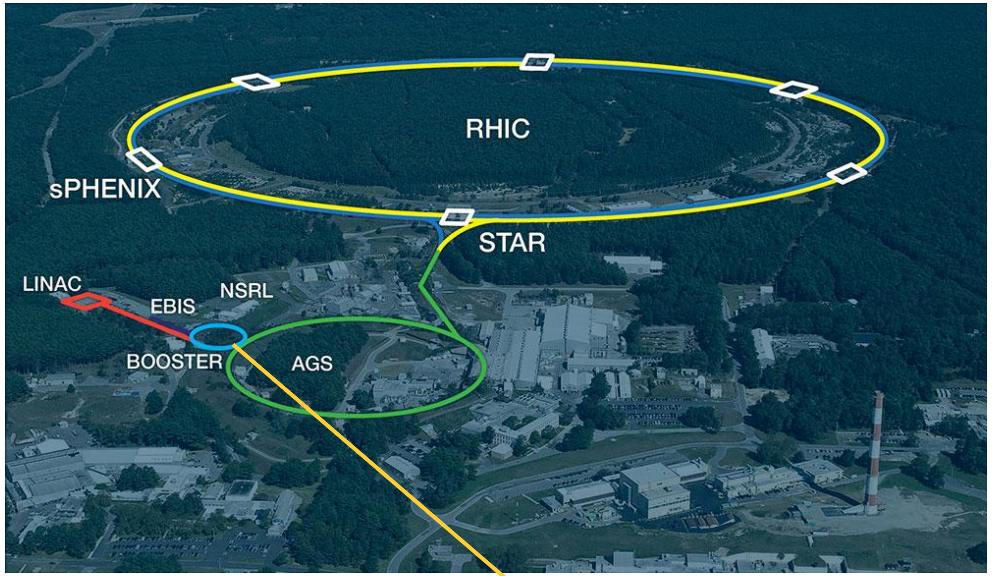
Machine learning assisted Bayesian calibration of an accelerator digital twin from orbit response data

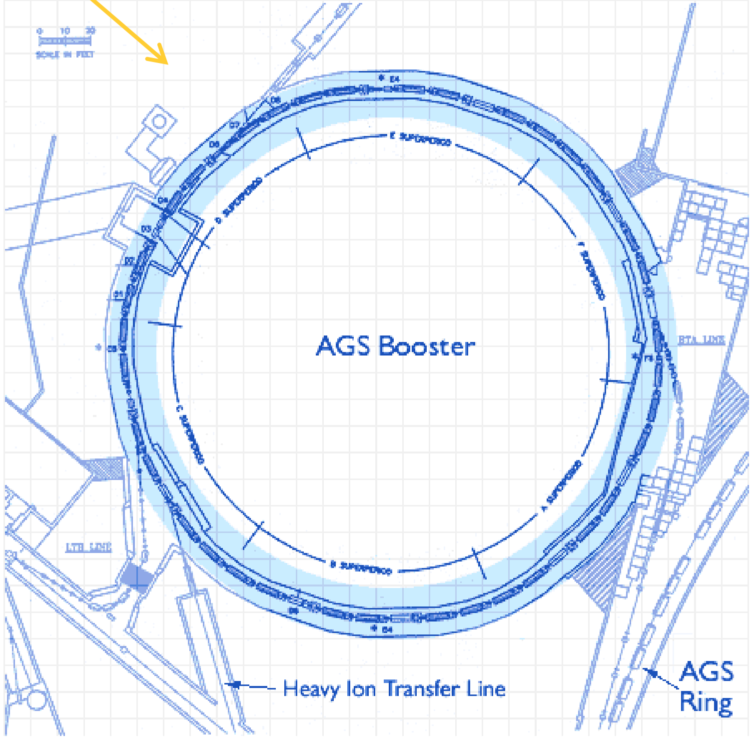
Christopher Kelly (Computational Science Dept.)
In collaboration with Cornell CLASSE, BNL Applied Math & Collider Accelerator Depts.

AI4EIC meeting
7/28/26

AGS Booster

- Responsible for accelerating beam from 200MeV to 1.5GeV for injection into AGS.
- Also provides heavy-ions for NASA Space Radiation Lab.
- **Accurate beam control in Booster required for providing high-quality beams to RHIC and to EIC.**
- Beams are controlled via 24 horizontal and 24 vertical quadrupole “corrector” magnets.
- Beam positions are read out live via 15 horizontal and 18 vertical beam position monitors (BPMs).

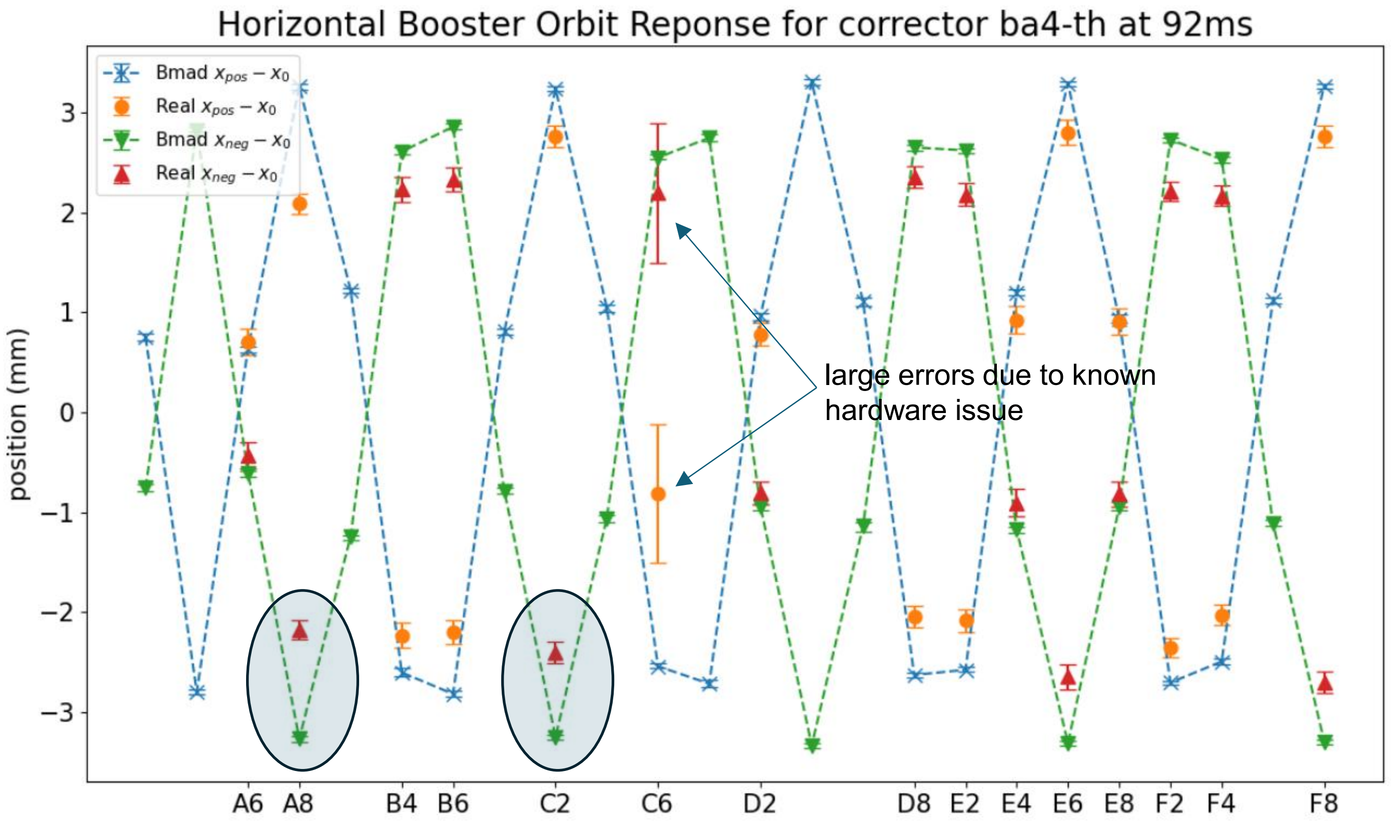




Brookhaven
National Laboratory

Digital Twin

- Digital twins are used to plan and control operations and to design data collection campaigns.
- The response of the orbit to changing the corrector magnets is the focus of modeling.
  - More reliably obtained than absolute beam positions due to error cancellation from e.g. inaccurate knowledge BPM locations.
- Booster's **Bmad digital twin** is highly accurate for orbit responses but **discrepancies of several percent remain of unknown origin.**
- We employed **Bayesian Uncertainty Quantification (UQ)** to probe for the origin of these discrepancies.



Comparison between measured and simulated horizontal orbit responses with corrector A4 set to ±22A

Bayesian Inference

Inference is a UQ method that probabilistically "fits" model parameters using measurement data.

"**Likelihood**": probability of data $\mathcal{D}$ given params Θ

Essentially fit function + model of statistical ("aleatoric") errors on measurements

"**Priors**": probability distributions describing prior expert knowledge of model *parameters* **(quantifies assumptions)**

"**Posterior**": (output) distribution of the parameters given the data

Bayes' Theorem

$$P(\Theta|\mathcal{D}) = \frac{P(\mathcal{D}|\Theta)P(\Theta)}{P(\mathcal{D})}$$

("Evidence" not needed for MCMC)

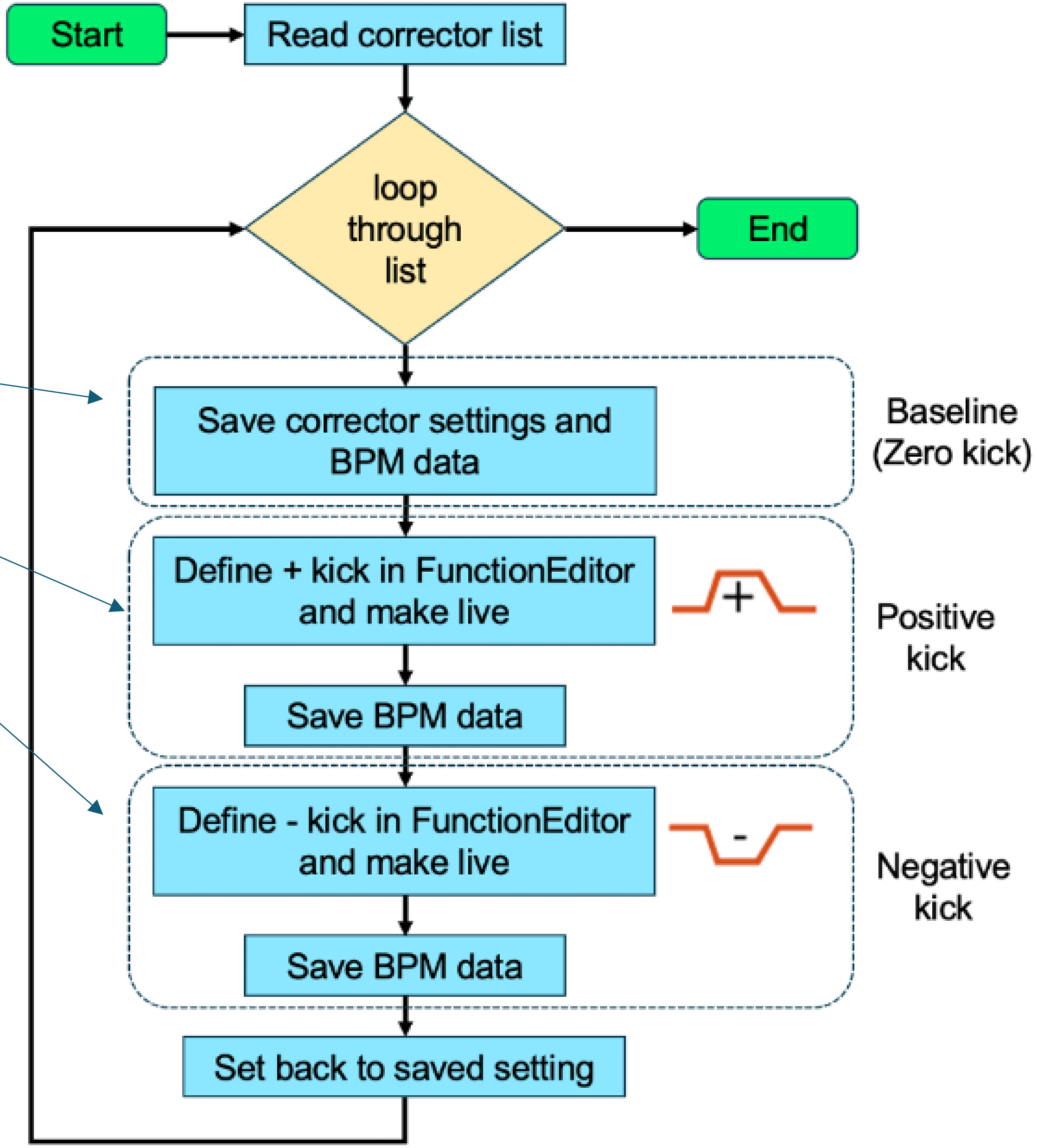
- "Inference" workflow:
  1. Identify (or introduce) a set of parameters in the digital twin model to infer
  2. Decide on prior distributions for the "fit" parameters using best available knowledge (absent data)
  3. Collect a set of data: complete orbit responses for different correctors current settings.
  4. Sample posterior for params Θ via Bayes' theorem using Markov Chain Monte Carlo
  5. Study the posterior distributions of each parameter (assess errors, correlations, etc)
  6. Sample from posterior and likelihood function for specific current settings gives the "posterior predictive distribution" : **orbit response predictions from calibrated model with uncertainties**

Use Julia "Turing" package

Brookhaven National Laboratory

Dataset

- Fix static magnet currents (up to power-supply fluctuations)
- Baseline setting with correctors set to 0A
- Vary each of the 24+24 corrector magnet currents in turn from 0A to +22A, -22A
- Record orbit for all BPMs before and after perturbation
- 2-5 repeat measurements for each setting
- Construct orbit responses by taking the orbit difference between arbitrary initial and final setting.
    - Many possibly combinations



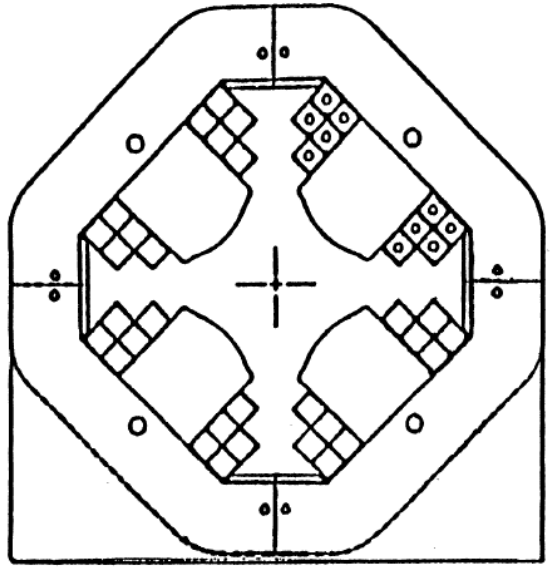
Brookhaven National Laboratory

Parametrizing the unknown

- Uncertainties in modeling of the 24(H)+24(V) static quadrupole magnets are expected to dominate.
- However, many possible causes for discrepancies that are hard to directly quantify, e.g.
  - magnet misalignments,
  - inaccuracies in power supply modeling
  - magnet nonlinearities
  - stray fields
- Instead of parametrizing each source, we introduce catch-all magnet-dependent multiplicative parameters on the magnet strength.
- Likelihood function uses the simulated result (with vars) for the central value.
- For aleatoric errors, assume BPM measurement normally distributed

$$\epsilon = \sqrt{2}\sigma_{\mathrm{BPM}}$$

Response is combination of two measurements!



$$(k'_{1,H})_i = var_i^{(x)} k_{1,H}$$

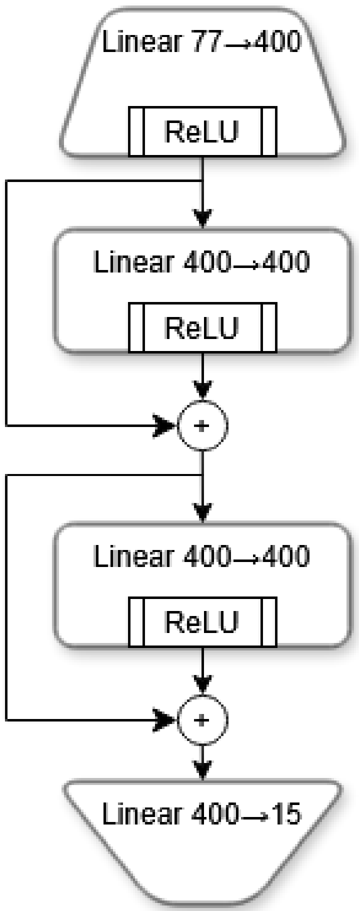
$$(k'_{1,V})_i = var_i^{(y)} k_{1,V}$$

i indexes the quadrole magnet

Brookhaven National Laboratory

Surrogate Model

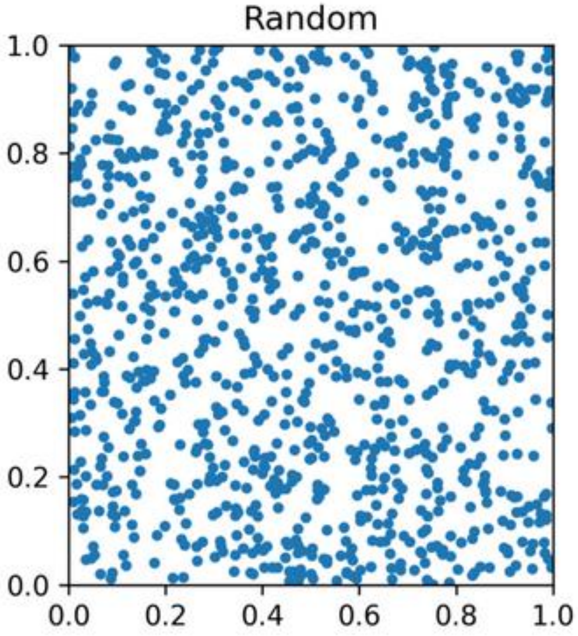
- Sampling the 48-dimensional posterior distribution is challenging and needs an efficient algorithm: “hybrid (aka Hamiltonian) Monte Carlo”
- This requires many simulation evaluations **including derivatives**
- **SciBmad will offer differentiable simulations but still under development.**
- Instead, replace direct Bmad sim with machine-learned surrogate model
  - **Require highly accurate surrogate with MSE validation losses ~10^{-4} to be << BPM error (0.15mm)**
  - Train separate model for each plane
  - 77 input features per plane (5 static magnet currents + 24 in-plane correctors + 48 vars)
  - 15 (H), 18 (V) output features (absolute BPM beam position)
  - 335,415 params residual neural network (RESNET)

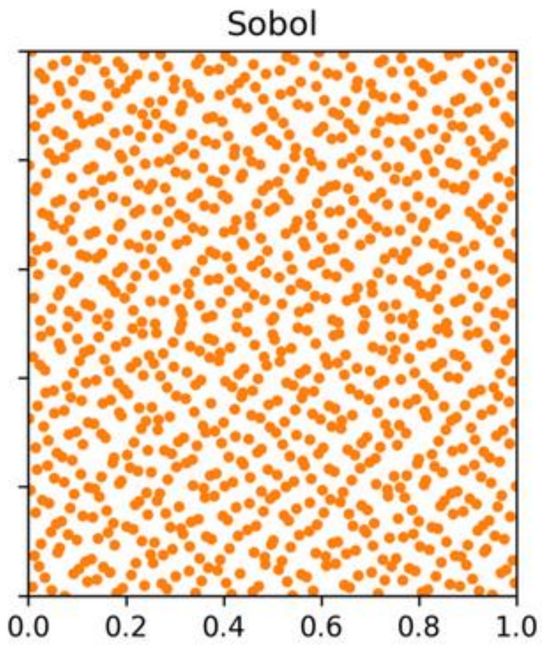


The model architecture employed for this study

Surrogate data

- Use a Sobol pseudo-random uniform sampling scheme to sample the 77 input dimensions (for each plane)
  - More uniform than uniform-random
- Use data-derived sampling ranges for static/corrector magnet currents.
- Corrector currents sampled in windows around [0A, +22A, -22A]
- For the “vars”, sample around the null value (1.0)
- Want to sample as wide a range of possible, however:
  - Found increasing orbit instability / sim. failures as var range was made larger
    - Possibly due to unfortunate combinations of large vars, or
    - Orbit resonance effects
  - **Had to restrict var range to 0.94 – 1.06 for sufficiently precise model**
  - Requires truncation of prior distributions to [0.94,1.06] to restrict sampling to regions where surrogate is reliable.
- Data subsampling required to restrict to vicinity of Booster tunes (x : 4.83±0.05, y: 4.665±0.05)
- **2M samples generated, 820-870K surviving tune cut**
- Achieved validation loss 3×10^{-4} (x), 1×10^{-4} (y)
- Corresponds to ±0.057mm, ± 0.017mm vs 0.15mm BPM error





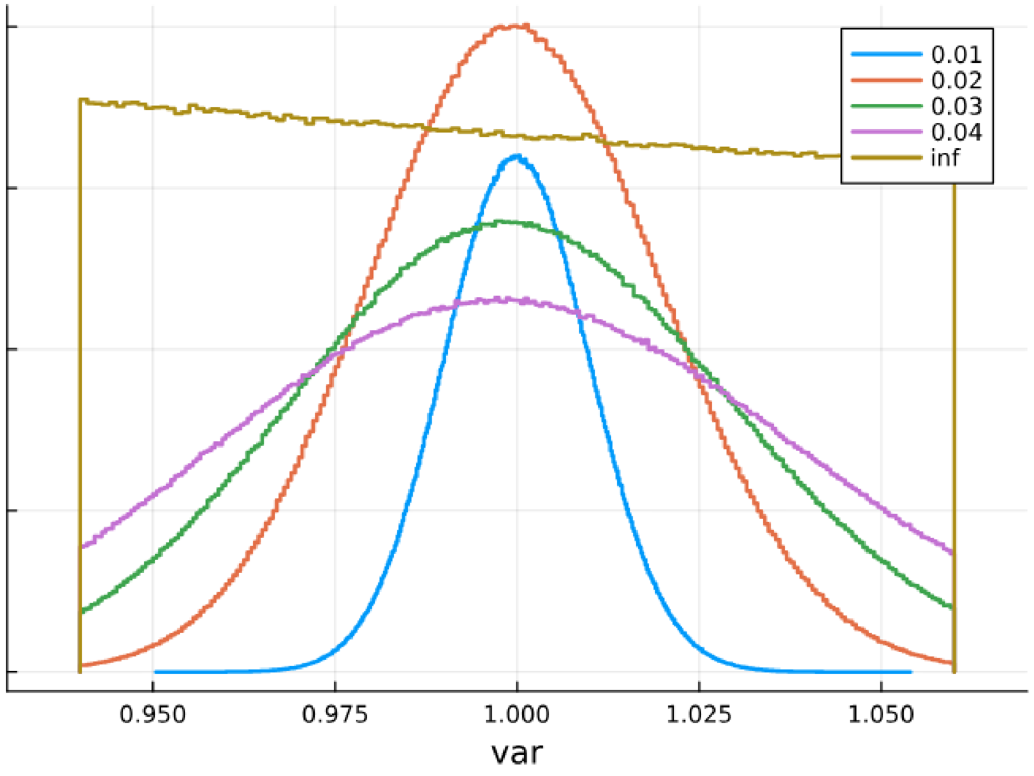
Brookhaven National Laboratory

Prior distributions

$$\text{var}_i \sim \text{Lognormal}(1 - \sigma^2/2, \sigma^2)$$

$$\overline{\text{var}_i} = 1, \quad \text{Var}(\text{var}_i) = \sigma^2 + \mathcal{O}(\sigma^4)$$

- Digital twin is not far wrong so vars are most likely close to 1 (null result).
- We don't know in advance whether they are expected to be higher or lower than 1 so we take an unbiased distribution.
- For a multiplicative parameter, the log-normal distribution is unbiased.
- But how wide should this distribution be? We don't really know other than that we don't expect the discrepancies to be more than a few %.
- We performed a **prior sensitivity analysis** to study how the width of the prior affects the results.
- Caveat: Surrogate model training data bounded in [0.94,1.06]
  - need to truncate prior distributions to ensure we don't probe outside the validity bounds of the surrogate



Study $\sigma \in \{0.01, 0.02, 0.03, 0.04, \infty\}$

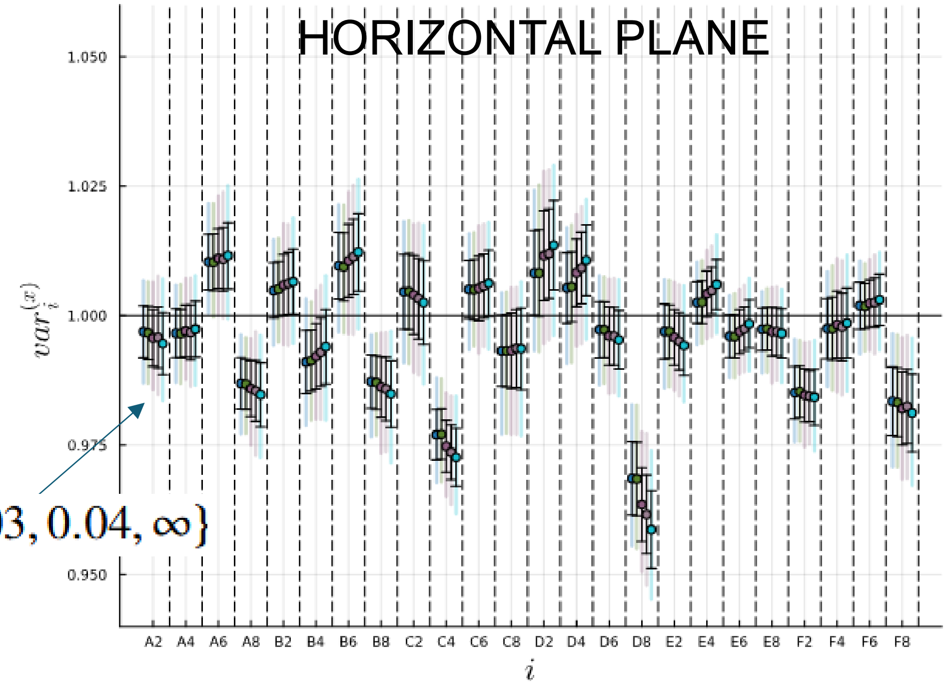
Brookhaven National Laboratory

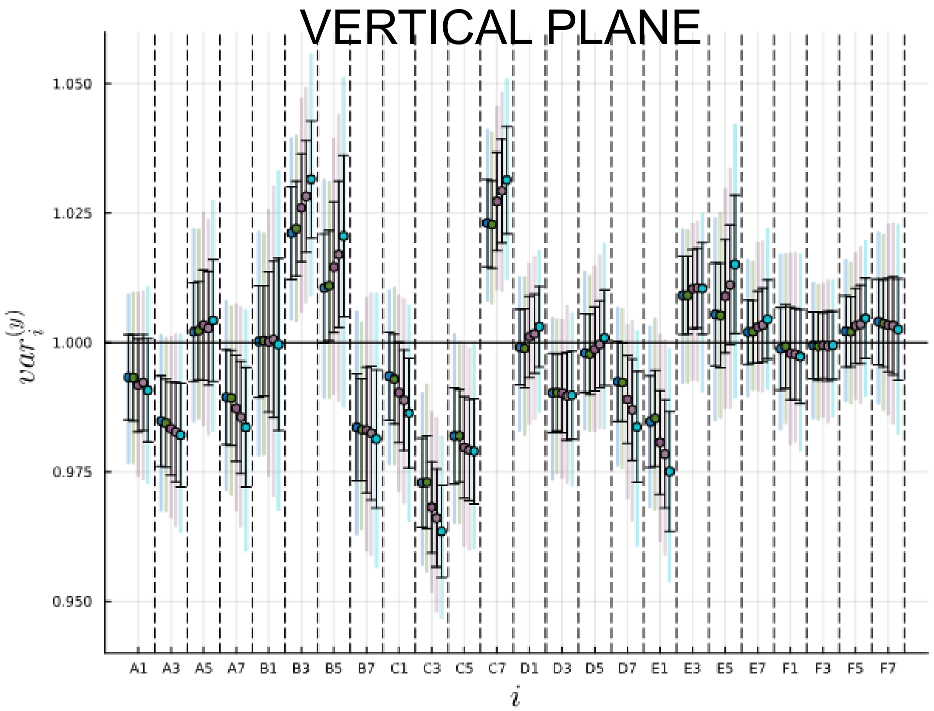
Prior sensitivity study

- Construct orbit responses where each corrector in turn is perturbed from 0A to ±22A
- Randomly chosen samples from among repeat measurements.
- Repeat inference for different σ values.
- Little prior dependence observed
- **Indicates posterior is well-constrained by data**
- Minor trends to larger vars may be explained by flatter directions in parameter space.
- (Prior-dependence is not a “bug” *per se*; priors parametrize assumptions and posterior should always be interpreted as conditional on those assumptions.)
- **Adopt widest (least informative) prior not overly affected by the truncation: σ=0.03.**
- **Some vars, notably D8, are quite far from unity!**

left-to-right

$\sigma \in \{0.01, 0.02, 0.03, 0.04, \infty\}$

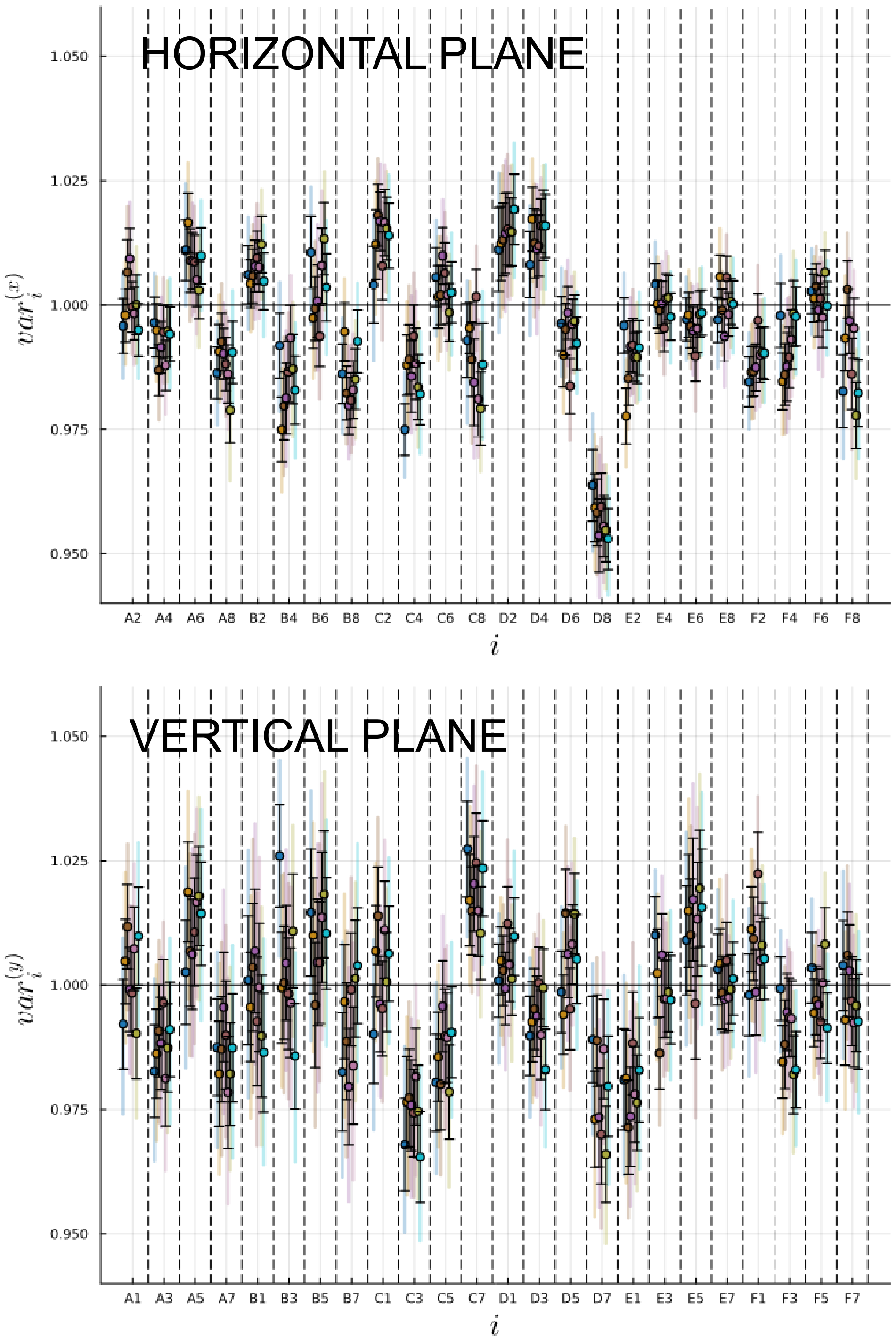




Brookhaven National Laboratory

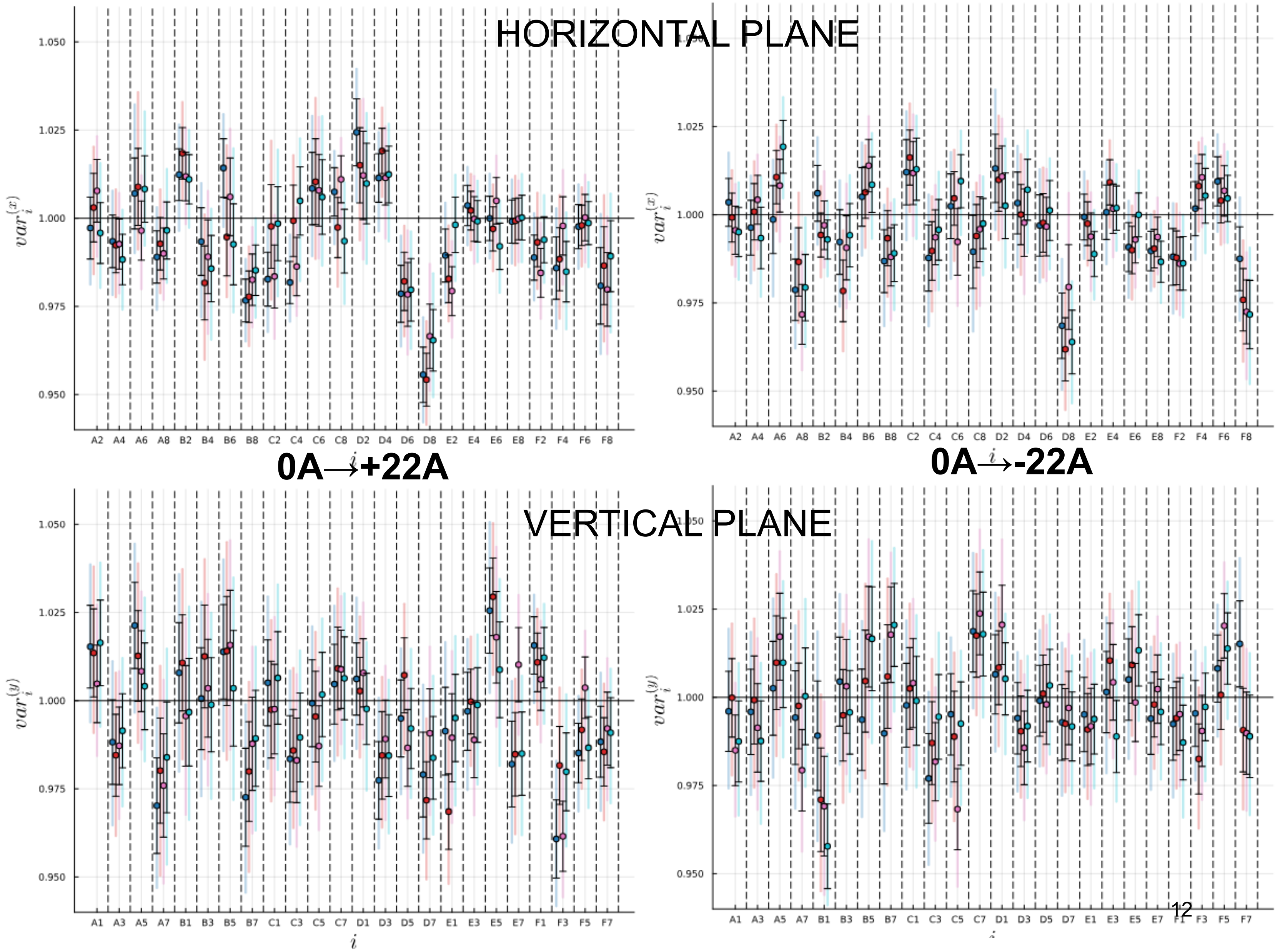
Stability study I

- So far just looked at orbit responses for a single random sample from among repeat measurements.
- To test the stability of our results we repeated this 7 more times.
- Observed strong consistency between repeat measurements.



Stability study II

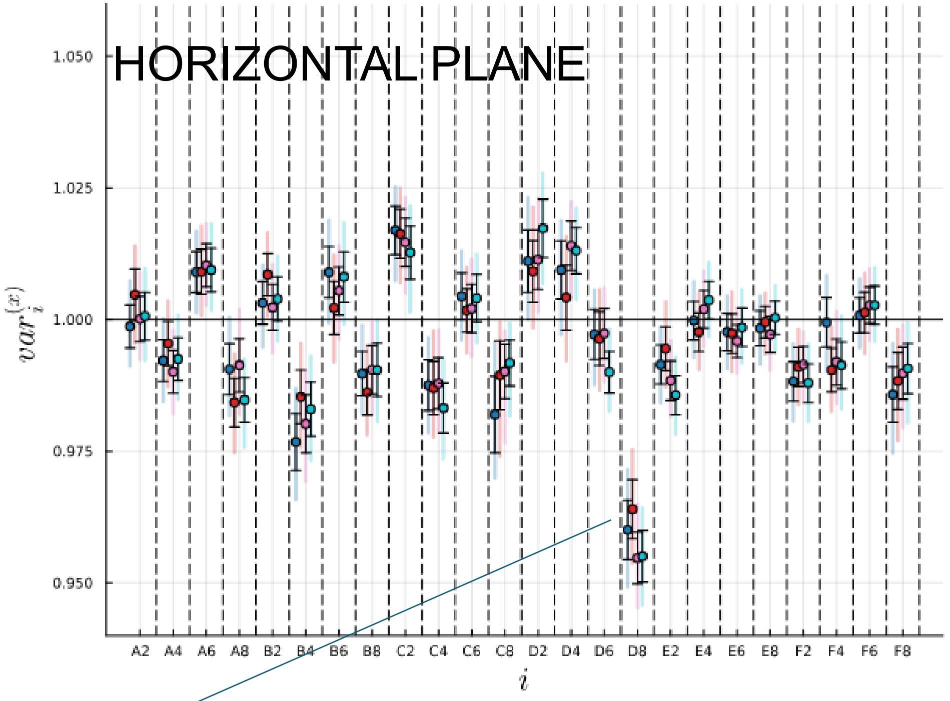
- Also looked at a subset of the data where the perturbed corrector current is **either** +22A or -22A.
- Consistent results observed for the two cases.

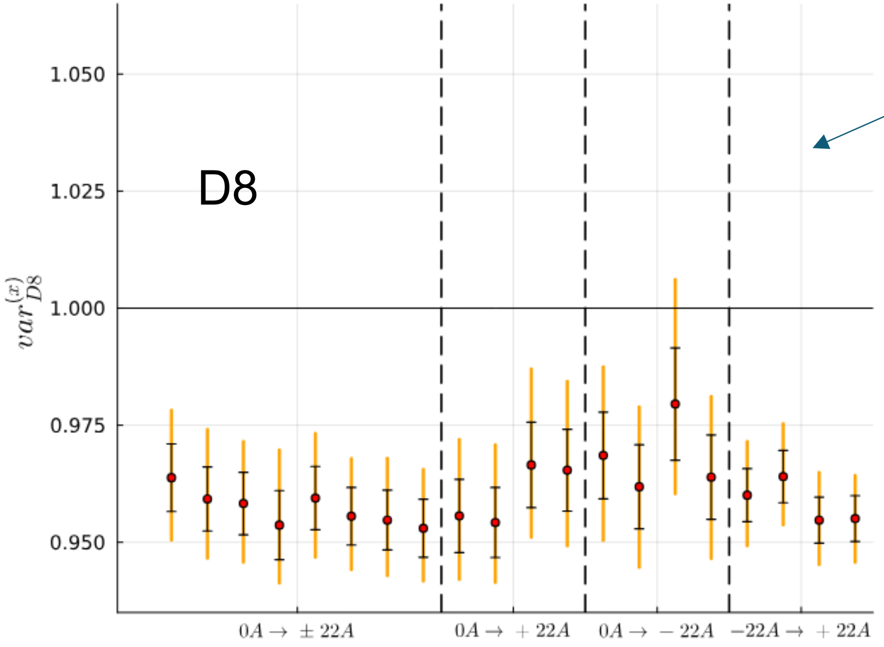


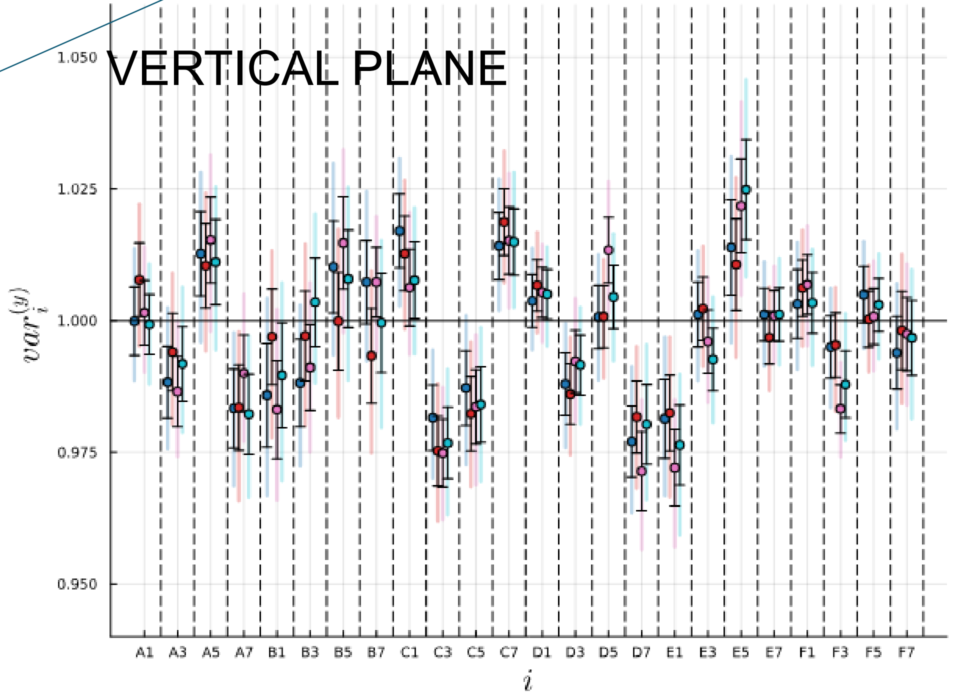
Brookhaven National Laboratory

Stability study III

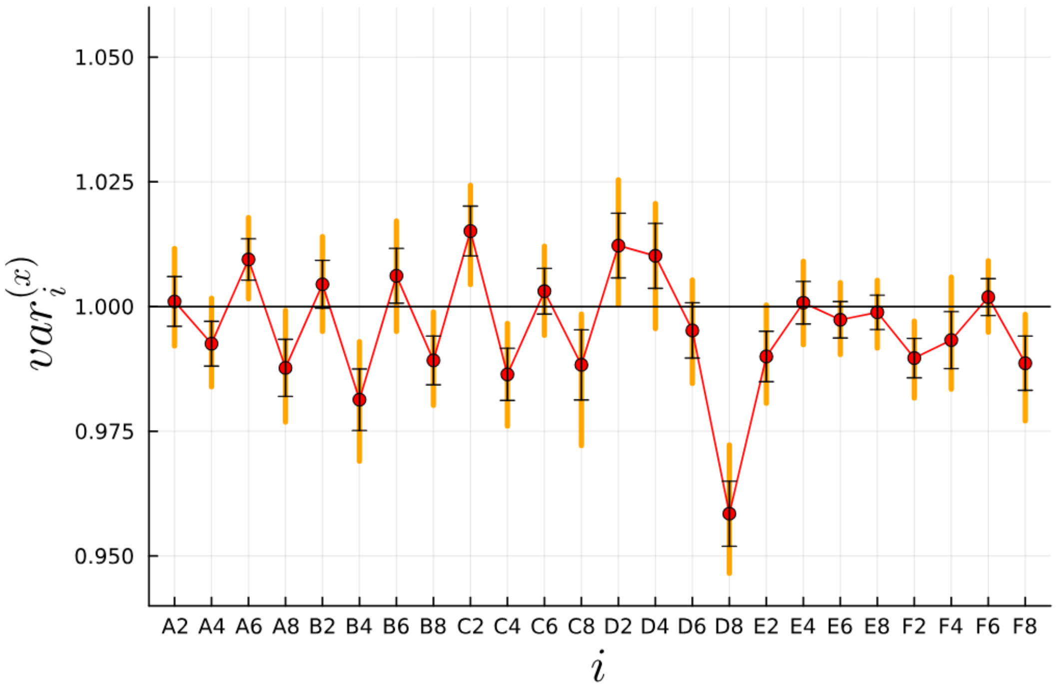
- Also considered data where the corrector is perturbed from -22A to +22A
- **Inferred results found to have much smaller uncertainties**
- **Likely because the digital twin error is amplified by transitioning between these extreme current values, making it more informative.**

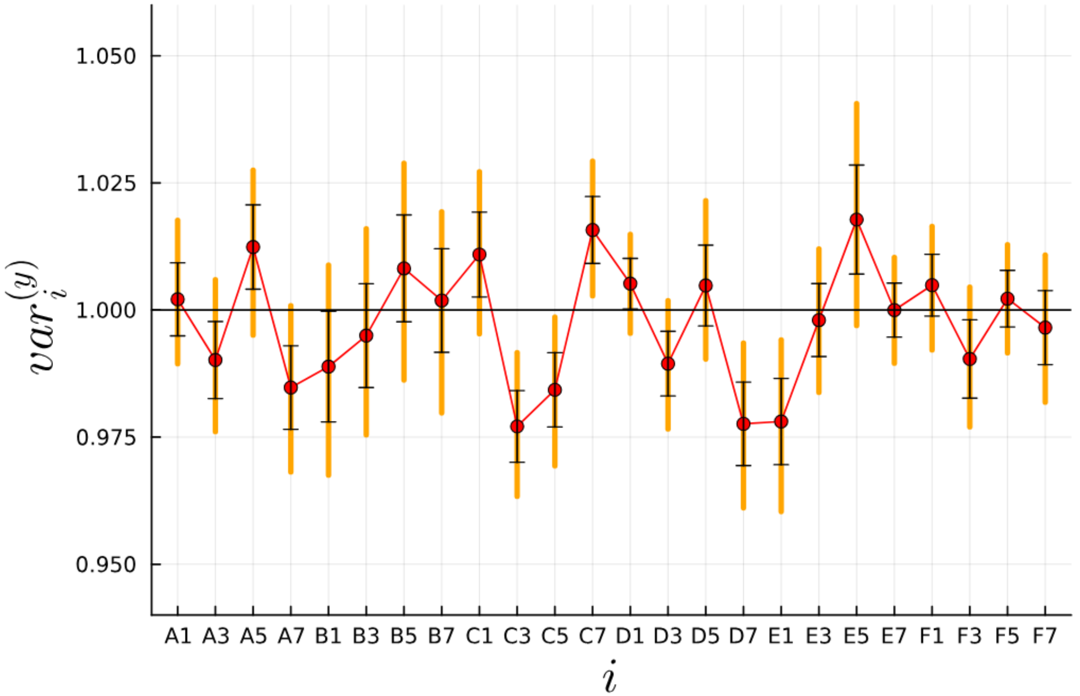






Inference result



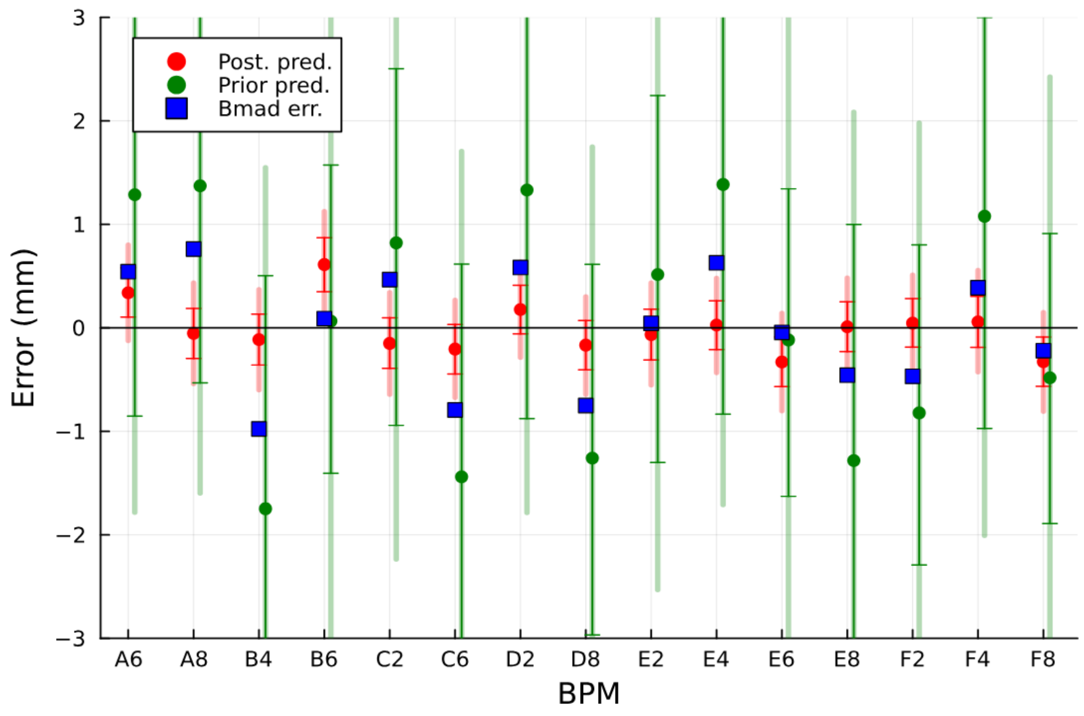


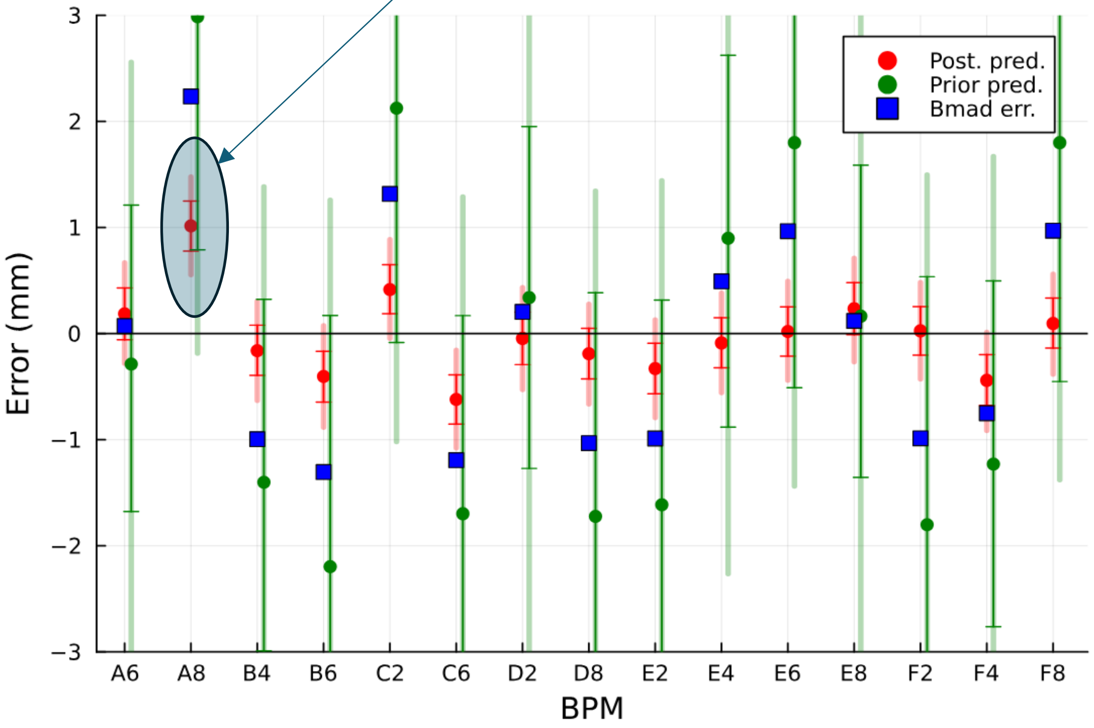
- Most results cluster around null result to within 2.5% percent and are generally consistent will null within uncertainties (with some minor tensions).
- D8 is notable exception: $0.958(7)_{(65\%\ \text{confidence})}\ (20)_{(95\%)}$
- **Evidence that quadrupole D8 is behaving unexpectedly**
- (Origin subject to future research)

Posterior/prior predictions

- Two example orbit responses in the -22A→+22A subset

Why is this one showing residual model-data disagreement?

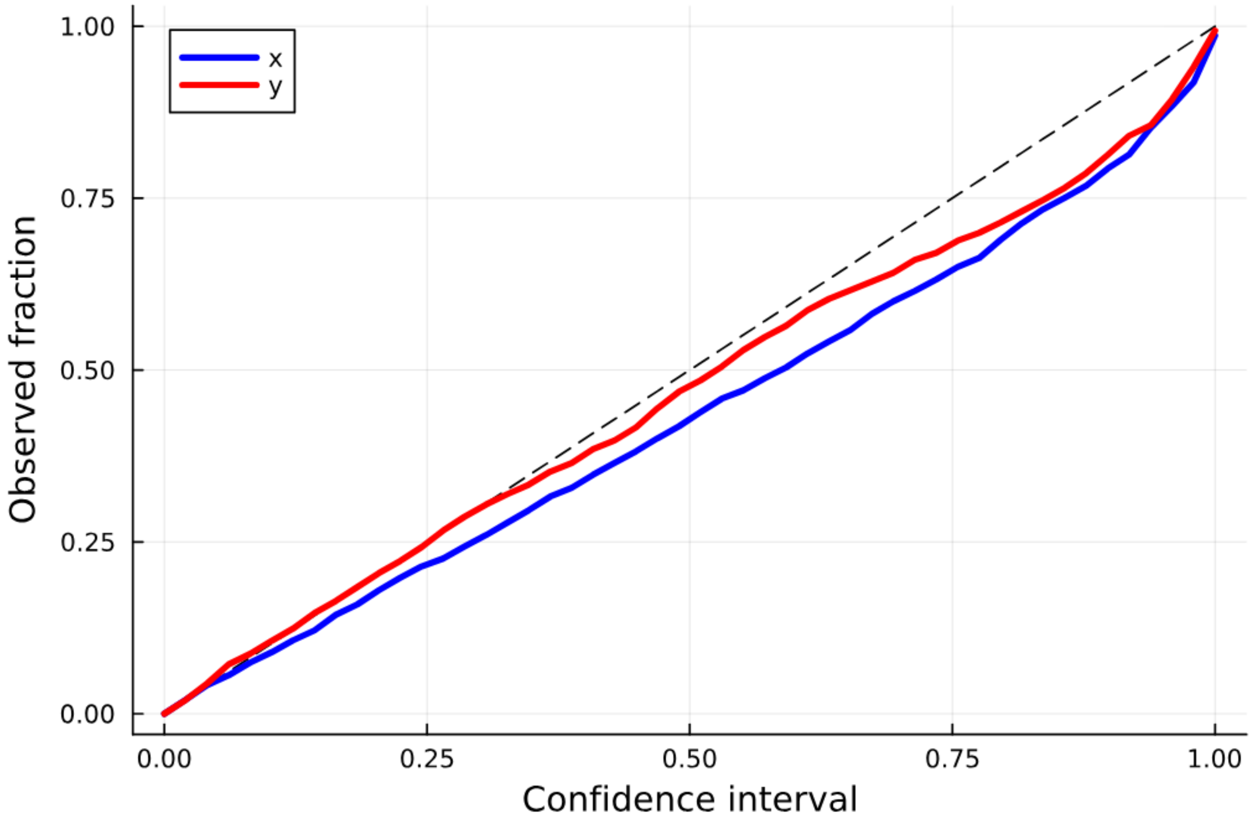




- Significant improvement in agreement between model and data

Debugging the statistics I

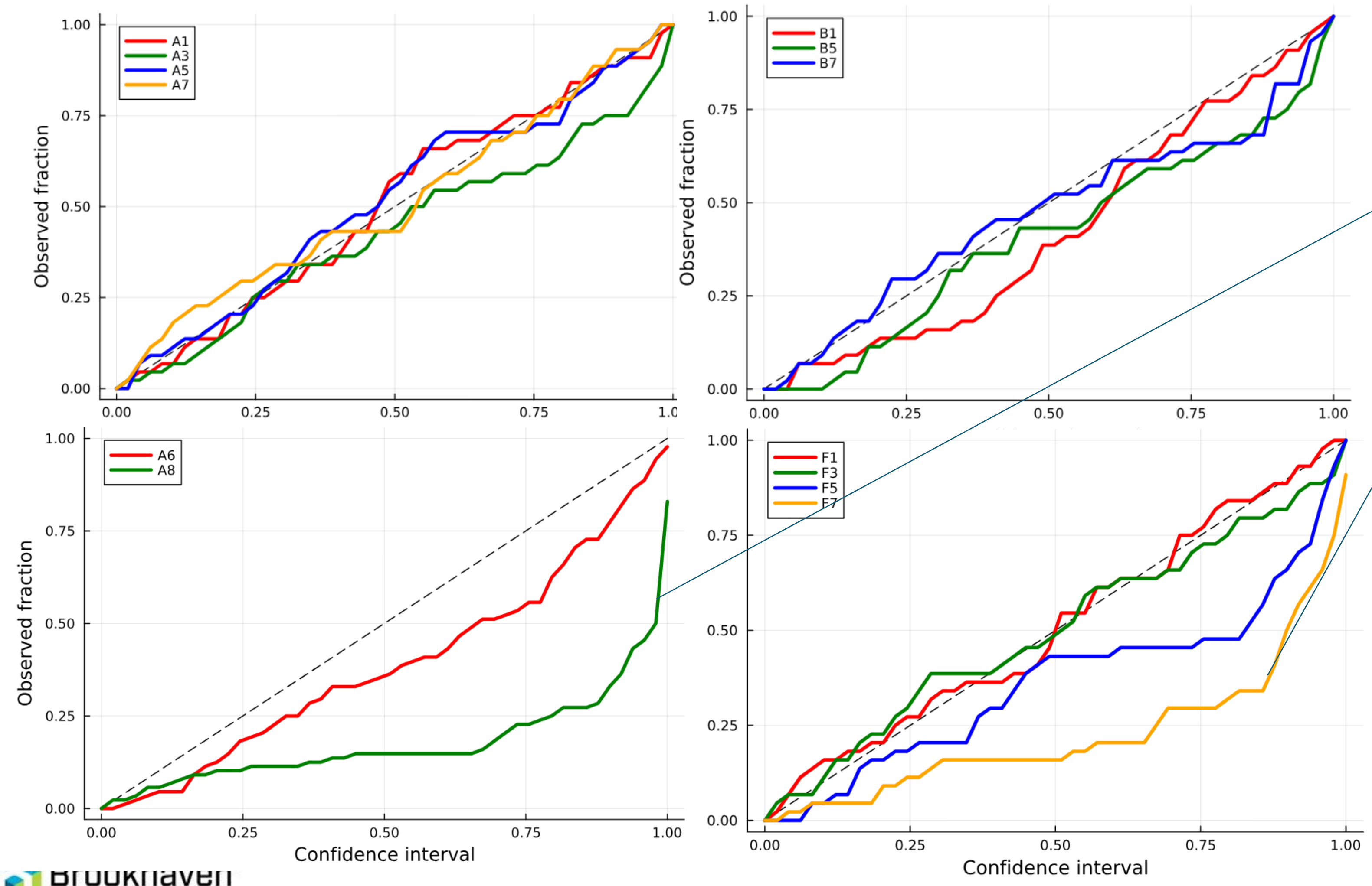
- Residual disagreements between the data and the calibrated digital twin could indicate that the optimized parameters do not have sufficient "power" to parametrize all error sources.
- To test this we examined the **calibration curves**
- These plots compare the actual to expected fraction of data points within a given confidence window around our calibrated model.
- Overall we find our model is well calibrated, only minor overconfidence (underestimation of uncertainties)



Brookhaven National Laboratory

Debugging the statistics II

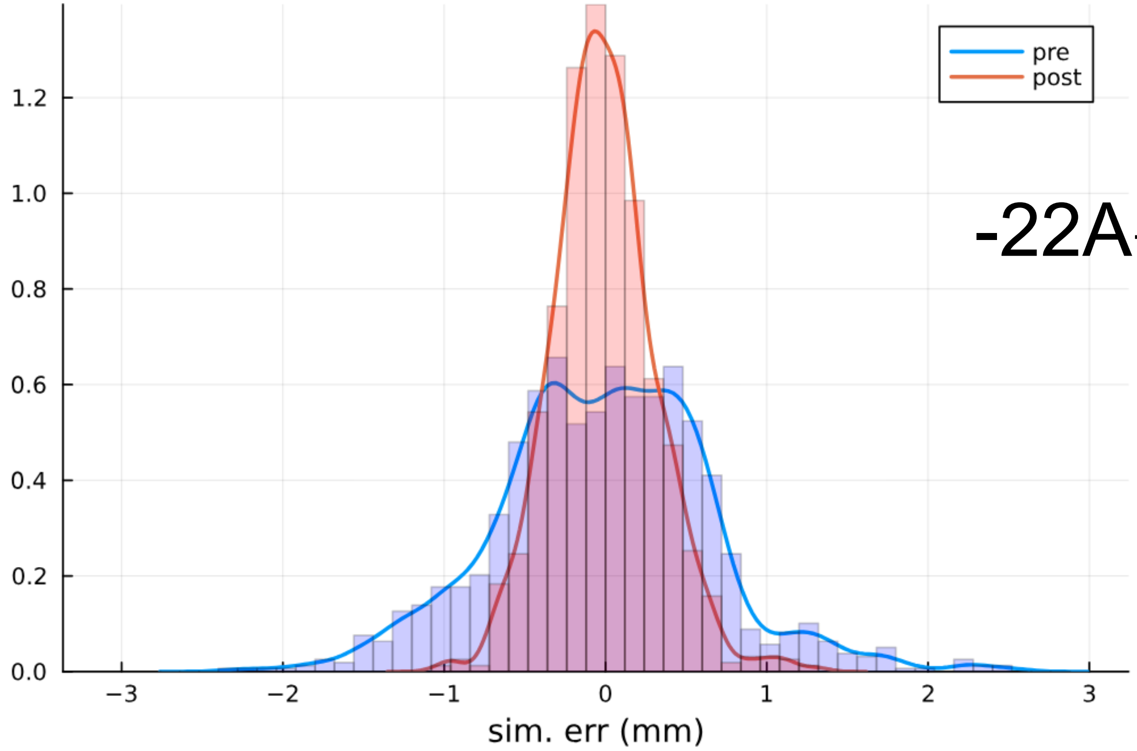
- Some examples broken down by ring segment



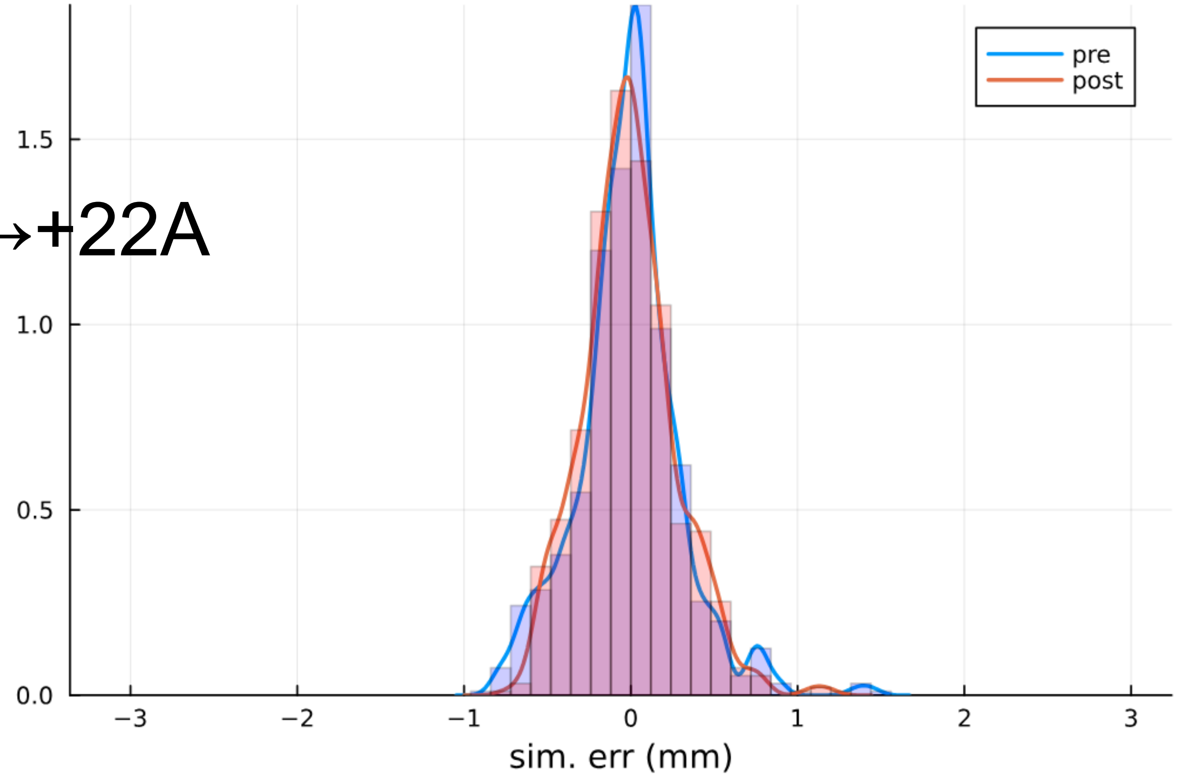
- A8 and F7 are atypical, showing significant overconfidence.
- Explains anomaly on previous slide.
- Intend to investigate further.

Point estimates

- Posterior predictions still use surrogate – possibility of additional error
- See if incorporating **central values** (point estimates) of var posteriors improved Bmad Digital Twin



-22A→+22A

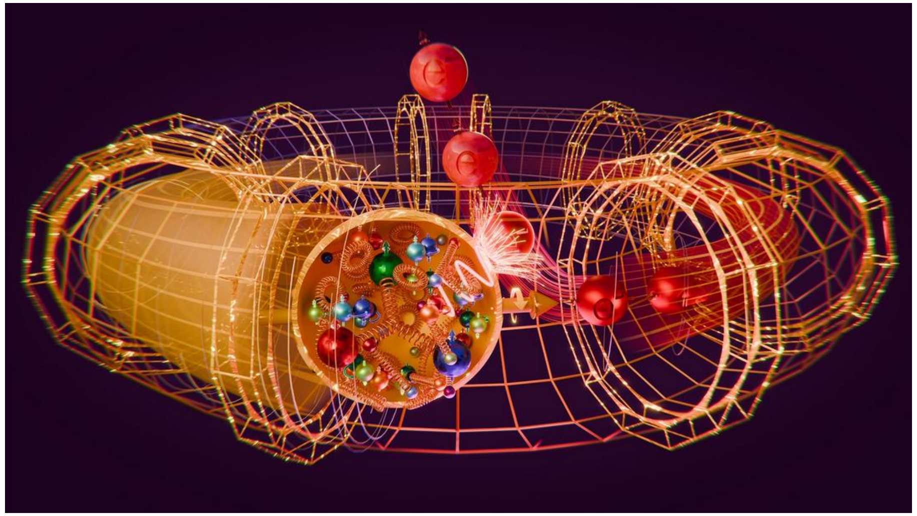


- Digital Twin error for -22A→+22A data improved by 2x for horizontal plane!

Brookhaven National Laboratory

Conclusions

- Introduction of magnet-dependent parameters constrained by data shows evidence of localized phenomenon that may help identify root cause.
- Including just the point-estimates significantly improves modeling errors.
- Bayesian UQ demonstrated as useful approach for calibrating and improving accelerator Digital Twins.
- **AI/ML played a vital role in providing a fast, differentiable surrogate model of the simulation**
- Future research could be to include additional data
  - e.g. with different static magnet currents
  - or with more complicated corrector setups
  - Bayesian Optimal Experimental Design could be used to identify an optimal experiment to improve uncertainties
- Uncertainty-aware Digital Twin model could be used for improved beam control
  - Allow for assessing trade-offs of control actions with knowledge of uncertainty of modeling



Brookhaven National Laboratory