# Fitting a line to data with large intrinsic spread

Sundar Srinivasan

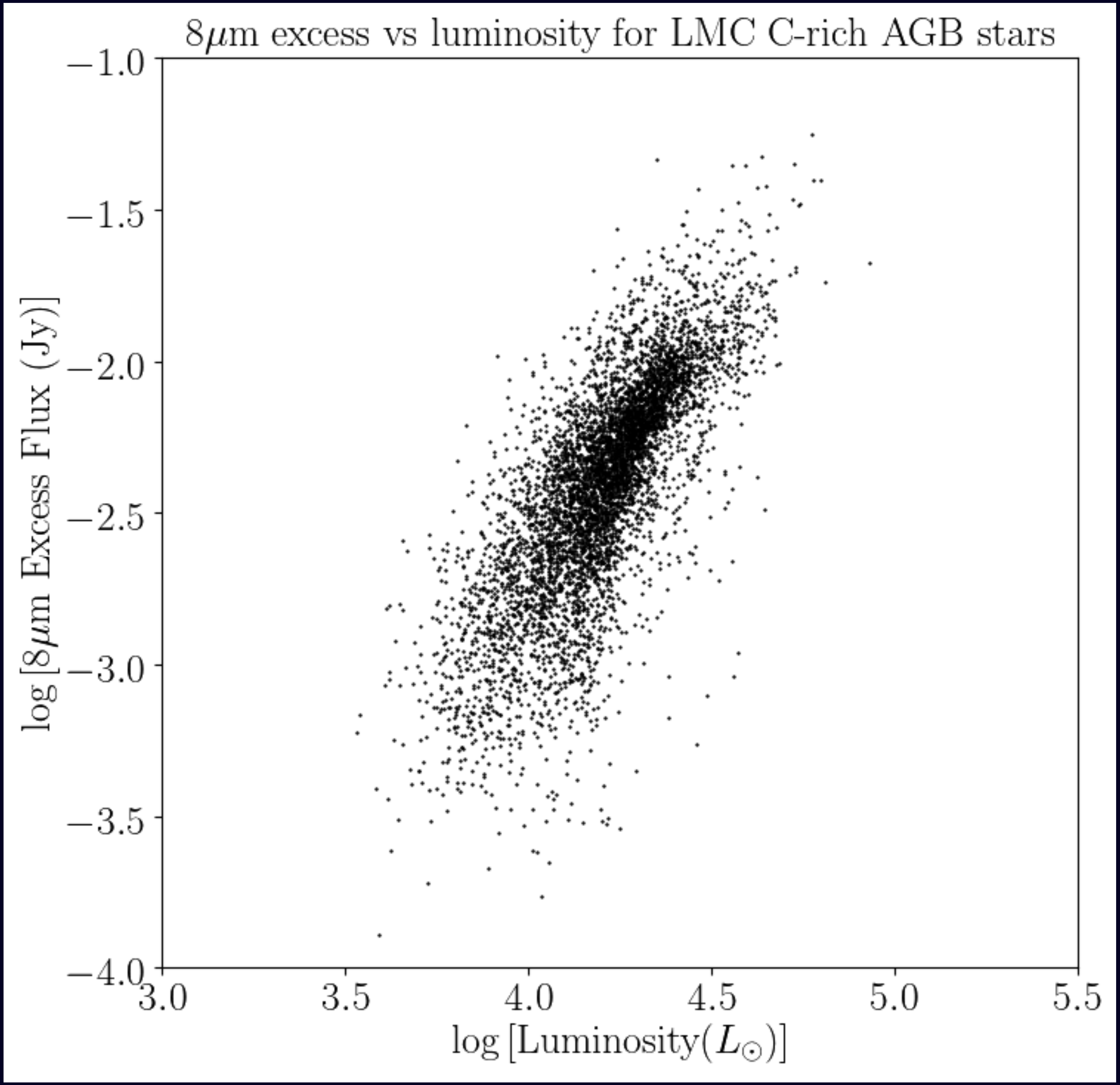
Institute of
Radio Astronomy and Astrophysics
National Autonomous University of Mexico

DAWGI Meeting, 17 September, 2019

# Gist

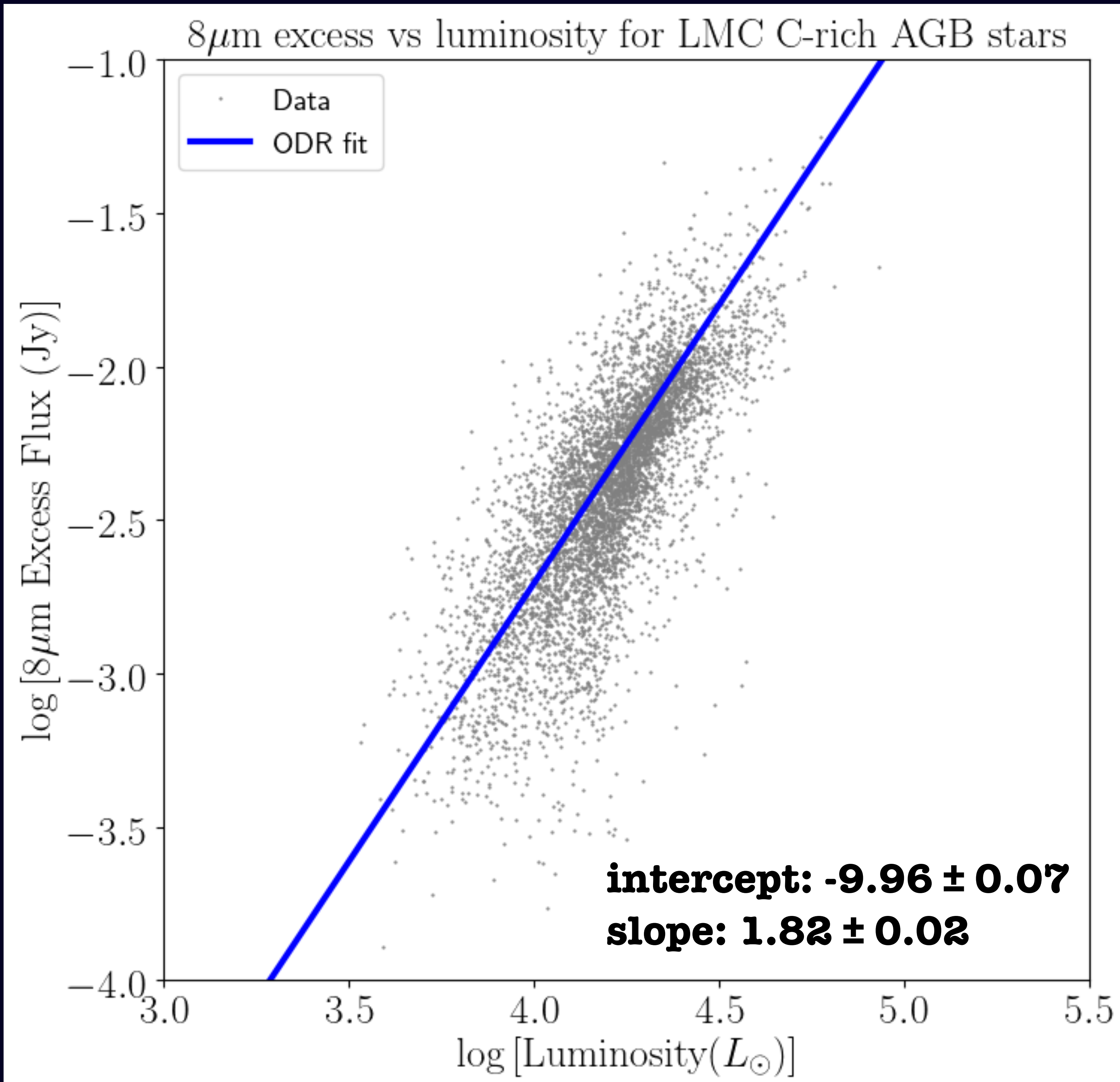
- If typical uncertainties are much smaller than intrinsic spread in the data, standard fitting routines
  1) produce ridiculously low parameter uncertainties, and
  2) produce fits biased towards outliers, if any.
- Hogg et al. (2010) discuss some ways to account for intrinsic variation.
- I'll discuss one example here.

# Data




- 8 μm excess fluxes for C-rich AGB stars in the LMC. Data available **here.**
- Small uncertainties (typically few % along either dimension).
- Intrinsic spread about 1 dex (orthogonal to mean relation). Dusty AGB stars have a large $T_{eff}$ range, and are LPVs.

# Try orthogonal distance regression (ODR)



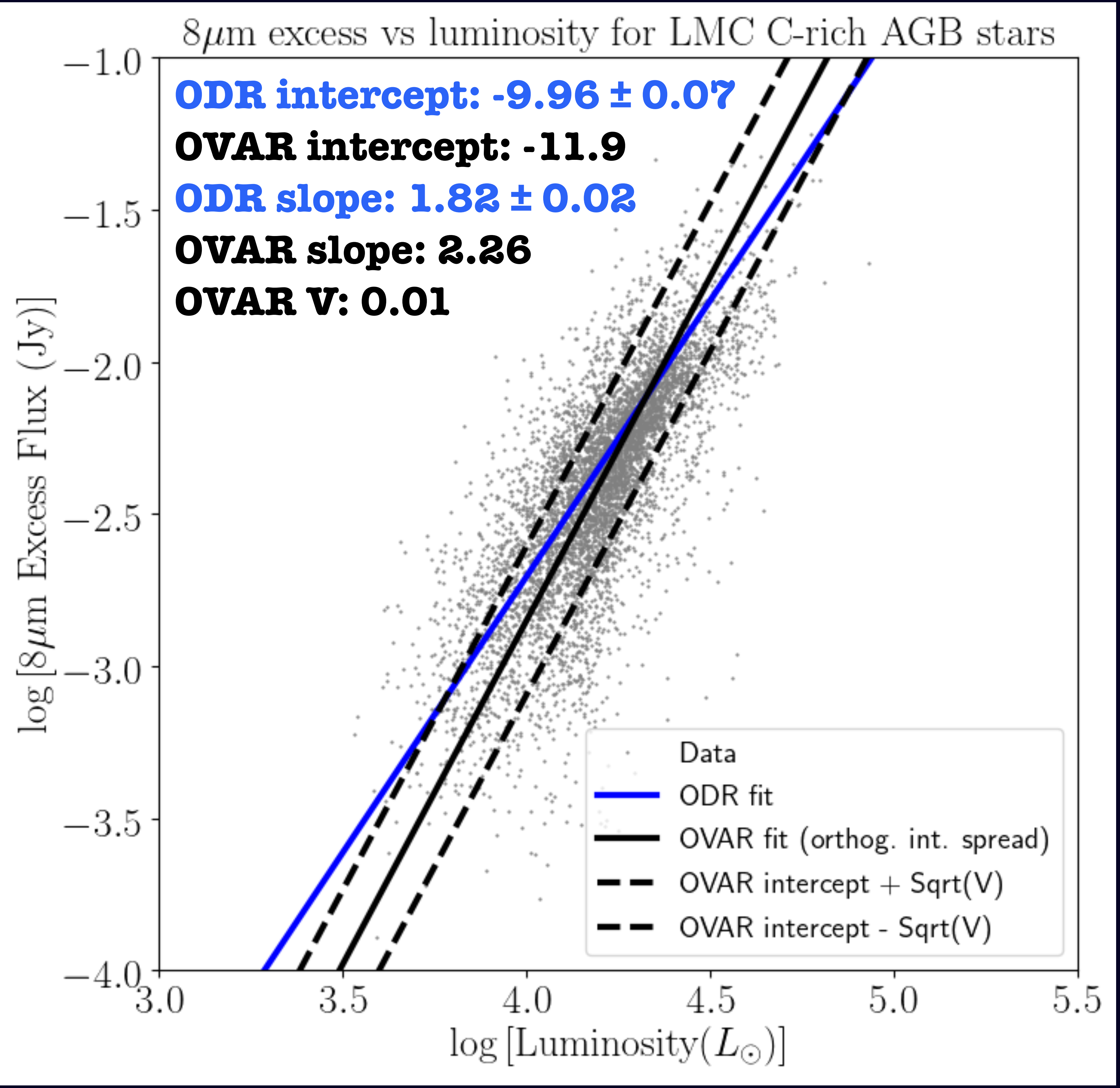

- Initial guesses obtained from simple linear fit.
- 2D uncertainties, “error ellipse” around each point.
- Intrinsic spread not treated.
- Problem 1: when uncertainties << intrinsic spread, parameter uncertainties too low.
- Problem 2: Not robust!

# Assume: intrinsic spread orthogonal to linear relation

$$\ln \mathscr{L} = K - \sum_{i=1}^{N} \frac{1}{2} \ln(\Sigma_i^2 + V) - \sum_{i=1}^{N} \frac{\Delta_i^2}{2\left[\Sigma_i^2 + V\right]}$$

- An extra parameter, V, can be added to the total variance.
- Initial guesses for slope and intercept can be obtained from a simple linear fit.
- Adjust V until the reduced $\chi^2$ until it is about 1. Use the corresponding V as an initial guess.

# ODR vs OVAR




- Better fit!
- More robust due to presence of V.
- V = 0.01 ==> 10% extra spread orthogonal to best-fit line.
- Enough to explain most of spread in data!
- Uncertainties in parameters much more realistic (e.g., dashed lines for intercept).

# References

- Python script available here.
- Hogg et al. (2010), in particular their sections 7 and 8.