Discussion: Deep Space – Deep Learning in Astronomy

A statistician's perspective

Yang Chen
Department of Statistics, University of Michigan

Joint Statistical Meetings 2026
Monday, August 3, 2026 | CC-255

Invited Paper Session
Boston, Massachusetts

Three departures from classical prediction

Rare-event scarcity

Generate strong-flare examples from a learned distribution Q.

data design: $P_n \longrightarrow P_n + Q_m$

Kevin Jin

Measurement entanglement

Separate physical variation from instrument systematics.

$$X \longrightarrow \{\phi_{\text{phys}}(X), \phi_{\text{inst}}(X)\}$$

Daniel Muthukrishna

Structured labels

Replace flat classification with a taxonomy-aware objective.

$$\ell_{\text{flat}} \longrightarrow \ell_{\mathcal{T}}$$

V. Ashley Villar

Common theme: scientific structure is being inserted into the learning problem. The statistical question is whether that structure targets the intended scientific risk.

Common Themes from a Statistical Perspective

Mixed Data Types

- Kevin Jin
  - multi-channel solar images (tensor and/or matrix time series)
  - vector time series (SHARPs)
  - different temporal resolutions (sunspot progression, X-ray flux, SHARPs, images)
- Asley Villar
  - multi-modality data (spectrocopic, photometric)
  - time series
  - with filters for one of the modes
- Daniel Muthukrishna
  - heterogeneous or multi-instrument
  - time series with systematic distortions

Unique features of astrostatistics problems

Science-driven statistical modeling: what are the benefits?

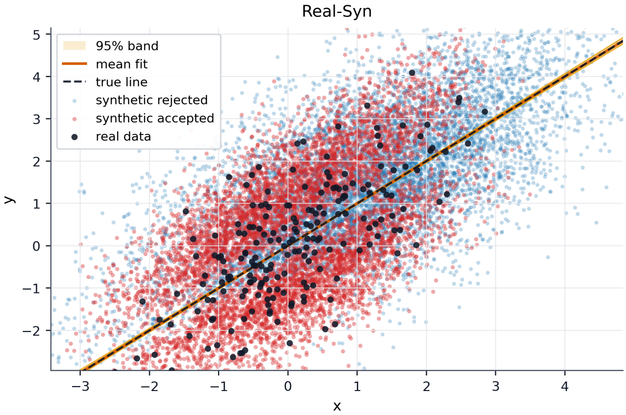
- Kevin Jin
  - machine learning: prediction problem with unabalanced samples
  - physics: power-law distribution of peak flux
- Asley Villar
  - machine learning: classification or hierarchical clustering
  - physics: taxonomy information $\rightarrow$ hierarchical cross-entropy
- Daniel Muthukrishna
  - machine learning: foundation model, dual encoder
  - physics: physical and instrumental factors disentangled

Synthetic augmentation: a game of bias and variance

Let P denote the real target distribution and Q the learned generator. A useful decomposition is

$$R_P(\widehat{f}_{n,m}) - R_P(f_P^*) \approx \underbrace{V(n,m)}_{\text{estimation variance}} + \underbrace{B^2(P,Q,m)}_{\text{generator-shift bias}} + \underbrace{A(\mathcal{F})}_{\text{approximation error}} .$$

- A careful use of bias-variance tradeoff.
- Theoretical goal: a detectable "safe augmentation" boundary, i.e., when does variance reduction exceed generator-induced bias?
- A combination of scientific knowledge and statistical property.
- Question: how it generalizes to operational settings with real-time decision making? Can a formal hypothesis testing be combined?



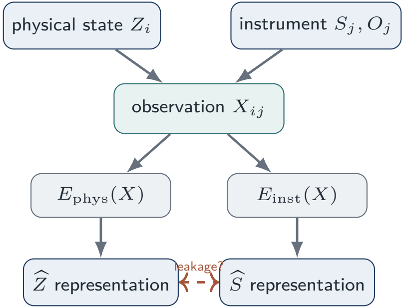
Causal foundation models: model the observation mechanism

Repeated observations supply two kinds of contrast:

- same object, different instruments
  $\Rightarrow$ physical invariance;
- same instrument, different objects
  $\Rightarrow$ instrument invariance.

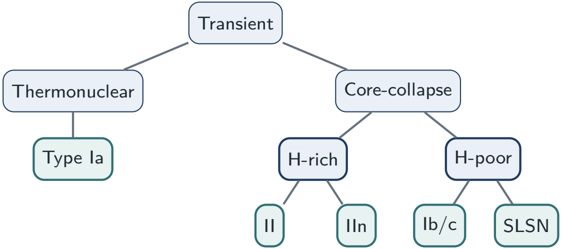
The dual encoders operationalize these invariances through structured contrastive losses.

Question: What happens with missing data of different types (systematic, missing-at-random)?



Audenaert, Muthukrishna, Gregory, Hogg and Villar (2025), "Causal Foundation Models: Disentangling Physics from Instrument Properties."

Hierarchical classification: make the label geometry explicit



- Uses broad and fine labels in one model.
- Produces predictions at multiple resolutions.
- Encourage errors within a related branch.

Fuzzy physics vs. noisy labels

A genuine continuum may require a latent or mixture model; annotator ambiguity may require soft labels or an explicit label-error model.

Question for Ashley:
How should hierarchy weights and class weights be selected when the downstream scientific costs are asymmetric and the taxonomy itself has fuzzy boundaries?

Villar, de Soto and Gagliano (2023), "Hierarchical Cross-entropy Loss for Classification of Astrophysical Transients"; Villar et al. (2026), host-galaxy classification with hierarchical labels.

Questions for discussion

Synthetic data

Can we estimate, from held-out real data, the largest augmentation budget that remains safer than target-only fitting?

Causal representation

Which overlap and structural assumptions make the physical/instrument split identifiable, and how can real data falsify it?

Hierarchical decisions

Can the loss simultaneously encode taxonomy, operational utility and calibrated probabilities under changing class prevalence?

What statistics contributes

A precise target, an explicit observation design, identifiable assumptions, falsification diagnostics, calibrated decisions and uncertainty.

The opportunity is statistical methods whose scientific claims are auditable.