



Artificial
Intelligence for
the Electron Ion
Collider (AI4EIC)
2025



Machine-Learning–Accelerated Bayesian Uncertainty Quantification for Digital Twin Modeling and Control of the AGS Booster

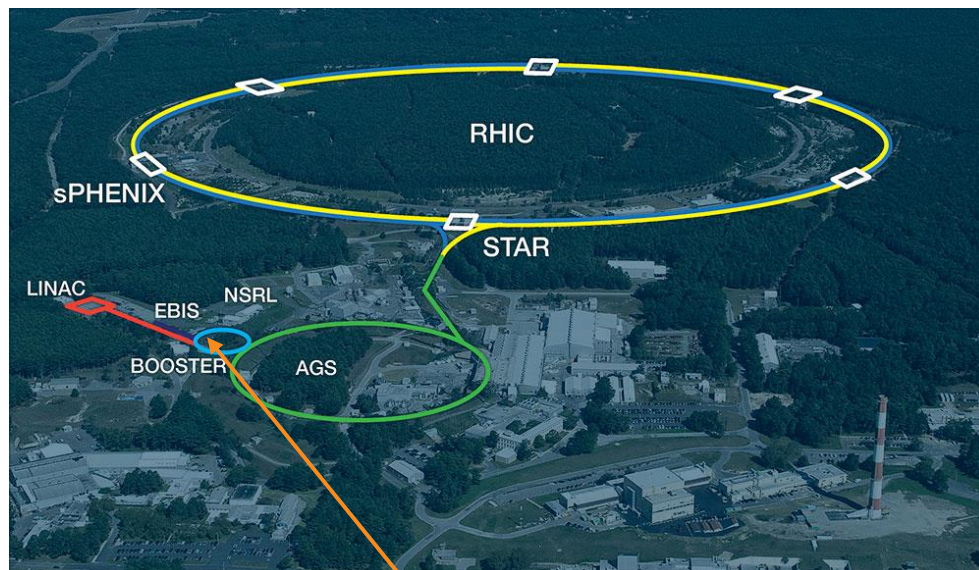
Christopher Kelly (Computational Science Dept.)

In collaboration with Cornell CLASSE, BNL Applied Math & Collider Accelerator Depts.

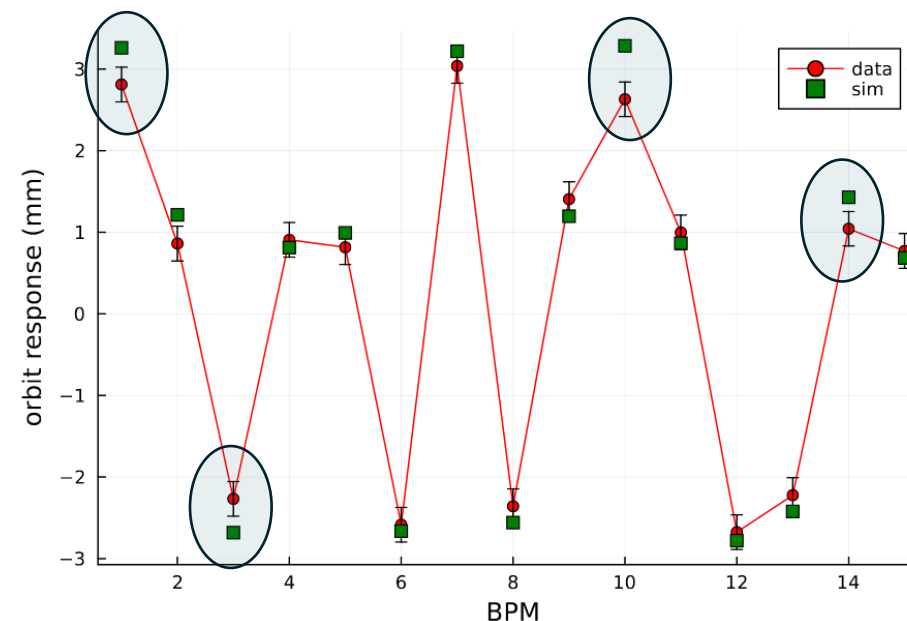
10/27/25



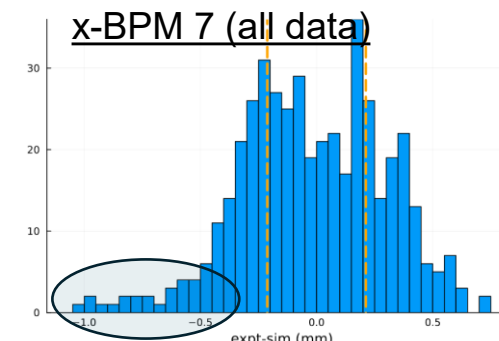
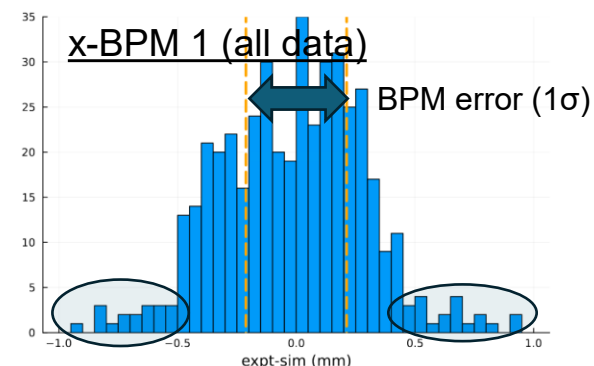
RHIC AGS Booster and Digital Twin



- A precise Digital Twin for AGS Booster would enable more reliable beam control, improving beam quality for RHIC and EIC.
- Comparison of existing Digital Twin (Bmad) to real beam orbit data shows discrepancies of unknown origin.
- Use Bayesian Uncertainty Quantification (UQ) to study the discrepancies.



Example orbit response to changing the strength of one of the corrector quadrupole magnets plotted versus the beam position monitor (BPM) index, showing deviation between simulation and real data,



Bayesian UQ

“Likelihood”: probability of data $\mathcal{D}$ given params Θ

normal dist. centered on simulation

$\sigma = \sqrt{2} \times \text{bpm err.}$ (sqrt 2 due to orbit difference!)

“Priors”: expert knowledge of 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 to infer
2. Decide on prior distributions and parameters
3. Collect a set of data: complete orbit responses for different current settings (notably correctors)
4. Sample posterior for params Θ via Bayes' theorem using Markov Chain Monte Carlo | Use Julia “Turing” package
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 with uncertainties

Parametrizing the unknown

- Many possible causes for discrepancies that are hard to directly quantify, e.g.
 - magnet misalignments,
 - power supply transfer function parameters
 - magnet nonlinearities
 - stray fields
- Rather than attempt to parametrize each source we introduce catch-all “nuisance parameters”
- **Allow static quadrupole strengths to vary by multiplicative “var” factors**

$$T'_i(I_{\text{quad}}, I_{\text{dipo}}) = \text{var}_i \times T(I_{\text{quad}}, I_{\text{dipo}}) \quad (\text{transfer function})$$

- Priors: lognormal centered on 1.0 (null)
- Prior width hard to guess due to unknowns: assume a few percent

```
function kqh_var(idipo, iqhc, var, qhstr1=0, qhstr2=0, quad="A2")
    klhc0 = 0.001818 #+ offset
    klhc1 = 9.080E-4  #* multf
    klhc2 = 6.657E-9
    klhc3 = 7.225E-12
    klhc4 = 3.239E-15
    klhc5 = 5.07E-19

    pl1 = ["A2", "A8", "C2", "C8", "E2", "E8"]
    ml1 = ["B2", "B8", "D2", "D8", "F2", "F8"]
    pl2 = ["A4", "B6", "C4", "D6", "E4", "F6"]
    ml2 = ["A6", "B4", "C6", "D4", "E6", "F4"]

    if quad in pl1
        iqh = idipo + ckc * (iqhc + bdot * iqhbd) + ckc2 * (qhstr1)
    elseif quad in ml1
        iqh = idipo + ckc * (iqhc + bdot * iqhbd) + ckc2 * (-qhstr1)
    elseif quad in pl2
        iqh = idipo + ckc * (iqhc + bdot * iqhbd) + ckc2 * (qhstr2)
    elseif quad in ml2
        iqh = idipo + ckc * (iqhc + bdot * iqhbd) + ckc2 * (-qhstr2)
    end

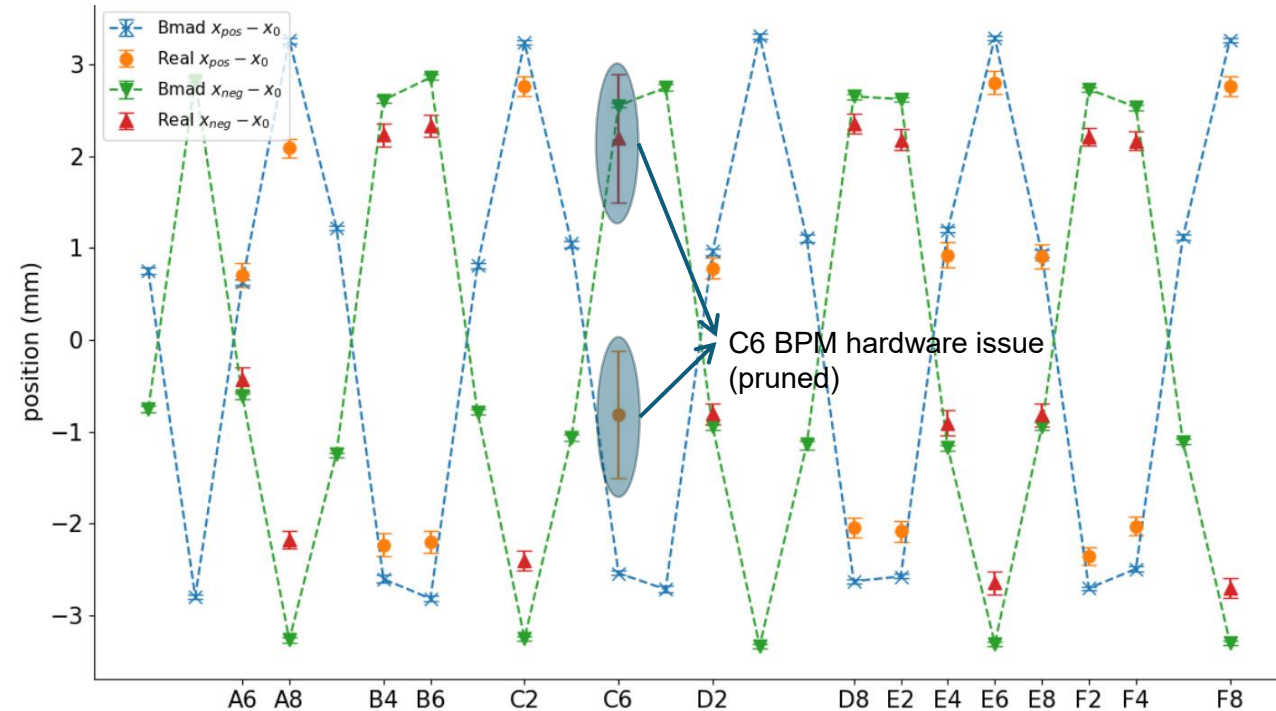
    bllh = klhc0 + iqh * klhc1 + iqh^2 * klhc2 - iqh^3 * klhc3 + iqh^4 * klhc4 - iqh^5 * klhc5
    ckh = (1 - 0.00004179 * (bdot / bdipo(idipo))) * 1.00 * bllh / (brho(idipo) * lenqh) * var

    return ckh
end
```

Dataset

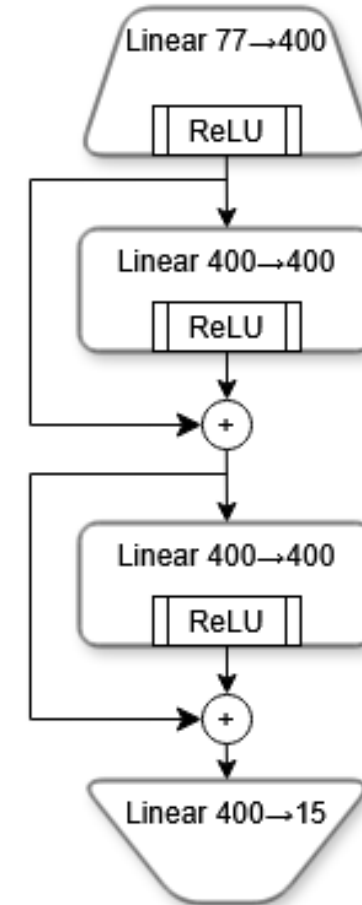
- Fix static magnet currents (up to power-supply fluctuations)
- Baseline setting with correctors set to 0A
- Vary each of the 24+24 corrector magnet currents in turn from 0A to +22A, -22A
- Record orbit for all BPMs before and after perturbation
- 2-5 repeat measurements for each setting
- Construct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

Horizontal (x) orbit response
Corrector A4 0A→±22A



Surrogate Model

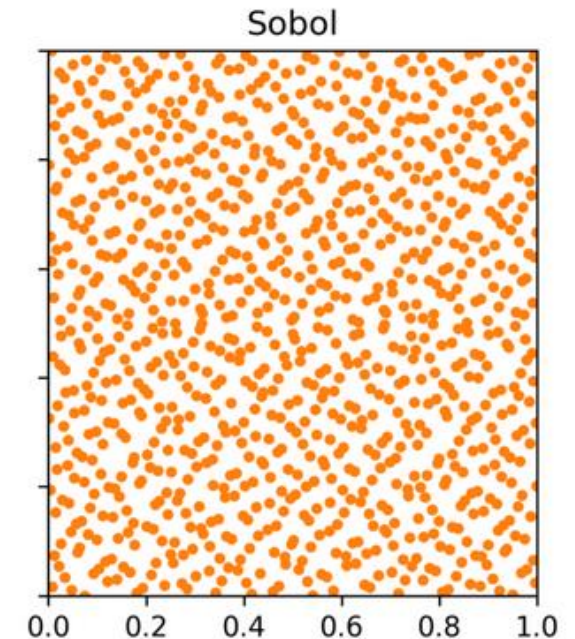
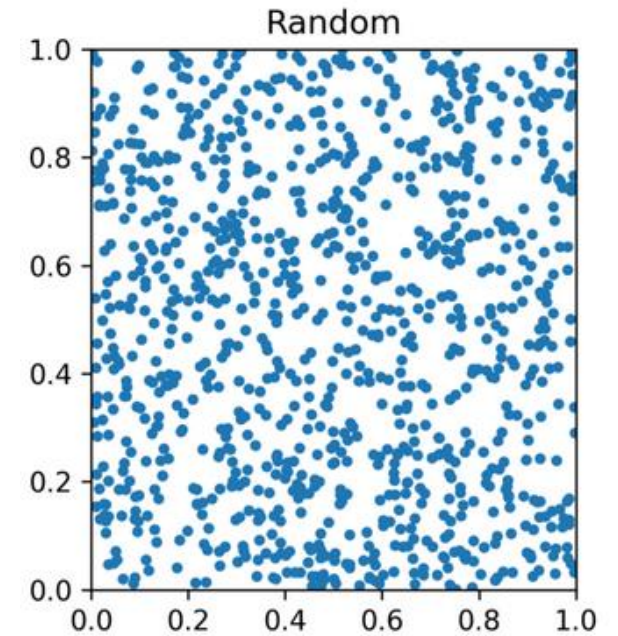
- Sampling the 48-dimensional parameter space is challenging and needs an efficient algorithm: “hybrid (aka Hamiltonian) Monte Carlo”
- This requires many simulation evaluations **including derivatives**
- Replace direct Bmad sim with **machine-learned surrogate model**
 - Require highly accurate surrogate with MSE validation losses $\sim 10^{-4}$ to be $\ll$ BPM error (0.15mm)
 - Train separate model for each plane
 - 77 input features and 15 (x), 18 (y) output features
 - **335,415 params residual neural network (RESNET)**



The model architecture employed for this study

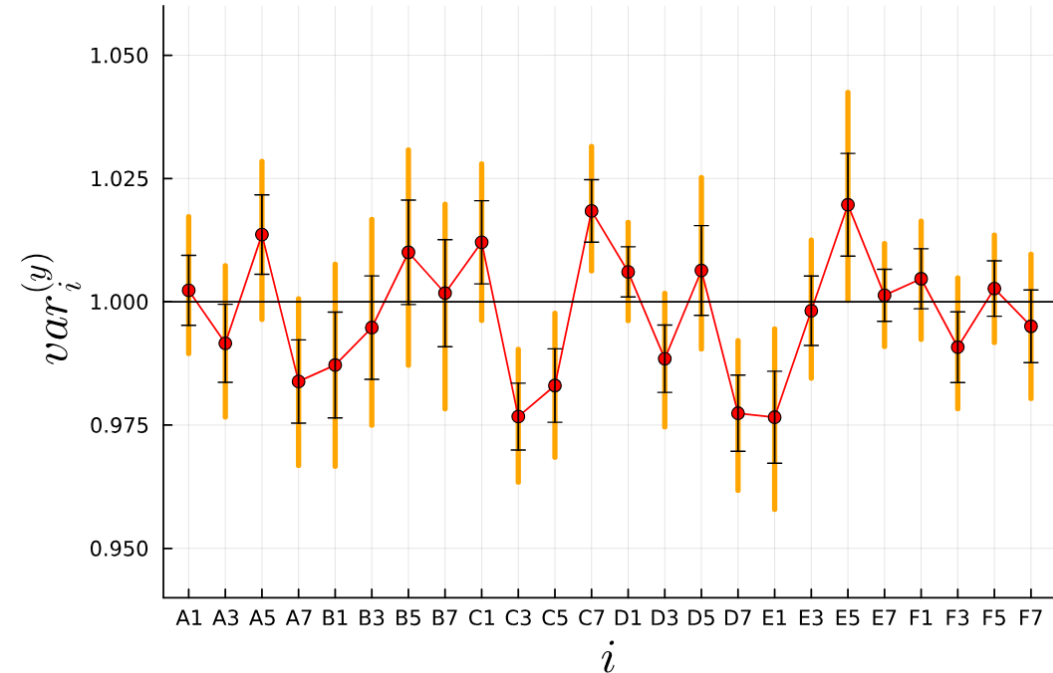
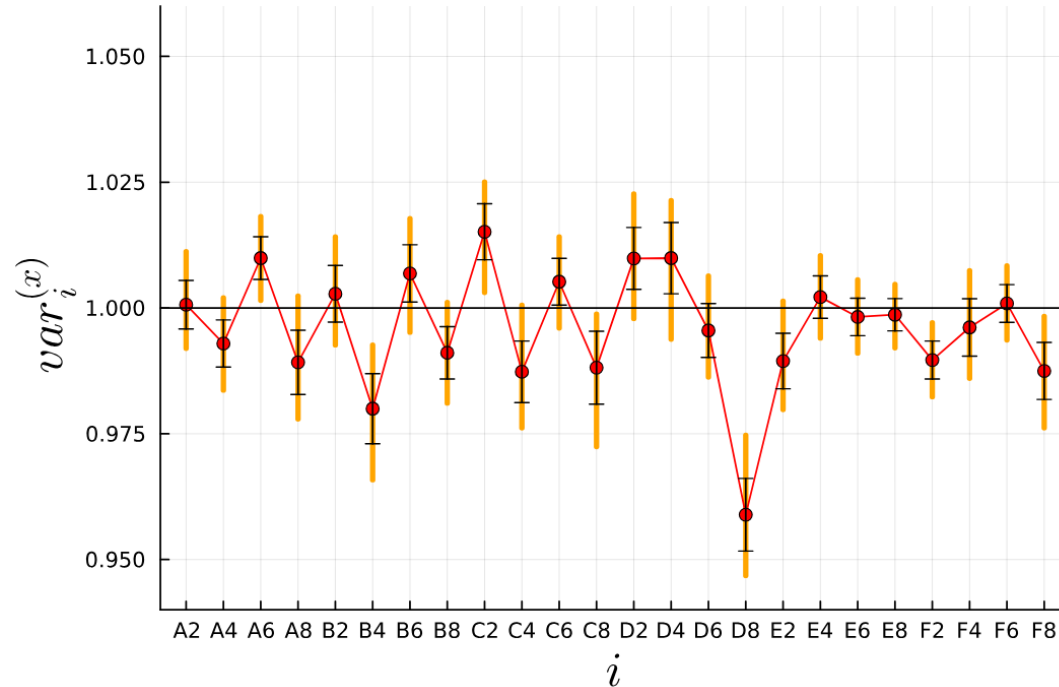
Surrogate data

- 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 ranges for static/corrector magnet currents
- Choosing ranges for the 48 “vars” nuisance parameters most challenging
 - 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 $\pm 0.057\text{mm}$, $\pm 0.017\text{mm}$ vs 0.15mm BPM error



Inference result

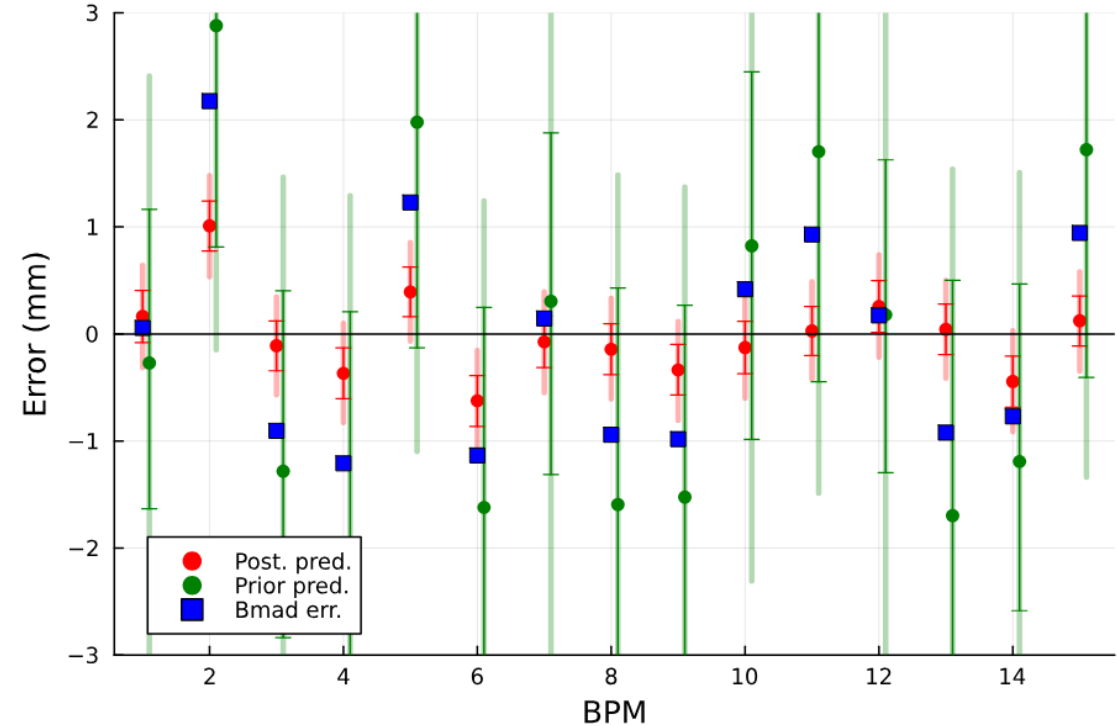
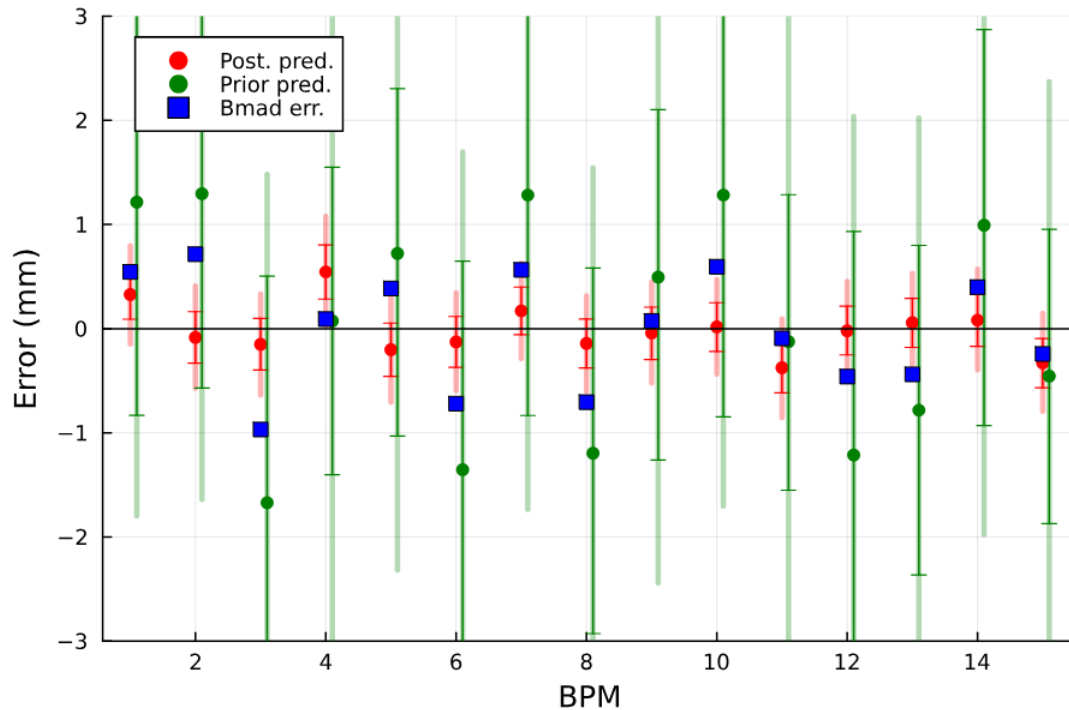
- Explored different prior widths
 - Settled on $\sigma=0.03$ as least constraining prior not overly distorted by truncation.
- Analyzed multiple data subsets with different corrector configurations
 - Smallest uncertainties for subset with orbit responses between -22A and +22A because this amplifies the digital twin error



- Most results cluster around null result to within a few %
- **D8 is notable exception**
 - Although close to prior lower bound, results strongly suggest data favors a low value – hint of real, physical origin?

Posterior/prior predictions

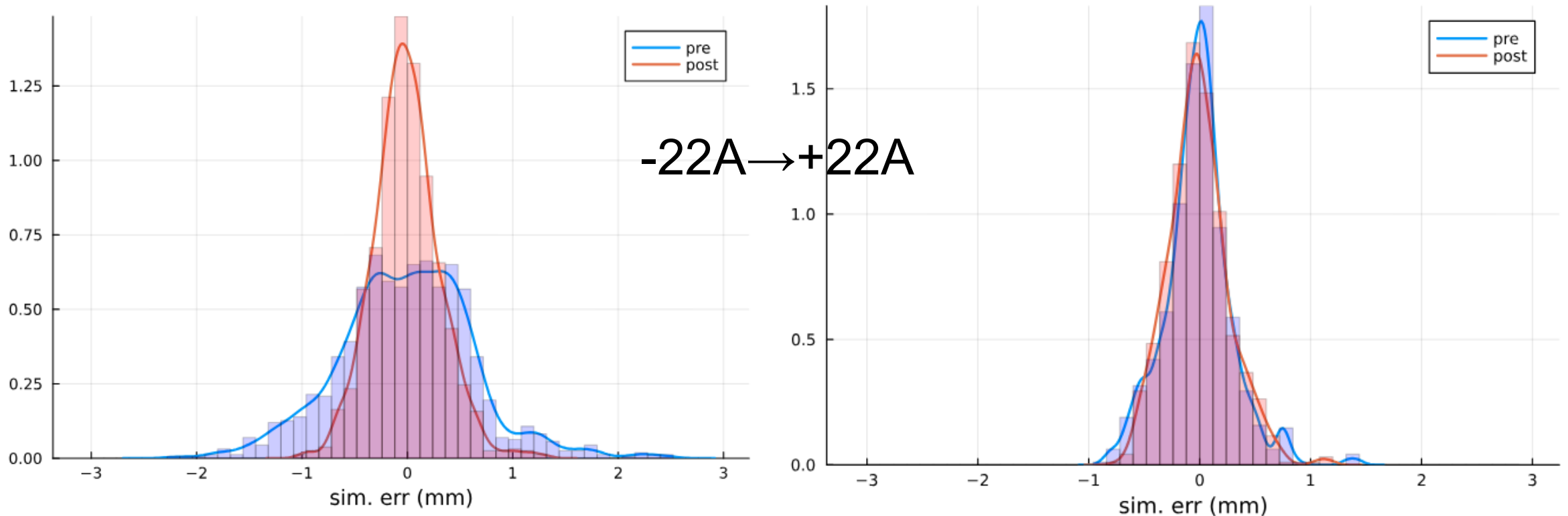
- Two example orbit responses in the $-22\text{A} \rightarrow +22\text{A}$ subset



- General improvement in agreement between model and data

Point estimates

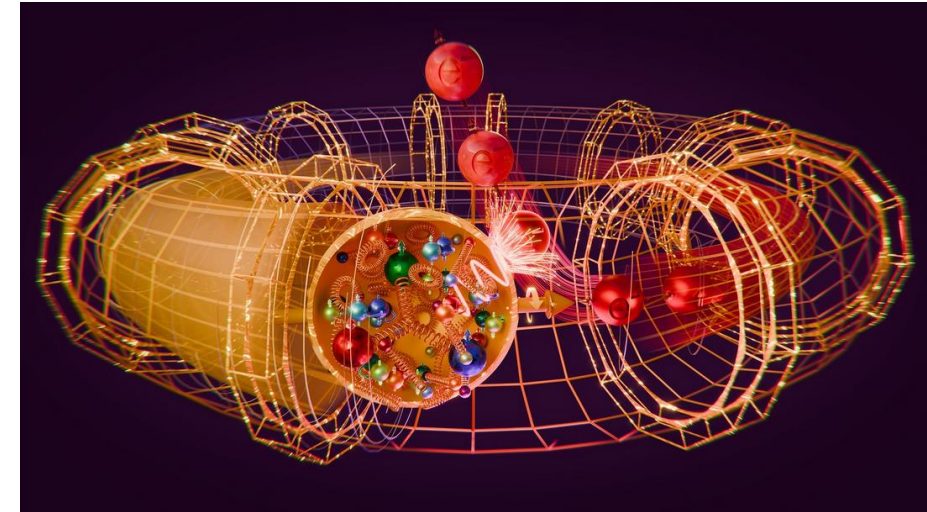
- Posterior predictions still use surrogate – possibility of additional error
- See if incorporating **central values** (point estimates) of var posteriors improved Bmad Digital Twin



- Digital Twin error for -22A → +22A data considerably improved for horizontal plane!

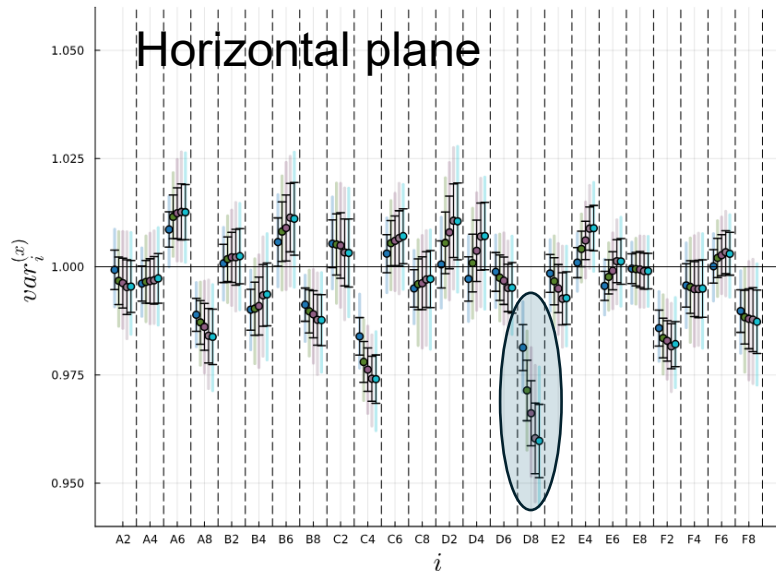
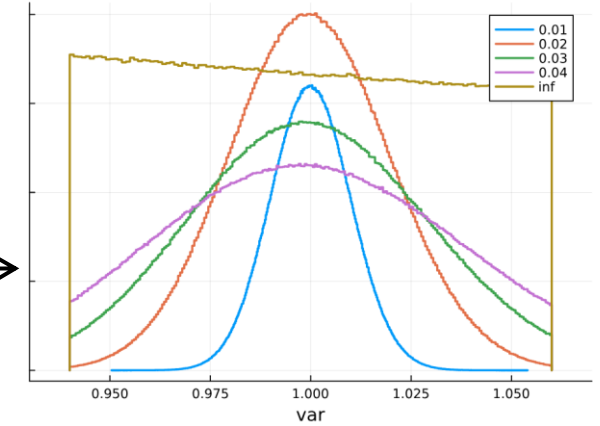
Conclusions

- Introduction of magnet-dependent nuisance 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 to quantifying error sources in 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

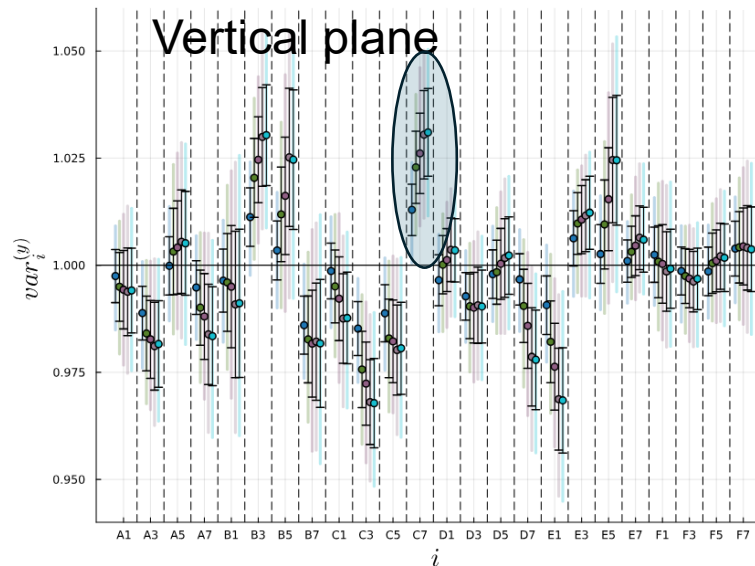


Prior sensitivity

- Lognormal prior width hard to estimate: how sensitive is our inference to the choice?
 - Perform a prior sensitivity study: $\sigma \in \{0.01, 0.02, 0.03, 0.04, \infty\}$
- Apply hard bound on prior distributions at $[0.94, 1.06]$ to account for surrogate validity
 - Distorts prior distributions for large $\sigma \geq 0.04$
- Use data subset where each corrector in turn kicked from 0A to +22A and -22A.



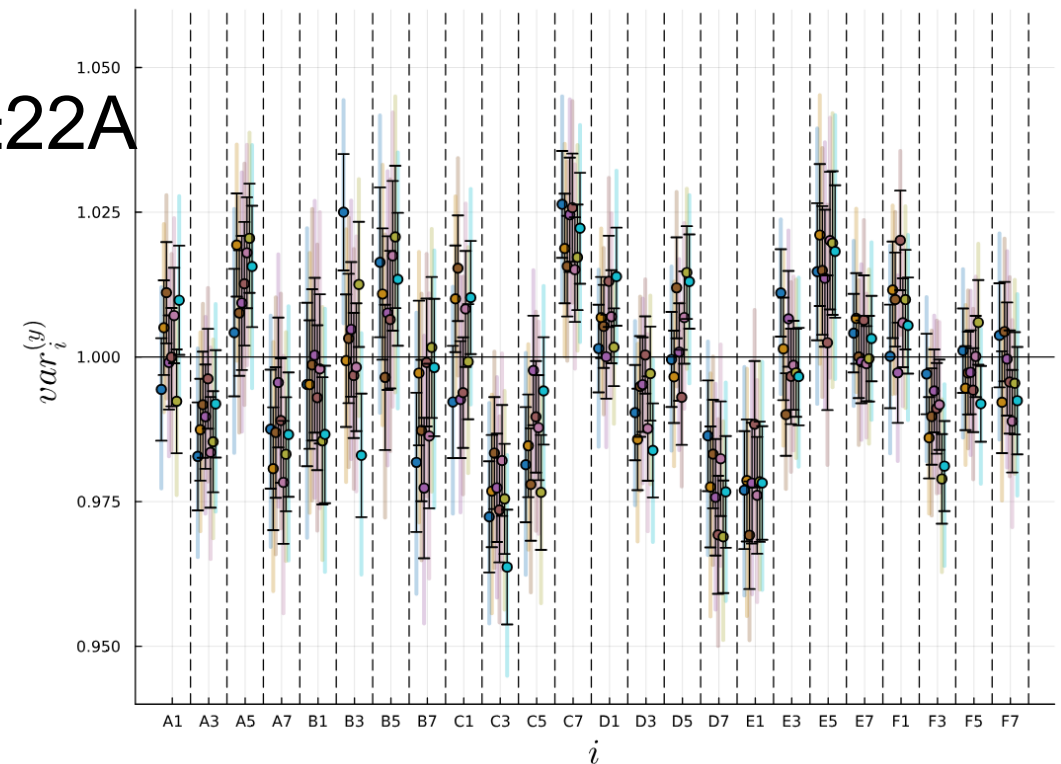
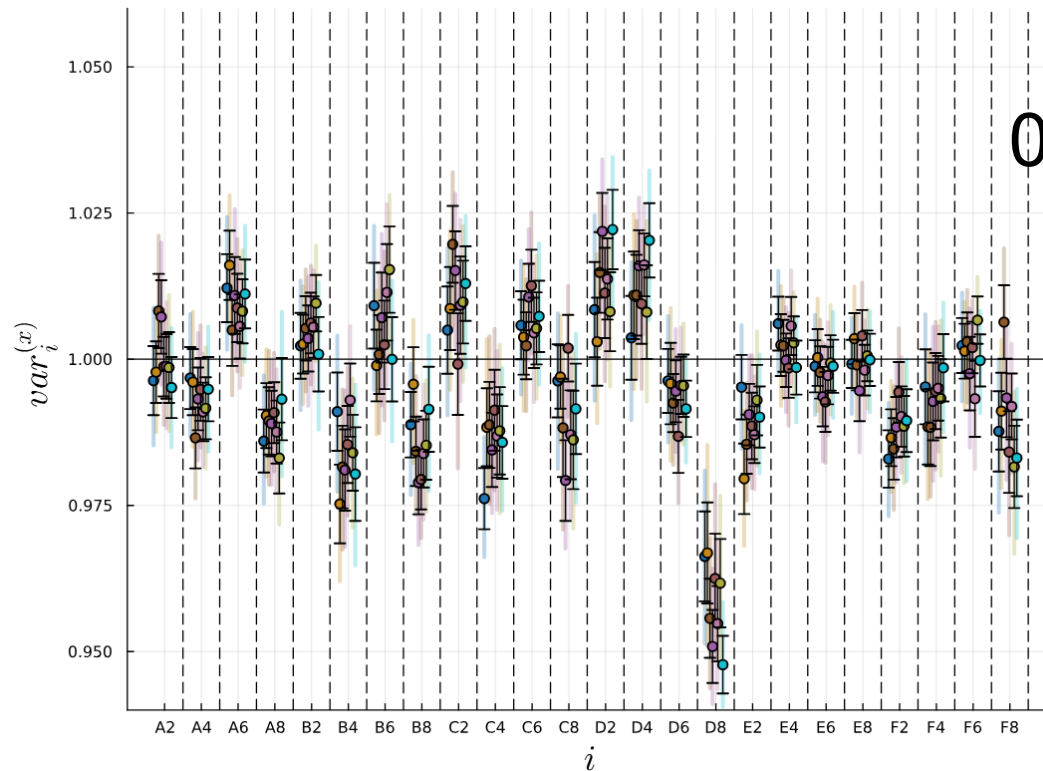
Error bars:
Inner: 68% conf
Outer: 95% conf



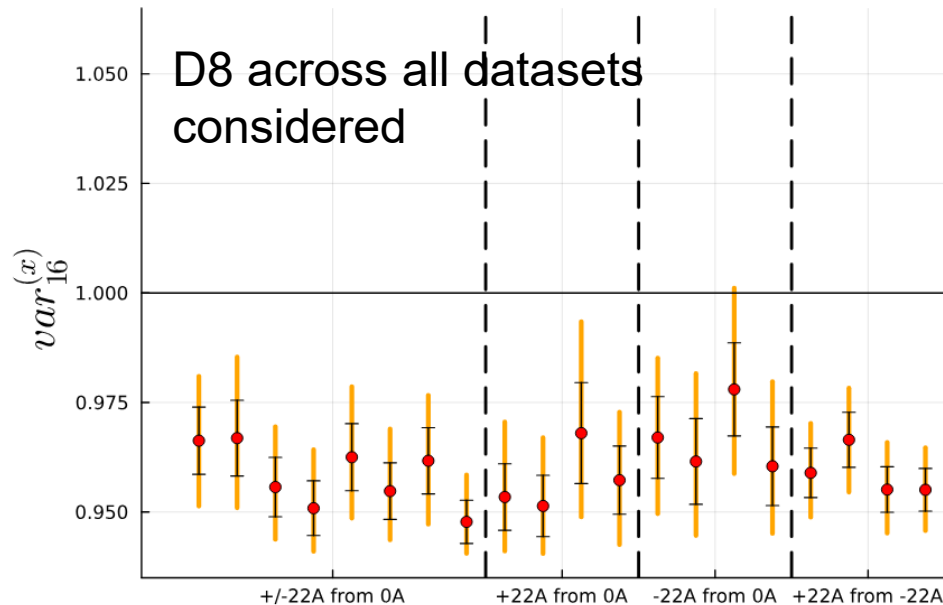
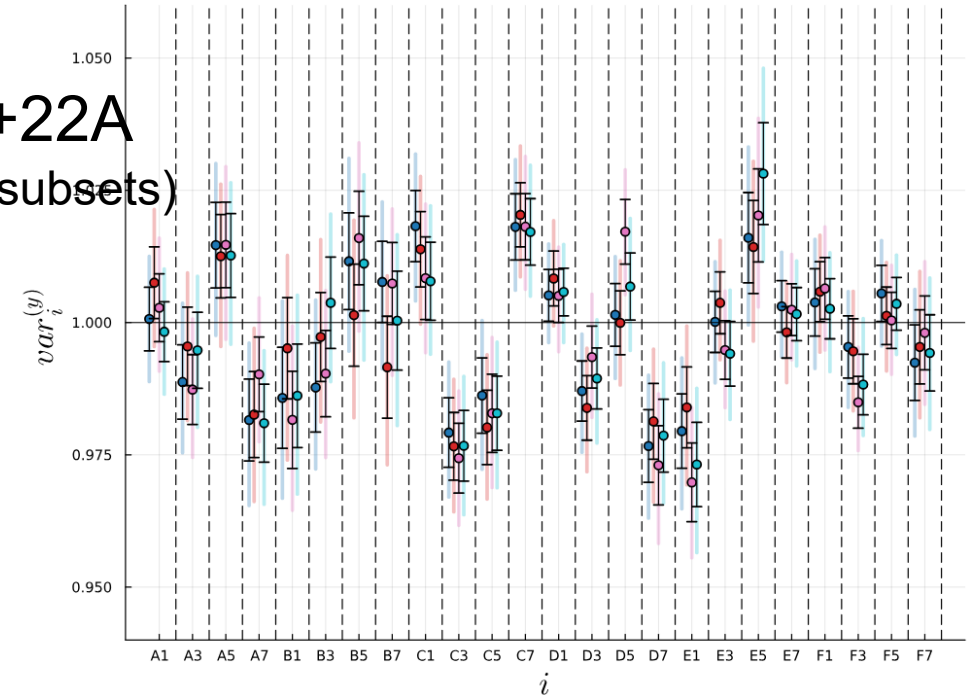
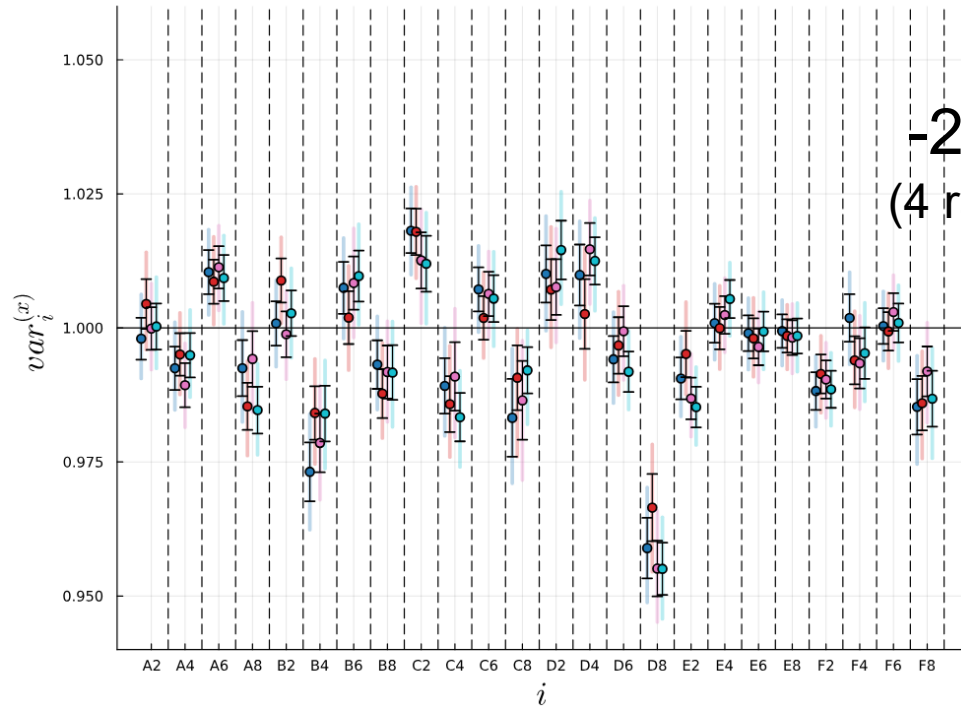
- Results generally stable under relaxing priors
- Stronger dependence on assumptions for some vars hints at near-flat directions in parameter space
- Some vars, particularly x-var “D8”, remain inconsistent with null result and move further away as priors relaxed – hints at real, physical cause?
- Choose $\sigma=0.03$ for final results as least constraining prior not overly distorted by truncation

Dataset dependence

- “Pooling” all data together for inference makes analysis more susceptible to unmodeled systematic effects between different subset classes.
- Instead, we repeat the analysis for different data subsets to test for consistency
 - Randomly draw from repeat measurements for fixed corrector settings
 - **Consider multiple choices of initial and final current setting for the perturbed corrector**

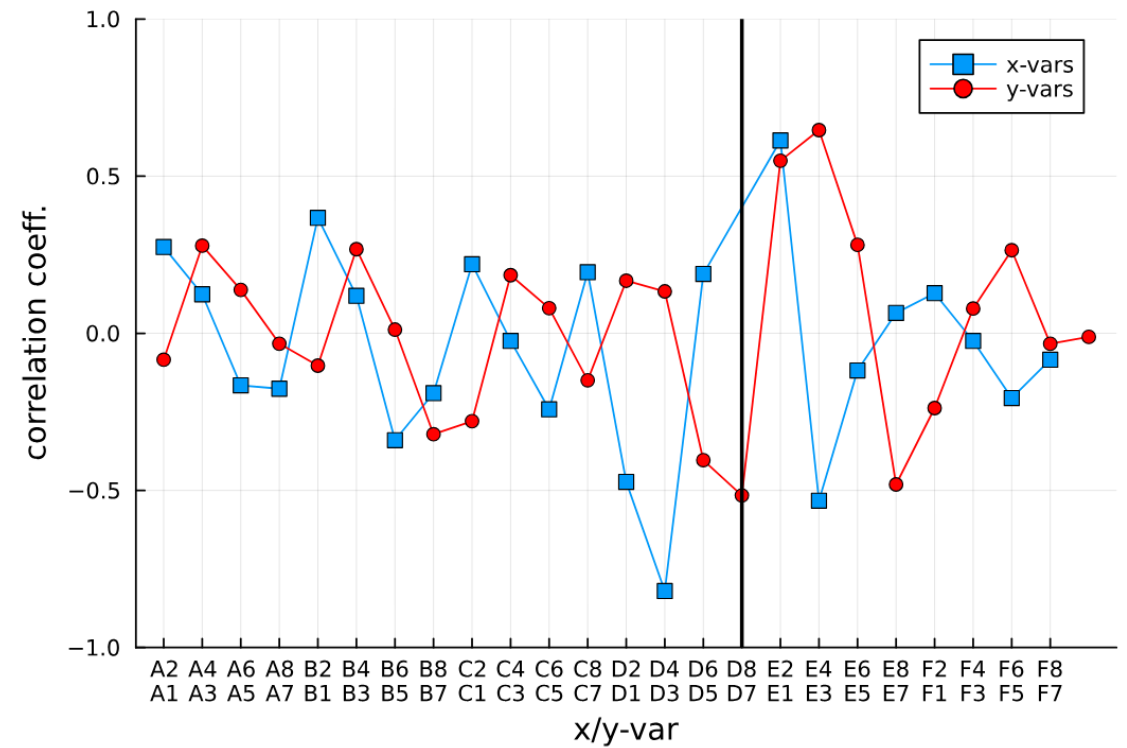
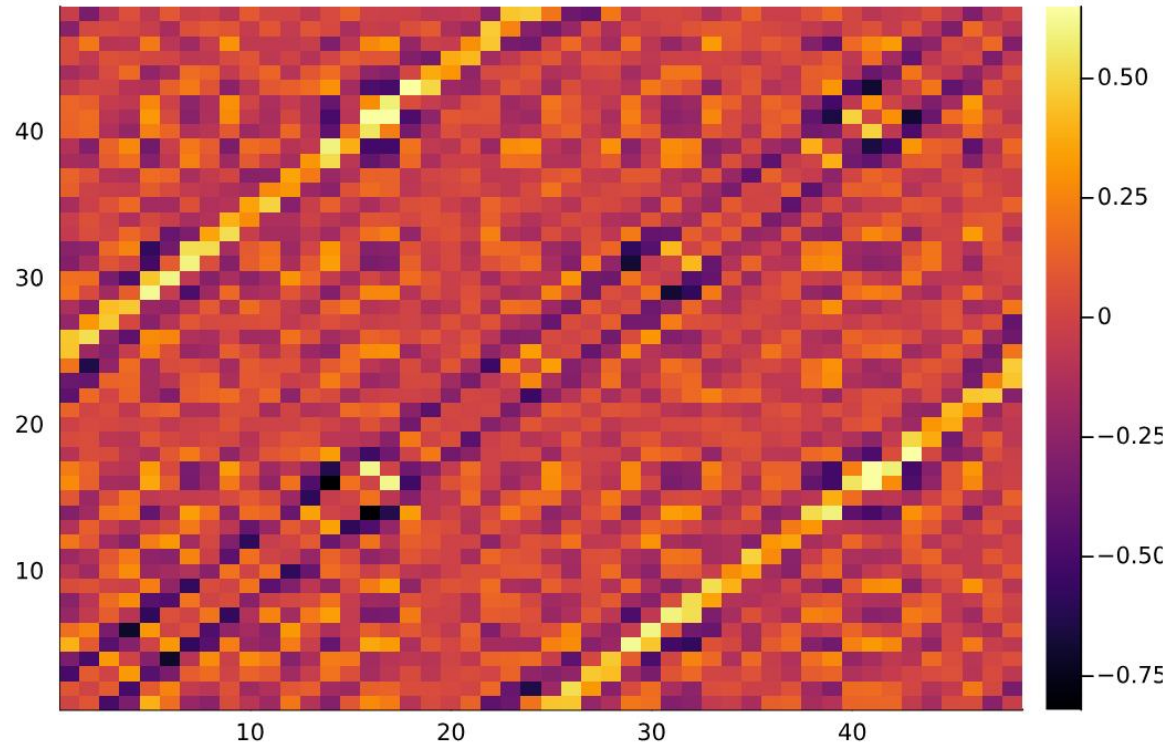


Excellent consistency



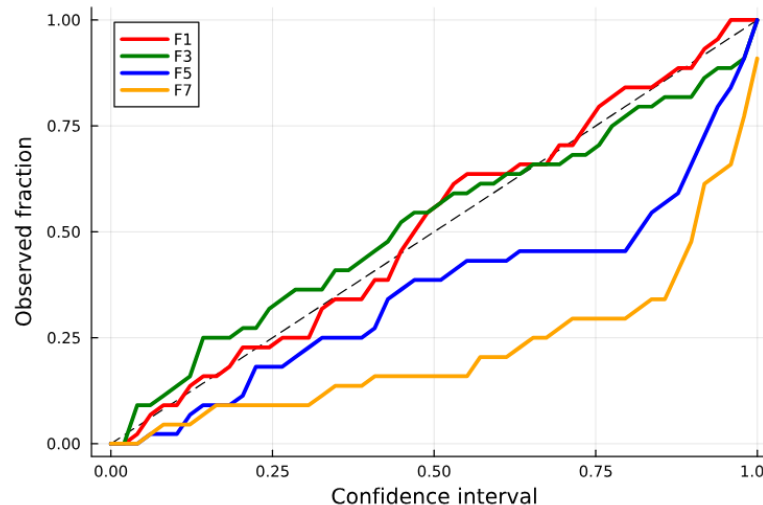
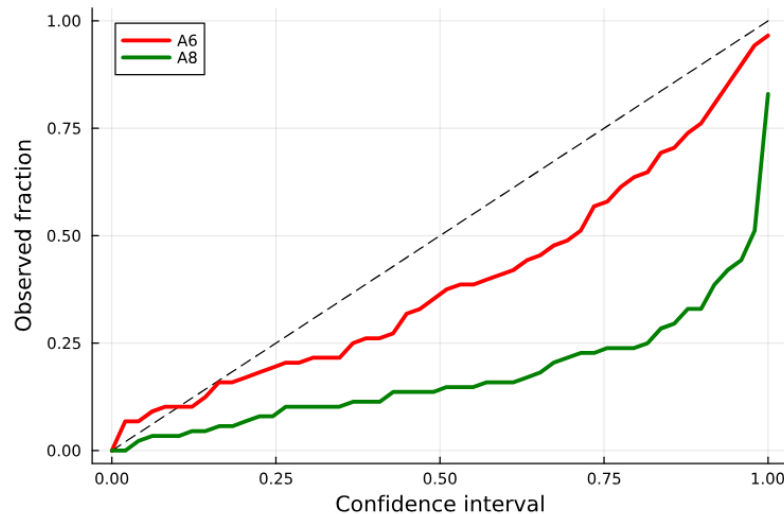
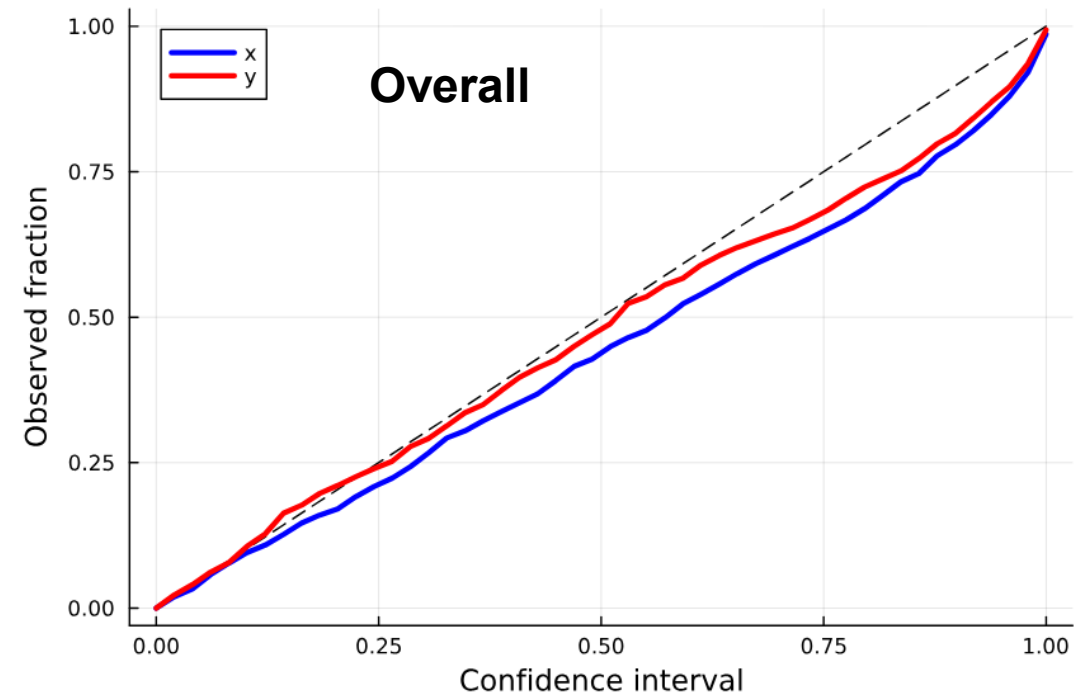
- Uncertainties significantly smaller for -22A → +22A subsets
- Likely because Digital Twin errors amplified by including two extreme corrector values

Correlation structure



Calibration curve

- Question: are the remaining larger-than-errorbar differences in the posterior predictions due to unmodeled behavior?
- Compute fraction of data points within different confidence bands of the corresponding posterior predictive distribution
- Find data distribution is generally well described by our error bars



- Evidence of underestimation of errors (overconfidence) for some BPMS
 - Investigate in future