An open dataset from the Large Plasma Device for machine learning and profile prediction

Phil Travis[1], Troy Carter[1,2]

1. UCLA Department of Physics and Astronomy
2. Oak Ridge National Laboratory

phil@physics.ucla.edu

UCLA · BaPSF BASIC PLASMA SCIENCE FACILITY · U.S. DEPARTMENT OF ENERGY Office of Science · NSF

GitHub repository
https://github.com/physicistphil/lapd-isat-predict

1 Motivation: develop tools to infer trends autonomously

Trend inference can be done using machine learning

- Long-term goal: self-optimizing fusion reactors
- Short-term goal: infer trends in Isat in various random mirror-like configurations for operations optimization

Releasing an open dataset for ML applications

- Collected first-of-its-kind diverse dataset
- LAPD configurations were randomly generated and sampled from a set of actuators
- A free, open dataset is useful for benchmarking ML architectures and for use in plasma physics education

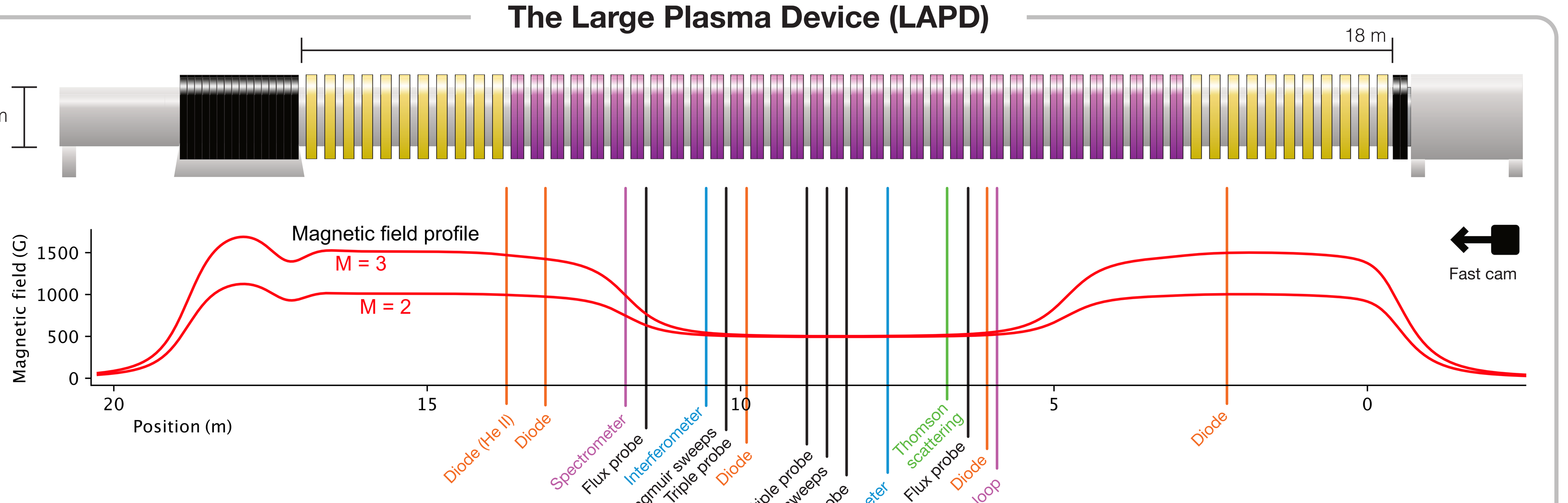
Short term: infer trends in how to operate the LAPD

- Built a machine learning model for Isat at any point in the machine
- Mirror machine, discharge voltage, and gas puff duration trends agree with intuition
- Can optimize for strongest and weakest axial variation of Isat
  - Trend generally agrees

Next: infer / optimize transport

- Many more diagnostics were recorded, including cross-field particle flux

2 The Large Plasma Device (LAPD)



- 19.7m long, 1m diameter
- T_e ~ 0.1 - 30 eV
- n_e ~ 10^{12} - 5 x 10^{13} cm^{-3}
- Flexible magnetic geometry — good for mirror studies
- **Many diagnostics** (permanent or moveable)
- **1 Hz shot rate** – up to 31 million shots per year (have 15m+ shots recorded right now)
- **Data-rich environment** — ideal for ML

3 Collecting a comprehensive diagnostic set

Machine state information (MSI)

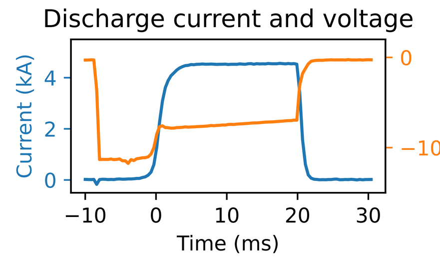
- Discharge current, voltage
- Total gas pressure
- RGA partial pressures
- Axial magnetic field
- Heater current, voltage

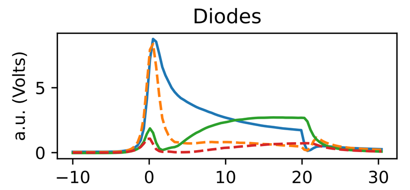
Fixed diagnostics

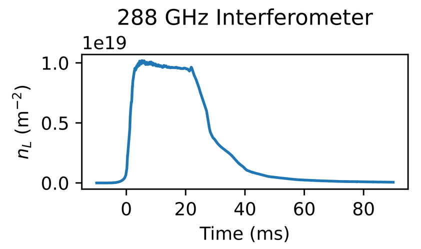
- Thomson scattering
- Interferometers (x2)
- Light diodes (x5)
- Fast framing camera
- Diamagnetic loop
- Monochromator (DR2 only)

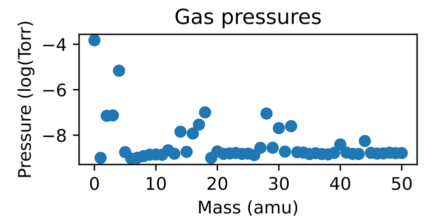
Probe diagnostics

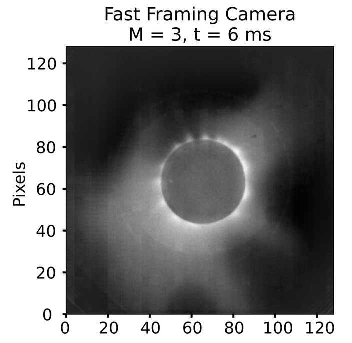
- Langmuir sweeps – Te
- Triple probes
  - **Isat**, Te, Vf
- Flux probes
  - **Isat**, Vf x2
- **Isat ~ n_e √Te**

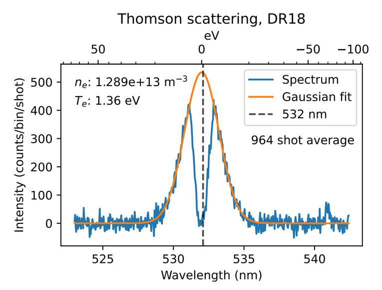












- Two run weeks: **DR1** (Feb 2023), **DR2** (Apr 2024)
- DR2 differences
  - Added monochromators (667, 707, 587 nm)
  - Four different probe positions
  - Two additional gas puff durations (5 ,10 ms in addition to 38 ms)

Random machine configurations

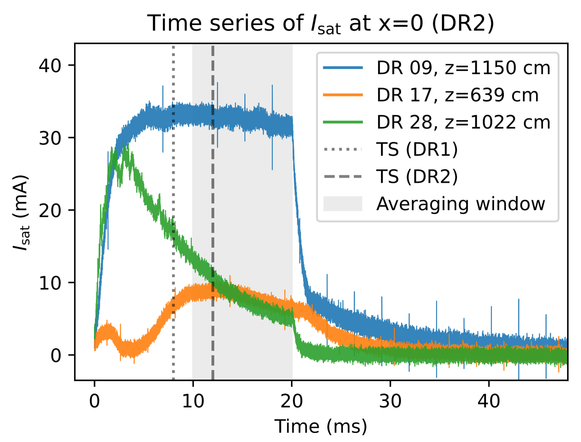
- Used Latin-hypercube sampling (LHS) for efficient coverage of machine actuator space
- Actuators changed
  - Source field
  - Mirror field
  - Midplane field
  - Discharge voltage
  - Gas puff voltage
  - Gas puff duration

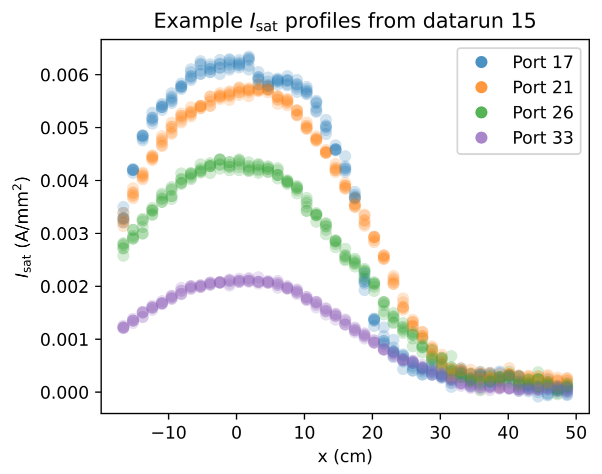
4 Training data: examples, balance, and bias

Description

- Inputs (12 total)
  - Actuators and machine condition
  - Probe position
- Outputs: time-averaged Isat (over 10-20 ms)
- 67 datarun; 59 train, 8 test
  - Conditions for most dataruns are **randomly selected**
- Total shot (example) count: **131,550**
- **Complete input range covered**
  - Via LHS
  - 272,160 possible combinations
- However, the input ranges are not covered evenly, e.g.,
  - 82V gas puff
  - 112V, 150V discharge voltages
- Only 6 runs where the gas puff duration was less than 38 ms
- **Test set cannot hit all conditions**
  - Took care to make a representative set

Examples





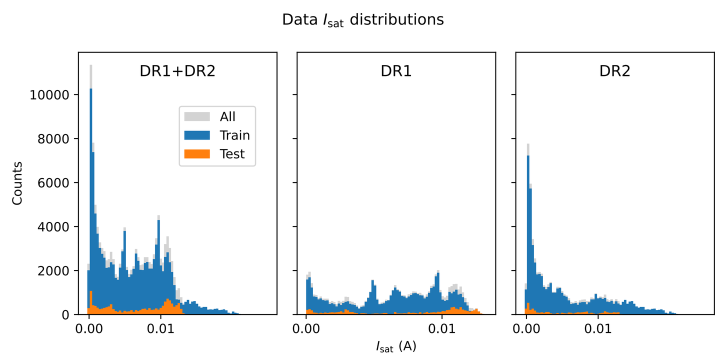
Distribution

Data breakdown by class and dataset (percent)

B source (G)	Train	Test	All	B mirror (G)	Train	Test	All	B midplane (G)	Train	Test	All
500	4.77	0	4.29	250	4.30	8.41	4.72	250	8.25	21.01	9.55
750	3.34	12.61	4.29	500	30.49	8.41	28.23	500	43.80	8.41	40.19
1000	43.13	78.99	46.78	750	6.68	16.81	7.72	750	6.62	52.19	11.27
1250	12.59	0	11.30	1000	28.85	57.97	31.82	1000	26.36	5.78	24.26
1500	19.23	0	17.27	1250	3.34	4.20	3.43	1250	9.24	0	8.30
1750	1.91	0	1.71	1500	26.34	4.20	24.08	1500	5.73	12.61	6.43
2000	15.03	8.41	14.35								

Gas puff voltage (V)	Train	Test	All	Discharge voltage (V)	Train	Test	All	Axial probe position (cm)	Train	Test	All
70	12.11	16.81	12.59	70	12.22	8.41	11.83	639	12.48	8.41	12.06
75	6.68	0	6.00	80	5.25	0	4.72	828	17.07	36.28	19.03
80	11.46	8.41	11.15	90	2.86	8.41	3.43	859	12.48	8.41	12.06
82	41.49	57.97	43.17	100	3.34	8.41	3.86	895	33.01	30.10	32.71
85	14.13	0	12.69	110	8.77	0	7.87	1017	12.48	8.41	12.06
90	14.13	16.81	14.40	112	20.62	0	18.52	1145	12.48	8.41	12.06
				120	3.82	8.41	4.29				
				130	0.95	0	0.86				
				140	2.86	8.41	3.43				
				150	39.30	57.97	41.20				

Gas puff duration (ms)	Train	Test	All	Vertical probe position (cm)	Train	Test	All
38	94.27	91.59	94.00	≈0	36.26	46.08	37.26
< 38	5.73	8.41	6.00	≠ 0	63.74	53.92	62.74



- Expect predicted Isat to be biased to lower values
- Significant difference in Isat distribution between DR1 and DR2

5 Predicting Isat using a dense neural network

Architecture: dense NN

- Dense NN
- 4 layers
- 256 wide
- 5-network ensemble
- Leaky ReLU

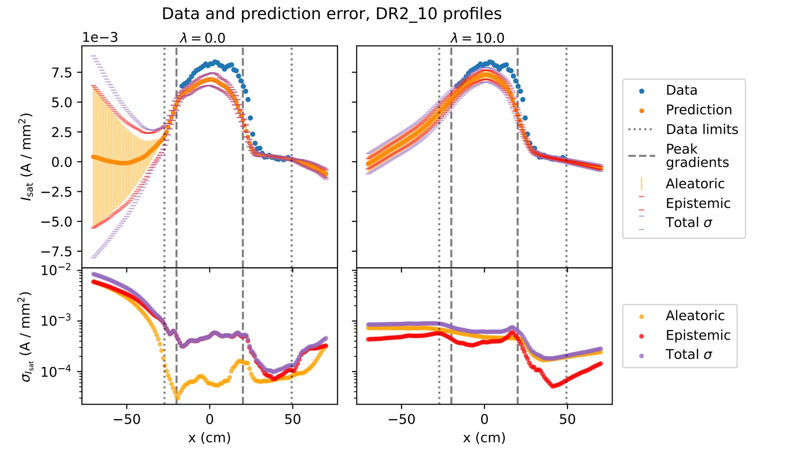
activations

- β-NLL loss
- Adam optimizer
- Epoch^{-1} learning rate decay schedule
- Gradient clipping

$$\mathcal{L}_{\beta\text{-NLL}} = \frac{1}{2}\left(\log\sigma_i^2(\mathbf{x}_n) + \frac{(\mu_i(\mathbf{x}_n) - y_n)^2}{\sigma_i^2(\mathbf{x}_n)}\right)\mathrm{StopGrad}\left(\sigma_i^{2\beta}\right)$$

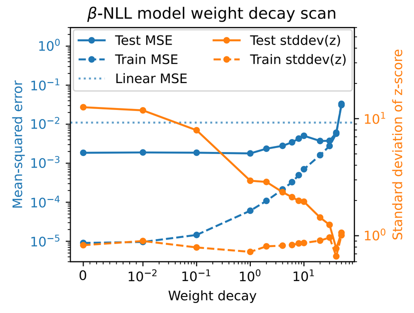
β-NLL loss function

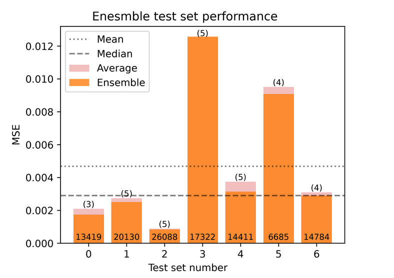
- β-NLL loss did not appear to improve calibration
  - but helped stabilize training
  - may have acted as a mild regularizer



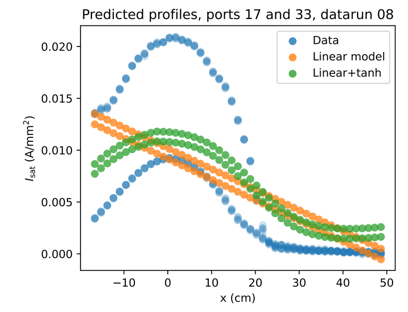
- DR2_10 (a test set datarun) is slightly on the worse side of the predictive ability of the model
- **Model uncertainty grows when predicting outside training data envelope**

Model calibration and validation

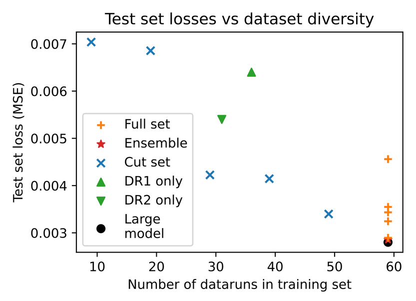




- Can use weight decay to calibrate uncertainty
  - **but relative uncertainty becomes less useful**
- Test set performance changes with test set
  - Hand picked (set 0)
  - Randomly selected (sets 1-6)
- Median RMSE is used as a baseline when plotting predictions with uncertainty



- Trained linear, linear-like models as a baseline
- A nonlinear model is required → NN is a good fit
- **Performance can be improved by:**
  - **Increased diversity** (measured by number of dataruns)
  - Providing more information (run set and top gas puff flag)
  - **Ensembles**
  - Using larger models



6 Inferring trends using the learned model

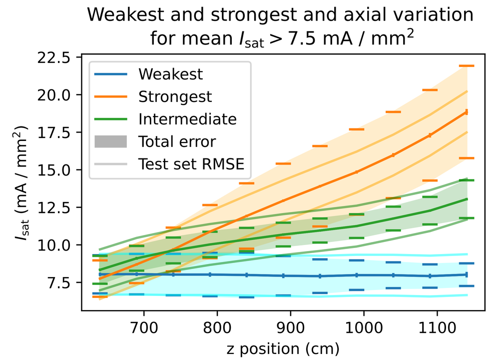
Making predictions

- **Cheap models → comprehensive search is tractable**
- Computed 127M different machine configurations, x5 models
  - Takes 151 seconds on an RTX 3090

Optimizing axial variation

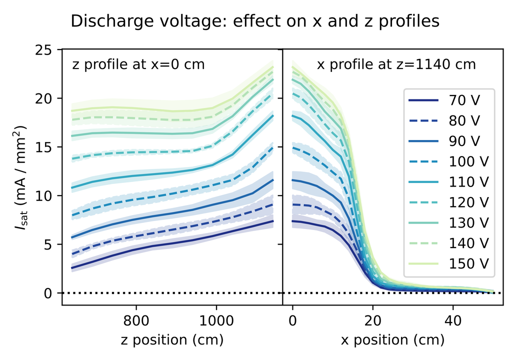
$$\text{Inputs} = \underset{\text{Inputs}\neq z}{\arg\min}\ \underset{z}{\mathrm{sd}}\left(I_{\mathrm{sat}}|_{x=0}\right)$$

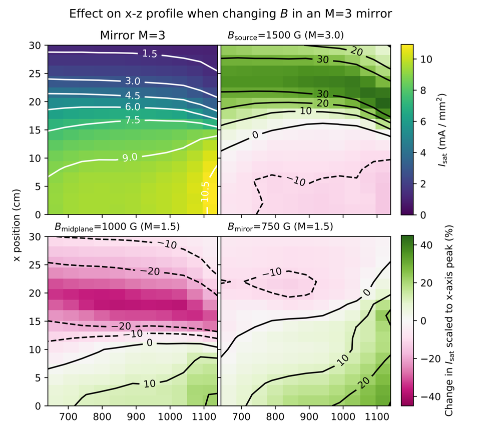
Input or actuator **Red** = tested	Weakest Isat = any	**Weakest** Isat > 7.5	**Strongest** Isat > 7.5
Source field	750 G	1000 G	500 G
Mirror field	1000 G	750 G	500 G
Midplane field	250 G	250 G	1500 G
Gas puff voltage	70 V	75 V	90 V
Discharge voltage	130 V	150 V	150 V
Gas puff duration	5 ms	5 ms	38 ms
Run set flag	on	on	on
Top gas puff flag	on	off	off



- Can find conditions needed strongest and weakest axial variation on-axis
- Uncertainty is very large

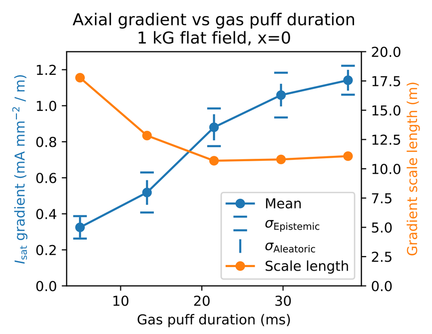
Checking predictions with intuition

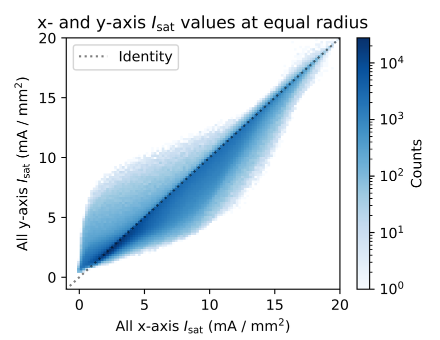




- Isat increases with discharge voltage
- Mirror width is modified as expected with changes in B

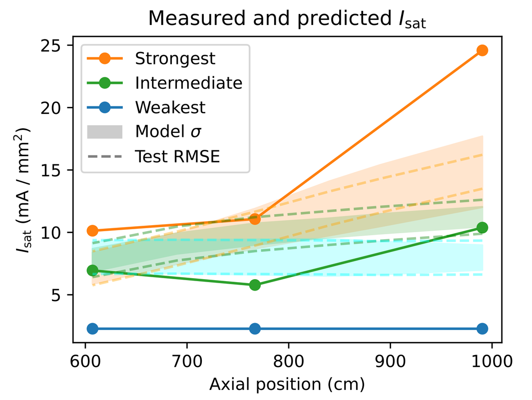
Inferring trends





- Model suggests:
  - Axial gradient scale length increases with a decrease in gas puff duration
  - Intrinsic x-y asymmetry in the data
- Take with a grain of salt (lack of diversity)

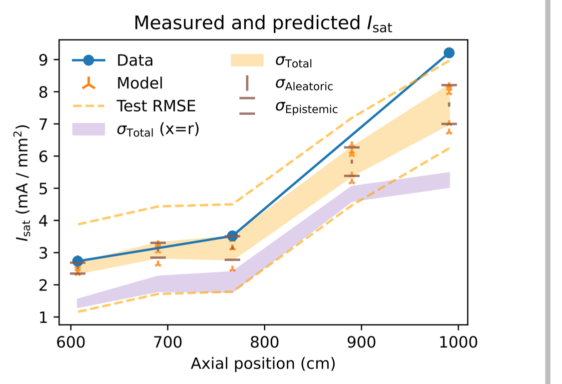
7 Validating the model



Axial variation measured by SD

Case	Predicted	Measured
Strongest	3.42	6.6
Intermediate	1.32	1.95
Weakest	0.04	0.0*

- Strongest and intermediate axial variation are in a range covered by model
- Weakest case is in a new regime not seen by the model
- Probe at 600 cm is beyond model training data
- **Isat calibration is suspect**
  - * preliminary interferometer and Te measurements suggest flat axial profile in weakest case
  - Further work: better Isat calibration
- **Predicting a single diagnostic requiring an absolute calibration is risky**



- **Measured Isat values matched model predictions off-axis**
  - But single shot with low discharge current (odd cathode state)
  - Isat calibration state unknown

8 Summary and future work

- A free, open, diverse dataset was created for machine learning purposes
- A machine learning (ML) model was trained on Isat measurements to infer trends
- Inferred trends agree with intuition
- ML model finds cases with strong and weak axial variation, but absolute values do not agree
- More diagnostics and time-series data will be processed and used as training data
- Additional data collection may be necessary

Acknowledgements

The authors would like to thank the following individuals for diagnostics setup, run support, and mentorship:

- Jia Han
- Tom Look
- Lukas Rovige
- Chris Niemann
- Shreekrishna Tripathi
- Steve Vincena
- Yhoshua Wug
- Zoltan Lucky
- Pat Pribyl
- Tony Peebles
- Ryan Buckley
- Walter Gekelman
- Jacob Bortnik

References

- Thomson scattering: S. Ghazaryan, M. Kaloyan, W. Gekelman, Z. Lucky, S. Vincena, S. K. P. Tripathi, P. Pribyl, and C. Niemann. Thomson scattering on the large plasma device. Review of Scientific Instruments, 93(8):083514, August 2022.
- Uncertainity quantification and estimation:
  - Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. On Calibration of Modern Neural Networks. http://arxiv.org/abs/1706.04599, August 2017. arXiv:1706.04599 [cs].
  - Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles. http://arxiv.org/abs/1612.01474, November 2017. arXiv:1612.01474 [cs, stat].
  - D.A. Nix and A.S. Weigend. Estimating the mean and variance of the target probability distribution. In Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94), pages 55–60 vol.1, Orlando, FL, USA, 1994. IEEE.
  - Maximilian Seitzer, Arash Tavakoli, Dimitrije Antic, and Georg Martius. On the Pitfalls of Heteroscedastic Uncertainty Estimation with Probabilistic Neural Networks. http://arxiv.org/abs/2203.09168, April 2022. arXiv:2203.09168 [cs, stat].
  - Matias Valdenegro-Toro and Daniel Saromo. A Deeper Look into Aleatoric and Epistemic Uncertainty Disentanglement. 2022. arXiv:2204.09308 [cs.LG].
- Gradient clipping: Prem Seetharaman, Gordon Wichern, Bryan Pardo, and Jonathan Le Roux. AutoClip: Adaptive Gradient Clipping for Source Separation Networks. http://arxiv.org/abs/2007.14469, July 2020. arXiv:2007.14469 [cs, eess, stat].

Work was performed at the Basic Plasma Science Facility (BaPSF), UCLA, supported by DOE FES with additional funding from the NSF