RECENT PROGRESS ON ML-BASED OPTIMIZATION AND ANOMALY PREDICTION AT APS

NIKITA KUKLEV (ADVANCED PHOTON SOURCE, ANL, USA)
Ihar Lobach
Yine Sun
Michael Borland
Katherine C. Harkay
Hairong Shang
Greg Fystro
Jonathan Edelen (RadiaSoft)

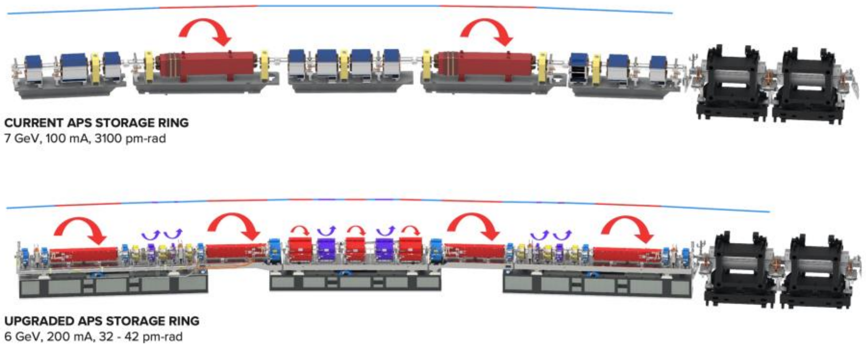
ICFA ML 2022
November 3, 2022

U.S. DEPARTMENT OF ENERGY
Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.
Argonne NATIONAL LABORATORY

APS

APS is undergoing a major upgrade (APS-U)

- New storage ring
- Refurbishment of injector complex
- More beamlines

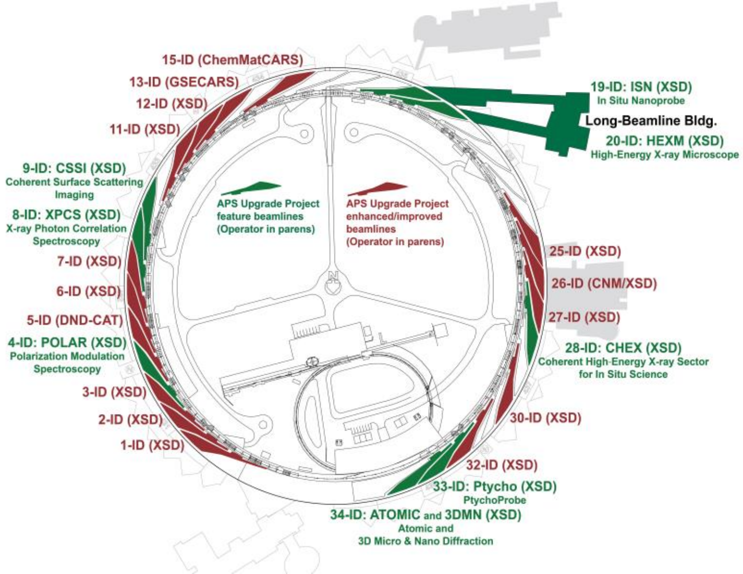


Shutdown April 2023, user operation April 2024

- **Only 3 months** of 24x7 beam commissioning

New opportunities and new challenges

- More challenging physics
- New diagnostics and increased data rates
- A chance to demonstrate ML tools at a critical moment



ML @ APS

Several simultaneous efforts underway

SUPPORT INFRASTRUCTURE	ML RESEARCH	PRODUCTION TOOLS
Data storage and analytics	Optimization algorithms	Expert-usable CLI/scripts
Control, I/O, configuration libraries	Anomaly detection and prediction	Operator-usable GUI tools
	Surrogate models	

OUTLINE

- **Data logging and analytics**
- Anomaly detection/prediction (I. Lobach)
- Adaptive optimization

APS DATA ANALYTICS

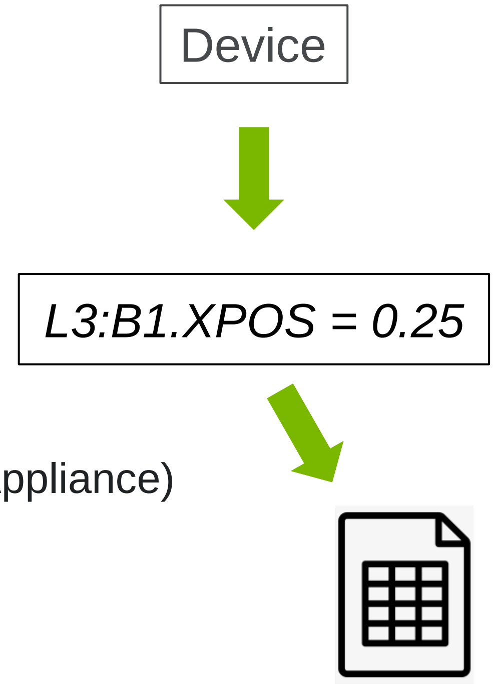
What kind of data does accelerator produce?

Time series of various types

How is data stored?

Read from control system into specialized tools

- APS – monthly compressed per-PV files (sddslogger)
- Many other labs – yearly compressed per-PV files (Archiver Appliance)
- Custom high speed loggers

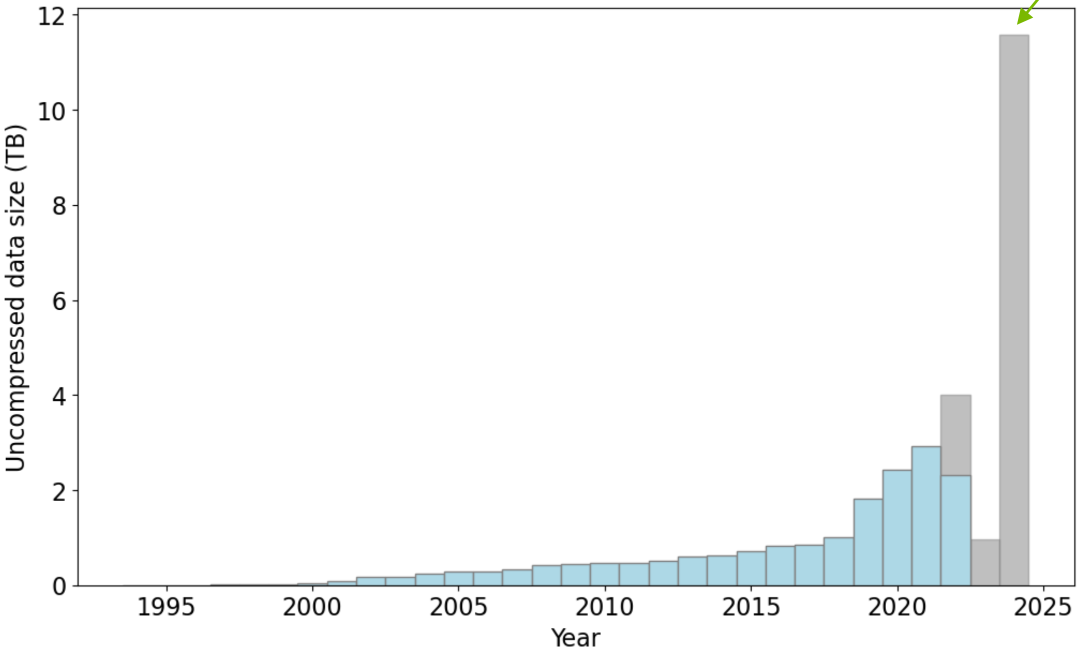


APS DATA ANALYTICS

How much data is there?

So far, not ‘too much’, but growing **A LOT**

New devices, even more ML data



APS DATA ANALYTICS

Current systems well suited to archival storage - why is this a problem for ML?

Different access patterns:

- Broad queries – thousands of PVs
- Analytical postprocessing (min, max, median, …)

Stricter requirements:

- Near-realtime availability
- Fault-tolerant

Solution?

Ads!

Ad · https://www.aps.anl.gov/

Shop XRays® at Argonne® - Official Site

Shop the Black Friday Sale and Save on **APS 7GeV electrons**. Bulk deals available **with no downpayment**. Save Now and Get a **2nC** bunch for a 1nC Price, with a free reinjection on us!

Simulation codes

Rated #1 in storage ring simulation for 10 years

Magnetic measurements

Experts on standby to tell you all the ways you messed up

SOLUTION FOR ML DATA

User data collection created new ‘Big Data’ analytics/time-series databases!

- Column oriented data storage
- Specialized for **fast data ingest, high compression, large mathematical queries**
- Scales to PB of data
- Open-source and (mostly) free

influxdb Timescale ClickHouse

SOLUTION FOR ML DATA

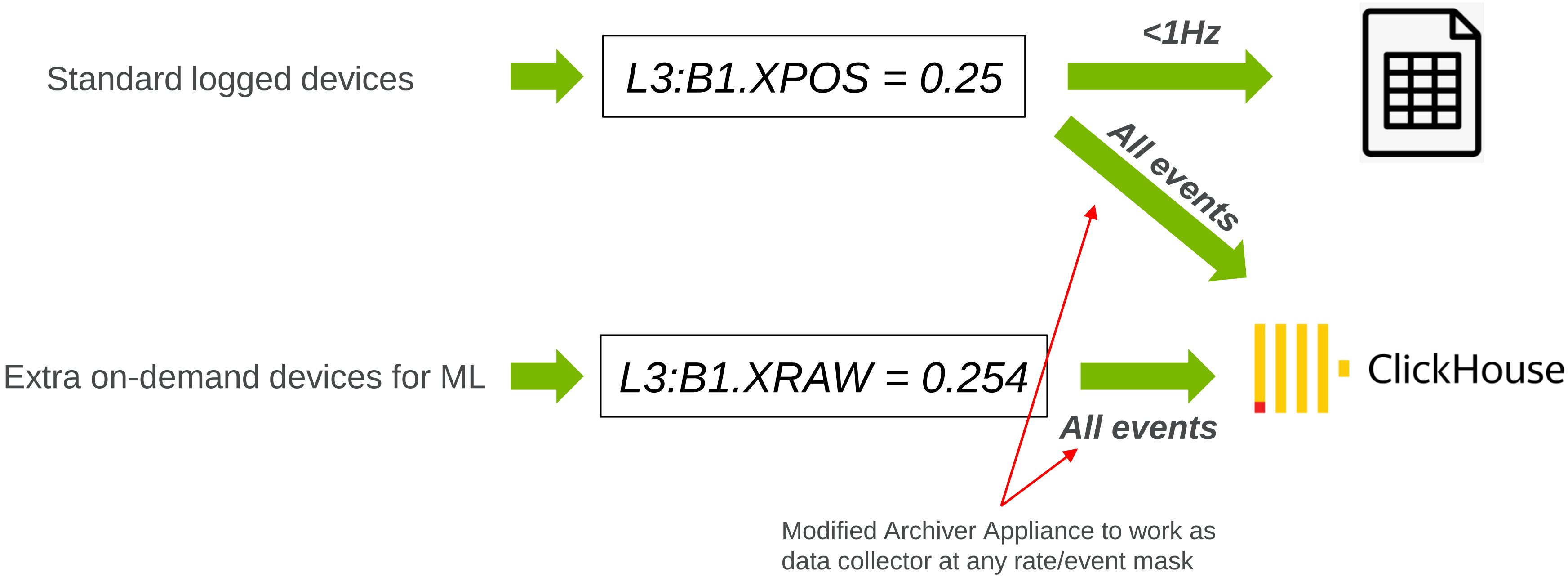
Benchmarked several popular projects on historical data (ask me for details)

- Current choice – **ClickHouse**
  - VERY fast and has SQL support
- How fast is fast? **~20x over existing scripts**
- Example:

```
mcrtest.aps4.anl.gov :) SELECT median(value) from merge('tsdata', 'l3*')

SELECT median(value)
FROM merge('tsdata', 'l3*')

Query id: eaf83ca5-3913-4e31-8193-1821471d3c96

┌──────median(value)─┐
│ 5.974925356446823e-8 │
└────────────────────┘

1 row in set. Elapsed: 176.484 sec. Processed 60.72 billion rows, 485.78 GB (344.07 million rows/s., 2.75 GB/s.)
```

INTEGRATION

Integrated side-by-side with existing tools to minimize disruptions



OUTLINE

- Data logging and analytics
- **Anomaly detection/prediction** (I. Lobach)
- Adaptive optimization

Anomaly Detection and Classification at APS

- Several proof-of-principle experiments for use cases of ML in the Injector Complex, using intentional-perturbations data for training and testing
  https://napac2022.vrws.de/papers/tuye4.pdf
  - Neural network classifier for sources of poor transmission efficiency
  - Autoencoder for anomaly detection in the Particle Accumulator Ring (PAR)
  - β-variational autoencoder for clustering of poor-performance data

In this presentation:

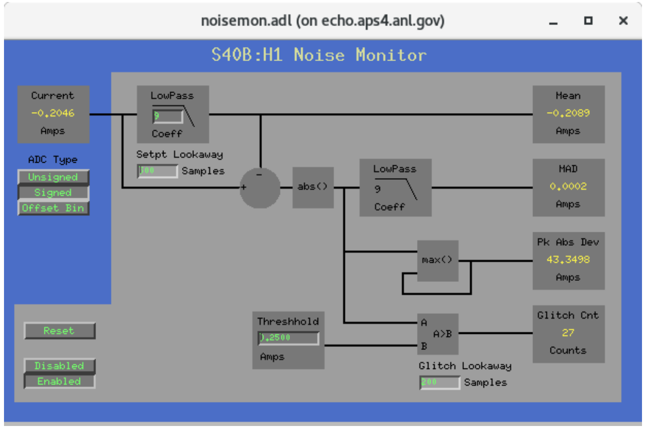
- Program for anomaly detection in the magnet power supplies in the Storage Ring (APS and APS-U)
  https://napac2022.vrws.de/papers/tupa29.pdf

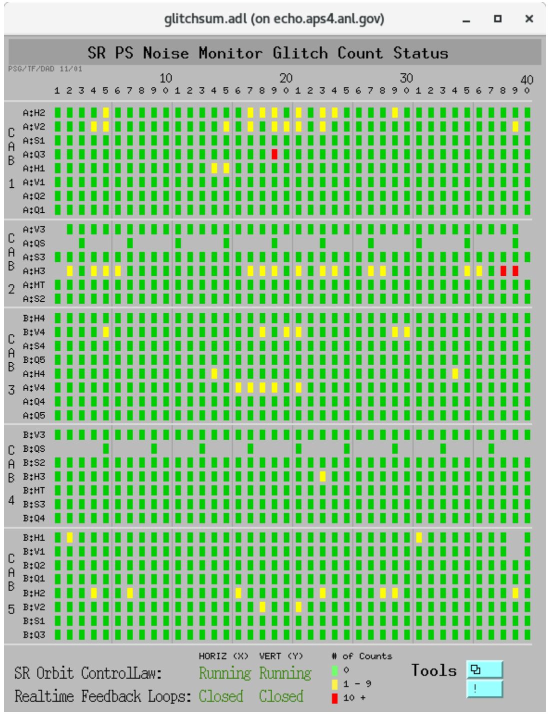
Anomaly Detection in Storage Ring Power Supplies

- Motivation: the detected anomalies may help predict trips or (at least) create a priority list of PSs for maintenance. In turn, this reduces downtime.
- Available historical data:

APS Storage Ring has 40 sectors, each has A and B subsectors. We consider quadrupoles (Q), sextupoles (S), horizontal (H) and vertical (V) correctors (overall, 1320). Example of a magnet name: S40B:H1

We analyzed the APS run history from 2001 to 2022:
Found 629 ring fills that ended with “Int. Dump: End of Period”
Found 149 ring fills that ended with a PS trip, glitch, fault, or problem





Process Variable	Explanation
CurrentAI	PS current
MagTempAI	Magnet temperature
CapTempAI	PS capacitor temperature
IGBTTempAI	PS transistor temperature
OutVoltageAI	PS voltage

1 point per 64 seconds starting from 2008

PS Current Noise Monitor Data

Instantaneous Mean Current

MeanMin, MeanMean, MeanMax in a 10 min window

Instantaneous Mean Absolute Deviation of Current

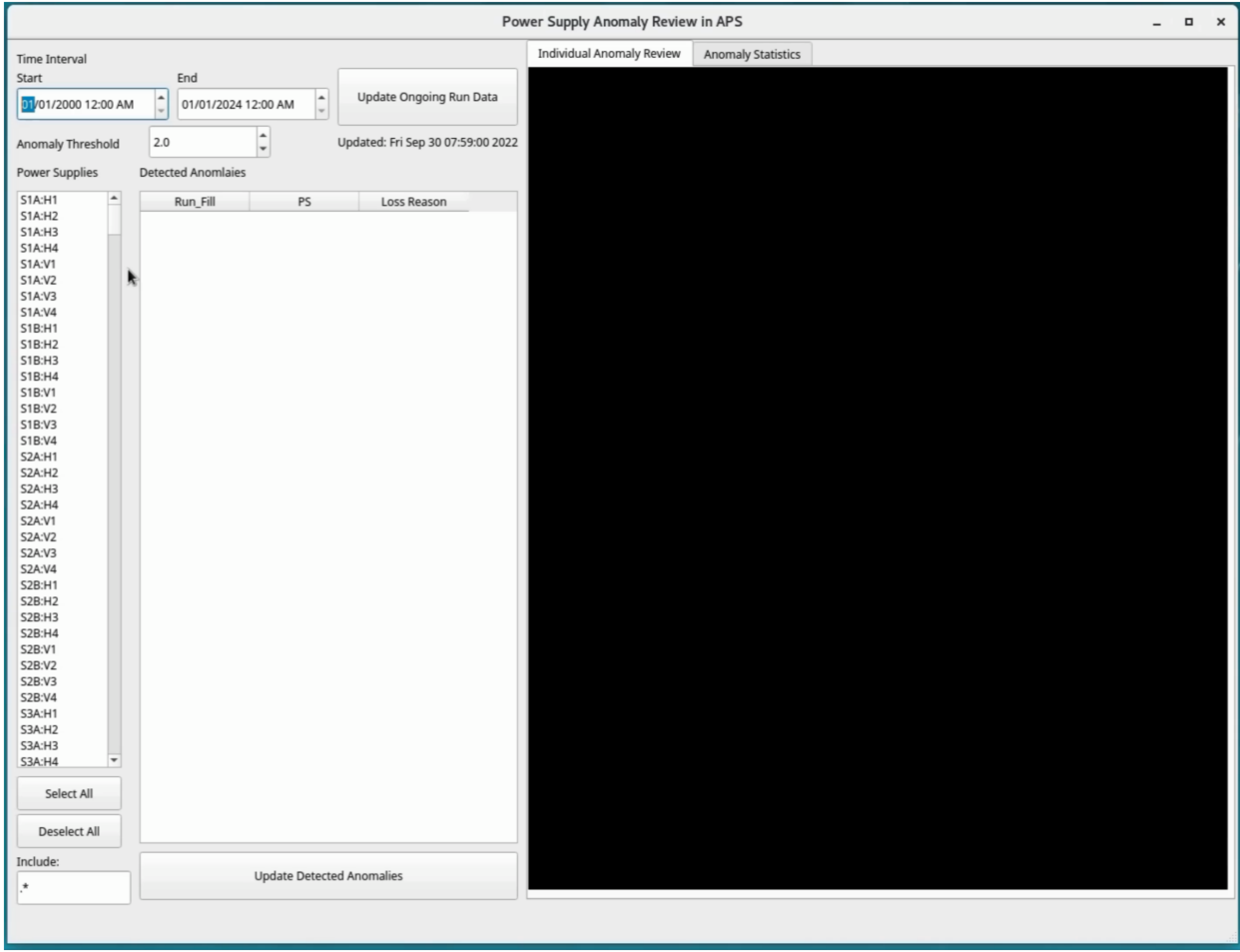
MADMin, MADMean, MADMax in a 10 min window

1 point per 10 minutes starting from 2001

Argonne NATIONAL LABORATORY

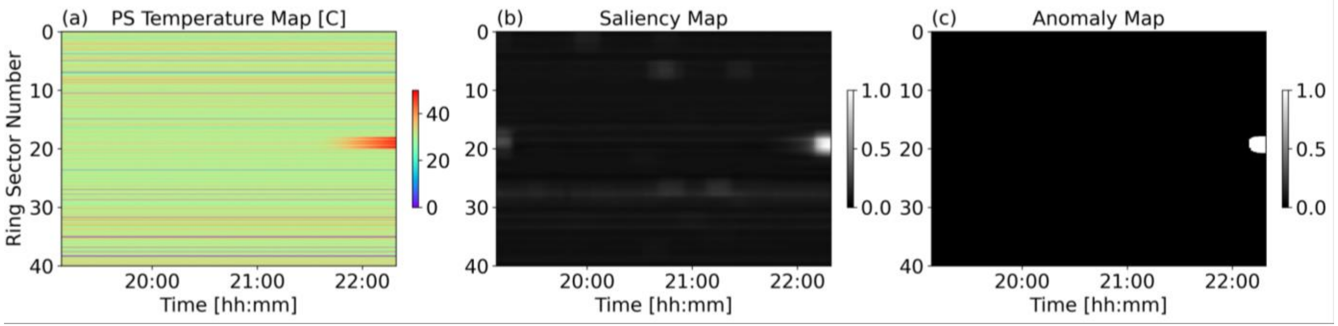
Tool to review past detected anomalies

- Can be used with any anomaly detection method, based on an anomaly score and a threshold.
- Currently, only H and V corrector power supplies
- Using current noise monitor data (MADMax-based approach, but also considered autoencoders, LSTMs)
- Model is retrained before every APS run, on the previous 10 runs (~3 years)
- Anomaly Statistics tab presents the number of fills with anomalies (not the number of anomalous data points)
- https://pypi.org/project/PyQt5/



Argonne NATIONAL LABORATORY

Anomalies in Power Supply Temperatures in the Storage Ring

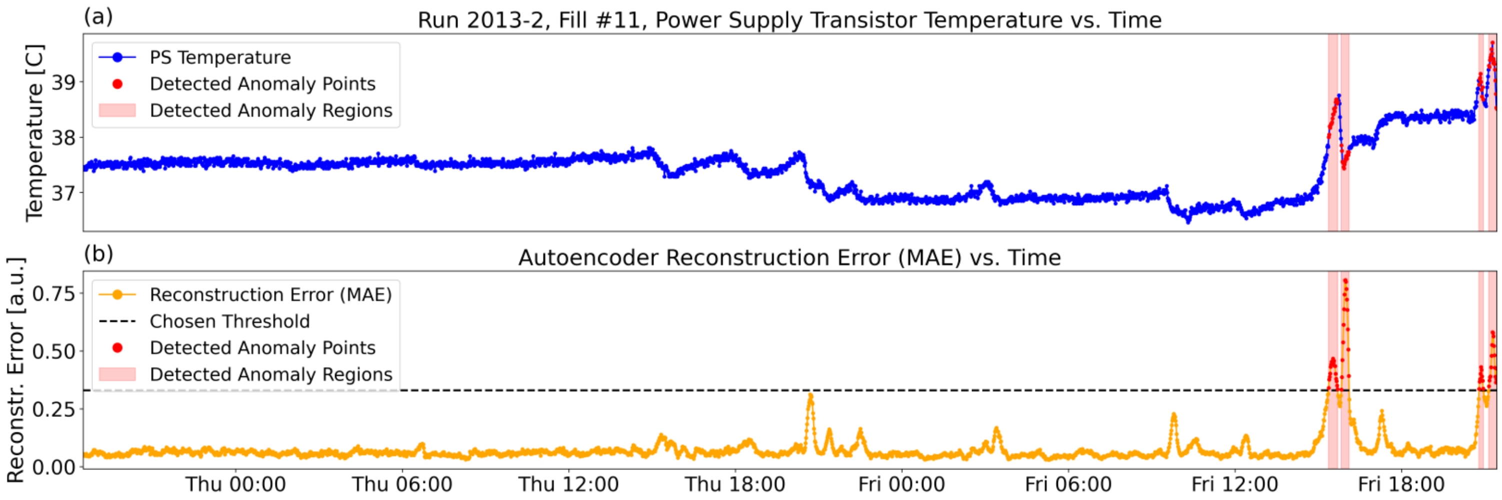


- Detection of anomalies in power supply temperature maps is similar to object detection in images. Therefore, we have used object detection methods. Further, our intern (Alexandre Sannibale) is considering Convolutional Neural Networks for PS temperature maps.
- The presented anomaly was related to a stuck mixing valve for the cooling water. There is one mixing valve per every two neighboring sectors in APS. Therefore, we can see rising temperatures in sectors 19 and 20.

Autoencoder for Temperature of a Single Power Supply (in a time window)

- One can use several contiguous temperature values of a single PS as input for an autoencoder. Such autoencoders can be sensitive to unusual temperature behavior in time (even if the absolute value of temperature is not high enough to trigger an alarm on high temperature)

- Chosen autoencoder architecture:
  20 → 10 → 5 → 10 → 20
- The autoencoder was trained on 10 preceding reference data files.
- This approach did not produce any false positives for this magnet's PS for the entire observation interval from 2008 to 2022.



This fill ended because of a trip in this PS (S16B:S2)

*Threshold = 1.5

Argonne NATIONAL LABORATORY

Anomaly Detection Performance / Outlook

- Certain anomalies in PS temperature maps (e.g., stuck mixing valve) can be detected with 100% precision and zero false positive rate.
- Autoencoders may be able to detect anomalies invisible to less sophisticated algorithms. The temperature anomaly in S16B:S2 was only detected due to its unusual behavior in time. The value of the temperature was not too high.
- We made progress in anomaly detection in the PS Current Noise Monitor data. However, only up to 20% of trips were successfully predicted, the false positive rate was not insignificant. Still, anomalies were up to 500 times more likely in the data leading to a trip, than in the normal-operation data.
- Jonathan Edelen and Ihar Lobach are working on improving the performance with more advanced algorithms, such as LSTM recurrent neural networks, which will be using all available process variables in one model (PS current, voltage, temperatures).
- In APS-U, we will have PS current data at 22 kHz. We have the freedom to do anything with these data. This will likely further improve performance
- Also, in APS-U, the PSs will have unique labels. In APS, they are labeled by magnet names.

OUTLINE

- Data logging and analytics
- Anomaly detection/prediction (I. Lobach)
- **Adaptive optimization**

ACCELERATOR OPTIMIZATION

Many approaches with various tradeoffs

Humans | Classic methods | Artificial Intelligence | Machine Learning | Deep Learning

Scientists
Engineers
Operators

Nelder-Mead Simplex
RCDS
Extremum Seeking

Genetic Algorithm
Particle Swarm

Bayesian Optimization

Reinforcement Learning
Deep Kernel Learning

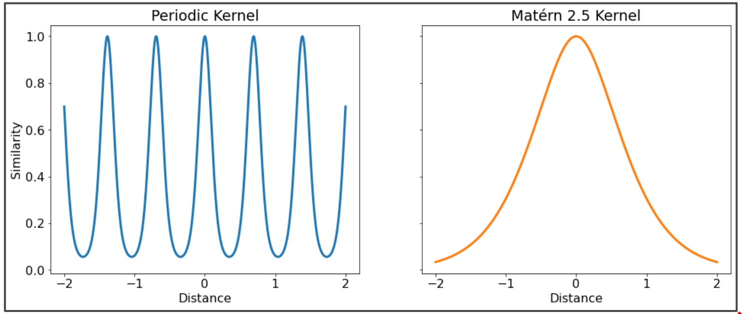
EXISTS | EXISTS | EXISTS | THIS TALK

BAYESIAN OPTIMIZATION

INPUTS

$\mathbf{x}$

MACHINE BLACK BOX



$$\mathbf{y} = f(\mathbf{x}) + \varepsilon$$

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$$

GAUSSIAN PROCESS (GP) MODEL

$$p(\mathbf{f} \mid \mathbf{y}, \mathbf{x})$$

$$\mathrm{UCB}(\mathbf{x}) = \mu(\mathbf{x}) + \sqrt{\beta} * \sigma(\mathbf{x})$$

MAXIMIZE ACQUISITION FUNCTION

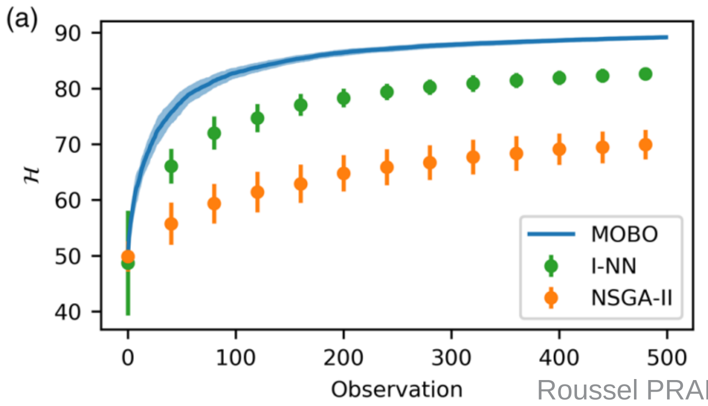
BAYESIAN OPTIMIZATION

Bayesian optimization (BO) a promising method for ***expensive problems***

- Model-based and can encode expert knowledge
- Interpretable and scalable

Previous work showed good performance in *time-invariant* tasks

See also Duris PRL 2020



Roussel PRAB 2021

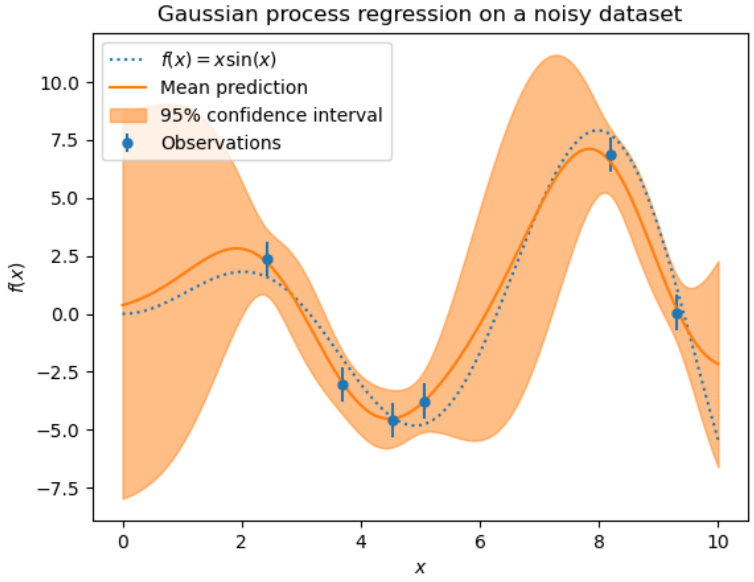
ADAPTIVE BAYESIAN OPTIMIZATION

Many accelerators have **time-dependent performance: *f(t,x)***

- External factors (temperature, etc.)
- Device drift / degradation

A challenge for conventional BO

- Without time model, **drift appears as noise**
- Convergence to average *suboptimal* state
- Common solution – run local optimizer after BO
- Can drift be modelled explicitly?



More details in: https://napac2022.vrws.de/papers/thxd4.pdf

ADAPTIVE BAYESIAN OPTIMIZATION KERNEL

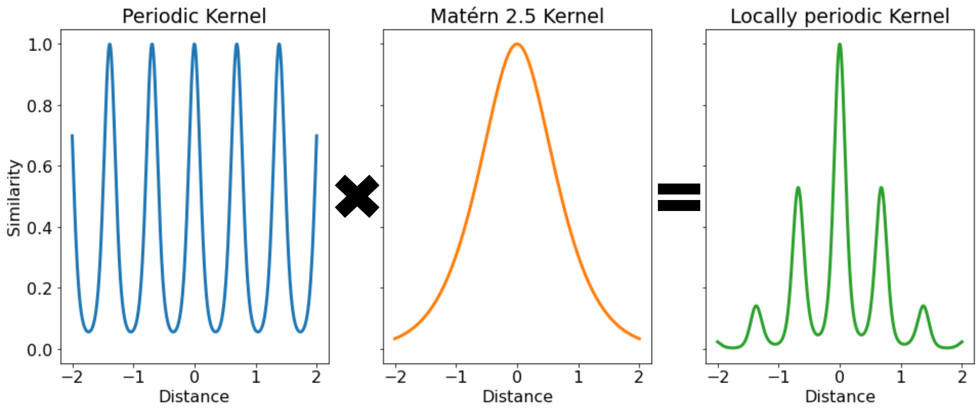
Consider time as another ‘input’ - has **very different properties**

- Periodic on various timescales (minutes, hours, days)
- Overall linear/polynomial trends

Can compose sub-kernels along any subspace - **what is the right one?**

K1 + K2 = LOGICAL **OR**

K1 * K2 = LOGICAL **AND**



ISOTROPIC KERNEL

Simplest choice – **isotropic** kernel

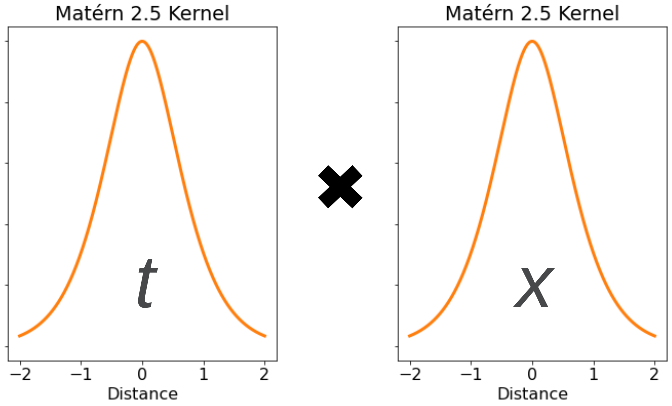
- 'Just add a dimension'

Advantages:

- Robust and simple to implement

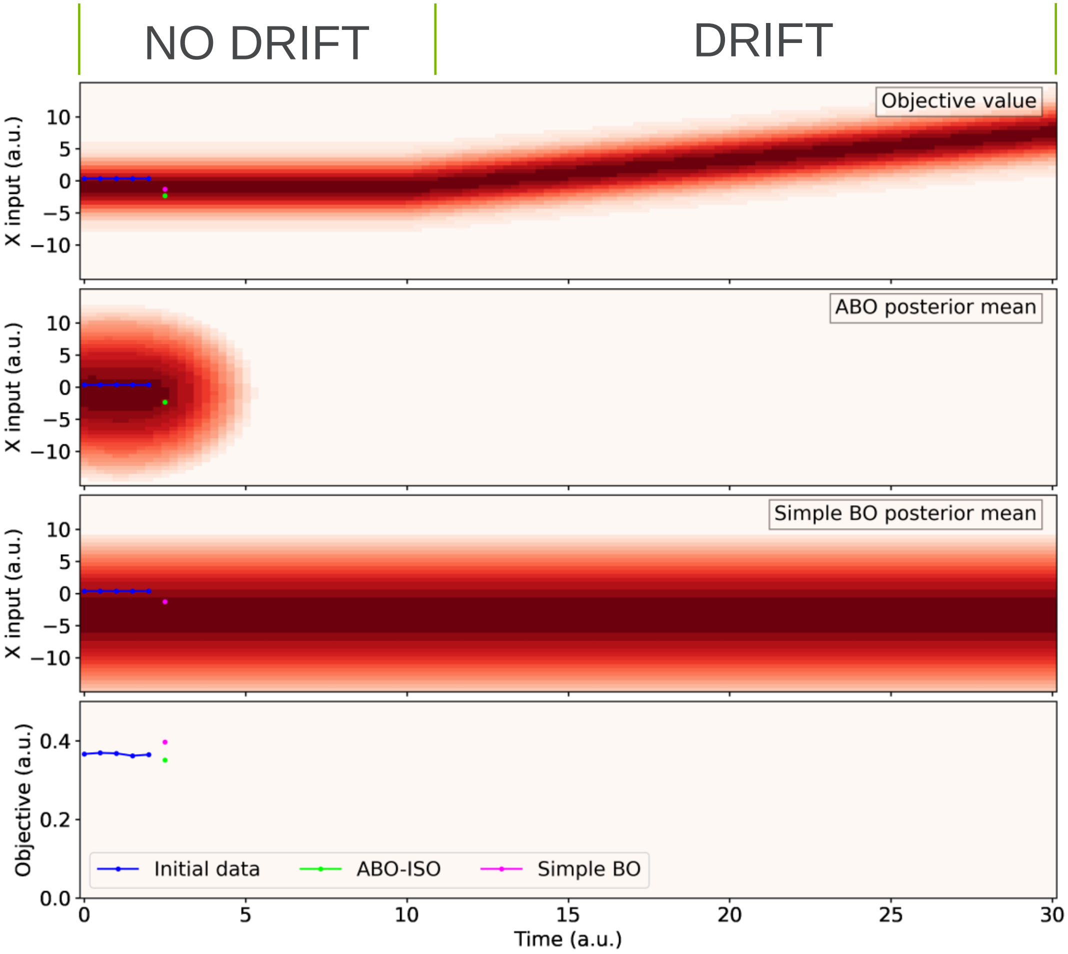
Disadvantages:

- No global structure
- Lags behind changes



ISOTROPIC KERNEL

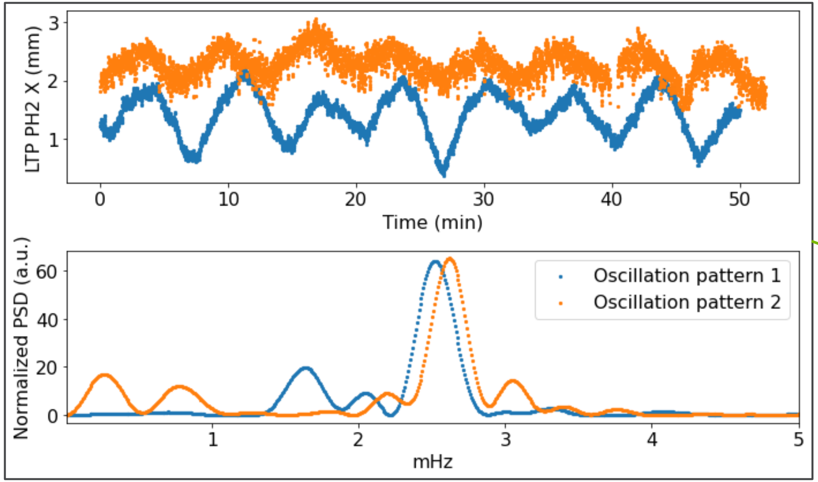
Simulation – corrector drift

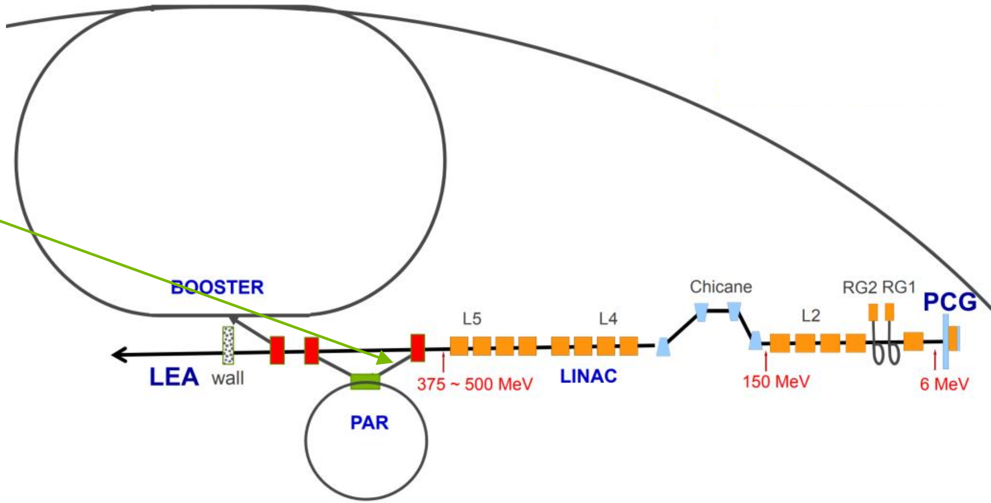


ADAPTIVE ML MOTIVATION @ APS

APS injector supplies beam to storage ring and linac extension area

- **Proportional feedback** used to compensate drifts but has **high jitter**
- Drift spectrum varies day to day - requires **time-aware** and **time-adaptive** control
- Have repeating patterns – want to exploit long-range correlations





ADAPTIVE BAYESIAN OPTIMIZATION KERNEL

Our choice: **Spectral Mixture Kernel**

- Expressed as spectral density of several Gaussians

kernel → $$k(\tau) = \int S(s) e^{2\pi i s^T \tau} ds,$$

$$S(s) = \int k(\tau) e^{-2\pi i s^T \tau} d\tau.$$

spectral density

- Can approximate **any** stationary kernel
  - No need to explicitly specify like periodic/RBF

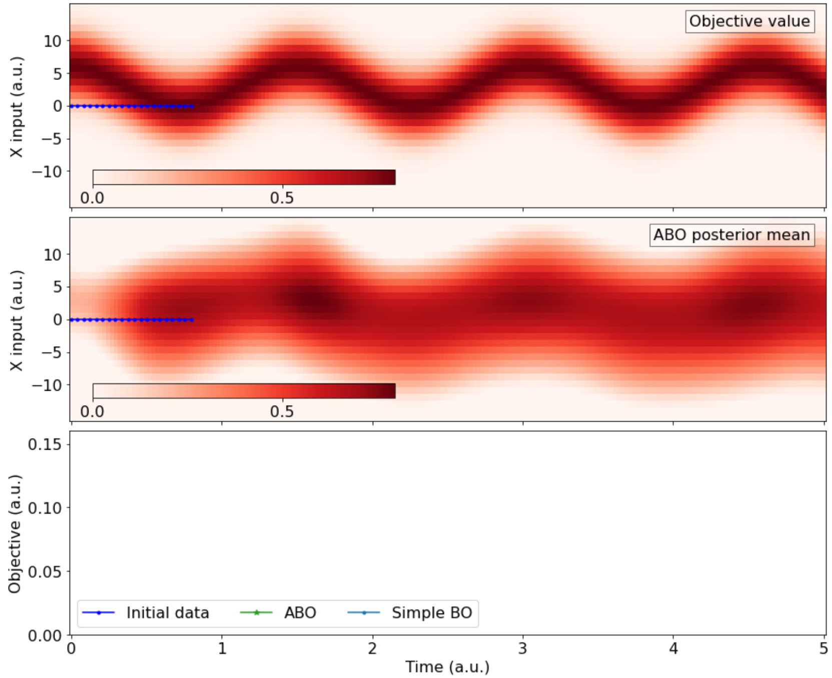
$$S(s) = \sum_{i=1}^{Q} w_i^2 \left[\mathcal{N}(s|\mu_i, \sigma_i^2) + \mathcal{N}(s|-\mu_i, \sigma_i^2) \right].$$

- Starting point for **Deep Kernel Learning**

SM ABO model $k_{ABO}(t, t', x, x') = (k_{SM}(t, t') + k_I(t, t')) \times \sigma^2 k_{Mt}(\mathbf{x}, \mathbf{x}')$

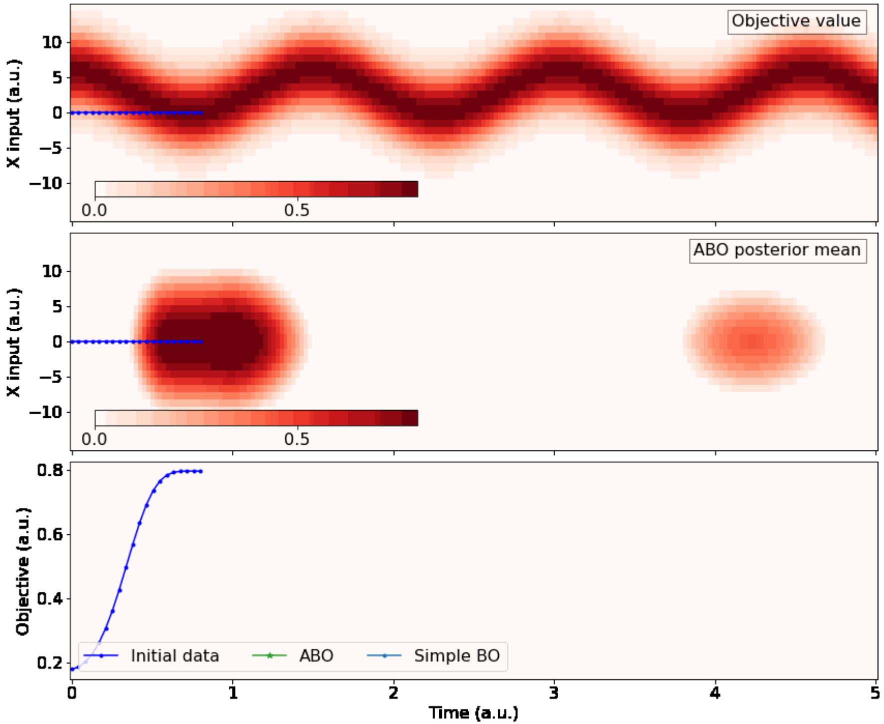
RESULTS - SIMULATION

Example: sinusoidal signal with noise + initial steady state sampling



RESULTS - SIMULATION

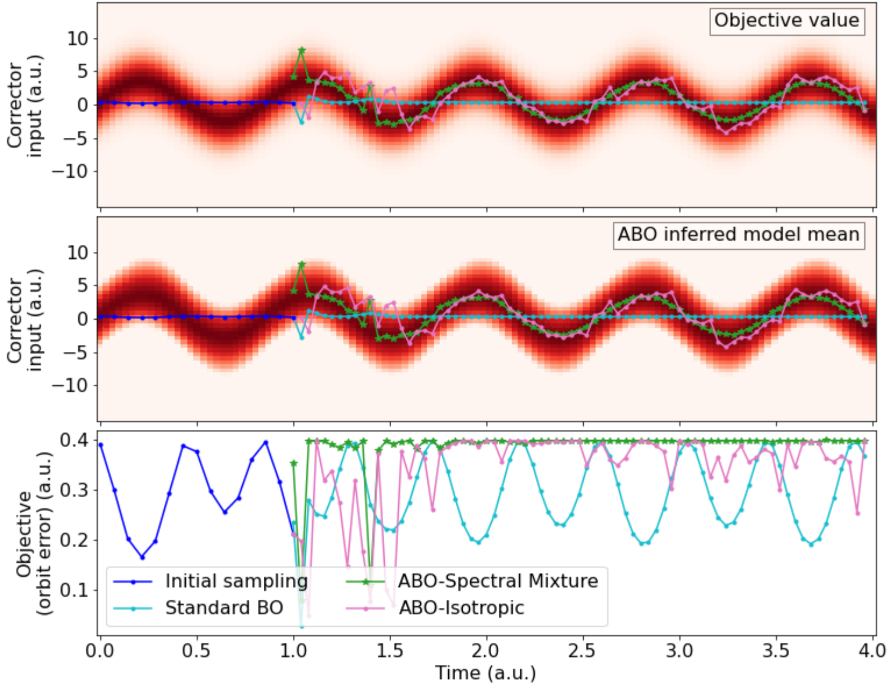
Example: sinusoidal signal with noise + initial steady state sampling

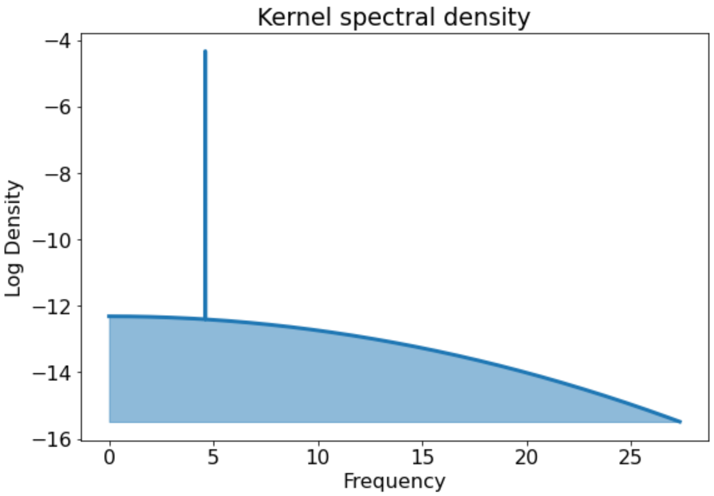


RESULTS - SIMULATION

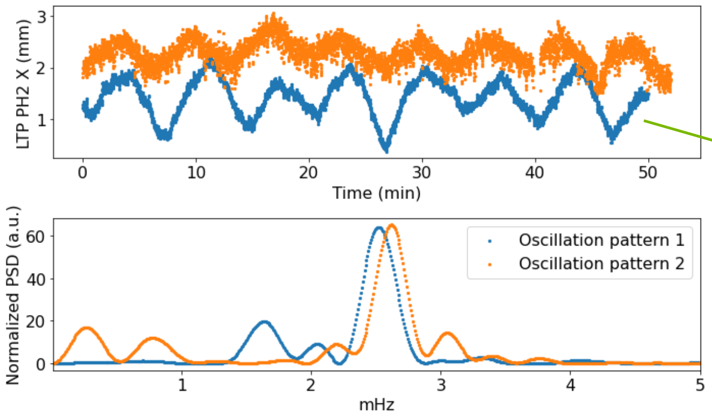
ABO finds correct oscillation within 1 period

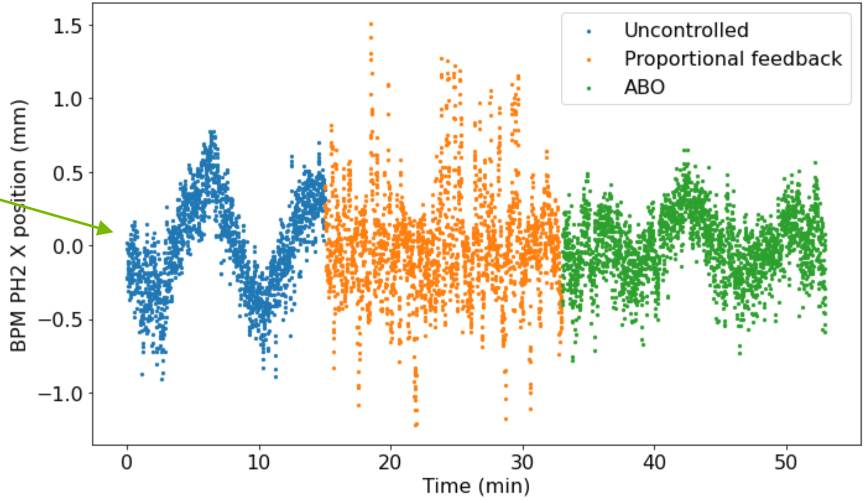
- Kernel density reflects broad noise + oscillation frequency





RESULTS - EXPERIMENT





Several tests in APS linac – trajectory MSE objective

- 10s cycle limit
- Train with last 20 minutes of history

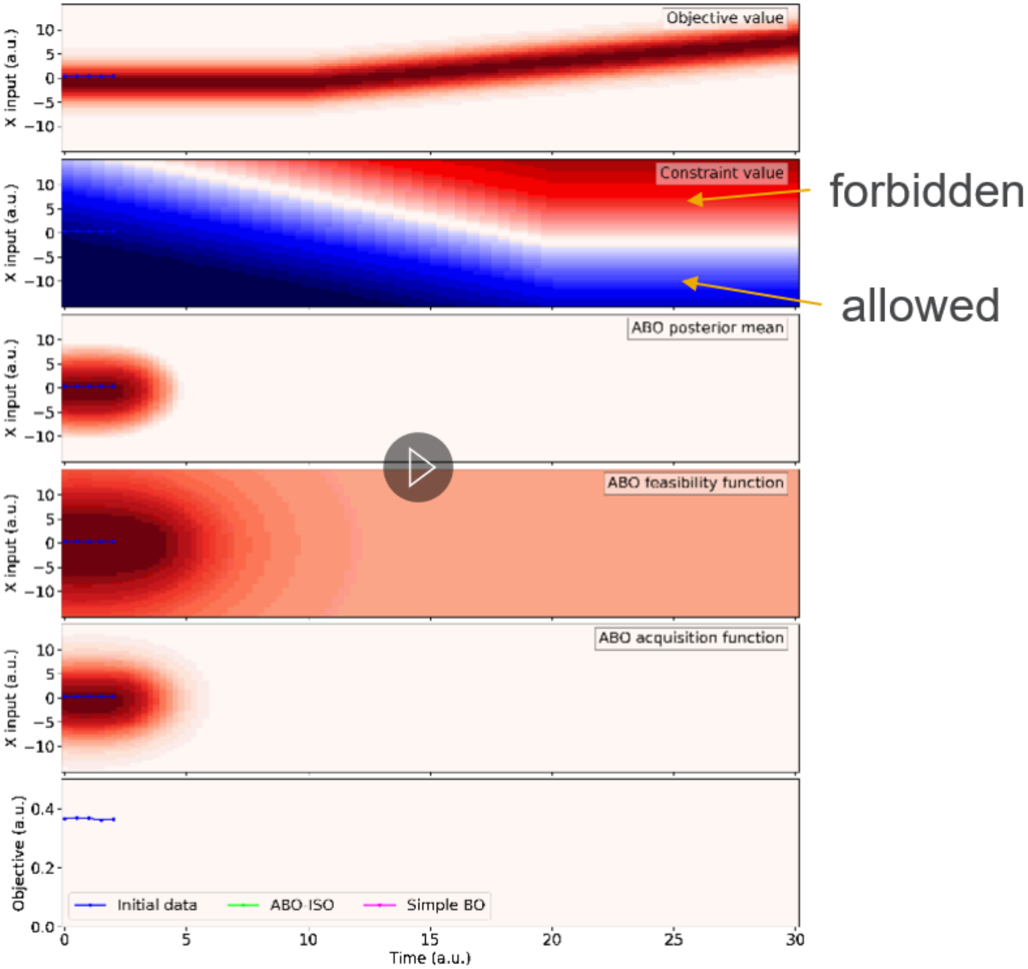
Overall **2x jitter** improvement (0.21/0.36/0.33)!

CONSTRAINTS

Time awareness can be added to **constraint models**

Gives anomaly **prediction and avoidance** capability!

Can predict ‘**lifetime**’ – how long is model capable of holding optimum

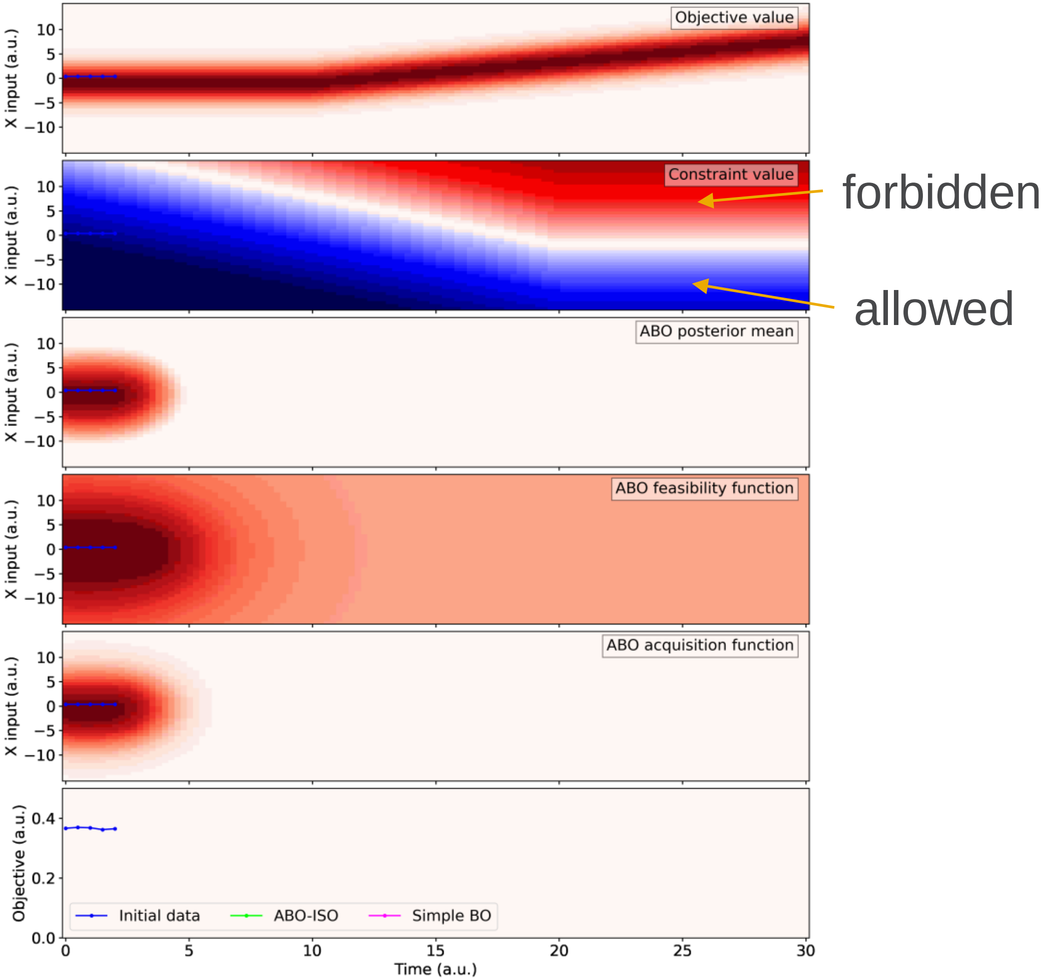


CONSTRAINTS

Time awareness can be added to **constraint models**

Gives anomaly **prediction and avoidance** capability!

Can predict ‘**lifetime**’ – how long is model capable of holding optimum

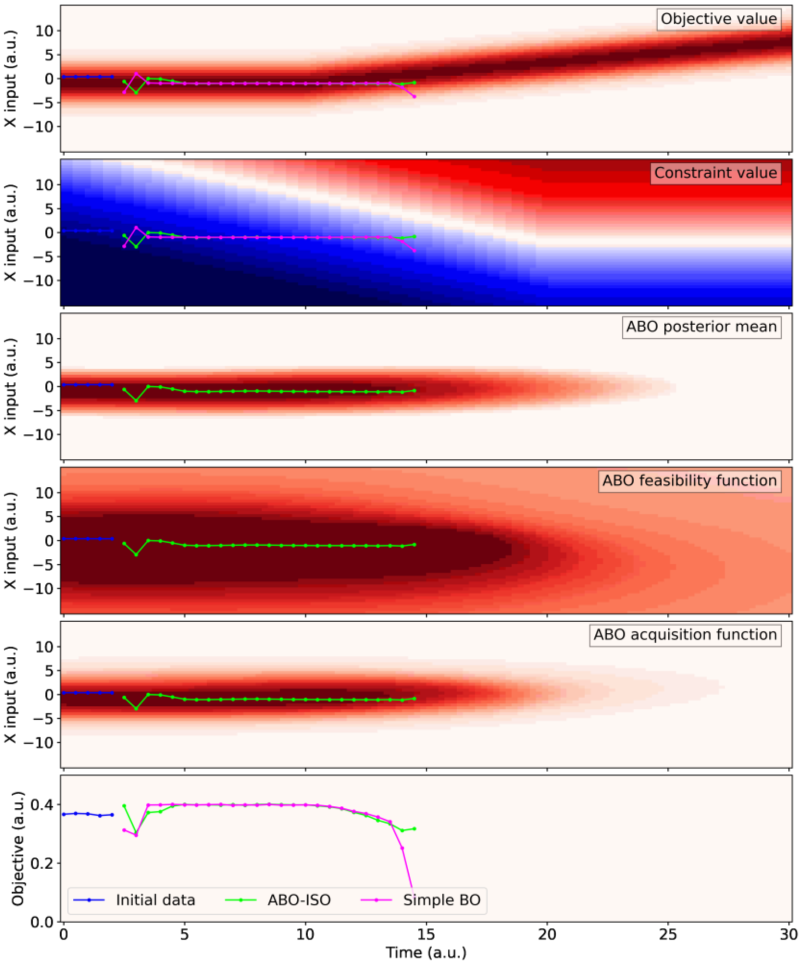


CONSTRAINTS

Time awareness can be added to **constraint models**

Gives anomaly **prediction and avoidance** capability!

Can predict ‘**lifetime**’ – how long is model capable of holding optimum

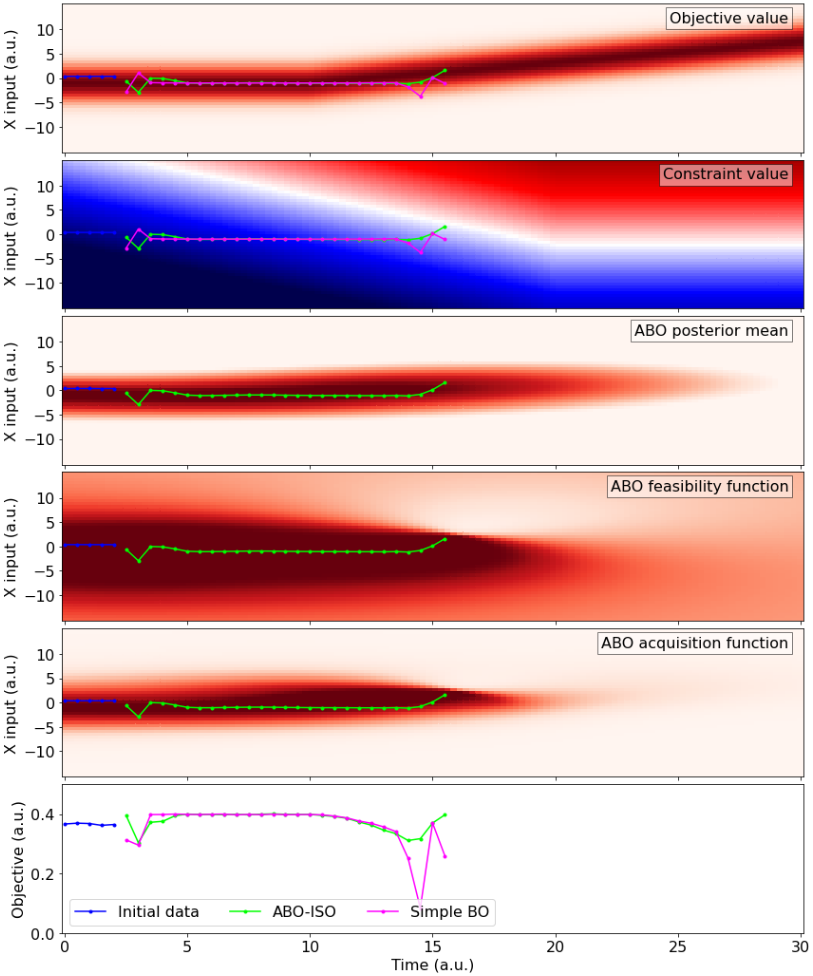


CONSTRAINTS

Time awareness can be added to **constraint models**

Gives anomaly **prediction and avoidance** capability!

Can predict '**lifetime**' – how long is model capable of holding optimum



CONTINUOUS USE CONSIDERATIONS

ABO has several other advantages for **continuous 'feedback'** over BO or PID:

- Graceful handling of failed points
  - Retries will have new time and give different candidates
- Can control systems with time-lag
- Automatic adaptability to local conditions

CONTINUOUS USE CONSIDERATIONS

ABO also inherits several BO disadvantages:

- Poor performance scaling (N^3)
  – Advanced techniques extend practical limit to >10k points [arXiv:1803.06058, see GPyTorch LOVE method]

- Eventually need to cut data
  – Simple solution: circular data buffer
  – Our approach: time-biased bandpass subsampling

- No hard robustness/convergence guarantees
  – Can use ***lengthscale rate of change*** as a stability heuristic [arXiv:1803.03432]

OPERATIONAL IMPLEMENTATION

We are making ML libraries to work with operational APS systems

- **APSopt** – algorithms + SDDS toolkit command line interface
  - Migrate general logic to **Xopt** for collaborative use soon™

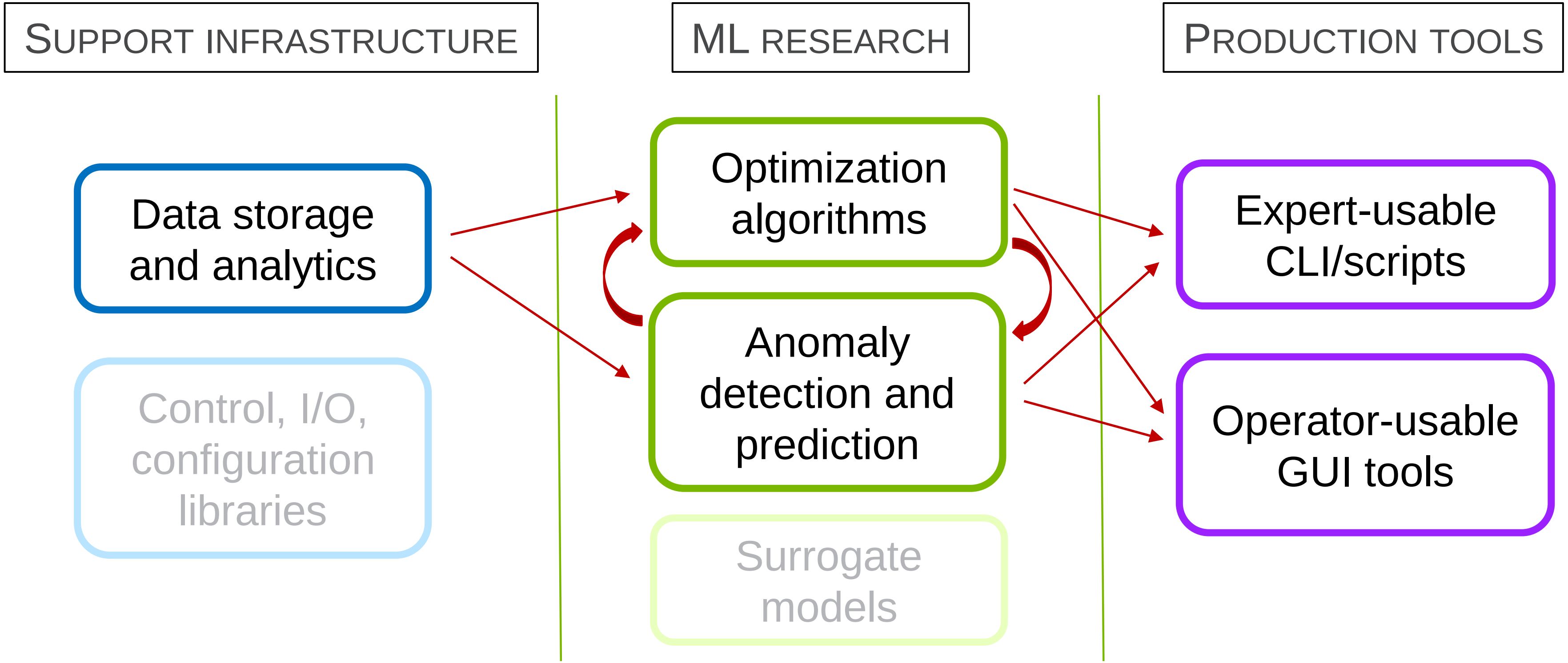
- **pySDDS** – Python SDDS format reader/writer

- **pybeamtools** – soft IOCs, surrogate models, and archiver/DB interface

Can make end-to-end virtual accelerators, test with real data, and deploy operationally

ML @ APS

Several simultaneous efforts underway



CONCLUSIONS

APS is preparing for APS-U

- Tight fixed schedule
- New known and unknown challenges
- ML tools can help in key areas

We are developing an ***interconnected production-level*** ML ecosystem:

- Shared data infrastructure
- Anomaly detection/prediction
- Adaptive optimization
- GUIs and control room tools

Focusing on internal use, but with portability and open-source in mind.

If you have things that drift/trip but shouldn't, reach out!


THANK YOU!