Making the Sum Greater than its Parts

Joshua S. Speagle (沈佳士)

Department of Statistical Sciences

David A. Dunlap Department of Astronomy & Astrophysics

Dunlap Institute for Astronomy & Astrophysics

Data Sciences Institute

UNIVERSITY OF TORONTO

Learning from “Big Data”

- Rapid technological advances have enabled us to collect data on hundreds of millions of astronomical sources (and counting).

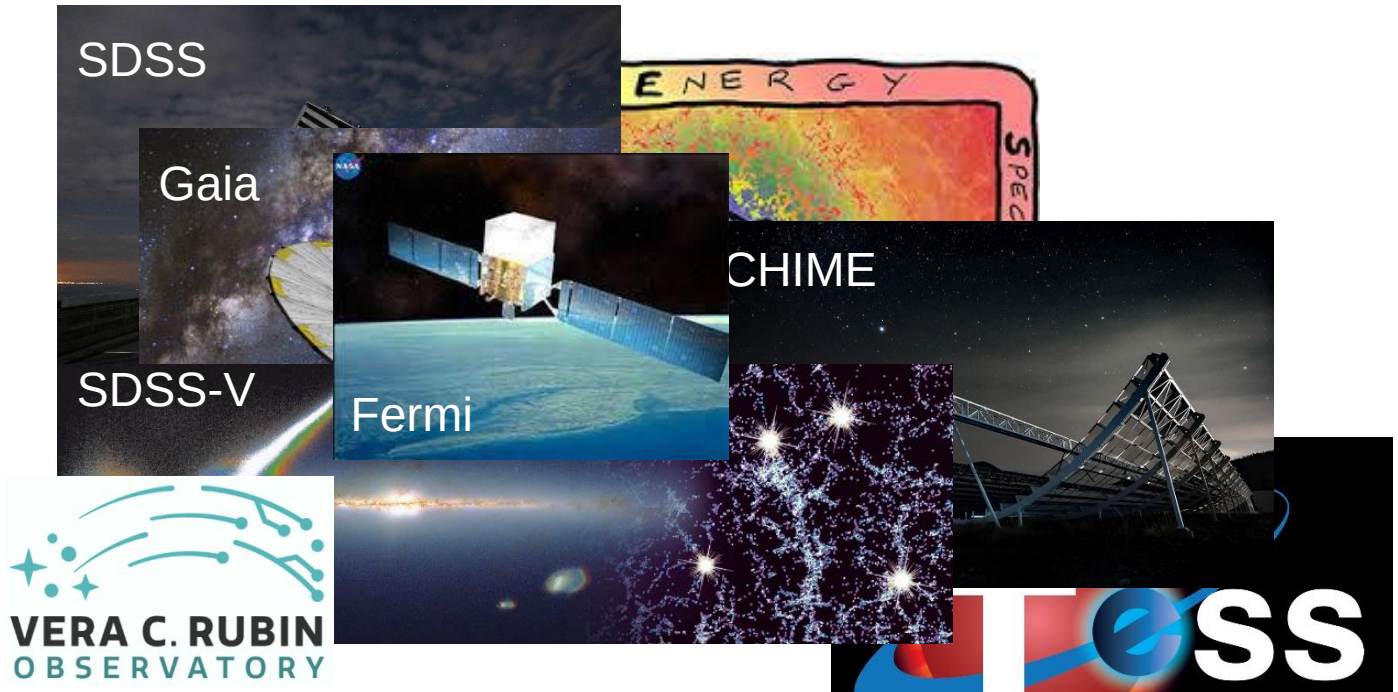
Past: Small populations (~10^5)
→ general trends

Present: Large populations (~10^7)
→ distributions, rare objects

Future: Enormous populations (~10^9)
→ details, substructure, rare populations
→ time resolution

Looking Forward to the “Datapocalypse”

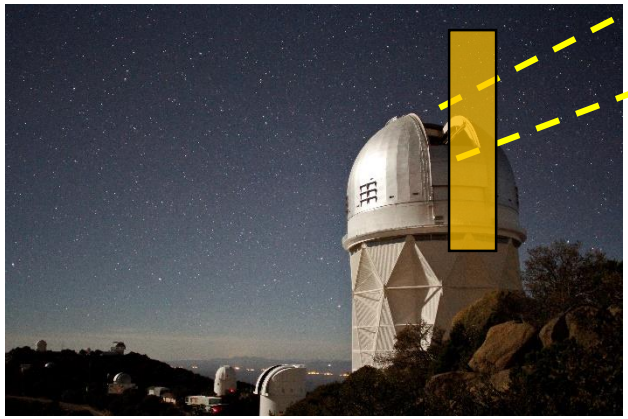
- We currently/will soon have data for **>10^9 sources** that will:
  - Cover large portions of the electromagnetic spectrum.
  - Encompass everything from photometry, spectroscopy, time series, etc.



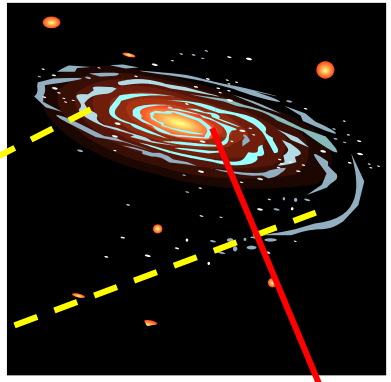
??? → **Science!**

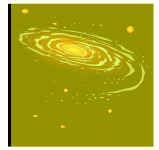
Different Datasets Complement Each Other

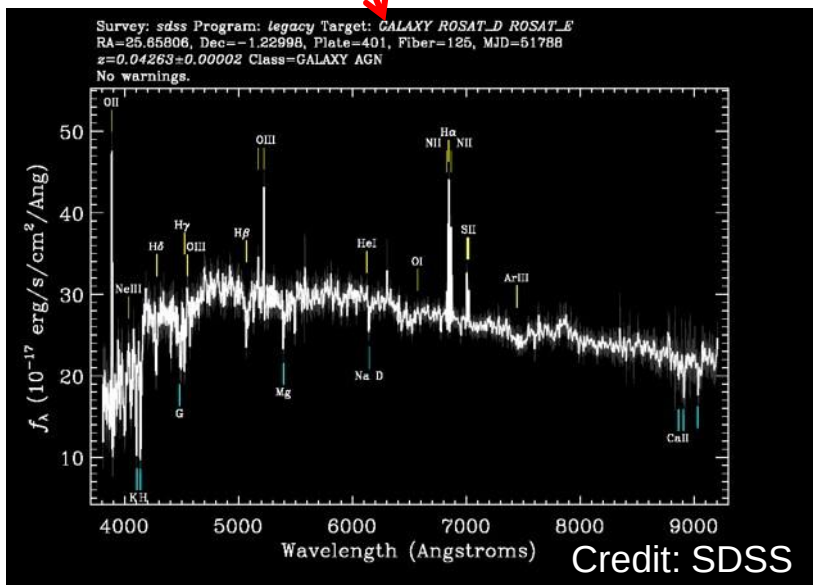
- **Spectroscopy**: Detailed, slow, costly



?



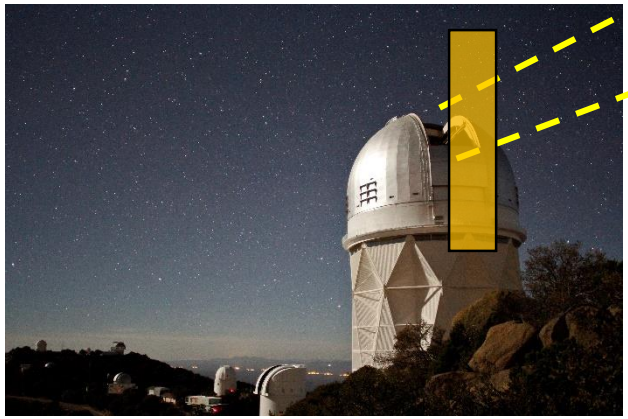




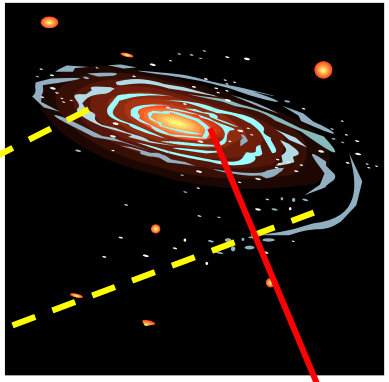
Credit: SDSS

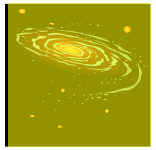
Different Datasets Complement Each Other

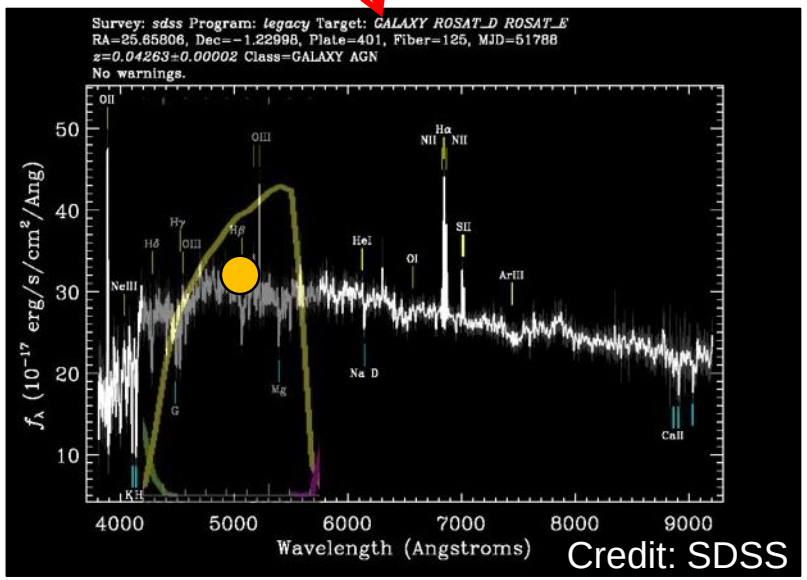
- **Spectroscopy**: Detailed, slow, costly
- **Photometry**: Crude, fast, cheap



?



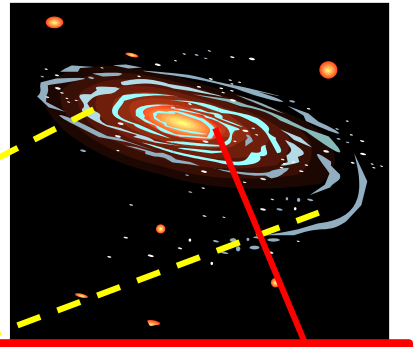




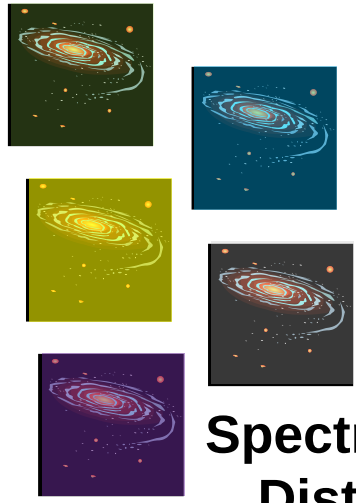
Credit: SDSS

Different Datasets Complement Each Other

- **Spectroscopy**: Detailed, slow, costly
- **Photometry**: Crude, fast, cheap
  - ~100x more objects

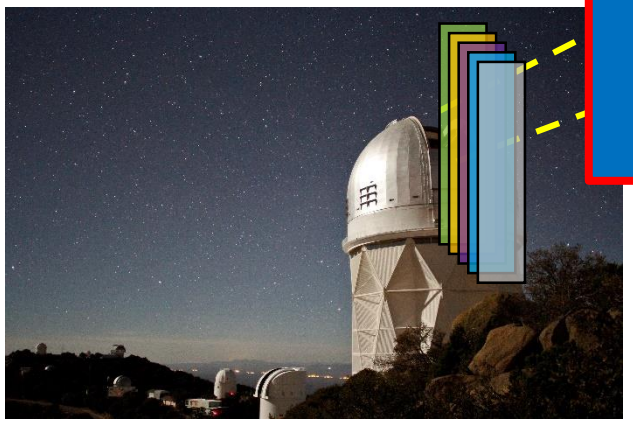


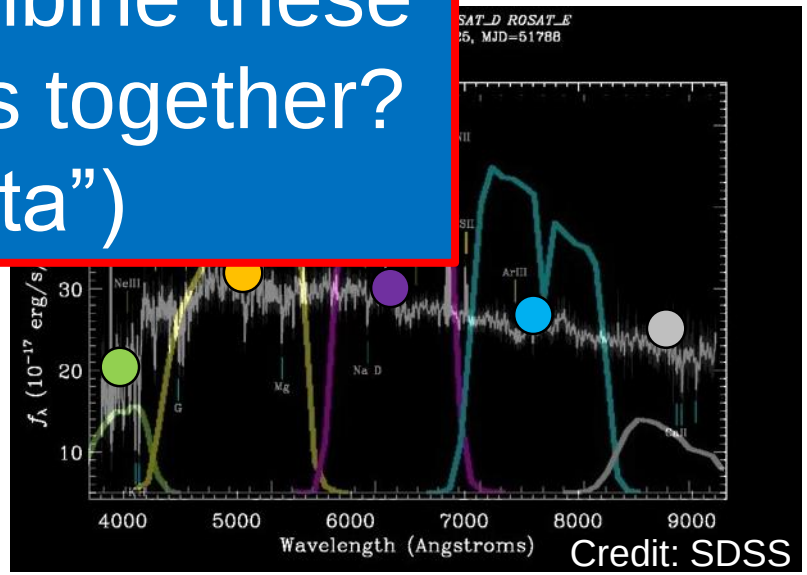
?



Spectral Energy Distribution (SED)

How **can** we combine these different datasets together? (“wide data”)





Credit: SDSS

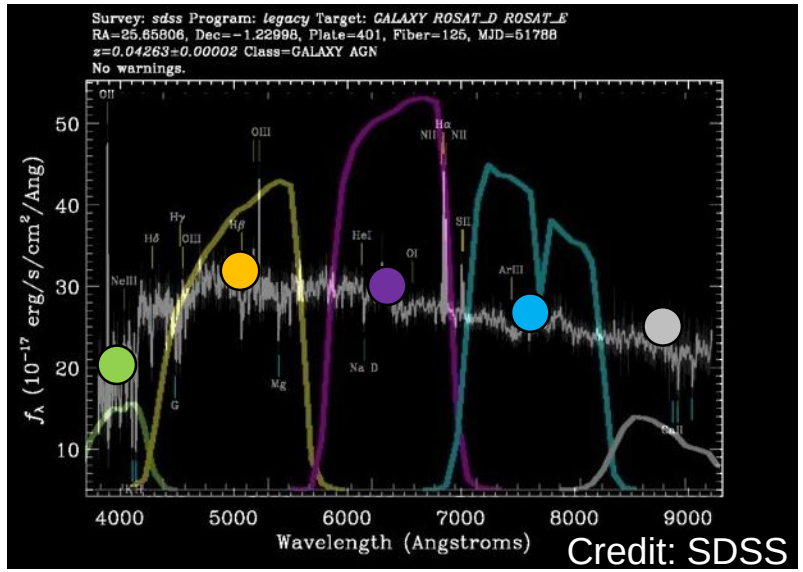
The details matter!

- Often **conceptual/practical problems** when trying to combine datasets or properly model sources come down to **unaccounted systematics**.
  - “Canary in the coal mine”

- Properly modeling and propagating these to other downstream tasks is becoming an ever-larger challenge.
  - Precision cosmology, population studies, background subtraction, etc.

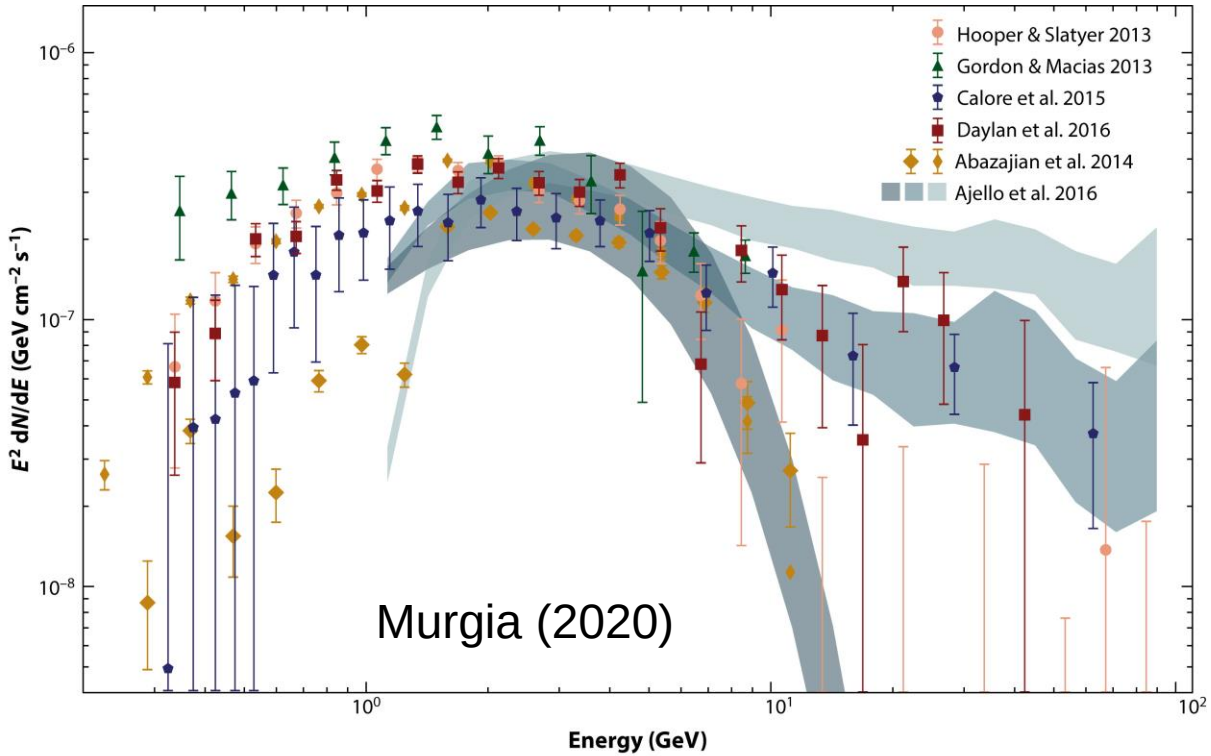
The details matter!

- Example: Spectrophotometry
  - Naïve $\chi^2_{\mathrm{spec}} + \chi^2_{\mathrm{phot}}$ often discounted because $N_{\mathrm{spec}} \gg N_{\mathrm{phot}}$.
  - But if we believe our data this is exactly what we want!
  - Skepticism arises because most approaches ignores different systematics present in spectral data and photometry, which make $N_{\mathrm{spec,eff}} \ll N_{\mathrm{spec}}$.



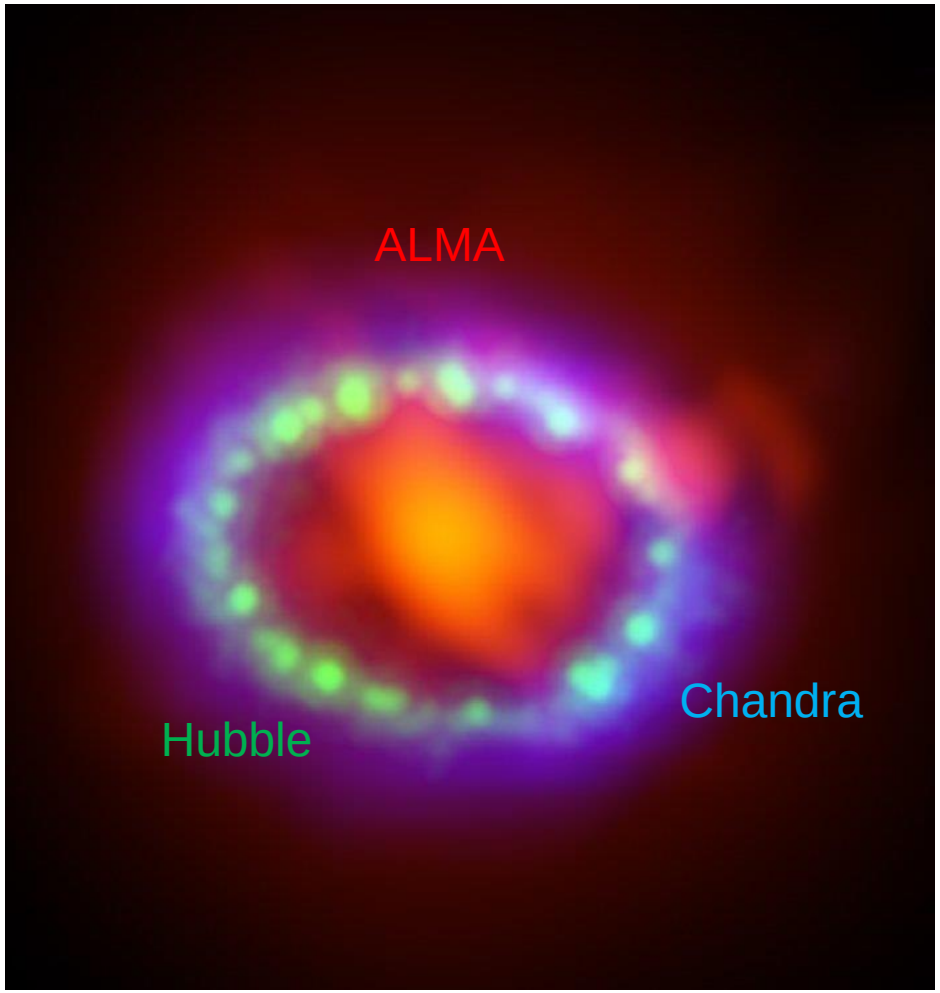
The details matter!

- Example: Galactic Center gamma-ray excess
  - Result *extremely* sensitive to background subtraction, modelling
  - Confirming origins involve multiwavelength observations



Case Study: SN 1987a

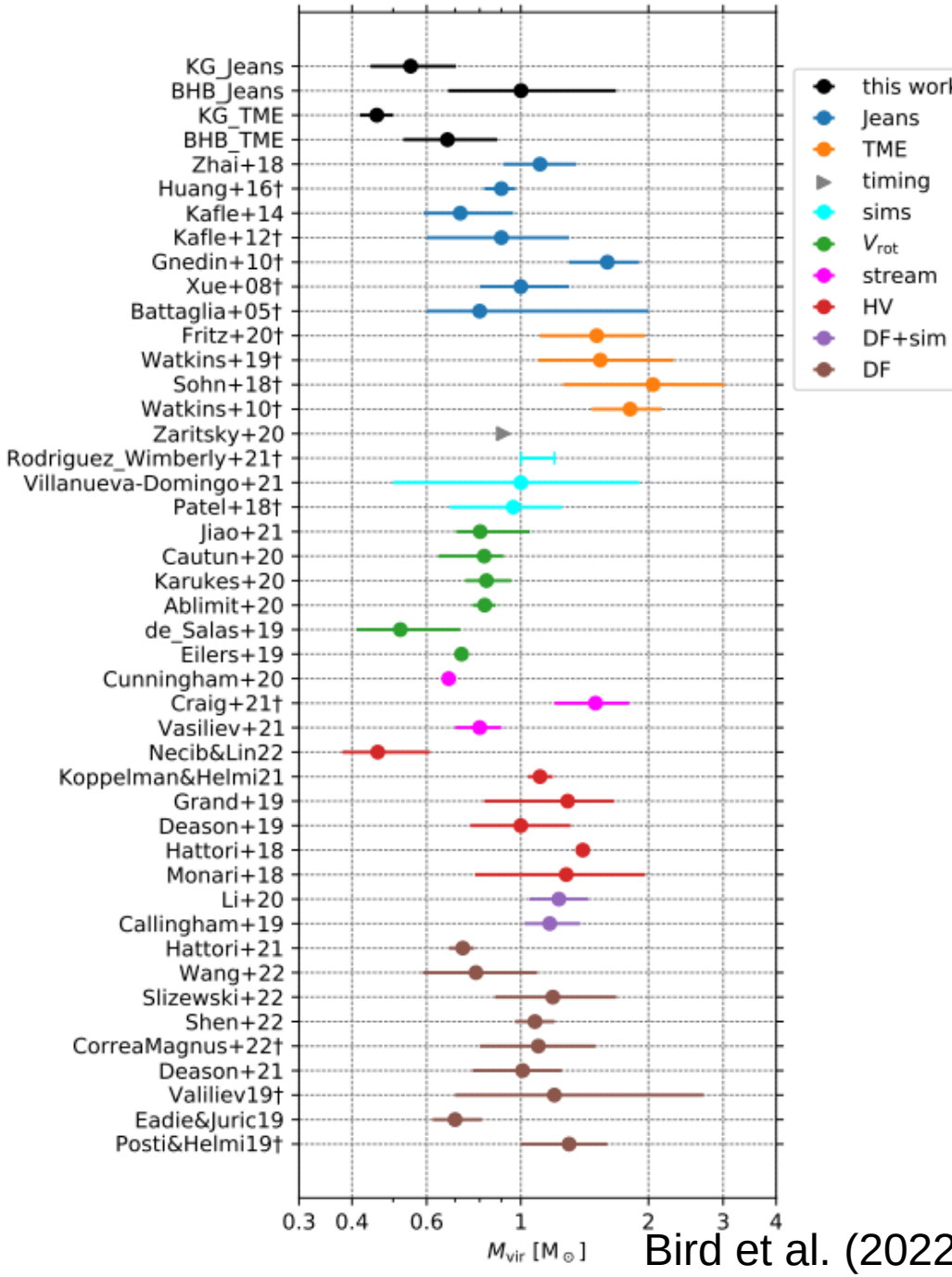
- A panchromatic success story
  - Before detection in **optical** and **ultraviolet** imaging surveys, observed in **neutrinos**.
  - Associations with possible progenitor derived from previous **imaging** data from other surveys.
  - Evolution of the circumstellar material relied on associated **light curve** and time-resolved **spectroscopy** of the ejecta across a wide range of wavelengths, including **X-rays**.
  - Searches for possible stellar remnant have relied on **radio** data.



ALMA/NASA/ESA

Promising Avenues

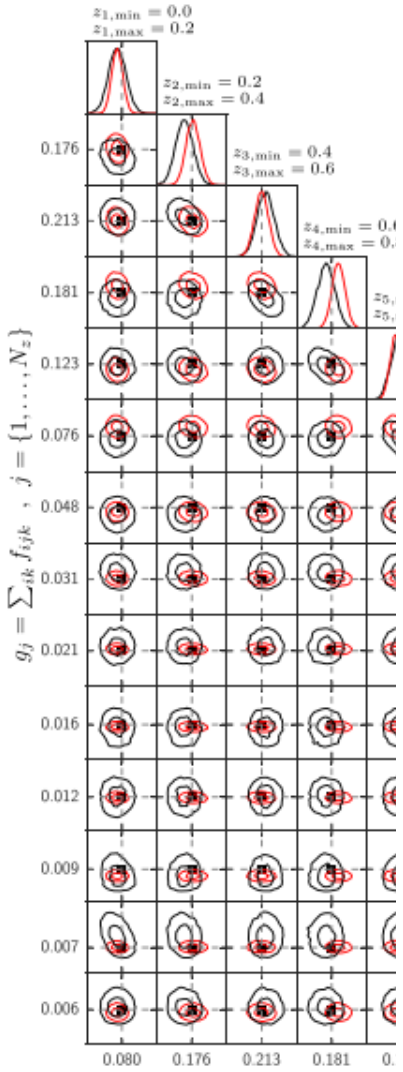
- Multi-messenger astronomy (GWs, TDEs)
- Cosmological tensions (H0, S8)
- Milky Way mass (GCs, halo, streams)

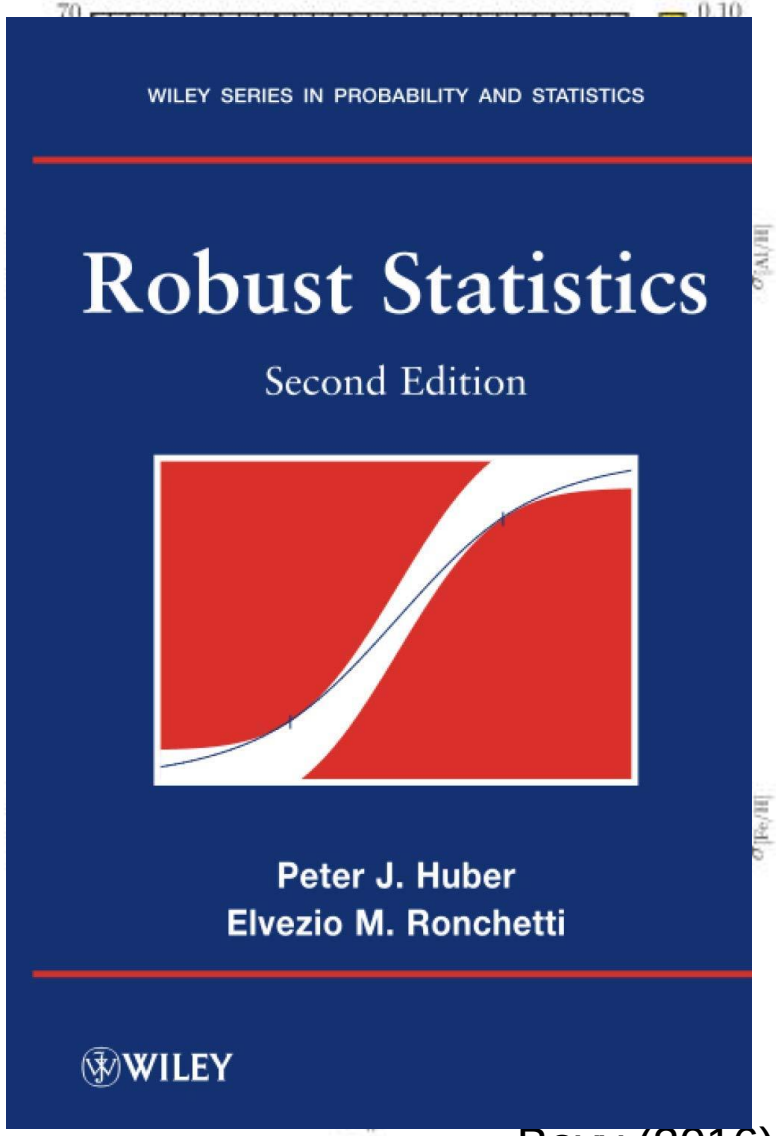


Bird et al. (2022

Looking Ahead: Focus Areas

- **Hierarchical Modelling**
  - “Self-calibration”
- **Likelihood-free Inference**
  - “Simulation-based inference”
- **Robustness**
  - “Proceed with caution”





Bovy (2016)

Key Takeaways

- The upcoming era of **multi-messenger, panchromatic, all-sky astronomy** is poised to be incredibly exciting!

- But will inevitably also be an **incredibly messy**.

- Need to build **robust frameworks for data integration** that go **beyond AIML**.