Batteries & Supercaps

Research Article
doi.org/10.1002/batt.202200228

Chemistry Europe
European Chemical Societies Publishing

www.batteries-supercaps.org

One-Shot Active Learning for Globally Optimal Battery Electrolyte Conductivity**

Fuzhan Rahmanian,[a, b] Monika Vogler,[a, b] Christian Wölke,[c] Peng Yan,[c] Martin Winter,[c, d, e] Isidora Cekic-Laskovic,[c] and Helge S. Stein*[a, b]

Non-aqueous aprotic battery electrolytes need to perform well over a wide range of temperatures in practical applications. Herein we present a one-shot active learning study to find all conductivity optima, confidence bounds, and relating formulation trends in the temperature range from −30 °C to 60 °C. This optimization is enabled by a high-throughput formulation and characterization setup guided by one-shot active learning utilizing robust and heavily regularized polynomial regression. Whilst there is an initially good agreement for intermediate and low temperatures, there is a need for the active learning step to improve the model for high temperatures. Optimized electrolyte formulations likely correspond to the highest physically possible conductivities within this formulation system when compared to literature data. A thorough error propagation analysis yields a fidelity assessment of conductivity measurements and electrolyte formulation.

Introduction

High-conductivity electrolytes in secondary batteries are of paramount importance for ensuring high performance and reliability of each battery cell chemistry.[1] In specialty applications such as aerospace or stationary storage in remote locations, bespoke electrolytes are however necessary.[2] High or low temperatures make the electrolyte a limiting performance factor,[1–4] e.g., in electric vehicles which suffer from relatively narrow optimal temperature windows of 15 °C to 35 °C.[1] Many studies[3–5] have thus been conducted to evaluate lithium-ion battery (LIB) electrolytes at low temperatures in respect to their conductivity. There exists only a limited number of electrolyte studies that consider wide temperature ranges[6–9] as recently reviewed by Lin et al.[10] Emblematic are the studies of Smart et al.[6] and Fan et al.[7] that both evaluate a limited number of formulations between −60 °C to 20 °C and −125 °C to 70 °C, respectively. The studies by Dave et al.[11,12] consider a wide range of electrolyte formulations but within a narrow range of temperatures. Utilizing an existing dataset[13,14] spanning a wide range of formulations and temperatures, we aim to perform as few as possible additional experiments to discover formulations with maximum conductivity for a wide range of temperatures. This is performed in a workflow called one-shot active learning. This means that a machine learning algorithm is used to suggest the most promising subsequent experiment for improving the outcome and model. Besides aiming to discover optimal electrolyte formulations for a range of temperatures, we also seek to evaluate whether there exists a globally optimal electrolyte. Although, conductivity optima can be predicted from the existing dataset using a machine learning model, we believe that more physically meaningful predictions can be obtained upon re-training of the model from one-shot active learning suggestions. From the post shot model, we then seek to deduce insights on the effects of different parameters on the conductivity values, which could not be generated solely based on the initially available dataset.

Compared to previous deployments of machine learning[15] in the field of battery electrolyte optimization,[11] we investigate whether an improvement in conductivity may already be achieved through a single iteration cycle. This approach is mostly analogous to the workflow of Attia et al.[16] for fast charging protocol optimization, as herein we are using a high-throughput electrolyte formulation robot and a machine learning based optimizer, that were not integrated and in fact run at two different locations asynchronously. This enabled us to deploy active learning without requiring the experimental equipment to have a direct interface to our active learning infrastructure, potentially allowing a greater adoption of this research paradigm.[17,18] This one-shot active learning study aims to find optimally conducting electrolyte formulations at temperatures ranging between −30 °C and 60 °C with as few

[a] F. Rahmanian, M. Vogler, Prof. Dr. H. S. Stein
Helmholtz Institute Ulm, Applied Electrochemistry, Helmholtzstr. 11, 89081 Ulm, Germany
E-mail: helge.stein@kit.edu
[b] F. Rahmanian, M. Vogler, Prof. Dr. H. S. Stein
Karlsruhe Institute of Technology, Institute of Physical Chemistry, Fritz-Haber-Weg 2, 76131 Karlsruhe, Germany
[c] Dr. C. Wölke, P. Yan, Prof. Dr. M. Winter, Dr. I. Cekic-Laskovic
Helmholtz-Institute Münster (IEK-12), Forschungszentrum Jülich GmbH, Corrensstraße 46, 48149 Münster, Germany
[d] Prof. Dr. M. Winter
MEET Battery Research Center, University of Münster, Corrensstraße 46, 48149 Münster, Germany
[e] Prof. Dr. M. Winter
Helmholtz-Institute Münster (IEK-12), Forschungszentrum Jülich GmbH, Corrensstraße 46, 48149 Münster, Germany

[**] A previous version of this manuscript has been deposited on a preprint server (https://doi.org/10.26434/chemrxiv-2022-1z8gn).

Supporting information for this article is available on the WWW under https://doi.org/10.1002/batt.202200228

extra measurements as necessary. Opposed to applying machine learning algorithms to conclude from existing datasets, active learning[18,19] is integrated in the data acquisition process with the idea of improving the model through intelligent suggestion of additional measurements. Usually, optimization loops in materials science[12,20] are run over several iterations, the approach herein, however, aims to only perform a single iteration to achieve an improvement in conductivity and potentially reduction in uncertainty. The existing dataset of lithium hexafluorophosphate ($LiPF_6$) in ethylene carbonate (EC), ethyl methyl carbonate (EMC) and propylene carbonate (PC) was totaling 80 electrolyte formulations with measured conductivities at −30 to 60 °C.[13,14] The suggestion of new formulations was fully exploitative,[20] i.e., requested formulations were selected solely based on their predicted conductivity at a respective temperature with complete neglect of model uncertainty. Active learning in fully exploitative mode has been shown to significantly increase the so-called "enhancement factor" by Rohr et al.[20] The enhancement factor describes the increase in probability of finding an optimum given a fixed budget of experiments. There are other research modes[20] not explored in this study. However, a recent study by Flores et al.[13] focused on the "understanding driven" research mode. Their symbolic regression approach[13] works well for high temperatures but fails for highly concentrated liquid electrolytes at low temperatures, indicating a change in the physicochemical behavior. Organizationally, this study is the human-in-the-loop version of the fully autonomous active learning study presented by Rahmanian et al.[21]

Results and Discussion

Pre-shot model training

The dataset DS1 used herein is the same underlying the study presented by Flores et al.[13] using the formulation and characterization setup reported by Krishnamoorthy et al.[14] The herein presented one-shot active learning approach is model free, meaning that we do not utilize any physics or chemistry knowledge except correct pose of the input (formulation) and output (conductivity) and a compartmentalization of the problem by temperature.

The global trends of electrolyte conductivity, captured by our model M1, are shown in Figure 1, which illustrates the conductivity (σ) over r_{LiPF_6} and r_{PC} at −30 °C, −10 °C, 20 °C, and 60 °C (additional temperatures see S2). For all considered temperatures the R^2 score is approximately 0.73–0.80, which indicates a good fit. However, the degree of the polynomial used for the fit is higher for the high temperatures compared to low temperatures. The orange datapoints in Figure 1 indicate the formulations covered by dataset DS1.

Overall, conductivity is strongly correlated with temperature as expected from Debye-Hückel-Onsager (DHO) theory,[13,22] however this theory is only valid for dilute solutions. Consequently, we observe low conductivity for $r_{LiPF_6} > 0.8$ or $r_{LiPF_6} < 0.1$. In general, we observe the maximum conductivity shifting towards higher conducting salt concentrations at higher temperatures as it was reported by Landesfeind et al.[23] and Ding et al.[24,25] for various electrolyte formulations. The lowest overall measured conductivity is 194 mS cm^{-1} at −30 °C. Conductivity is showing a generally less pronounced dependence on r_{PC} than on r_{LiPF_6}. This observation correlates with the concentration-conductivity relationship that is primarily dependent on conducting salt concentration.[25]

Going from low to high temperatures, the system seems to allow for higher r_{PC} and r_{LiPF_6} while yielding a high conductivity which is in good agreement with established theory.[25,26] The model M1 also seems to prefer little presence of PC at low temperatures for higher conductivity.[25,26] Our finding is in good agreement with Ding et al.[26] who report, similar trends with temperature.[27] They discuss the higher EC and PC contents by an increase in the dielecteric constant and consequently higher conductivity.[27,28] At 20 °C, a narrow global optimum at relatively high $r_{PC} \approx 0.35$ is observed. The plot of the conductivity corresponding to 60 °C shows a very small region with high conductivity around $r_{PC} \approx 0.35$ and $r_{LiPF_6} \approx 0.38$, and additionally a maximum at $r_{PC} \approx 0.3$ and very high $r_{LiPF_6} \approx 1.2$. All but the −10 °C optima exist near unsampled formulations. Based on the prediction of the trained model M1, 10 samples with highest predicted conductivity for each temperature were selected and reported to the experimentalists. The requested and considered formulations can be found in the https://github.com/BIG-MAP/electrolyte_optimization_one_shot_active_learning repository.

One-shot predictions and measurements

Utilizing the above results obtained from M1 (see Figure 1), we predict 10 top percentile formulations at every temperature, resulting in a total of 100 electrolyte formulations. These formulations were communicated to the experimentalists omitting the predicted conductivity. The experimentalists randomly selected 24 formulations from these 100 suggestions. These selected formulations correspond to optimization temperatures of −30 °C, 20 °C and 60 °C. Conductivity measurements were conducted for the selected formulations covering all the temperatures between −30 °C and 60 °C in steps of 10 °C analogously to the generation of the dataset DS1. The data obtained from the measurement of these 24 formulations constitute dataset DS2. Figure 2 compares the M1 predicted vs. measured conductivities for the 24 newly measured formulations. There is a small deviation between the requested and measured formulations due to slight imperfections in the formulation process. Hence, Figure 2 shows the conductivity prediction at the actually formulated composition. Inaccuracies occurring during the solid and liquid dispensing processes are technical in nature and are negligible given the fidelity assessment presented in the section "Interactions and method fidelity". The error bar illustrates the conductivity error by reporting the maximum and minimum values among the repeated measurements.

Research Article
doi.org/10.1002/batt.202200228
Chemistry Europe
European Chemical Societies Publishing

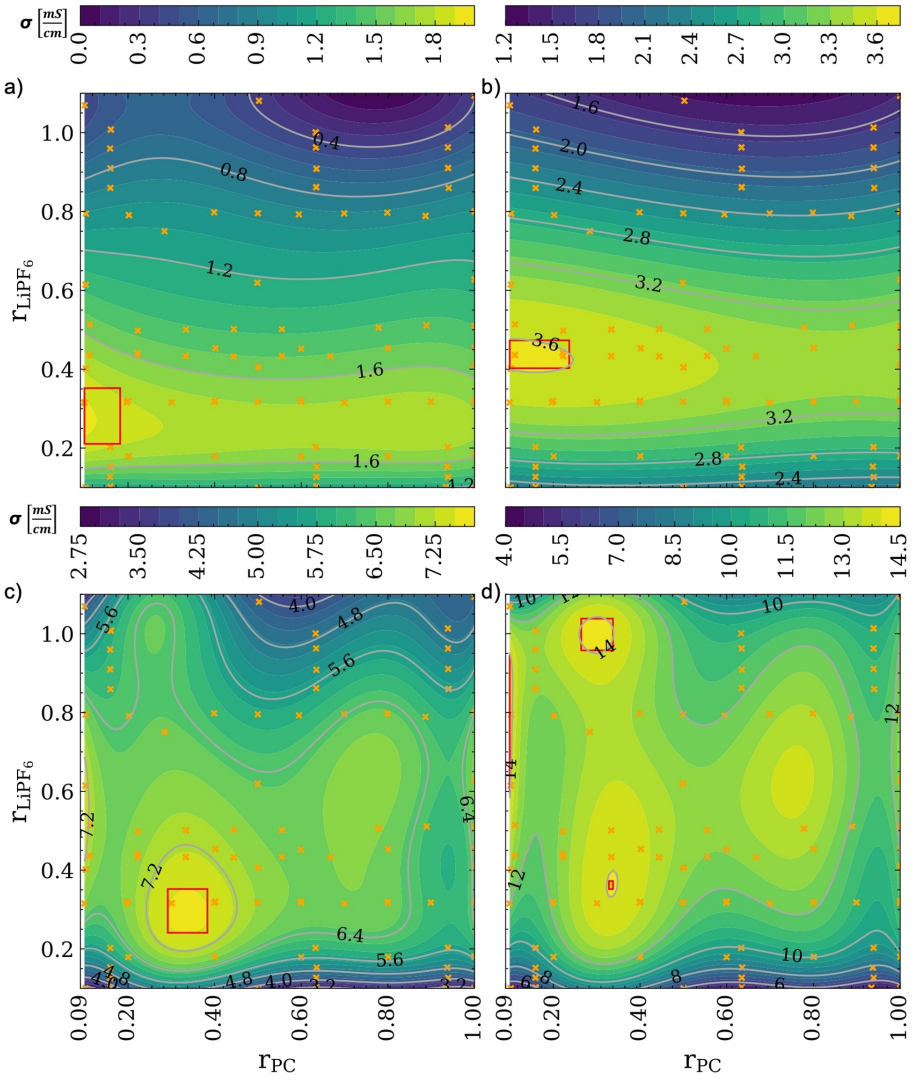

Figure 1. Trends in electrolyte conductivity at a) −30 °C, b) −10 °C, c) 20 °C, d) 60 °C as obtained from model M1. Orange data points represent the r_{PC} and r_{LiPF_6} position of formulations, which were experimentally measured. There is an overall incremental trend for higher r_{LiPF_6} from −30 °C to −10 °C and narrow optima in electrolyte conductivity at higher r_{PC} at unsampled formulations for the higher temperatures. The red boxes in the plots represent the range of formulations corresponding to the top percentile of the conductivity as obtained from the predictions of the model M1.

Figure 2(a) shows predicted and measured conductivities at a temperature of −30 °C. Formulations predicted to be optimally conducting electrolytes at −30 °C are verified experimentally to be among the best conducting at this temperature. The evaluation metric here should therefore not be the exact value prediction, as all originate from a narrow distribution of only 1% variance, which is significantly lower than the experimental noise, i.e., all points are virtually indistinguishable. All but one formulation optimized for −30 °C fall within the top percentile, i.e., the success rate is 87%, see also Figure 4. The formulations optimized for 20 °C and 60 °C exhibit a significantly lower conductivity at −30 °C, also with relatively large deviations between M1 prediction and measurement. This data suggests that there exists no electrolyte with a globally optimal conductivity. Differences in performance between formulations optimized for the temperature of interest and those not optimized for this temperature can amount up to 100%.

The measured conductivities for the requested formulations are added as prior knowledge and the model is retrained using dataset DS3.

Post-shot model refinement

After one-shot active learning and the Bayesian hyperparameter tuning[29] as described in the methods section, the models are significantly improved. The predicted trends for low temperatures changed only marginally, whereas the improvements for temperatures of 20 °C and 60 °C are significant, as shown in Figure 3. Additional temperatures can be found in S3. Together with the low temperature trends there is now a coherent trend across temperatures suggesting higher r_{PC} and r_{LiPF_6} for optimal conductivity at elevated temperatures.[23,26] Also, the range of formulations, for which the maximum

Research Article
doi.org/10.1002/batt.202200228
Chemistry Europe
European Chemical Societies Publishing

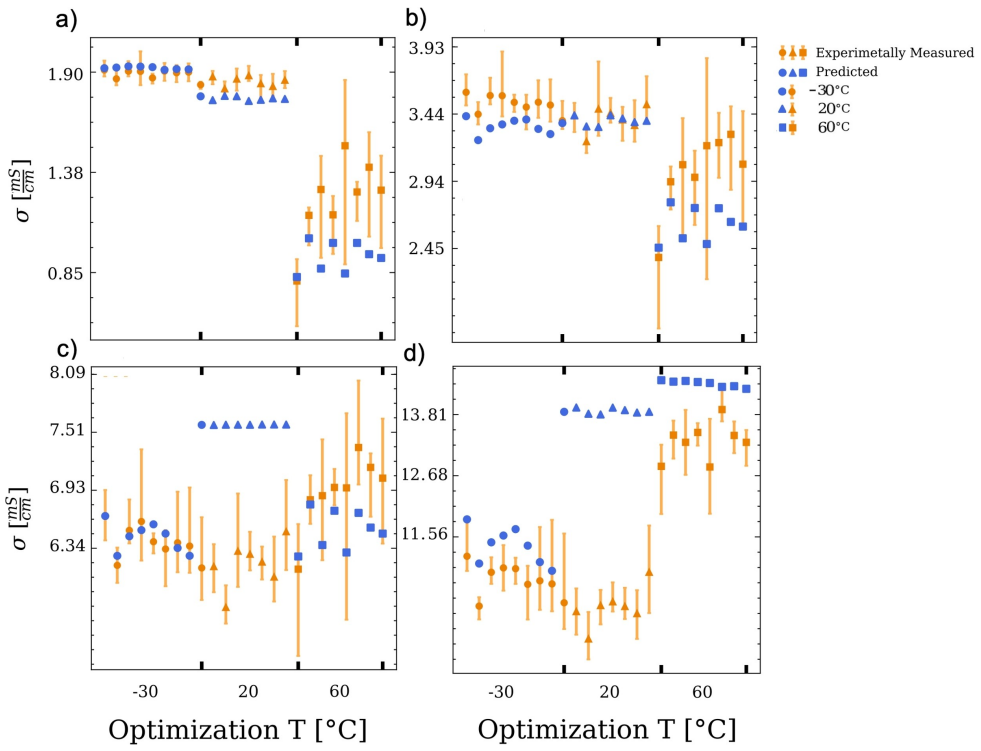

Figure 2. Comparison between measured and predicted conductivity values at a) −30 °C, b) −10 °C, c) 20 °C, and d) 60 °C for the formulations selected based on the predictions of model M1. Orange points represent the mean values of measured conductivities with error bars relating to the min/max spread from repeated measurements. The high accuracy for low temperature predictions is best observed in a) where the formulations predicted to be best at 60 °C and the ones optimized for conductivity at 20 °C perform worse at −30 °C. Overall, this suggests, that there exists no globally optimal electrolyte and performance can vary by up to a factor of two.

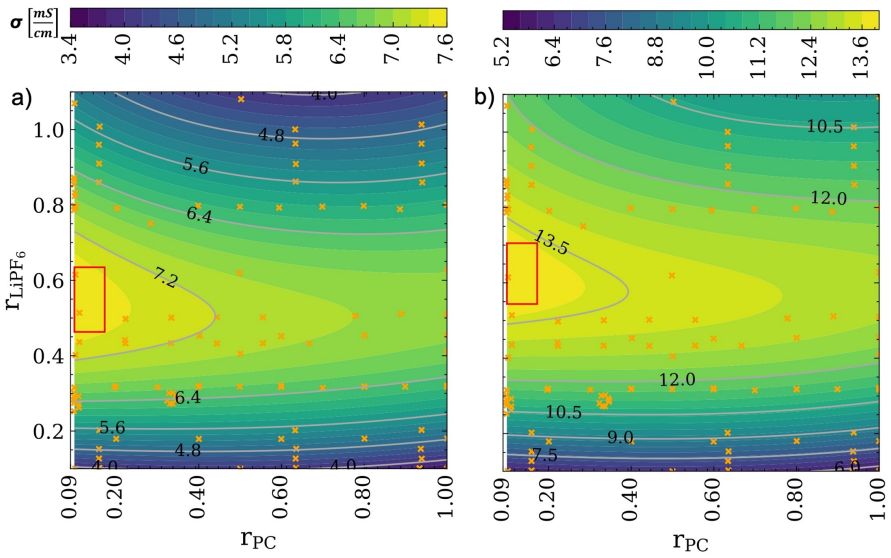

Figure 3. Trends in electrolyte conductivity after one-shot active learning for a) 20 °C, and b) 60 °C as predicted by M2. The red boxes correspond to the top percentile, which was obtained from M1. The selected formulations which were suggested by M1 and used for conductivity determination (DS2), were added to the training set as an additional prior knowledge (DS3). Model parameter tuning and uncertainty measurement were implemented at this stage of active learning (see S4). Trends for additional temperatures can be seen in the S3.

conductivity was predicted by M1, changed significantly with M2. Therefore, the region of the formulations constituting the top percentile obtained from M1, represented as a red rectangle in Figure 3, is far from the top percentile of the improved model M2.

The improvement of the model's predictions towards a more physically meaningful trend highlights the significance of

Research Article
doi.org/10.1002/batt.202200228
Chemistry Europe
European Chemical Societies Publishing

active learning in model refinement as we only added an additional 30% of data points to the dataset whilst qualitatively improving the model. Comparing these results to the symbolic regression model by Flores et al.[13] reveals significant differences in the mass ratios required for maximum conductivity.

Figure 4 shows the conductivity as predicted by M1 and measured experimentally for the formulations comprised in dataset DS2 at the temperature for which the respective formulation was optimized. Furthermore, the top percentiles obtained from M1 (red areas) and M2 (green areas) are shown. Overall, the optimization and prediction worked best at low temperatures. The measured conductivity values are close to the predicted ones at −30 °C. Hence, the measured values are within the top percentile of M1. Unsurprisingly, the top percentile does not change significantly after the one-shot active learning and the formulations are also within the top percentile of M2. From the plots corresponding to 20 °C and 60 °C, a poorer performance is observed. At 20 °C, the model significantly overestimates the highest conductivity. The measured conductivity values at 20 °C remain below the predictions and below the top percentile. The results for 60 °C reveal a less severe deviation between the predictions and the measured conductivity values. However, the measured values are still below the predicted ones and below the top one percentiles of both, M1 and M2. This can be understood based on the results obtained from Figure 3. Therefore, the success rate is only about 12% as only one formulation is within the pre-shot top percentile. The significant changes in the position of the conductivity maxima upon one-shot active learning result in the formulations contained in dataset DS2 not being the highest conducting ones anymore. The range of the top percentile in conductivity does not change severely, however the formulations corresponding to these conductivity values differ strongly. This indicates that M1, which is trained solely on DS1, is not fit well for temperatures around 20 °C and above. Based on the differences between Figure 3 and Figure 4, it can be assumed, that the quality of the models significantly increases through active learning.

The drastic improvement of the model becomes even more obvious upon plotting the temperature maxima with the spread of the top percentile as displayed in Figure 5. Before the learning shot, the optima followed no physically meaningful or interpretable trend whereas after adding the extra data contained in DS2, the very fine trends in optima towards higher r_{LiPF_6} and slightly more r_{PC} become obvious. Uncertainty quantification was performed using the jackknife plus[30,31] strategy resulting in an average 95% prediction interval of $3*10^{-1}$ mS cm^{-1} (see S4). However, the incorporation of the model agnostic prediction technique allows the measurement of aleatoric and epistemic uncertainty at any point. Comparing the results for electrolyte conductivity found by our one-shot active learning approach to literature such as Ding et al.[24] at 60 °C, and −30 °C suggests that the herein reported maxima correspond to the globally maximum conductivity in this system, which is approximately 12 mS cm^{-1} and 1.9 mS cm^{-1}, respectively. In another study, Landesfeind et. al.[23] indicate global maxima of 4.7 mS cm^{-1}, 7.6 mS cm^{-1} and 9.25 mS cm^{-1} at −10 °C, 20 °C, and 30 °C, respectively. Their results are in agreement with our findings.

Interactions and method fidelity

Through the availability of a machine learning model M2 that accurately and precisely predicts the trends in conductivity for all temperatures, an assessment of confounding inputs and method fidelity can be pursued. The model has two inputs: r_{PC} and r_{LiPF_6}, and through the polynomial nature an analytical derivation is facile. The post-shot regularized polynomial equation [Eq. (1)] for conductivity (σ) post hyperparameter tuning is:

$$\sigma = c_0 + c_1 r_{PC} + c_2 r_{LiPF_6} + c_3 r_{PC}^2 + c_4 r_{PC} r_{LiPF_6} + c_5 r_{LiPF_6}^2 + c_6 r_{PC}^3 + c_7 r_{PC}^2 r_{LiPF_6} + c_8 r_{PC} r_{LiPF_6}^2 + c_9 r_{LiPF_6}^3 \quad (1)$$

i.e., a polynomial of degree 3 with the individual parameters shown in Table 1. Some coefficients change drastically with temperature whilst others barely change. Upon careful comparison to Equation (1), one can see that those coefficients corresponding to a conducting-salt-ratio-only term, scale almost exponentially whilst all others, i.e., solvent-ratio-only and

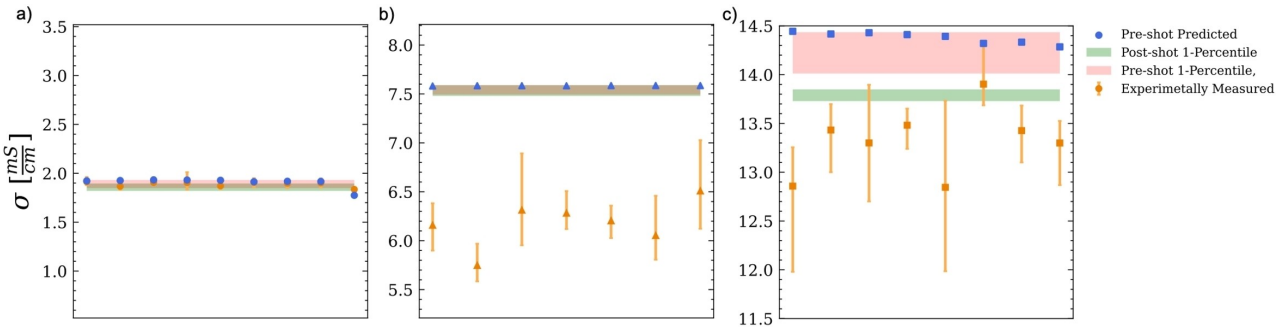

Figure 4. Predicted and measured conductivities for the formulations contained in dataset DS2 optimized regarding conductivity at the temperatures a) −30 °C, b) 20 °C and c) 60 °C. Each subfigure shows the data for the formulations optimized at the respective temperature. Predictions originating from M1 are shown. Additionally, the range of conductivities spanned by the top percentile as predicted by M1 (red areas) and M2 (green areas) are shown.

Research Article
doi.org/10.1002/batt.202200228
Chemistry Europe
European Chemical Societies Publishing

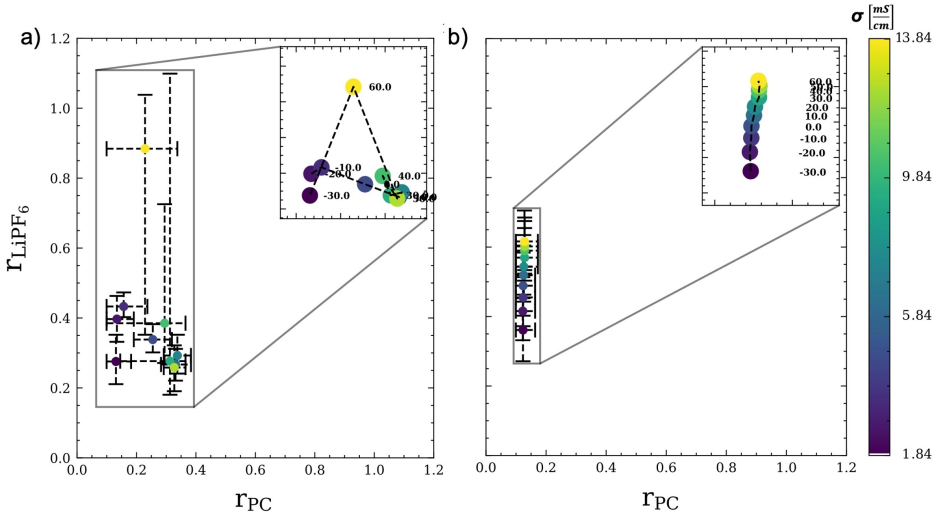

Figure 5. Trends of maximum conductivity a) before (M1) and b) after (M2) one-shot active learning and model optimization. Each point corresponds to the mean conductivity value of the top percentile obtained from the respective model trained for the temperature of interest. The error bars represent the spread of r_{LiPF_6} and r_{PC} within the top percentile. Before introduction of the additional 24 electrolyte formulations (DS2), the trends are neither physically nor qualitatively interpretable. Overall, higher r_{LiPF_6} is needed at higher temperatures to reach the optima, with a minutely higher r_{PC} from 20 °C onward.

Table 1. Polynomial coefficients incorporating ridge regularization after one-shot active learning for T = −30 °C to 60 °C.

T [°C] / $c \times 10^{-3}$	c_0	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9
−30	1.0	−0.3	5.6	0.9	−2.3	−9.7	−0.6	1.4	0.6	4.0
−20	1.2	−0.3	8.9	0.9	−3.2	−13.6	−0.7	1.9	0.8	5.1
−10	1.4	1.0	12.1	0.1	−3.9	−17.1	−0.3	2.4	0.9	6.1
0	1.5	0.4	15.9	−0.1	−4.9	−21.2	−0.4	3.3	0.9	7.3
10	1.5	1.2	20.4	−0.9	−5.9	−26.1	−0.4	4.2	0.7	8.8
20	1.6	1.7	25.0	−1.4	−6.8	−31.0	−0.3	5.1	0.5	10.5
30	1.6	2.7	29.7	−3.2	−6.9	−35.8	0.4	5.6	0.0	12.2
40	1.6	3.0	35.1	−3.2	−7.3	−41.9	1.0	6.2	−0.6	14.6
50	1.6	3.2	41.3	−3.2	−7.6	−49.5	0.0	6.4	−0.7	17.6
60	1.6	3.0	47.4	−2.5	−6.9	−57.2	−0.7	6.6	−1.8	21.1

solvent-conducting-salt-ratio terms scale sigmoidal with temperature (see S6b). These interaction coefficients allow for further research into the relationship governing the solvation shell properties upon electrolyte solvent variation.[32]

A long-lasting debate of how precise the electrolyte formulation needs to be answered using model M2. An error propagation estimation can be done when the gradient of a function and the uncertainty of the underlying input is known. From the herein reported measurements, we know the uncertainty of the conductivity and we can easily calculate the gradient of the conductivity w.r.t. the formulation. Here, we take the median uncertainty of the conductivity measurements ($\Delta\sigma_{exp} = 0.3527$ mS cm^{-1}) and divide it by the largest gradient of conductivity w.r.t. to formulation (both uni- and bivariate) at every temperature (Figure 6, Table 2) to obtain a conservative estimate of the maximally allowed formulation error [Eq. (2)] that would be on the same order of magnitude like the measurement noise. Unsurprisingly one can have larger errors in solvent-to-co-solvent ratios as in conducting-salt-to-solvent ratios. Interesting, however, is that an error of about 10 % in the solvents is acceptable for most temperatures. Dosing of the conducting salt should however be as precise as possible as at high temperatures the error should not exceed 1.5 %.

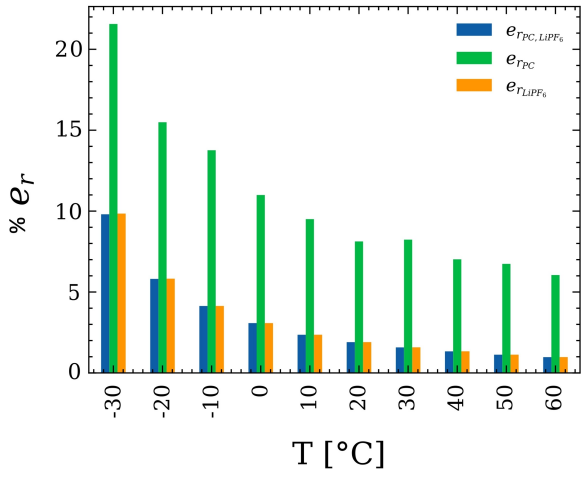

Figure 6. The maximum formulation error calculated by Equation (2) with the median $\Delta\sigma_{exp}$ of approximately 0.352 mS cm^{-1} with respect to uni- and bivariate combination of r_{LiPF_6} and r_{PC} between −30 °C and 60 °C.

Research Article
doi.org/10.1002/batt.202200228
Chemistry Europe
European Chemical Societies Publishing

Table 2. The maximum norm of the predicted conductivity gradient.

T [°C] / max($\partial\sigma$) [$\frac{mS}{cm}$]	$\partial\sigma_{r_{PC},r_{LiPF_6}}$	$\partial\sigma_{r_{PC}}$	$\partial\sigma_{r_{LiPF_6}}$
−30	3.598	1.636	3.582
−20	6.071	2.277	6.057
−10	8.526	2.563	8.523
0	11.454	3.208	11.455
10	14.955	3.711	14.947
20	18.535	4.341	18.516
30	22.345	4.283	22.294
40	26.462	5.024	26.404
50	31.307	5.234	31.248
60	36.173	5.831	35.999

$$(\alpha(\frac{\partial\sigma}{\partial r_{PC}})^2+\beta(\frac{\partial\sigma}{\partial r_{LiPF_6}})^2)^{0.5}\cdot e_{r_{PC,LiPF_6}} \leq \Delta\sigma_{exp}$$
$$for \qquad \alpha,\beta \in \{(0,1),(1,0),(1,1)\} \tag{2}$$

Conclusion

This study shows the utility of active learning to improve model accuracy and precision on complex data with few examples. The pre-shot model M1 significantly underfit the data such that obtained trends did not follow a physically meaningful trend. After one-shot active learning, the discovered model M2 produced smooth optima across the temperatures under investigation even though temperature was not a parameter in model training. Obtained trends in the optima suggest that for low temperatures, the conducting salt concentration should be minimized whilst for higher temperatures the salt concentration should be increased. We find that a globally optimally conducting electrolyte does not exist as those optimal at low temperatures perform poorly at higher temperatures. Those electrolytes optimized for near room temperature show approximately 20% less conductivity at low and about half the conductivity at high temperatures compared to the formulations optimized for the respective temperature range. Through the availability of an easily differentiable model M2, we can discuss electrolyte solvent-conducting salt interactions and find mostly sigmoidal or exponential temperature trends hinting at two different mechanisms. The differentiable model M2 also allows an elucidation of maximally allowed formulation errors which lie at approximately 10% for the solvent composition and 1.5% for the conducting salt ratio at most temperatures. Through the conservative choice of a low degree polynomial model, we were able to obtain optima and interpretable insights translatable to existing physicochemical laws such as the DHO theory, however at high salt concentrations.

We believe, our approach can be transferred to novel electrolyte systems e.g., for Na-ion batteries. In our opinion, this could accelerate knowledge generation when starting from small datasets and unravel complex interrelations early in the research process.

Experimental

Workflow

The overall idea of this study is the optimization through a one-shot active learning iteration. To this end a pre-existing dataset DS1 was utilized to pre-train a model M1. From M1, a set of optimally conducting electrolytes was suggested to the experimentalists. The experimentalists measured the conductivity of the newly suggested formulations and reported the results back to the machine learning team. The newly measured data collected in the dataset DS2 was merged with dataset DS1 to obtain an extended dataset DS3. The dataset DS3 is then used to retrain the model to provide refined trends with formulation and temperature. In the following we will refer to the retrained model as M2. The predicted conductivities obtained from M2 are also used to understand error propagation.

Further details about the model training can be found in the section "Model training and one-shot active learning". A schematic of this study's workflow is shown in Figure 7. Summarizing, there are three stages in this pipeline: 1) model training, 2) formulation suggestions and measurement, 3) retraining and refinement of the model and uncertainty quantification.

Description of the initial dataset DS1 and Measurements

The initial dataset DS1 used herein to pre-train the model M1, totals 80 distinct electrolyte formulations measured at Helmholtz-Institute Münster for general purpose, using their automated formulation and characterization setup described in detail in Krishnamoorthy et al.[14] The formulations reported in DS1 contain ethylene carbonate (EC), propylene carbonate (PC), and ethyl methyl carbonate (EMC) in a solvent/co-solvent mixture and lithium hexafluorophosphate ($LiPF_6$) as the conducting salt. The data reported in DS1 covers temperatures between −30 °C and 60 °C, at increments of 10 °C as described by Krishnamoorthy et al.[14] Conductivity measurements were repeated 5 to 7 times. For each datapoint the electrolyte formulation, conductivity and measurement temperature were recorded. Across all formulations, the ratio of (EC + PC) : EMC was fixed either at 3:7 or 1:1 by weight and the concentration of $LiPF_6$ was varied between 0.2 and 2.1 mol kg^{-1}.

We express the uncertainties for the experimental values by the min/max spread of the individual measurements. The mass ratios of PC and $LiPF_6$ were normalized and referenced as $r_{PC} = \frac{PC}{(PC+EC)}$ and $r_{LiPF_6} = \frac{LiPF_6}{(PC+EC)}$ for using them as inputs for model training and one-shot active learning. The (EC + PC):EMC ratio was not considered during model training as it is not an independent variable.

Model training and one-shot active learning

The dataset size poses the challenge of finding well performing models that are simple and interpretable.[33] We therefore settle on polynomial regression[34] for our study. Contrary to Flores et al.[13] we do not consider temperature as a parameter in model training and train our model independently for each temperature. The basic model is a strongly regularized polynomial regressor aiming to avoid multicollinearity,[35,36] i.e., linear correlations among the input parameters, which would negatively affect the estimates of the coefficients in the regression model.[37–40] The polynomial regression,[34] ridge regularization,[41] and in step two for optimization purposes hyperparameter tuning are performed. All of the machine learning steps were performed using the scikit-learn library[42–44] available for Python. From the fitted polynomial model

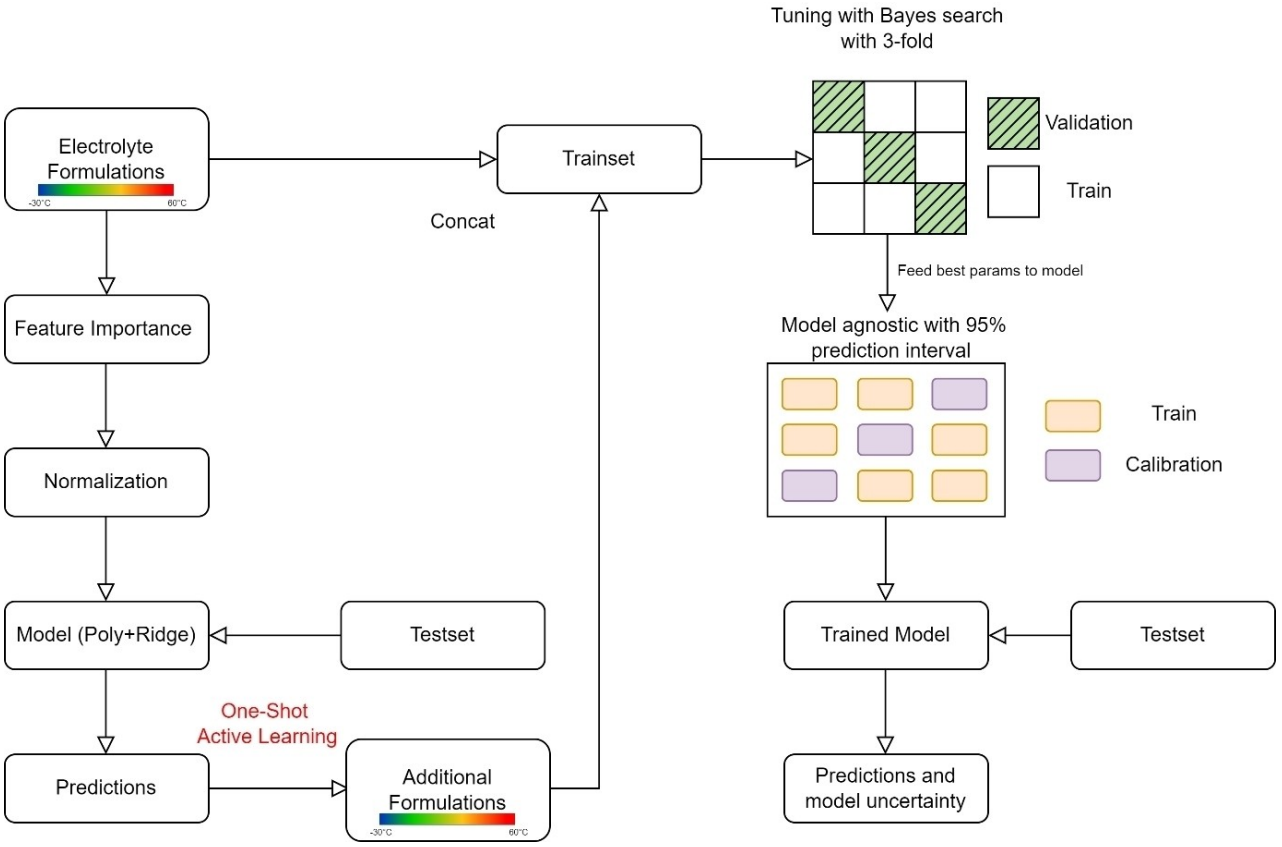

Figure 7. Schematic diagram of this study's pipeline consisting of initial model training, suggesting formulation to experimentalists for measurement of requested formulations and retraining the model by one-shot active learning with uncertainty quantification.

a fine subsampling is performed comprising 10^4 formulation ratios at a fixed grid spacing of 1 ratio-%. From this fine subsampling, the 10 formulations corresponding to the maximum predicted conductivity, for each temperature were reported to the experimentalists resulting in a total of 100 suggested formulations. A subset of 24 formulations was chosen by the experimentalists covering all suggestions for −30 °C, 20 °C and 60 °C. Subsequent to the formulation and conductivity measurements of the new formulations, the model was retrained on dataset DS3. For hyperparameter tuning we performed a Bayesian search[29,42] with a threefold cross validation (details see S5). The best parameters are then fed to our model. This search uses ridge regularized polynomial models to favor low polynomial degrees. To assess the model uncertainty for both aleatoric and epistemic uncertainty after the learning shot (and the possible necessity for a second learning shot) we build a pipeline using the so called model agonistic prediction interval estimator (MAPIE).[45] This estimator uses the jackknife plus[31] library to estimate the uncertainty[46] of the model for a 95 % prediction interval, i.e. a newly predicted value has a probability of 95 % to lie within this prediction interval.

Acknowledgements

This work contributes to the research performed at CELEST (Center for Electrochemical Energy Storage Ulm-Karlsruhe) and was funded by the German Research Foundation (DFG) under Project ID 390874152 (POLiS Cluster of Excellence). This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957189. Open Access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are openly available in github at https://github.com/BIG-MAP/electrolyte_optimization_one_shot_active_learning

Keywords: active learning · batteries · electrolyte · machine-learning · optimization

[1] S. Ma, M. Jiang, P. Tao, C. Song, J. Wu, J. Wang, T. Deng, W. Shang, *Prog. Nat. Sci. Mater. Int.* **2018**, *28*, 653–666.
[2] A. Gupta, A. Manthiram, *Adv. Energy Mater.* **2020**, *10*, 2001972.
[3] J. Jaguemont, L. Boulon, Y. Dubé, *Appl. Energy* **2016**, *164*, 99–114.
[4] S. Panchal, J. Mcgrory, J. Kong, R. Fraser, M. Fowler, I. Dincer, M. Agelin-Chaab, *Int. J. Energy Res.* **2017**, *41*, 2565–2575.
[5] D. Hubble, D. Emory Brown, Y. Zhao, C. Fang, J. Lau, B. D. McCloskey, G. Liu, *Energy Environ. Sci.* **2022**, *15*, 550–578.

[6] M. C. Smart, B. V. Ratnakumar, K. B. Chin, L. D. Whitcanack, *J. Electrochem. Soc.* **2010**, *157*, A1361.
[7] X. Fan, X. Ji, L. Chen, J. Chen, T. Deng, F. Han, J. Yue, N. Piao, R. Wang, X. Zhou, X. Xiao, L. Chen, C. Wang, *Nat. Energy* **2019**, *4*, 882–890.
[8] Q. Zhou, S. Dong, Z. Lv, G. Xu, L. Huang, Q. Wang, Z. Cui, G. Cui, *Adv. Energy Mater.* **2020**, *10*, 1903441.
[9] Y. Yang, Y. Yin, D. M. Davies, M. Zhang, M. Mayer, Y. Zhang, E. S. Sablina, S. Wang, J. Z. Lee, O. Borodin, C. S. Rustomji, Y. Shirley Meng, *Energy Environ. Sci.* **2020**, *13*, 2209–2219.
[10] X. Lin, G. Zhou, J. Liu, J. Yu, M. B. Effat, J. Wu, F. Ciucci, *Adv. Energy Mater.* **2020**, *10*, 2001235.
[11] A. Dave, J. Mitchell, K. Kandasamy, H. Wang, S. Burke, B. Paria, B. Póczos, J. Whitacre, V. Viswanathan, *Cell Rep. Phys. Sci.* **2020**, *1*, 100264.
[12] A. Dave, J. Mitchell, S. Burke, H. Lin, J. Whitacre, V. Viswanathan, *ArXiv211114786 Cs* **2021**.
[13] E. Flores, C. Wölke, P. Yan, M. Winter, T. Vegge, I. Cekic-Laskovic, A. Bhowmik, **2022**, DOI 10.26434/chemrxiv-2022-nmmd4.
[14] A. N. Krishnamoorthy, C. Wölke, D. Diddens, M. Maiti, Y. Mabrouk, P. Yan, M. Grünebaum, M. Winter, A. Heuer, I. Cekic-Laskovic, **2022**, DOI 10.26434/chemrxiv-2022-vbl5d.
[15] A. D. Sendek, E. D. Cubuk, B. Ransom, J. Nanda, E. J. Reed, in *Transit. Met. Oxides Electrochem. Energy Storage*, John Wiley & Sons, Ltd, **2022**, pp. 393–409.
[16] P. M. Attia, A. Grover, N. Jin, K. A. Severson, T. M. Markov, Y.-H. Liao, M. H. Chen, B. Cheong, N. Perkins, Z. Yang, P. K. Herring, M. Aykol, S. J. Harris, R. D. Braatz, S. Ermon, W. C. Chueh, *Nature* **2020**, *578*, 397–402.
[17] T. Lombardo, M. Duquesnoy, H. El-Bouysidy, F. Årén, A. Gallo-Bueno, P. B. Jørgensen, A. Bhowmik, A. Demortière, E. Ayerbe, F. Alcaide, M. Reynaud, J. Carrasco, A. Grimaud, C. Zhang, T. Vegge, P. Johansson, A. A. Franco, *Chem. Rev.* **2022**, *122*, 10899–10969.
[18] T. Mueller, A. G. Kusne, R. Ramprasad, in *Rev. Comput. Chem.*, Wiley Hoboken, NJ, **2016**, pp. 186–273.
[19] Z. Wang, Z. Sun, H. Yin, X. Liu, J. Wang, H. Zhao, C. H. Pang, T. Wu, S. Li, Z. Yin, X. Yu, *Adv. Mater.* n.d., *n/a*, 2104113.
[20] B. Rohr, H. S. Stein, D. Guevarra, Y. Wang, J. A. Haber, M. Aykol, S. K. Suram, J. M. Gregoire, *Chem. Sci.* **2020**, *11*, 2696–2706.
[21] F. Rahmanian, J. Flowers, D. Guevarra, M. Richter, M. Fichtner, P. Donnely, J. M. Gregoire, H. S. Stein, *Adv. Mater. Interfaces* **2022**, 2101987.
[22] L. Onsager, *Trans. Faraday Soc.* **1927**, *23*, 341–349.
[23] J. Landesfeind, H. A. Gasteiger, *J. Electrochem. Soc.* **2019**, *166*, A3079.
[24] M. S. Ding, T. R. Jow, *ECS Trans.* **2009**, *16*, 183.
[25] M. S. Ding, T. R. Jow, *J. Electrochem. Soc.* **2003**, *150*, A620.
[26] M. S. Ding, Q. Li, X. Li, W. Xu, K. Xu, *J. Phys. Chem. C* **2017**, *121*, 11178–11183.
[27] M. S. Ding, T. R. Jow, *J. Electrochem. Soc.* **2004**, *151*, A2007.
[28] R. Naejus, D. Lemordant, R. Coudert, P. Willmann, *J. Chem. Thermodyn.* **1997**, *29*, 1503–1515.
[29] J. Wu, X.-Y. Chen, H. Zhang, L.-D. Xiong, H. Lei, S.-H. Deng, *J. Electron. Sci. Technol.* **2019**, *17*, 26–40.
[30] A. F. Bissell, *J. Appl. Stat.* **1977**, *4*, 55–64.
[31] R. F. Barber, E. J. Candès, A. Ramdas, R. J. Tibshirani, *Ann. Stat.* **2021**, *49*, 486–507.
[32] D. J. Siegel, L. Nazar, Y.-M. Chiang, C. Fang, N. P. Balsara, *Trends Chem.* **2021**, *3*, 807–818.
[33] X. Cheng, B. Khomtchouk, N. Matloff, P. Mohanty, *ArXiv180606850 Cs Stat* **2019**.
[34] E. Ostertagová, *Procedia Eng.* **2012**, *48*, 500–506.
[35] K. M. Marcoulides, T. Raykov, *Educ. Psychol. Meas.* **2019**, *79*, 874–882.
[36] R. Johnston, K. Jones, D. Manley, *Qual. Quant.* **2018**, *52*, 1957–1976.
[37] J. M. Wooldridge, *Introductory Econometrics: A Modern Approach*, Cengage Learning, **2015**.
[38] E. Vittinghoff, D. V. Glidden, S. C. Shiboski, C. E. McCulloch, in *Regres. Methods Biostat. Linear Logist. Surviv. Repeated Meas. Models* (Eds.: E. Vittinghoff, D. V. Glidden, S. C. Shiboski, C. E. McCulloch), Springer US, Boston, MA, **2012**, pp. 69–138.
[39] G. James, D. Witten, T. Hastie, R. Tibshirani, *An Introduction to Statistical Learning: With Applications in R*, Springer US, New York, NY, **2021**.
[40] S. Menard, *Applied Logistic Regression Analysis*, SAGE, **2002**.
[41] G. C. McDonald, *WIREs Comput. Stat.* **2009**, *1*, 93–100.
[42] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. Vanderplas, A. Joly, B. Holt, G. Varoquaux, *ArXiv13090238 Cs* **2013**.
[43] G. Hackeling, *Mastering Machine Learning with Scikit-Learn*, Packt Publishing Ltd, **2017**.
[44] O. Kramer, in *Mach. Learn. Evol. Strateg.* (Ed.: O. Kramer), Springer International Publishing, Cham, **2016**, pp. 45–53.
[45] "MAPIE - Model Agnostic Prediction Interval Estimator – MAPIE 0.3.2 documentation," can be found under https://mapie.readthedocs.io/en/latest/index.html, **n.d.**
[46] E. Hüllermeier, W. Waegeman, *ArXiv191009457 Cs Stat* **2020**.

Manuscript received: May 19, 2022
Revised manuscript received: July 13, 2022
Version of record online: August 23, 2022