learning standpoint but also seems grounded on the physical-chemical mechanisms underpinning electrolyte transport. Our work opens multiple avenues to further pursuing the data-driven discovery of accurate surrogate models capable of bridging the existing gap[38] in understanding concentrated electrolyte solutions. To start with, atomistic descriptors can be incorporated to generalize to solvent mixtures other than PC/EC/EMC and salt chemistries beyond conventional Li-ion technology.[40] In addition, using other promising SR algorithms[53] and implementing domain-knowledge constraints in the feature selection step[54] could alleviate the issue with expression consistency and yield physically-sound expressions; i.e., rigorous to known boundary conditions (e.g., $\boldsymbol{\kappa}(\boldsymbol{c}=\mathbf{0})=\mathbf{0}$) and asymptotic behaviour on key limits (e.g., $\lim_{\boldsymbol{c}\to\mathbf{0}} \boldsymbol{\kappa} \propto \boldsymbol{c}$). These constraints will have to be carefully balanced, given our observations of the data-intensive nature of constrained models.

2.4 Code and model availability

Real-time prediction (with visual user interface) of our discovered expression (https://big-map.github.io/SR-electrolytes/). The code and data are available at https://github.com/BIG-MAP/SR-electrolytes

3 Local and global descriptors for battery degradation with Gaussian process

Accurate battery models act as a digital twin of batteries and help in expeditious design optimization as well as help us explore the interplay of electrochemical phenomena. Here we present a descriptor analysis of a pseudo-two-dimensional battery model coupled with a capacity fade model based on solid electrolyte interphase formation and the corresponding irreversible charge loss. The proposed method is based on training an inexpensive differentiable surrogate Gaussian process regression model on observed input-output pairs and analyzing the surrogate model to learn about the key parameters that define the time trajectory. The proposed method identifies the most sensitive input parameters globally, and the same method can also be used to explore local sensitivities around specific sets of inputs providing insights into governing electrochemical processes. The automated framework discovers a strong correlation between the growth of the solid electrolyte interphase and the irreversible charge loss, especially at low current rates. Analyzing descriptors to which the

model is most sensitive locally can help optimize batteries at a system level. Input sensitivity analysis is a well-utilized XAI method for creating understanding from large-scale models.

3.1 Background

Sensitivity analysis (SA), in general, studies how variation in the inputs of a given system affects its outputs and can be done at different levels of complexity and sampling cost ranging from inexpensive screening for irrelevant inputs to a detailed exploration of the effect of inputs along with their entire range of variation[55]. Local SA methods focus on how a single set of input parameters affects the output; for example, the partial derivative of the model at a specific point can be estimated by slightly varying the input and observing the output (the one-at-a-time method). In contrast, global SA methods aim to understand the behavior of the overall system, or at least within a specified range of values for each input parameter. Global SA can be achieved by fitting a surrogate model on observed pairs of inputs and outputs and then analyzing the surrogate model in place of the original system, which is especially useful when the surrogate model is cheaper to evaluate and more straightforward to analyze than the original system. An accurate surrogate model can subsequently be used for detailed local SA in specific input space regions and suggest new design parameters. In this work, we present an SA of a pseudo-two-dimensional (P2D) battery model coupled with a capacity fade model based on the formation of a solid electrolyte interphase and the corresponding irreversible charge loss (ICL) for Li-ion batteries. The aim is to apply SA using a surrogate model to identify the global descriptors causing SEI formation and related ICL and to explore local sensitivities and interactions between descriptors in critical parts of the input space.

Gaussian processes are our choice of surrogate model in SA because it is highly flexible, provides uncertainty estimates, and can be applied even with small datasets. Meanwhile, approximate, sparse implementations enable an analysis of large datasets[56,57]. Automatic relevance determination (ARD) of GPs with isotropic kernels can be used to evaluate feature importance. Still, recent work has shown how ARD overestimates the importance of features with nonlinear effects on the output. At the same time, SA is based on derivatives of the GP output with regards to the inputs is robust to a mix of linear, nonlinear, and noise features[58]. Measuring sensitivity by derivatives makes intuitive sense since changing an input parameter is expected to change the output significantly; it is important to consider that input parameter carefully. SA methods based on analyzing derivatives of GP regression models have previously been applied for analyzing neuroimages[59] global ocean chlorophyll prediction[60] showing promising results.

3.2 Data and method setup

The P2D model for the formation of the solid electrolyte interphase (SEI) used in this study is based on the reduction of the solvent near the surface of the negative electrode during charging. The electrolyte considered in the model is a mixture of ethyl carbonate/ethyl methyl carbonate (EC/EMC) with $LiPF_6$ salt. Hence, we assumed that the main product forming the SEI layer is Li_2CO_3. During the discharge of the Li-ion half-cell, a constant negative current is applied to oxidize Li-ions at the Li metal and intercalate them into the graphite. This results in a gradual decrease in the potential of the cell until when almost all the available intercalating sites have been occupied (at a state of charge of 1). At this point, the cell reaches the minimum cut-off potential of 0 V, and then a constant positive current is applied to de-intercalate the Li ions from the graphite and deposit them in the Li metal. The de-intercalation or reduction process proceeds until a cut-off potential of

1.5 V. At this potential; all the Li ions are assumed to be de-intercalated and reduced, resulting in an equal intercalation and de-intercalation time. However, the intercalation time is longer than that of the de-intercalation, indicating that not all the intercalated Li ions were extracted. The trapped Li-ions are consumed to form the solid electrolyte interphase (SEI). The simulation was performed with the parameters presented in Figure 6. The SEI product's formation increases the SEI layer's thickness on the graphite anode, as presented in Figure 7. The SEI layer was assumed to be homogeneous and only grows during the intercalation process; thus, its thickness was constant during the de-intercalation process. Owing to the formation of the SEI, the charge capacity (capacity obtained during the intercalation process) was higher than that of the discharge capacity (capacity obtained during the de-intercalation process). The difference between the charge and discharge capacity is the irreversible charge loss (ICL) and is mainly attributed to the formation of the SEI layer. The SEI thickness and the ICL obtained from the simulation are comparable to what has been obtained experimentally[61,62]. There are several factors that affect the formation of the SEI at the surface of the graphite electrode during charging owing to the competition between the numerous reduction processes. The rates of the electrochemical reaction for the (de-)lithiation process and the SEI formation are greatly influenced by both the intrinsic properties of the reactants, such as the exchange current densities, reductive potential, and reduction activation energy, and the operating condition parameters such as the temperature, concentration of the electrolyte and the reduction current rate[63].

d	Parameter	Description	Minimum	Maximum	Nominal	Unit
1	i_app	Applied current	0.13	6.5	1.3	A
2	rp_neg	Particle radius	1×10^{-6}	1.1×10^{-5}	5.5×10^{-6}	µm
3	Eeq_side	Equilibrium potential (SEI)	0	0.4	0.4	V
4	kappa_film	SEI film conductivity	1×10^{-6}	2.4×10^{-4}	2.4×10^{-4}	S/m
5	epsl_neg	Porosity of anode	0.23	0.4	0.3	-
6	Dl_elect	Electrolyte diffusion coefficient	1.5×10^{-10}	7.5×10^{-10}	3.75×10^{-10}	m2/s
7	Ds_neg	Diffusion coefficient of Li in Anode	1×10^{-14}	1×10^{-13}	3.6×10^{-14}	m2/s
8	i0ref_neg	Anode exchange current density	0.36	3.6	0.96	A/m2
9	E_min	Minimum cut-off voltage	0.0	0.1	0.05	V
10	i0_SEI	SEI exchange current density	8×10^{-8}	1.5×10^{-6}	4.5×10^{-7}	A/m2
11	csmax_neg	Maximum Li ion concentration in anode	2.9×10^{4}	3.3×10^{4}	3.15×10^{4}	mol/m3
12	cl_0	Initial electrolyte concentration	1000	1200	1150	mol/m3
13	t_plus	Transference number	0.25	0.43	0.363	-
14	i0ref_metal	Li metal exchange current density	50	100	100	A/m2
15	sigma_neg	Anode conductivity	50	100	100	S/m

Figure 6. List of P2D-SEI model input parameters for sensitivity analysis

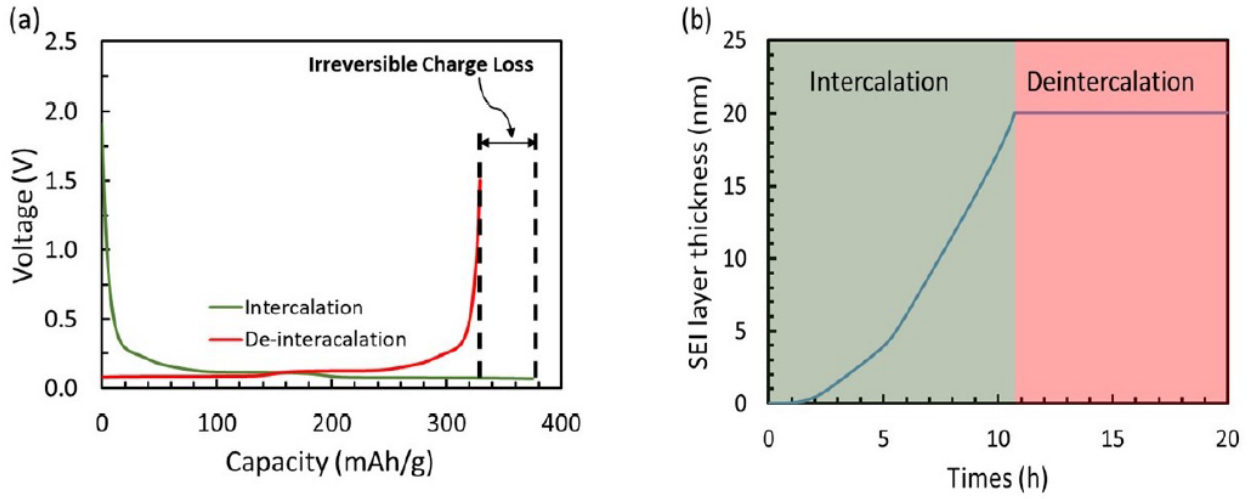

Figure 7. a) Simulated voltage profile for the interaction and de-intercalation process of a Li-ion half-cell with a graphite anode and Li metal and b) related SEI layer thickness.

A dataset of input-output pairs was generated by sampling 30,000 sets of input parameters uniformly at random within the input ranges defined in Figure 6. The corresponding output values were computed by querying the P2D SEI model resulting in a labelled dataset of 23,074 examples since not all input configurations converged. The labelled dataset was then split into a training dataset of 20,000 examples and a validation dataset of the remaining reserved for model evaluation. Since both of the target outputs consist of non-negative values by definition, we transform them by the base 10 logarithm (log) to put them on an unbounded scale and thus simplify the modelling and avoid unphysical, negative predictions. To enable direct comparison of the results for each input parameter and to further simplify modelling, all inputs were normalized to values between zero and one using the minimum and maximum values, and the outputs were normalized to zero mean and unit variance using statistics derived from the training data before training the surrogate model.
As the surrogate model, we apply a sparse GP regression model from the Pyro probabilistic programming library for Python (Python 3.9.6, Torch 1.9.0, Pyro 1.8.0). We also applied a Bayesian linear regression model as a baseline model, likewise implemented using Pyro. Both models were trained with stochastic variational inference (SVI) using the ADAM optimizer with a learning rate of 0.01. The Bayesian linear regression baseline achieved a good fit on the log SEI thickness with an R^2 = 0.96 indicating a robust linear relationship between the input parameters and this output. However, the Bayesian linear regression model achieved a lower R^2 = 0.70 on the log ICL. The GP regression model achieved R^2 = 0.99 on the log SEI thickness and an R^2 = 0.98 on the log ICL, indicating a good fit of the data in both cases. The distribution of the GP regression model predictions on the validation dataset is presented in Figure 8. The predictive performance of the surrogate GP regression model indicates that it is able to capture the dynamics of the original system, which enables us to analyse it in place of the P2D-SEI model in the subsequent SA.

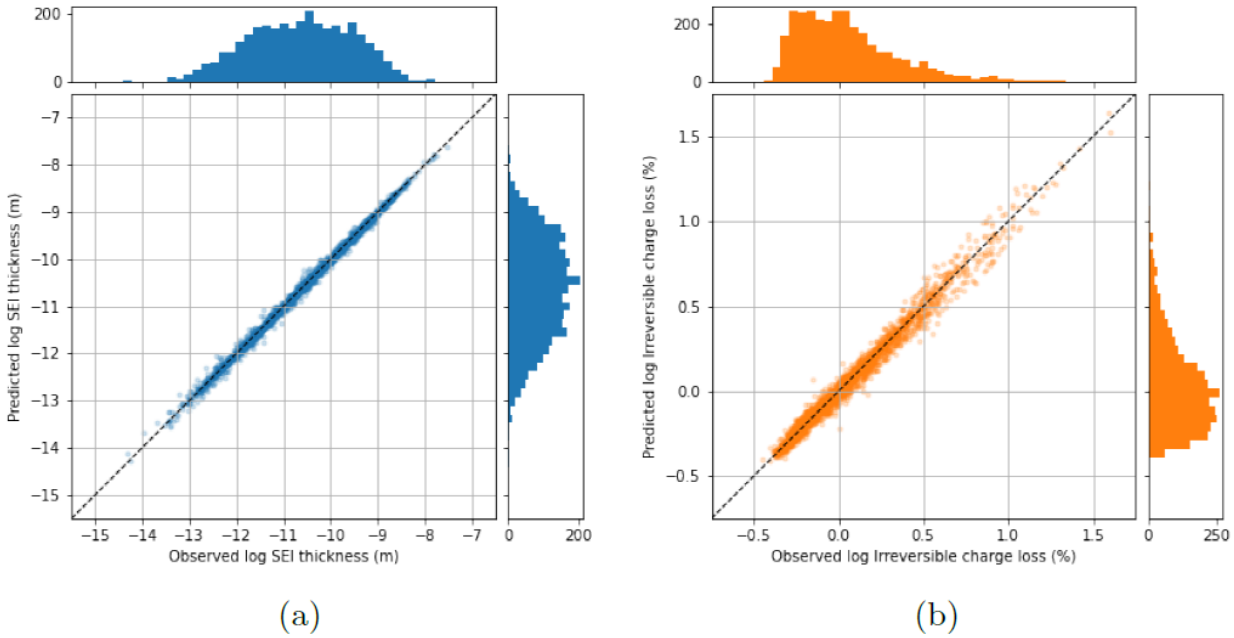

Figure 8. Scatter plots of observed and predicted log SEI thickness (a) and log ICL (b) on the validation dataset using the GP regression surrogate model. In both cases, the observed and predicted values are well correlated.

3.3 Global descriptors

To quantify and compare the sensitivity of each input parameter, the global sensitivities with respect to log SEI thickness and log ICL were computed using the surrogate GP regression model on the validation dataset, and the results are presented in Figure 9. Based on these results, the applied current density, the equilibrium potential, the minimum cut-off potential, and the exchange current density were identified as the most sensitive input parameters to the SEI thickness. The other 11 parameters showed little or negligible sensitivity to the log SEI thickness. Out of these four parameters, two are directly related to the SEI formation rate (SEI model-based input parameters), the equilibrium potential, and the exchange current density of SEI. Based on the Tafel equation used to describe the rate of the SEI formation, a variation in the exchange current density and the equilibrium potential of the SEI causes a linear and exponential variation in the SEI thickness, respectively, and hence the different degree of sensitivity. The increase in the SEI thickness due to changes in the exchange current density can be related to the overcharge of graphite electrodes for cells with excess cyclable lithium owing to either higher than desired initial mass ratio or lower than expected lithium loss during the formation period[64]. While the other two parameters are related to the operating conditions of the cell: the applied current density and the minimum cut-off potential. An increase in the applied current density exponentially accelerates the increase in the thickness of the SEI owing to an increase in the rate of SEI formation while the minimum cut of potential dictates the amount of lithium inserted into the graphite at a given current density and thus the quantity of lithium consumed to form the SEI. The global SA on the log ICL identified six input parameters with high average sensitivity and three additional input parameters showing significant effects. The six parameters were the applied current density, the particle radius of the graphite particles, the equilibrium potential for the SEI growth, the diffusion coefficient for lithium in the solid phase, the exchange current density for the (de-)intercalation of lithium in the graphite and the minimum cut-off potential. Interestingly, four of the most sensitive input parameters for the SEI thickness were a subset of those for the ICL. This corresponds well with the observation that

there is a strong correlation between the changes in the SEI thickness and the ICL. Sensitive input parameters such as the radius of the graphite particles, the porosity of the electrode, and exchange current density for lithium (de-)intercalation into the graphite are related to the Brunauer-Emmett-Teller (BET) specific surface area of the graphite anode, the solvated lithium intercalation, and the nature of the surface sites for electrolyte decomposition. Based on previous experimental reports, the BET-specific surface area and the solvated lithium intercalation contribute enormously to the ICL[65]. The nature of the graphite surface also plays a significant role in the extent of the ICL[66].

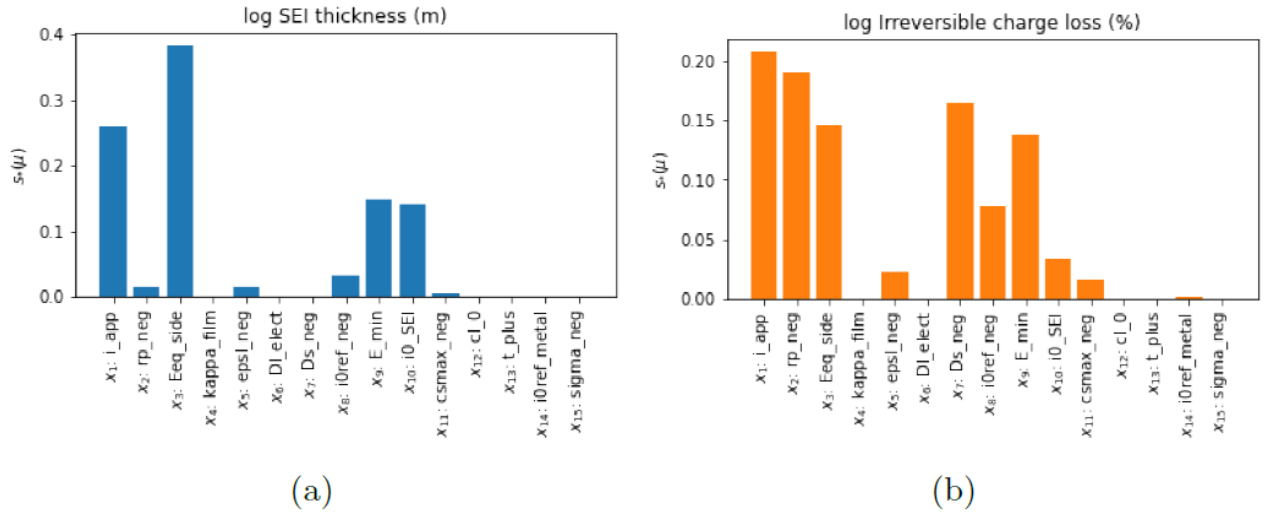

Figure 9. Normalized global sensitivities were computed on the validation dataset for the log SEI thickness (a) and log ICL (b).

3.4 Local descriptors

Figure 10 illustrates an example of the local SA of some of the sensitive input parameters identified in the global SA in the previous section, along with their entire range of variation, while keeping all other inputs fixed at their nominal values. The applied current density exhibits a nonlinear relationship with the SEI's growth, which is highly sensitive at low current densities. In principle, at low current densities, large amounts of Li ions are consumed to form the SEI due to the amount of time required to reach a given cut-off potential and the slow kinetics at the surface of the electrode. However, at high current densities, the kinetic reaction is swift, and the time required to reach the cut-off potential is less; thus, the growth of the SEI, even though it is at a fast rate, does not change significantly as observed in Figure 10a. As expected, the equilibrium potential for the SEI formation showed a strong linear relationship with the growth of the SEI with a constant sensitivity along with its range of input parameters. However, the sensitivity of the equilibrium potential for SEI formation was constant at lower values of the range of input parameters and increased at values between 0.25 and 0.45. This indicates that even though there is a linear increase in the SEI at equilibrium potential for the SEI values below 0.25, the amount of SEI formed is not sensitive to the ICL, and hence the other degradation mechanisms such as loss of active materials due to the variation in the volume fraction of the active material are the most sensitive mechanism. Figure 11 shows the interaction of the sensitive inputs along with their entire range of variation while keeping all other inputs fixed at their nominal values. The physical meaning of the variation of some of the relevant sensitive input parameters has been presented in Figure 11. The sensitivity of the amount of Li ions inserted into

the graphite during charging at a wide range of applied current density on the half-cell performance was studied by conducting simulations at a different end of charge voltage (EOCV) (Figure 11a). In principle, the charge and discharge capacity of the cell is expected to decrease with an increase in EOCV because the cell's capacity is directly proportional to the number of available Li-ions inserted into the graphite. However, the proportionality also depends on the amount of Li ions consumed to form the SEI layer. As presented in Figure 11, increasing the EOCV decreased both the SEI layer thickness and the ICL, respectively, at low current densities. This indicates that increasing the state of charge (SOC) of the graphite by decreasing the EOCV increases the amount of Li ions used in the formation of the SEI, resulting in an increase in the ICL. Nevertheless, an increase in the SOC increases the discharge capacity, and thus an optimum EOCV needs to be obtained to ensure a minimum increase in the resistance due to the increase in the thickness of the SEI layer and maximum achievable discharge capacity. Figure 11b presents the ICL as a function of the graphite particle size at various current densities and the corresponding SA. The particle size is inversely proportional to the BET surface area. Previous studies[65,66] indicate that an increase in the BET surface area results in an increase in the ICL due to high electrolyte decomposition and excessive lithium loss to the formation of the SEI at low current densities. This is in line with the results presented in Figure 11b where an increase in the particle size resulted in a decrease in the ICL and a linear decrease in the SEI thickness, respectively. However, this trend changed as the current density increased. High ICL was observed at high current densities for both the EOCV and the particle size and similar for all the other variable input parameters even though the SEI thickness decreased as expected. This can be attributed to the fact that the main factor causing the capacity loss at high current densities is the overpotential due to the ohmic resistance. In addition to the effect of the changes of EOCV on the ICL and SEI thickness, we also observed interesting phenomena when other input parameters such as the exchange current density for the (de-)lithiation and SEI formation and the diffusion coefficient of Li in the graphite creates local extrema, i.e., optimal design choices can be made.

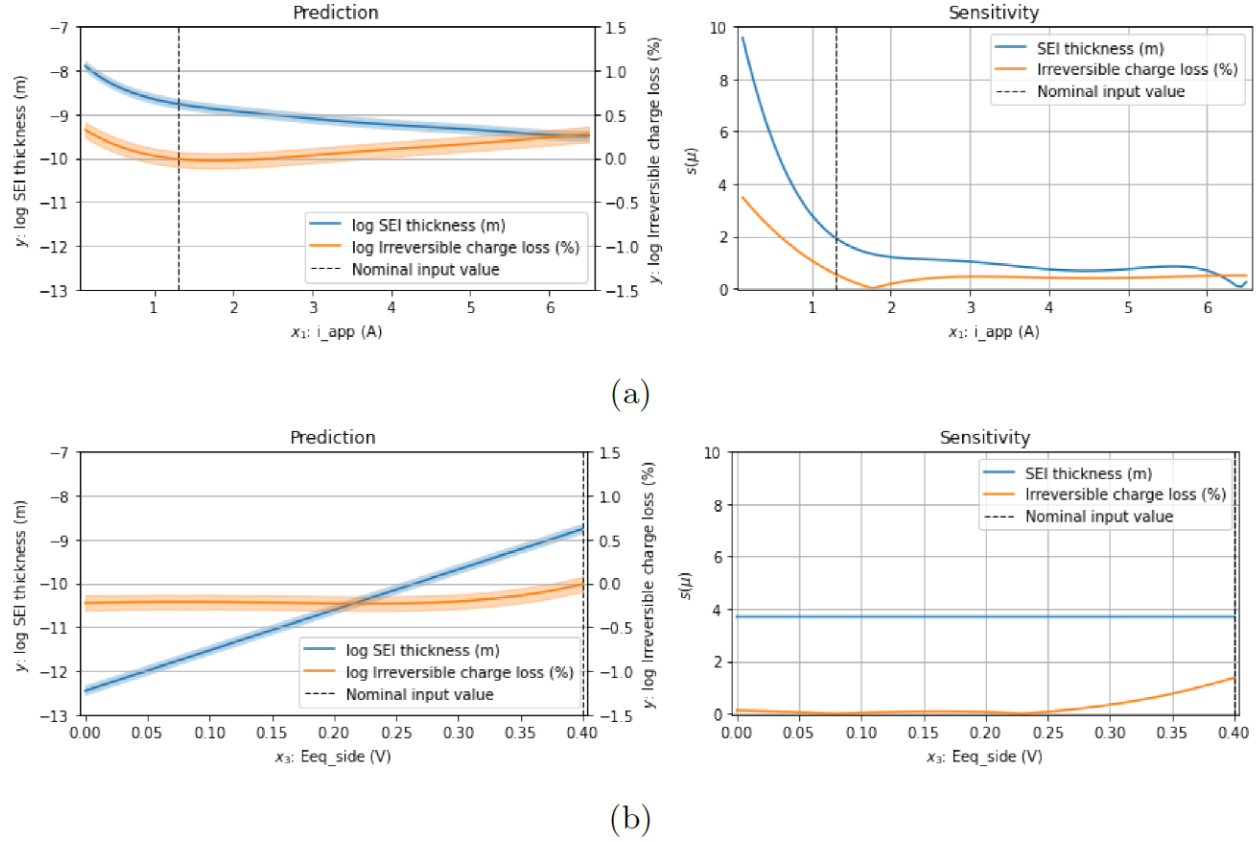

Figure 10. Local SA of sensitive input parameters and their entire range of variation while keeping all other inputs fixed at their nominal values. The left plots show the predicted outputs with uncertainty (two standard deviations), and the suitable plots show the corresponding local sensitivities.

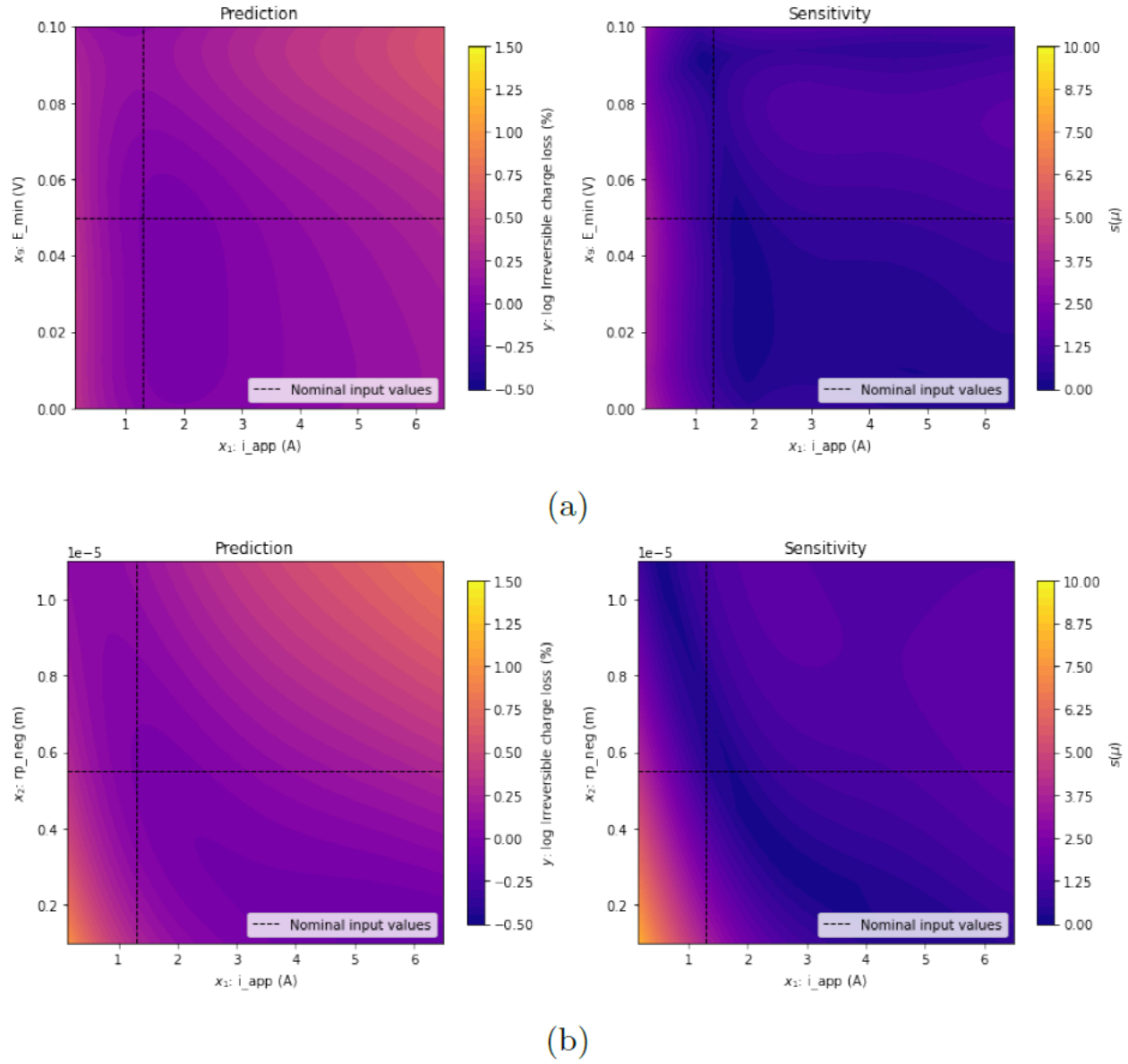

Figure 11. Local SA for the ICL output parameter of some of the relevant sensitive input parameters and their entire range of variation while keeping all other inputs fixed at their nominal values. The left plot shows the predicted log ICL, and the right plot shows the corresponding local sensitivity as a function of the inputs.

3.5 Code and model availability

Real-time prediction (with visual user interface) of local descriptors for any two descriptors is feasible and available at https://github.com/BIG-MAP/sa_p2d_sei_interactive. The code and data are available at https://github.com/BIG-MAP/sensitivity_analysis_tutorial

4 References

1. Ghiringhelli, L. M., Vybiral, J., Levchenko, S. V., Draxl, C. & Scheffler, M. Big Data of Materials Science: Critical Role of the Descriptor. *Phys. Rev. Lett.* **114**, 105503 (2015).
2. Cheng, L. *et al.* Accelerating Electrolyte Discovery for Energy Storage with High-Throughput Screening. *J. Phys. Chem. Lett.* **6**, 283–291 (2015).
3. Nørskov, J. K., Bligaard, T., Rossmeisl, J. & Christensen, C. H. Towards the computational design of solid catalysts. *Nature Chemistry* **1**, 37–46 (2009).
4. Zhao, Z.-J. *et al.* Theory-guided design of catalytic materials using scaling relationships and reactivity descriptors. *Nat Rev Mater* **4**, 792–804 (2019).
5. Melot, B. C. & Tarascon, J.-M. Design and Preparation of Materials for Advanced Electrochemical Storage. *Acc. Chem. Res.* **46**, 1226–1238 (2013).
6. Rong, Z. *et al.* Materials Design Rules for Multivalent Ion Mobility in Intercalation Structures. *Chem. Mater.* **27**, 6016–6021 (2015).
7. Sendek, A. D. *et al.* Holistic computational structure screening of more than 12 000 candidates for solid lithium-ion conductor materials. *Energy Environ. Sci.* **10**, 306–320 (2017).
8. Zhao, Q. *et al.* Identifying descriptors for Li+ conduction in cubic Li-argyrodites via hierarchically encoding crystal structure and inferring causality. *Energy Storage Materials* **40**, 386–393 (2021).
9. Chang, J. H. *et al.* On-the-fly assessment of diffusion barriers of disordered transition metal oxyfluorides using local descriptors. *Electrochimica Acta* **388**, 138551 (2021).
10. Bölle, F. T., Bhowmik, A., Vegge, T., Maria García Lastra, J. & Castelli, I. E. Automatic Migration Path Exploration for Multivalent Battery Cathodes using Geometrical Descriptors. *Batteries & Supercaps* **4**, 1516–1524 (2021).

11. Pande, V. & Viswanathan, V. Descriptors for Electrolyte-Renormalized Oxidative Stability of Solvents in Lithium-Ion Batteries. *J. Phys. Chem. Lett.* **10**, 7031–7036 (2019).
12. Ko, J. S. *et al.* Differentiating Double-Layer, Psuedocapacitance, and Battery-like Mechanisms by Analyzing Impedance Measurements in Three Dimensions. *ACS Appl. Mater. Interfaces* **12**, 14071–14078 (2020).
13. Nishijima, M. *et al.* Accelerated discovery of cathode materials with prolonged cycle life for lithium-ion battery. *Nat Commun* **5**, 4553 (2014).
14. Ouyang, R., Curtarolo, S., Ahmetcik, E., Scheffler, M. & Ghiringhelli, L. M. SISSO: A compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates. *Phys. Rev. Materials* **2**, 083802 (2018).
15. Wang, Y., Wagner, N. & Rondinelli, J. M. Symbolic regression in materials science. *MRS Communications* **9**, 793–805 (2019).
16. Singstock, N. R. *et al.* Machine Learning Guided Synthesis of Multinary Chevrel Phase Chalcogenides. *J. Am. Chem. Soc.* **143**, 9113–9122 (2021).
17. Ishikawa, A. *et al.* Machine learning prediction of coordination energies for alkali group elements in battery electrolyte solvents. *Physical Chemistry Chemical Physics* **21**, 26399–26405 (2019).
18. Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J. & Müller, K.-R. Explaining Deep Neural Networks and Beyond: A Review of Methods and Applications. *Proceedings of the IEEE* **109**, 247–278 (2021).
19. Friederich, P., Krenn, M., Tamblyn, I. & Aspuru-Guzik, A. Scientific intuition inspired by machine learning-generated hypotheses. *Mach. Learn.: Sci. Technol.* **2**, 025027 (2021).
20. Xu, W., Andersen, M. & Reuter, K. Data-Driven Descriptor Engineering and Refined Scaling Relations for Predicting Transition Metal Oxide Reactivity. *ACS Catal.* **11**, 734–742 (2021).

21. Diddens, D. *et al.* Modeling the Solid Electrolyte Interphase: Machine Learning as a Game Changer? *Advanced Materials Interfaces* **9**, 2101734 (2022).
22. Chen, C. *et al.* A Critical Review of Machine Learning of Energy Materials. *Advanced Energy Materials* **10**, 1903242 (2020).
23. Xu, K. Nonaqueous liquid electrolytes for lithium-based rechargeable batteries. *Chemical Reviews* **104**, 4303–4417 (2004).
24. Xu, K. Electrolytes and interphases in Li-ion batteries and beyond. *Chemical Reviews* **114**, 11503–11618 (2014).
25. Schmitz, R. W. *et al.* Investigations on novel electrolytes, solvents and SEI additives for use in lithium-ion batteries: Systematic electrochemical characterization and detailed analysis by spectroscopic methods. *Progress in Solid State Chemistry* **42**, 65–84 (2014).
26. Schmuch, R., Wagner, R., Hörpel, G., Placke, T. & Winter, M. Performance and cost of materials for lithium-based rechargeable automotive batteries. *Nature Energy* **3**, 267–278 (2018).
27. Seo, D. M. *et al.* Electrolyte Solvation and Ionic Association III. Acetonitrile-Lithium Salt Mixtures–Transport Properties. *Journal of The Electrochemical Society* **160**, A1061–A1070 (2013).
28. Ding, M. S. & Jow, T. R. Physicochemical Properties of Non-Aqueous Solvents and Electrolytes for Lithium Battery Applications. *ECS Transactions* **16**, 183–214 (2009).
29. Armand, M. *et al.* Lithium-ion batteries – Current state of the art and anticipated developments. *Journal of Power Sources* **479**, 228708 (2020).
30. Xu, K. & Von Cresce, A. Interfacing electrolytes with electrodes in Li ion batteries. *Journal of Materials Chemistry* **21**, 9849–9864 (2011).
31. Dufrêche, J.-F., Bernard, O., Durand-Vidal, S. & Turq, P. Analytical Theories of Transport in Concentrated Electrolyte Solutions from the MSA. (2005) doi:10.1021/jp050387y.

32. Atkins, P. & Paula, J. de. Molecules in motion. in *Phyiscal Chemistry* 747–790 (Oxford University Press, 2006).
33. Matsuda, Y., Morita, M. & Kosaka, K. Conductivity of the Mixed Organic Electrolyte Containing Propylene Carbonate and 1,2-Dimethoxyethane. *Journal of The Electrochemical Society* **130**, 101–104 (1983).
34. Martínez, L. MEASURING THE CONDUCTIVITY OF VERY DILUTE ELECTROLYTE SOLUTIONS, DROP BY DROP. *Química Nova* **23**, 341 (2018).
35. Duncan A., M. I. The conductance of aqueous solutions of electrolytes. *Journal of the Franklin Institute* **225**, 661–686 (1938).
36. Onsager, L. Report on a revision of the conductivity theory. *Transactions of the Faraday Society* **23**, 341 (1927).
37. Anderko, A., Wang, P. & Rafal, M. Electrolyte solutions: from thermodynamic and transport property models to the simulation of industrial processes. *Fluid Phase Equilibria* **194–197**, 123–142 (2002).
38. Kontogeorgis, G. M., Maribo-Mogensen, B. & Thomsen, K. The Debye-Hückel theory and its importance in modeling electrolyte solutions. *Fluid Phase Equilibria* **462**, 130–152 (2018).
39. Gilliam, R. J., Graydon, J. W., Kirk, D. W. & Thorpe, S. J. A review of specific conductivities of potassium hydroxide solutions for various concentrations and temperatures. *International Journal of Hydrogen Energy* **32**, 359–364 (2007).
40. Nilsson-Hallén, J., Ahlström, B., Marczewski, M. & Johansson, P. Ionic liquids: A simple model to predict ion conductivity based on DFT derived physical parameters. *Frontiers in Chemistry* **7**, 1–6 (2019).

41. Landesfeind, J. & Gasteiger, H. A. Temperature and Concentration Dependence of the Ionic Transport Properties of Lithium-Ion Battery Electrolytes. *Journal of The Electrochemical Society* **166**, A3079–A3097 (2019).
42. Lombardo, T. *et al.* Artificial Intelligence Applied to Battery Research: Hype or Reality? *Chem. Rev.* (2021) doi:10.1021/acs.chemrev.1c00108.
43. Bhowmik, A. *et al.* Implications of the BATTERY 2030+ AI-Assisted Toolkit on Future Low-TRL Battery Discoveries and Chemistries. *Advanced Energy Materials* **n/a**, 2102698.
44. Bhowmik, A. *et al.* A perspective on inverse design of battery interphases using multi-scale modelling, experiments and generative deep learning. *Energy Storage Materials* **21**, 446–456 (2019).
45. Weng, B. *et al.* Simple descriptor derived from symbolic regression accelerating the discovery of new perovskite catalysts. *Nat Commun* **11**, 3513 (2020).
46. Valøen, L. O. & Reimers, J. N. Transport Properties of LiPF6-Based Li-Ion Battery Electrolytes. *Journal of The Electrochemical Society* **152**, A882 (2005).
47. Flores, E., Åvall, G., Jeschke, S. & Johansson, P. Solvation structure in dilute to highly concentrated electrolytes for lithium-ion and sodium-ion batteries. *Electrochimica Acta* **233**, 134–141 (2017).
48. Dave, A. *et al.* Autonomous Discovery of Battery Electrolytes with Robotic Experimentation and Machine Learning. *Cell Reports Physical Science* **1**, (2020).
49. Hall, D. S., Self, J. & Dahn, J. R. Dielectric Constants for Quantum Chemistry and Li-Ion Batteries: Solvent Blends of Ethylene Carbonate and Ethyl Methyl Carbonate. *Journal of Physical Chemistry C* **119**, 22322–22330 (2015).

50. Seo, D. M., Borodin, O., Han, S.-D., Boyle, P. D. & Henderson, W. A. Electrolyte Solvation and Ionic Association II. Acetonitrile-Lithium Salt Mixtures: Highly Dissociated Salts. *Journal of The Electrochemical Society* **159**, A1489–A1500 (2012).
51. Seo, D. M. *et al.* Electrolyte Solvation and Ionic Association III. Acetonitrile-Lithium Salt Mixtures–Transport Properties. *Journal of The Electrochemical Society* **160**, A1061–A1070 (2013).
52. Ding, M. S. & Jow, T. R. Conductivity and Viscosity of PC-DEC and PC-EC Solutions of LiPF[sub 6]. *Journal of The Electrochemical Society* **150**, A620 (2003).
53. Broløs, K. R. *et al.* An Approach to Symbolic Regression Using Feyn. (2021).
54. Cornelio, C. *et al.* AI Descartes: Combining Data and Theory for Derivable Scientific Discovery. 1–26 (2021).
55. Iooss, B. & Lemaître, P. A Review on Global Sensitivity Analysis Methods. in *Uncertainty Management in Simulation-Optimization of Complex Systems: Algorithms and Applications* (eds. Dellino, G. & Meloni, C.) 101–122 (Springer US, 2015). doi:10.1007/978-1-4899-7547-8_5.
56. Titsias, M. Variational Learning of Inducing Variables in Sparse Gaussian Processes. in *Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics* 567–574 (PMLR, 2009).
57. Quiñonero-Candela, J. & Rasmussen, C. E. A Unifying View of Sparse Approximate Gaussian Process Regression. *Journal of Machine Learning Research* **6**, 1939–1959 (2005).
58. Paananen, T., Piironen, J., Andersen, M. R. & Vehtari, A. Variable selection for Gaussian processes via sensitivity analysis of the posterior predictive distribution. *arXiv:1712.08048 [stat]* (2019).

59. Rasmussen, P. M., Madsen, K. H., Lund, T. E. & Hansen, L. K. Visualization of nonlinear kernel models in neuroimaging by sensitivity maps. *NeuroImage* **55**, 1120–1131 (2011).

60. Blix, K., Camps-Valls, G. & Jenssen, R. Gaussian Process Sensitivity Analysis for Oceanic Chlorophyll Estimation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **10**, 1265–1277 (2017).

61. Shim, J. & Striebel, K. A. Effect of electrode density on cycle performance and irreversible capacity loss for natural graphite anode in lithium-ion batteries. *Journal of Power Sources* **119–121**, 934–937 (2003).

62. Lee, S.-H. *et al.* A new approach to surface properties of solid electrolyte interphase on a graphite negative electrode. *Journal of Power Sources* **247**, 307–313 (2014).

63. An, S. J. *et al.* The state of understanding of the lithium-ion-battery graphite solid electrolyte interphase (SEI) and its relationship to formation cycling. *Carbon* **105**, 52–76 (2016).

64. Arora, P., White, R. E. & Doyle, M. Capacity Fade Mechanisms and Side Reactions in Lithium-Ion Batteries. *J. Electrochem. Soc.* **145**, 3647 (1998).

65. Winter, M., Novák, P. & Monnier, A. Graphites for Lithium-Ion Cells: The Correlation of the First-Cycle Charge Loss with the Brunauer-Emmett-Teller Surface Area. *J. Electrochem. Soc.* **145**, 428 (1998).

66. Zaghib, K., Nadeau, G. & Kinoshita, K. Effect of Graphite Particle Size on Irreversible Capacity Loss. *J. Electrochem. Soc.* **147**, 2110 (2000).