Panel Session
Data analysis in the era of 100,000+ events

Koustav Chandra, Deep Chatterjee, Francesco Iacovelli, Ling Sun, Aaron Zimmerman (chair)

PAX-X and CE Symposium
July 2, 2025

Click to download as image

What are you most worried about (“online”)?

Scaling searches	Automation	Overlapping events
Low latency infrastructure	Data calibration and quality	
MMA followup	Computational cost	Choose your own fear

What are you most worried about (“offline”)?

Scaling PE	Automation	Scaling population analyses
Model accuracy	Data calibration and quality	Holistic population (cosmo, astro, EoS)
Data curation and release	Computational cost	Choose your own fear

🤕List of our headaches (Should be yours too)

- **Noise Modelling & Glitches**
  Now with bonus jump scares!
- **Too Many Waveforms, PE Methods, Searches**
  "Lost in parameter space 🧭🌀 — Send help (and maybe a flowchart)."
- **Waveform Systematics**
  "Still haunting us like a ghost ex 😬👻"
- **Calibration Uncertainty**
  "What even is the true strain? 🤷‍♂️📉 — If I don't know, it's definitely a problem!" & we want no human intervention (even if to do just astrophysics)
- **Computational Resources**
  "WE NEED MORE CLUSTER!!! 🖥️💥 — and by that, we mean 💰💰💰

Population analysis: holistic approach and scaling challenges

In the next-generation era the focus will shift more and more towards populations!

In the high-statistics regime we will be data-driven:

- Join astrophysical – cosmological – EoS – multimessenger – stochastic – testing GR analyses
- Use flexible models to capture features and correlations

but…

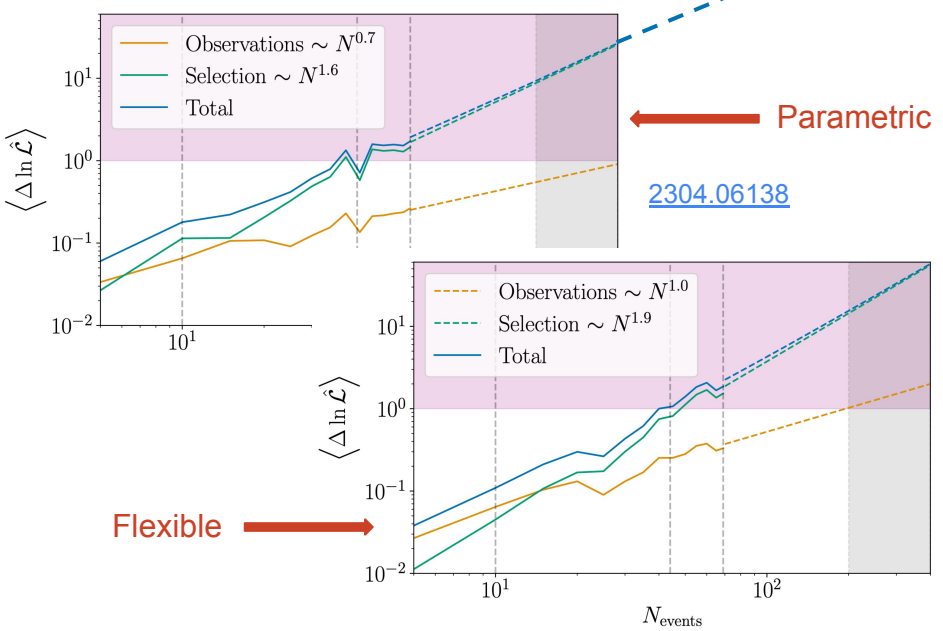
- Keep integrals (especially selection effects) under control if using the "standard" approach, or find alternatives (HMC, SBI,...)
- Handle propagation of systematic uncertainties to the population level

Observations

$$\mathcal{L}(\{\boldsymbol{\theta}\},\{\mathcal{D}\}|\boldsymbol{\lambda}) \propto e^{-N(\boldsymbol{\lambda})} \prod_{i=1}^{N_{\rm ev}} \mathcal{L}(\mathcal{D}_i|\boldsymbol{\theta}_i) \frac{\mathrm{d}N}{\mathrm{d}\boldsymbol{\theta}}(\boldsymbol{\lambda})$$

Selection

$$N(\boldsymbol{\lambda}) \equiv \int \mathrm{d}\mathcal{D}\mathrm{d}\boldsymbol{\theta} p(\mathcal{D}|\boldsymbol{\theta}) \frac{\mathrm{d}N}{\mathrm{d}\boldsymbol{\theta}}(\boldsymbol{\lambda})$$



Challenges in detector calibration

- We need more than one absolute reference, e.g., NCal, Multi-color calibration
- Need fully automated diagnostics and remediation systems to make calibration model changes (can integrate AI)
- How does the huge number of events impact calibration lines, monitoring, and line subtraction?
- Stacking is more likely to be impacted by the same systematics
- Cannot wait for careful investigation in the detectors – events work together with people for calibration

Challenges from low-latency perspective

- Need inter-observatory coordinated Mock-data challenge to prepare for 1K alerts
- Reduce human-in-the-loop
- Need to rely on external compute provided by facilities like ACCESS, NAIRR.

A few other thoughts

- How do we identify interesting events in the flood?
- How can we scale searches for rare events, e.g. lensed events?