D1.10 - Proceedings from Early Stage Research Seminar

VERSION

VERSION	DATE
1	24/2/2022

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.	1
WP LEADER	DTU
CONTRIBUTING PARTNERS	EMIRI
NATURE	Report
AUTHORS	G. Garcia, A. Bhowmik
CONTRIBUTORS	T. Vegge
CONTRACTUAL DEADLINE	M18
DELIVERY DATE TO EC	24/2-2022
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

Today, energy production and transport are evolving fast to meet challenging environmental targets and growing demand. The Achilles' heel is energy storage, which is incapable of providing both low-cost and high-performance solutions. The answer is not a simple evolution of existing batteries but disruptive technologies that must be discovered fast. The BIG-MAP vision is to develop a modular, closed-loop infrastructure and methodology to bridge physical insights and data-driven approaches to accelerate the discovery of sustainable battery chemistries and technologies. BIG-MAP's strategy is to integrate machine learning cohesively, computer simulations, and AI-orchestrated experiments and synthesis to accelerate battery materials discovery and optimization. The project will be a lever to create the infrastructural backbone of a versatile and chemistry-neutral European Materials Acceleration Platform, capable of reaching a 10-fold increase in the discovery rate of novel battery materials and interfaces.

This deliverable, included in the WP 1 'Project management, education, exploitation, and outreach', aims at reporting on 'D1.10 - Proceedings from Early Stage Research Seminar'. One of the main objectives of the WP1 is to create and promote a dissemination and training strategy for accelerated AI-based materials discovery to different groups of stakeholders throughout the whole battery chain value and to ensure high visibility and awareness for the project outcome by selecting the right type of communication channels and timing. In this regard, 2022 BIG-MAP has put in place a series of AI workshops to train young researchers in AI development. Specifically, this deliverable reports on the 1st BIG-MAP AI workshop organized by DTU on January 24-25th.

TABLE OF CONTENTS

1. Introduction and purpose

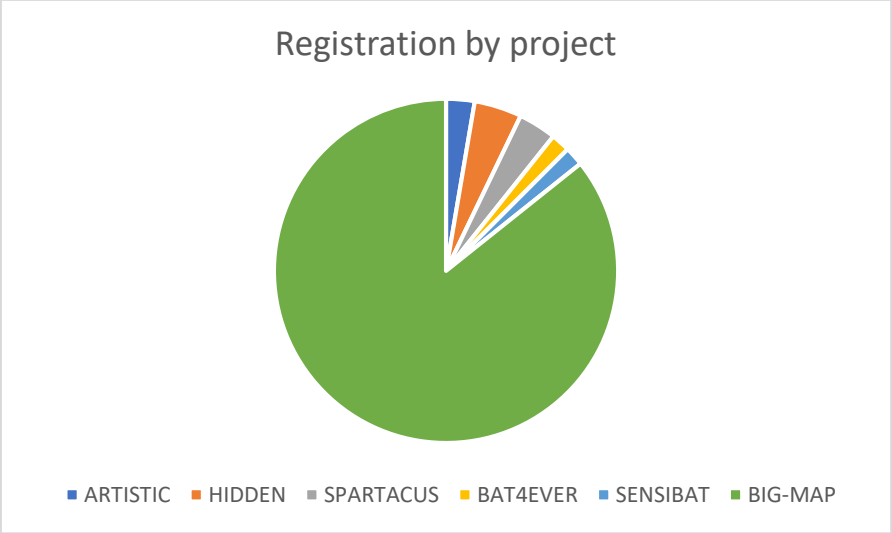
BATTERY 2030+ projects (especially BIG-MAP) target AI toolkits to accelerate the development of new sustainable high-performance batteries. As these models are being developed, the aim is to continually train the BATTERY 2030+ members in effectively using those and integrating AI models in their research practice. To that goal, BIG-MAP has launched a series of AI workshops where young researchers involved in AI development train others through lectures and hands-on tutorials. The first BIG-MAP AI Workshop was organised on January 24-25th as a two-day online event, and it counted with the participation of more than 90 attendees.

2. Organisation

Arghya Bhowmik from DTU was the person responsible for the organisation of this workshop as WP11 lead, with the administrative support from Anne Heglingegård, who is also from DTU. This workshop was promoted among the BIG-MAP community, and it was shared with the BATTERY 2030+ as well. The event had two main parts: the morning session included lectures imparted by the different experts. In the afternoon, participants were divided into smaller groups for a more practical exercise. The event was held mainly over Zoom but the afternoon session combined Zoom with Slack channels, which allowed participants to chat and even call to prepare the homework.

3. Registration and attendance

Among researchers affiliated with BATTERY 2030+ projects (BIG-MAP, HIDDEN, SPARTACUS, SENSIBAT, BAT4EVER), 114 members attended the workshop (with observed peak simultaneous 87 members being online) covering all seniority levels from graduate students to senior professors. Uniquely, the researchers performed the teaching and training – PhDs and postdocs while professors played the role of moderators and teaching assistants.



4. Content and programme

The 1st workshop (January 2022 edition) was conducted online on 24th and 25th January 2022. In the morning, there were three oral presentations in a single track attended by all. After lunch, three parallel hands-on sessions were executed by respective oral presenters and additional tutors. In the afternoon, people are divided into three groups to learn one tool in-depth and how to use it. Each oral presentation was a 30-40 min lecture followed by a discussion totalling 60 min.

Following topics were covered:

(a) Ioan-Bogdan Magdău (Cambridge University) Machine learning potentials. For more information, see Annex 1.

(b) Marc Duquesnoy (Université de Picardie Jules Verne) AI applied to battery manufacturing mesoscale data. For more information see Annex 2.

(c) Jonas Busk (Technical University of Denmark) Uncertainty and sensitivity with surrogate models. For more information see Annex 3.

(d) Eibar Flores (Technical University of Denmark) Discovering governing equations from data with AI. For more information, see Annex 4.

(e) Alexander Sougaard Tygesen (Technical University of Denmark) Multiscale modelling of battery electrodes using DFT and Cluster Expansion. For more information, see Annex 5.

(f) Fuzhan Rahmanian (Karlsruhe Institute of Technology) Orchestrating distributed materials acceleration platforms. For more information, see Annex 6.

5. Feedback

Disregarding positive feedback on usefulness etc., the main constructive feedback was – (a) having a handout that lists all requirements for the workshops beforehand would help save software set up time (b) having serial workshops spread over days so that people can attend many.

6. Conclusion

One of the main objectives of the BIG-MAP WP1 ‘Project management, education, exploitation, and outreach’ is to create and promote a dissemination and training strategy for accelerated AI-based materials discovery to different groups of stakeholders throughout the whole battery chain value and to ensure high visibility and awareness for the project outcome by selecting the right type of communication channels and timing. By organising this workshop, BIG-MAP contributed to a wider and more efficient use and integration of AI models in the research practice of the different projects coordinated by BATTERY 2030+

Annex – Presentations used in the AI Workshop

1) Machine learning potentials

UNIVERSITY OF CAMBRIDGE

Machine Learning Potentials
Challenges in EC/EMC Binary Solvent

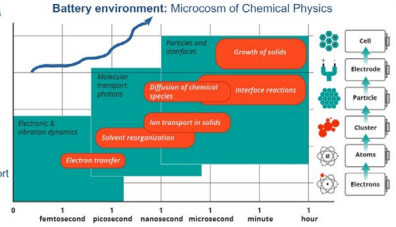
Ioan-Bogdan Magdău
Department of Engineering

24th January 2022 BIG-MAP AI Workshop

The Case of ML in BIG MAP

- Rapid **increase** of degrees of freedom (**DOFs**) with time/length scales
- **Prohibitive** to accurate methods (**ab initio**)
- Often can coarse-grain: **scales are decoupled**
- Sometimes, processes on the **smaller scales are important**: e.g. ion transport
- **ML** methods can **bridge these scales!**

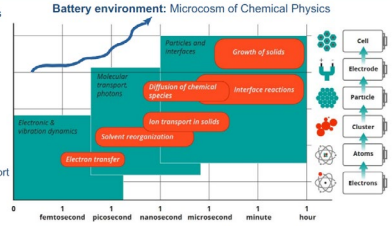
Battery environment: Microcosm of Chemical Physics



The Case of ML in BIG MAP

- Rapid **increase** of degrees of freedom (**DOFs**) with time/length scales
- **Prohibitive** to accurate methods (**ab initio**)
- Often can coarse-grain: **scales are decoupled**
- Sometimes, processes on the **smaller scales are important**: e.g. ion transport
- **ML** methods can **bridge these scales!**

Battery environment: Microcosm of Chemical Physics



ML Interatomic Potentials from Ab Initio Data

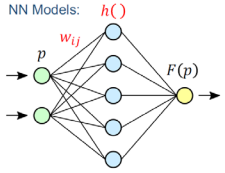
Linear Models:

$$F(p) = \sum_s d_s \, b_s(p)$$

Kernel Models:

$$F(p) = \sum_s d_s \, K(p, p_s)$$

NN Models: $h(\,)$



Mathematical functions: input atomic environments p and output observables $F(p)$ (energies, forces, stresses, dipoles, polarizabilities)

Learning: use train set of $\{p, F(p)\}$ to determine d_s, p_s, w_{ij} etc

---

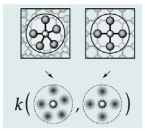
Gaussian Approximation Potential (GAP)

Kernel Model (Gaussian Process Regression):

$$F(p) = \sum_s d_s \, K(p, p_s)$$

Similarity Kernel

$$K(p, p_s) = (p \cdot p_s)^\zeta$$



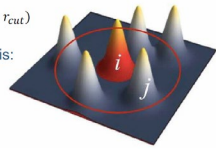
ζ: Kernel power (nonlinearity)
p_s: Sparse Point (selected environments: basis)
p: descriptor of atomic environments: Smooth Overlap of Atomic Positions (SOAP)

Volker L. Deringer, Albert P. Bartók, Noam Bernstein, David M. Wilkins, Michele Ceriotti and Gábor Csányi *Chem. Rev.* 2021, 121, 16, 10073−10141

Smooth Overlap of Atomic Positions (SOAP)

smooth atomic density for each atom i of species α:

$$\rho^{i\alpha}(\boldsymbol{r}) = \sum_j \delta_{\alpha\alpha_j} \exp\left(\frac{-|\boldsymbol{r} - \boldsymbol{r}_{ij}|}{2\sigma_\alpha^2}\right) f_{cut}(r_{ij} > r_{cut})$$



expand into radial and spherical harmonics basis:

$$\rho^{i\alpha}(\boldsymbol{r}) = \sum_{nlm} c_{nlm}^{i\alpha} R_n(r) Y_{lm}(\hat{r})$$

obtain power spectrum (SOAP descriptor):

$$p = p_{nn'l}^{i\alpha\alpha'} = \frac{1}{\sqrt{2l+1}} \sum_m c_{nlm}^{i\alpha} c_{n'lm}^{i\alpha'}$$

Symmetries of SOAP

permutational invariance: $p(\boldsymbol{r}_i) = p(\boldsymbol{r}_{\sigma\{i\}})$

translational invariance: $p(\boldsymbol{r}_i) = p(\boldsymbol{r}_i + \boldsymbol{T})$

rotational invariance: $p(\boldsymbol{r}_i) = p(\hat{\boldsymbol{R}}\,\boldsymbol{r}_i)$

The properties F, we are fitting, have similar invariances:

$$F(\boldsymbol{r}_i) = F(p(\boldsymbol{r}_i)) = F\big(p(\hat{\boldsymbol{R}}\,\boldsymbol{r}_i)\big) = F(\hat{\boldsymbol{R}}\,\boldsymbol{r}_i)$$

therefore this choice of descriptor makes fitting significantly easier

---

GAP Fitting (Afternoon Tutorial)

We solve the model by minimizing the loss:

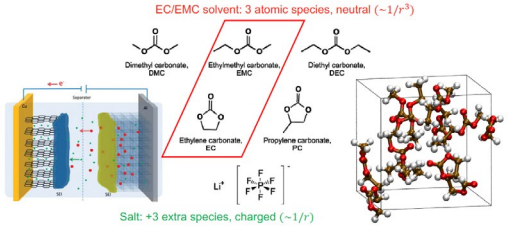
$$\mathcal{L} = \sum_{data} [F_{ab\ initio} - F_{GAP}(p)]^2 + \sigma^2 \sum_{s,s'}^{N_s} d_s K(p_s, p_{s'}) d_{s'}$$

Important hyper-parameters:

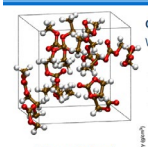
σ: regularization parameter → reduces the risk of overfitting
N_s: sparse points → number of distinct environment geometries used as references
$N_{\max}, l_{\max}$: resolution of radial / spherical harmonics basis
σ_a: width of Gaussians placed on atoms, width of Gaussian radial basis
r_{cut}: neighborhood cut-off for the atomic environment

Battery Electrolytes: Organic Carbonates

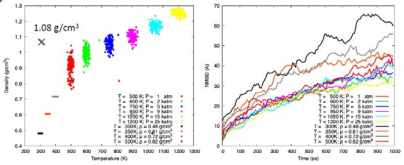
EC/EMC solvent: 3 atomic species, neutral ($\sim 1/r^3$)



Original Training Set



OPLS sampling, DFT calculations **PBE+G06**
Wide range of **densities and temperatures**, maintaining **diffusive** behavior

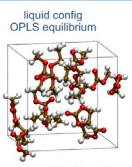


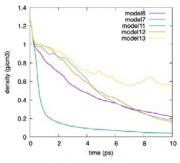
Binary Solvent
33% EC: 66% EMC

Target Density:
~1.08 g/cm³ at 300K

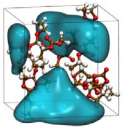
The Density Problem: Possible Issues

liquid config
OPLS equilibrium





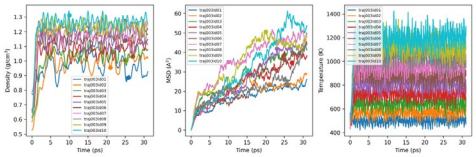
- liquid density collapses
- bubble formation



- OPLS configs not representative for true PES
- Model **pathologies** must be included in training
- Diversity of molecular **compositions** matters
- Intra- / Inter- difference in **scale** and dimensionality

Iterative Training (Gen 5)

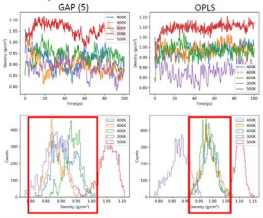
GAP-MD NPT (12 molecules): **500K → 1200K, 1bar → 25kbar**



Proxy for folding back: **density instabilities**

Testing on Larger System (48 molecules)

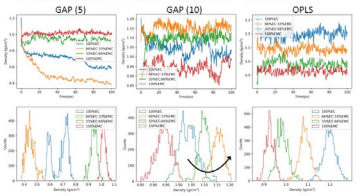
Testing at different temperatures:



- **GAP densities stable at all temperatures!**
- Densities not well reproduced at 400K
- Same density at 400K and 500K

Diversity of Molecular Compositions

Testing different molecular compositions: pure EC → pure EMC



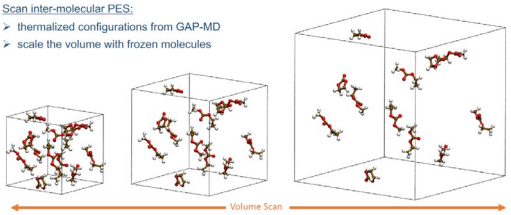
Add **compositions to iterative learning**

Gen 10: **stable densities** across compositions (better 33:66 predictions), **unphysical equilibrium values**

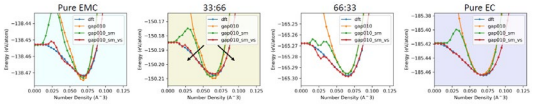
Volume Scans: Pure Inter- Contributions

Scan inter-molecular PES:

- thermalized configurations from GAP-MD
- scale the volume with frozen molecules



Volume Scans: Pure Inter- Contributions



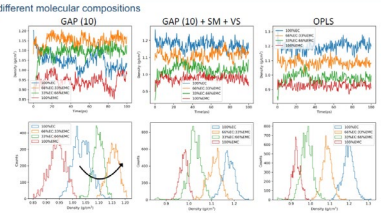
GAP (10)
- describes bottom of the well
- stable density for wrong reasons: underestimates density fluctuations

GAP (10) + Single Molecules (SM)
- reproduces the free molecule limit

GAP (10) + SM + Volume Scans (VS)
- reproduces the entire E(V) curve
- still some oscillatory behavior

Testing 48 Molecules GAP-MD

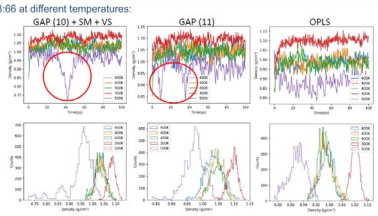
Testing different molecular compositions



Densities stack up in the **correct order!**

Testing 48 Molecules GAP-MD

Testing 33:66 at different temperatures:



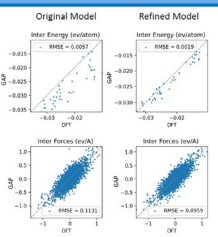
Unphysical density fluctuations 500K: shallower E(V). **Not improved** by iterative training.

Testing Inter-Molecular Interactions: Refining GAP Hyperparameters

Refined GAP model:

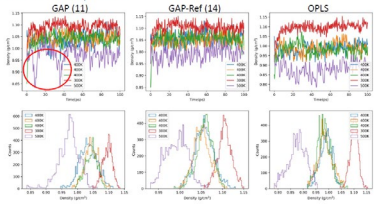
more sparse points → better representation

larger atom sigma → smoother radial basis



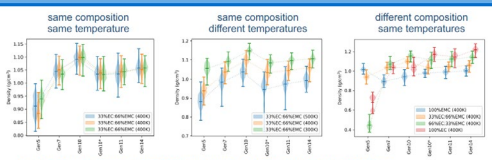
Testing 48 Molecules GAP-MD

Testing 33:66 at different temperatures:



Stable densities for the right reasons. **No unphysical fluctuations!**

Conclusions and Outlook



- NPT testing is crucial, these problems would be hidden in NVT / NVE ensembles
- Iterative training is a must for obtaining a robust potential
- For mixed molecular liquids, variety of molecular composition in the training set matters
- Large volume fluctuations need to be considered to capture subtle intermolecular interactions

Thank you for your attention! Join the GAP fitting tutorial in the afternoon!

2) AI applied to battery manufacturing mesoscale data

1st AI BIG-MAP Workshop

Marc Duquesnoy

Improving Li-ion battery (LIBs) requires a major focus on manufacturing process optimization. In this multi-variable problem, the analysis of experimental and synthetic datasets is crucial to unravel the dependencies between manufacturing parameters and countless electrode properties directly related to electrochemical performances, as demonstrated by the ARTISTIC Project[1]. In this presentation, I demonstrated the application of Artificial Intelligence (AI) to first automatically extract meaningful informations from experimental data to enlarge the manufacturing process knowledge[2], and second to develop AI tools bypassing physics-based models calculations to fast predict simulation results, leading to a drastic reduction of the computational cost[3]. The first study case was the application of un/supervised Machine Learning (ML) algorithms to automatically assess heterogenous coated electrodes based on mass loading and thickness measurements generated in the context of the ALISTORE-ERI Project. Also, I demonstrated in another study case the interest of functional data analysis to study the electrode slurry viscosity from mechanistic simulations, whose main goal is to mimic the manufacturing steps through 3D electrodes generation[4]. We took advantage of such a theoretical presentation in the afternoon to proceed to a training session with students on different experimental and synthetic datasets. The latter enabled students to have practical applications of the detailed ML algorithms on different types of fabrication parameters, from the slurry formulation to the calendering step. Lastly, we talked about the setting up of predictive ML models to provide to students an accurate approach to train supervised regression models, with diverse well-known methods.

[1] Duquesnoy, M. Lombardo, T., Chouchane, M., Primo, E. N., Franco, A. A. (2020) Data-Driven Assessment of Electrode Calendering Process by Combining Experimental Results, In Silico Mesostructures Generation and Machine Learning. Journal of Power Sources. 480, 229103

[2] Duquesnoy, M., Boyano, I., Ganborena, L., Cereijo, P., Ayerbe, E., Franco, A. A. (2020). Machine Learning-Based Assessment of the Impact of the Manufacturing Process on Battery Electrode Heterogeneity. Energy & AI. 100090

[3] Duquesnoy, M., Lombardo, T., Caro, F., Haudiquez, F., Ngandjong, A. C., Xu, J., ... Franco, A. A. (2022). Functional Data-Driven Framework for Fast Forecasting of Electrode Slurry Rheology Simulated by Molecular Dynamics. ArXiv preprint arXiv:2201.04394.

[4] Lombardo, T. Caro, F., Ngandjong A. C., Hook J-B., Duquesnoy, M. Delepine, J-C., Ponchelet, A., Doison, S., Franco, A. A. (2022) Batteries & Supercaps. [1]

3) Uncertainty and sensitivity with surrogate models

BIG-MAP AI School, January 2022

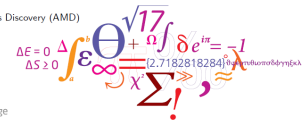
Sensitivity analysis with surrogate models

- Applied to a P2D battery degradation model

Jonas Busk (jbusk@dtu.dk)

Postdoc at Section for Autonomous Materials Discovery (AMD)

DTU Energy



DTU Energy

Department of Energy Conversion and Storage

Acknowledgements

- Joint work with:
  - Williams Agyei Appiah, DTU Energy, WP3
  - Arghya Bhowmik, DTU Energy, WP11
  - Peter Bjørn Jørgensen, DTU Energy, WP11

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957189.

Overview

- Part I: Motivation and background
  - What is sensitivity analysis?
  - Different methods.
  - Local vs. global sensitivity analysis.
  - Sensitivity analysis with surrogate models.
- Part II: Sensitivity analysis for P2D battery degradation model
  - Data collection.
  - Model checking.
  - Preliminary results.
- Introduction to exercises

Part I: Motivation and background

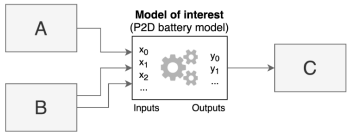
What is sensitivity analysis?

"The study of how uncertainty in the output of a system can be allocated to different sources of uncertainty in its inputs."

- Applications:
  - Identify inputs that have no significant effect on the output and can therefore be fixed or removed to simplify the system.
  - Analyse the effect of the inputs on the output to better understand the system.
  - Identify inputs that cause significant uncertainty in the output in order to reduce uncertainty in the system.

Sensitivity analysis in BIG-MAP

- Multi-scale modelling: Models feed into each other and uncertainty accumulates down stream.



- Sensitivity analysis can help reduce uncertainty in the overall system.

Sensitivity analysis methods

- Methods in increasing order of complexity / evaluation cost.
- **Screening**: Coarse sorting of the most influential inputs / identify non-influential inputs.
  - Useful for large number of inputs.
  - Vary each input one at a time and observe effect on output (one-at-a-time (OAT) method).
  - Simplify the model by ignoring unimportant inputs.
- **Importance measures**: Compute feature importance or sensitivity indices.
  - Measure correlations of sampled input-output pairs.
  - Fit a simple model and analyse model coefficients (linear regression, LASSO, random forest).
- **Deep exploration**: Measure effects of inputs along their entire range of variation.
  - Visual inspection of input-output pairs.
  - Fit a surrogate model that fits the data and has good prediction capabilities.
  - Analyse the model and use it to generate more data (e.g. predict unobserved inputs).

Iooss B., Lemaître P. (2015) A Review on Global Sensitivity Analysis Methods. In: Dellino G., Meloni C. (eds) Uncertainty Management in Simulation-Optimization of Complex Systems. Operations Research/Computer Science Interfaces Series, vol 59. Springer, Boston, MA. https://doi.org/10.1007/978-1-4899-7547-8_5

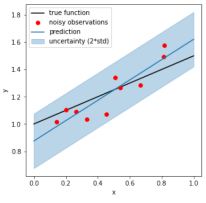
Sensitivity analysis methods

- Local vs. global sensitivity analysis methods.
- **Local sensitivity analysis** is concerned with how a single set of input parameters affects the output.
  - Apply small perturbations of inputs around their nominal value (one-at-a-time (OAT) method).
  - Estimate the partial derivative of an output wrt. the inputs at a specific point.
  - (Does not consider how the system behaves in other regions of input space.)
- **Global sensitivity analysis** aims to understand the behaviour of the overall system.
  - Overcome limitations of local methods by considering the whole range of variation of the inputs.
  - Can be achieved by fitting a (cheap) surrogate model on observed input-output pairs and analysing the surrogate model instead of the original system.
  - The surrogate model can then be used for getting local insights as well.

Iooss B., Lemaître P. (2015) A Review on Global Sensitivity Analysis Methods. In: Dellino G., Meloni C. (eds) Uncertainty Management in Simulation-Optimization of Complex Systems. Operations Research/Computer Science Interfaces Series, vol 59. Springer, Boston, MA. https://doi.org/10.1007/978-1-4899-7547-8_5

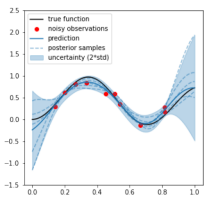
Sensitivity analysis with surrogate models

- Linear regression model:
  - $y(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + \epsilon$ where $\epsilon \sim \mathcal{N}(0, \sigma^2)$.
  - Or $p(y|\mathbf{x}, \theta) = \mathcal{N}(\mathbf{w}^T\mathbf{x}, \sigma^2)$.
  - Normalisation of the input data enables direct comparison of the regression weights.
  - Feature importance can be measured by the size of the regression weights or the absolute t-statistic: $t_d = \text{mean}(w_d)/\text{std}(w_d)$.
  - Test linear hypothesis.
- Pros: Simple to use and easy to interpret.
- Cons: Linear. Not very useful for local sensitivity analysis as gradients are the same everywhere.
- The linear model can serve as a useful baseline!



Sensitivity analysis with surrogate models

- Gaussian process (GP) regression is a popular choice of surrogate model for sensitivity analysis.
  - $f \sim GP(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$, where $m(\mathbf{x})$ is the mean function and $k(\mathbf{x}, \mathbf{x}')$ is the covariance function also known as the kernel.
  - Often used with the squared exponential kernel: $k_{SE}(x, x') = \sigma_f^2 \exp\left(-\frac{(x-x')^2}{2l_d^2}\right)$, where σ_f^2 is the variance parameter and l_d is the length scale parameter.
- Pros: Flexible and data efficient. Uncertainty estimates. Differentiable with some kernels.
- Cons: Risk of overfitting. Expensive to fit large datasets (although there are sparse GPs).



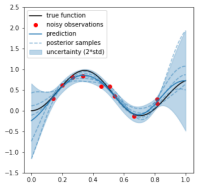
Carl Edward Rasmussen and Chris Williams, Gaussian Processes for Machine Learning, the MIT Press, 2006

Sensitivity analysis with surrogate models

- We want to measure the expected change in the output as a function of the input features.
  - Empirically computed as the root mean squared derivative of the function f w.r.t. the inputs:

$$s_j^f = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(\frac{\partial f(\mathbf{x}_n)}{\partial x_{n,j}}\right)^2}$$

  - Can be computed analytically for some kernels or with automatic differentiation.



Rasmussen PM, Madsen KH, Lund TE, Hansen LK. Visualization of nonlinear kernel models in neuroimaging by sensitivity maps. Neuroimage. 2011 Apr;55(3):1120-1131. DOI: 10.1016/j.neuroimage.2010.12.035. PMID: 21168511.

K. Blix, G. Camps-Valls and R. Jenssen, "Gaussian Process Sensitivity Analysis for Oceanic Chlorophyll Estimation," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 10, no. 4, pp. 1265-1277, April 2017, doi: 10.1109/JSTARS.2016.2641583.

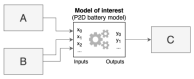
Sensitivity analysis with surrogate models

- Feature importance in GPs with Automatic Relevance Determination (ARD):
  - Squared exponential kernel (RBF) with separate length scale parameter for each input.
  - Feature importance measured by inverse length scale.
  - Has been shown to overestimate the importance of inputs with nonlinear effects.

  Paananen T, Piironen J, Andersen MR, Vehtari A. Variable selection for Gaussian processes via sensitivity analysis of the posterior predictive distribution. In The 22nd International Conference on Artificial Intelligence and Statistics 2019 Apr 11 (pp. 1743-1752). PMLR.
- Other methods for measuring feature importance:
  - Linear regression.
  - LASSO (least absolute shrinkage and selection operator).
  - Decision tree models (random forest).
- Sensitivity analysis with neural network surrogate model:
  - Differentiable.
  - Data hungry.

Part I summary

- *"Sensitivity analysis is the study of how uncertainty in the output of a system can be allocated to different sources of uncertainty in its inputs."*
- Applied to simplify, understand, or improve a system.
- Screening, importance, and exploration methods.
- Local vs. global methods.
- Sensitivity analysis with surrogate models:
  - Linear regression.
  - Gaussian process regression.



Part II: Sensitivity analysis for P2D battery degradation model

Sensitivity analysis for P2D battery degradation model

- Sensitivity analysis recipe:
  - Understand the data.
  - Fit a simple baseline model (Bayesian linear regression).
  - Fit a more flexible nonlinear model (Gaussian process regression).
  - Sensitivity analysis of the model.

Sensitivity analysis for P2D battery degradation model

- The dataset consists of inputs and outputs of a battery degradation model. More specifically, a pseudo-two-dimensional (P2D) model configured to simulate the formation of the solid electrolyte interphase (SEI) 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 assume that main product forming the SEI layer is Li_2CO_3 and it is formed according to the reaction:

$$S + 2Li^+ + 2e^- \rightarrow P$$

- where S is the solvent species and P is the product of the reaction between the solvent ant the Li ions. The growth of the SEI layer is assumed to be in one-dimension and to be controlled by the kinetics of the reaction occurring at the interphase.

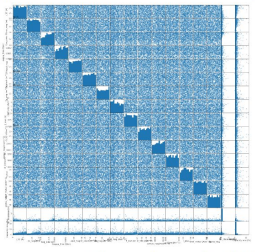
Sensitivity analysis for P2D battery degradation model

#	Parameter	Description	Low	High	Nominal value	Unit
0	I_1C	Applied current	0.13	6.5	1.3	A
1	rp_neg	Particle radius	1×10^{-6}	1.1×10^{-5}	5.5×10^{-6}	µm
2	Eeq_side	Equilibrium potential (SEI)	0	0.4	0.4	V
3	kappa_film	SEI film conductivity	1×10^{-6}	2.4×10^{-4}	2.4×10^{-4}	S/m
4	epsl_neg	Porosity of anode	0.23	0.4	0.3	-
5	Dl	Electrolyte diffusion coefficient	1.5×10^{-10}	7.5×10^{-10}	3.75×10^{-10}	m2/s
6	Ds_neg	Diffusion coefficient of Li in Anode	1×10^{-14}	1×10^{-13}	3.6×10^{-14}	m2/s
7	i0ref_neg	Anode exchange current density	0.36	3.6	0.96	A/m2
8	E_min	Minimum cut-off voltage	0.0	0.1	0.05	V
9	i0_SEI	SEI exchange current density	8×10^{-8}	1.5×10^{-6}	4.5×10^{-7}	A/m2
10	csmax_neg	Maximum Li ion concentration in anode	2.9×10^4	3.3×10^4	3.15×10^4	mol/m3
11	cl_0	Initial electrolyte concentration	1000	1200	1150	mol/m3
12	t_plus	Transference number	0.25	0.43	0.363	-
13	i0ref_metal	Li metal exchange current density	50	100	100	A/m2
14	sigma_neg	Anode conductivity	50	100	100	S/m

Table: Input parameters of P2D battery degradation model.

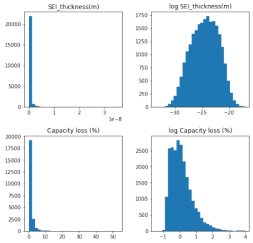
Sensitivity analysis for P2D battery degradation model

- Sets of inputs were sampled uniformly at random within the input ranges.
- Then the P2D model was used to compute the output labels of each row.

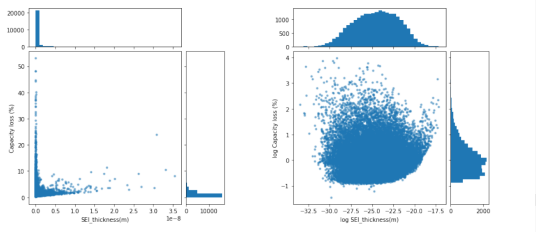


Sensitivity analysis for P2D battery degradation model

- Outputs of interest:
- **SEI thickness (m)**: Thickness of the solid electrolyte interphase (SEI).
- **Capacity loss (%)**: Loss of capacity due to SEI formation.



Sensitivity analysis for P2D battery degradation model

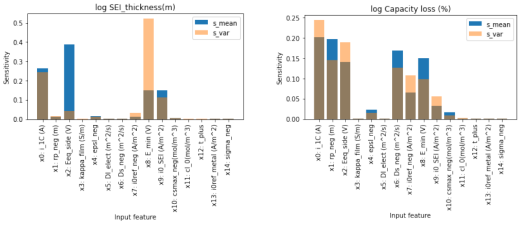


Preliminary results: Model evaluation

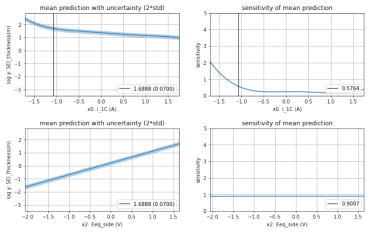
Coefficient of determination (R^2) values evaluating the model fit to the data.

Model	Data split	N	R^2 ↑	
			log SEI thickness	log Capacity loss
Bayesian linear regression	Training	10k	0.9581	0.7055
Bayesian linear regression	Validation	10k	0.9582	0.7022
GP regression w. RBF	Training	10k	0.9958	0.9728
GP regression w. RBF	Validation	10k	0.9956	0.9711

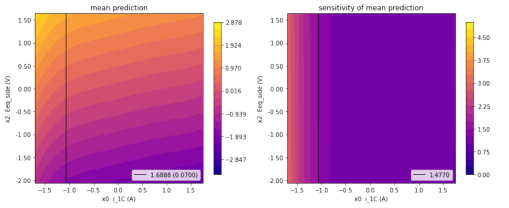
Preliminary results: Global sensitivity analysis of GPs



Preliminary results: SEI thickness, local 1D



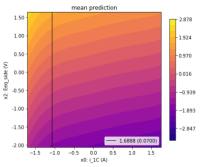
Preliminary results: SEI thickness, local 2D



Demonstration of interactive sensitivity analysis

Part II summary

- Demonstrated sensitivity analysis on a practical example of a P2D battery degradation model.
  - Data collection.
  - Model checking.
  - Preliminary results.
- Join the exercise session if you want to try out this analysis!



Introduction to exercises

Available at:

https://github.com/BIG-MAP/sensitivity_analysis_tutorial

4) Discovering governing equations from data with AI

Engineering better performing Li-ion battery electrolytes requires balancing compromises among multiple desirable properties, where the electrolyte conductivity is critical. In this multi-objective optimization problem, a fast and accurate model describing ionic transport on practical electrolyte formulations is highly desired. In this work, 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. We demonstrate the emergence of an expression outstanding for being simple, accurate, consistent, and generalizable. Notably, even if found from a purely statistical approach, the expression inherits functional aspects from established thermodynamic laws, indicating our model to be grounded on the physical mechanisms underpinning electrolyte transport. Our approach demonstrates the potential of using machine learning to find accurate and physically-sound surrogate models in complex systems without established physico-chemical theories.

5) Multiscale modelling of battery electrodes using DFT and Cluster Expansion

Modeling battery electrode materials using *ab initio* methods is just the first step of the way towards simulating the performance. In many cases, *ab initio* is too restricted by the length scales it allows. Multiscale methods allow for bridging simulations at different length scales. In the case of the Cluster Expansion method, we can simulate disordered materials starting from *ab initio* calculations which are limited to a few hundred atoms, and venture into the thousands and even millions of atoms. This allows us to explore a completely different set of properties, which only manifest on the longer length scales.

6) Orchestrating distributed materials acceleration platform

BIGMAP 1st AI-Workshop

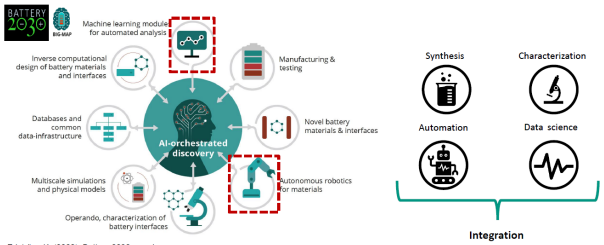
Fuzhan Rahmanian

PhD Candidate

Karlsruhe Institute of Technology (KIT), Institute of Physical Chemistry, Germany
Helmholtz Institute of Ulm (HIU), Germany

Orchestrating distributed materials acceleration platform

Integration is the precursor for acceleration

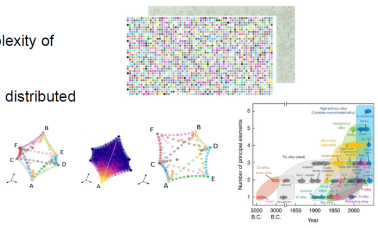


Edström, K. (2020). Battery 2030+ roadmap.

2 2/15/2022 Fuzhan Rahmanian Applied Electrochemistry, HIU & IPC, KIT

Material Acceleration Platforms (MAPs)

- Acceleration through complexity of materials optimization
- Necessity for orchestrating distributed research instruments



1. Brian Rohr, Helge S. Stein et al. Chem. Sci. 2020, 11, 2696-2706
2. Oh, H. S., Kim, S. J., Odbadrakh, K., Ryu, W. H., Yoon, K. N., Mu, S., ... & Park, E. S. (2019). Engineering atomic-level complexity in high-entropy and complex concentrated alloys. *Nature communications*, *10*(1), 1-8.
3. Helge S. Stein, Dan Guevarra, Aniketa Shinde, Ryan Jones, John M. Gregoire, Joel A. Haber, Functional mapping reveals mechanistic clusters for OER catalysis across (Cu-Mn-Ta-Co-Sn-Fe)Ox composition and pH space, Materials Horizons, 2019

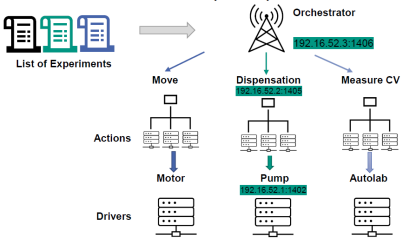
3 2/15/2022 Fuzhan Rahmanian Applied Electrochemistry, HIU & IPC, KIT

Limitation in the current MAP

- Capability of autonomous workflows on only single computer instrument pairing
- Complex software dependencies & heavyweight implementation
- Incompatibility of some platforms with the open-source development of version-controlled software
- Lack of data acquisition beyond FAIR guidelines.
- Lack of orchestrating multiple instruments

4 2/15/2022 Fuzhan Rahmanian Applied Electrochemistry, HIU & IPC, KIT

Hierarchical Experimental Laboratory Automation and Orchestration framework (HELAO)



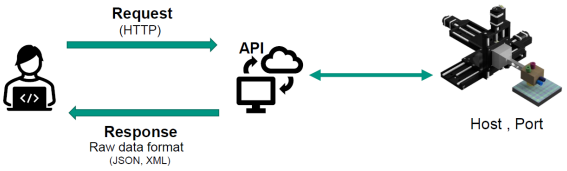
- Language: Python
- Communication: FastAPI
- Design: Hierarchical

Rahmanian, F., Flowers, J., Guevarra, D., Richter, M., Fichtner, M., Gregoire, J. M., & Stein, H. S. (2021). Enabling modular autonomous feedback-loops in materials science through hierarchical experimental laboratory automation and orchestration.
helgestein/helao-pub: hierachical automation of the natural sciences (github.com)

5 2/15/2022 Fuzhan Rahmanian Applied Electrochemistry, HIU & IPC, KIT

Why web-server communications? Why API?

Communication between various software components



6 2/15/2022 Fuzhan Rahmanian Applied Electrochemistry, HIU & IPC, KIT

Why web-server communications? Why API?

Asynchronous Web Server

- Immediate response to the API call
- Multiple async calls will be completed in some order and not at the same time

- Ease in sharing data
- Accessibility by any programming language
- Reusability of devices across laboratories

7 2/15/2022 Fuzhan Rahmanian Applied Electrochemistry, HIU & IPC, KIT

Why FASTAPI and how it works?

- Automatic data validation
- Auto documentation
- Extremely fast even with multithreading

FastAPI

8 2/15/2022 Fuzhan Rahmanian Applied Electrochemistry, HIU & IPC, KIT

What can FASTAPI do for us ? FastAPI

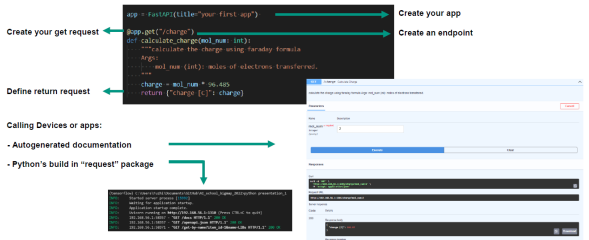
- Installation requirement: fastapi, uvicorn, pydantic
- Pydantic: data validation by using type annotation

- Creates request with an endpoint
  - **GET** : ask for something to do
  - **POST** : creates something new
  - **PUT**: update/modify something existed
  - **DELETE**: remove something

9 2/15/2022 Fuzhan Rahmanian Applied Electrochemistry, HIU & IPC, KIT

Tutorial 2

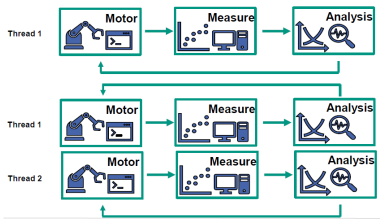
What can FASTAPI do for us? An Example



10 2/15/2022 Fuzhan Rahmanian Applied Electrochemistry, HIU & IPC, KIT

What can FASTAPI do for us? An Example

Sequential Vs. Parallel



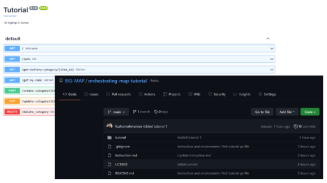
- Blocking
- Async & await
- FastAPI

What can FASTAPI do for us ? FastAPI

A simple live example: Our lab ☺

Tutorial 1:

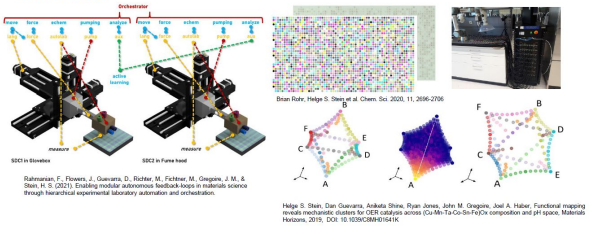
- Show the basic functionality of FastAPI
- How to talk with several servers
- How to write an asynchronous server



BIG-MAP/orchestrating-map-tutorial: This tutorial for orchestrating distributed materials acceleration platform was prepared for the BIG-MAP AI School held in January 2022 by Fuzhan Rahmanian and Jack Flowers (github.com)

How can we accelerate more ?

Active learning !



Rahmanian, F., Flowers, J., Guevarra, D., Richter, M., Fichtner, M., Gregoire, J. M., & Stein, H. S. (2021). Enabling modular autonomous feedback-loops in materials science through hierarchical experimental laboratory automation and orchestration.

Helge S. Stein, Dan Guevarra, Ankkita Shinde, Ryan Jones, John M. Gregoire, Joel A. Haber, Functional mapping reveals mechanistic clusters for OER catalysis across (Cu-Mn-Ta-Co-Sn-Fe)Ox composition and pH space, Materials Horizons, 2019, DOI: 10.1039/C9MH01641K

Active learning: Motivation

Supervised learning

Too much data to label

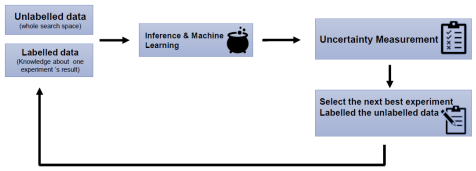
?

unsupervised learning

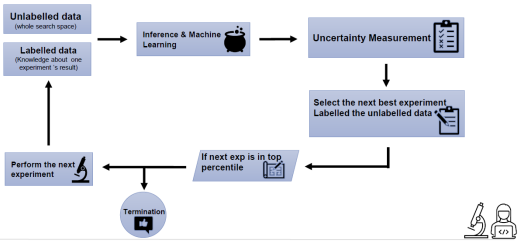
Not enough data

Active learning
Deep active learning
Reinforcement learning
Deep reinforcement learning
Semi-supervised learning

Active learning: How it works?



Active learning in laboratory experiments



AL: How to sample?

Uncertainty sampling

1. Least confidence $s(\theta, x) = 1 - \max_{y \in Y} p_\theta(y|x)$
2. Smallest margin $s(\theta, x) = p_\theta(y_m|x) - p_\theta(y_n|n)$
3. Entropy $s(\theta, x) = -\sum_{y \in Y} p_\theta(y|x) \log p_\theta(y|x)$

AL: How to sample?

Uncertainty sampling

Least confidence $s(\theta, x) = argmin\left(1 - \max_{y \in y} p_\theta(\hat{y}|x)\right) = argmax(p_\theta(y|x))$

$s(\theta, x_1) = 0.8$
$s(\theta, x_2) = 0.76$
$s(\theta, x_3) = 0.45$ ✓

	FCC	BCC	HCP
x_1	0.8	0.15	0.05
x_2	0.76	0.08	0.16
x_3	0.45	0.35	0.2

AL: How to sample?

Uncertainty sampling

Smallest margin $s(\theta, x) = p_\theta(y_m|x) - p_\theta(y_n|n)$

$s(\theta, x_1) = 0.8 - 0.15 = 0.65$
$s(\theta, x_2) = 0.74 - 0.16 = 0.58$
$s(\theta, x_3) = 0.45 - 0.35 = 0.1 = 0.55$ ✓

	FCC	BCC	HCP
x_1	0.8	0.15	0.05
x_2	0.76	0.08	0.16
x_3	0.45	0.35	0.2

AL: How to sample?

Uncertainty sampling

Entropy $s(\theta, x) = -\sum_{y \in Y} p_\theta(y|x) \log p_\theta(y|x)$

	FCC	BCC	HCP
x_1	0.8	0.15	0.05
x_2	0.76	0.08	0.16
x_3	0.45	0.35	0.2

$s(\theta, x_1) = -0.8 * \log(0.8) - 0.15 * \log(0.15) - 0.05 * \log(0.05) = 0.266$
$s(\theta, x_2) = -0.74 * \log(0.74) - 0.08 * \log(0.08) - 0.16 * \log(0.16) = 0.319$
$s(\theta, x_3) = -0.45 * \log(0.45) - 0.35 * \log(0.35) - 0.2 * \log(0.2) = 0.455$ ✓

This sampling can be overconfident ! But is easy !

AL: How to sample?

Query By Committee

- Train an ensemble of various models
- Uncertainty can be measured based on votings

Optimization is more accurate! But expensive and does not reduce the expected error!

Variance reduction

Optimization is successfull! But expensive and sometimes difficult to get!

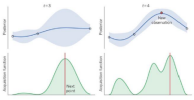
AL: How to sample?

More common for material science community

Lower confidence Bound

$$f(x) = \alpha\mu + (1-\alpha)\sigma^2 \qquad 0 < \alpha < 1$$

$$x^* = \arg\max\big(f(x)\big)$$

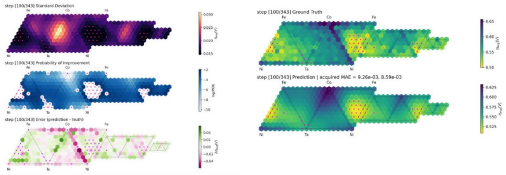


- Good when the objective function is unknown
- Good for noisy measurement
- Good when the whole search space evaluation is expensive

Rohr, B., Stein, H. S., Guevarra, D., Wang, Y., Haber, J. A., Aykol, M., ... & Gregoire, J. M. (2020). Benchmarking the acceleration of materials discovery by sequential learning. Chemical science, 11(10), 2696-2706.
Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., & De Freitas, N. (2015). Taking the human out of the loop: A review of Bayesian optimization. Proceedings of the IEEE, 104(1), 148-175.

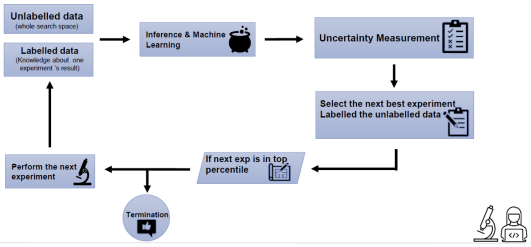
AL: How to sample?

Still a lot more … **POI, EI, Thompson**,..



Choice of aqf depends on the selected ML model

Active learning in laboratory experiments



Tutorial 3

Supervised Machine Learning

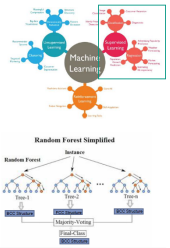
- Classification task: LR, DT, NB, RF, SVM, KNN, LDA,…
- Regression task: LR, NN, DT, SGD, Ensemble models,…
- Common python package: Scikit-learn

How to choose a right model ?

- Type of data, statistical analysis, bias, tuning, feature importance, ease of implementation,..

Tutorial 2&3:

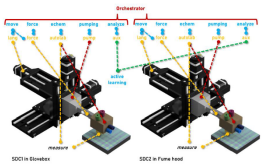
- Example for writing machine learning model (i.e. RF and GP) & how to hypertune
- Write an active learning algorithm with different uncertainty sampling



Accelerating experiments with active learning

Requirements:

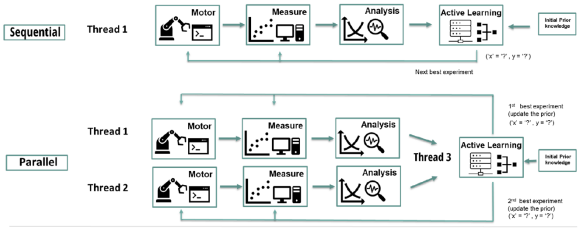
- Automatic data analysis
- ML algorithms
- Data accessibility (without human intervention)
- Possibility of parameter alteration for each experiment



Rahmanian, F., Flowers, J., Guevarra, D., Richter, M., Fichtner, M., Gregoire, J. M., & Stein, H. S. (2021). Enabling modular autonomous feedback-loops in materials science through hierarchical experimental laboratory automation and orchestration.

Accelerating experiments with active learning

Sequential Vs. Parallel



Accelerating experiments with active learning

Sequential Vs. Parallel