Gaussian Process Regression in Microlensing

Amber Malpas

Postdoctoral Scholar

Supervisor: B. Scott Gaudi

What is Gaussian Process (GP) Regression

- $\mathcal{L} = \prod_i^N \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(m-x_i)^2}{2\sigma_i^2}}$, $\quad \ln \mathcal{L} = -\frac{N}{2} \ln 2\pi - \sum \ln \boldsymbol{\sigma} - \frac{\chi^2}{2}$,

  "normal" likelihood, often using χ^2 for the simplified objective function

  assuming Gaussian-distributed noise and no time-dependent correlation.

- $\mathrm{F}(t) = physics(t) + GP(t) + white\ noise$

  For (single source) microlensing lightcurves, $physics(t) = F_S A + F_B$, where A is the magnification model based on the arrangement of the lens

- The GP allows for time-dependent error correlation through use of a covariance matrix (Σ):

$$\ln \mathcal{L} = -\frac{N}{2} \ln 2\pi - \ln \det \boldsymbol{\Sigma} - \frac{\mathbf{r}^T \boldsymbol{\Sigma}^{-1} \mathbf{r}}{2},$$

Equivalent likelihood with correlated errors; reduces to the "normal" $\mathcal{L}$ when $\boldsymbol{\Sigma}$ is diagonal.

where $\mathbf{r}$ is the residual vector, $\Sigma_{nm} = \sigma_n^2 \delta_{nm} + \kappa(t_n, t_m)$, and κ is the (physically motivated) GP kernel(s) that correlates the errors in the time-domain (the off-axis terms in the covariance matrix, Σ).

General Astrophysics Examples

- Gibson et al. (2012), Evans et al. (2015), Grunblatt et al. (2017) transit timing analysis.
- Brewer & Stello (2009), Barclay et al. (2015), Grunblatt, Howard & Haywood (2016), Czekala et al. (2017) used with radial velocity measurements.
- Used to model the background granulation noise in asteroseismic and helioseismic analyses (Harvey 1985; Huber at al. 2009; Michel et al. 2009; Kallinger et al. 2014; and others):

$$\kappa(t_i, t_j) = S_0 \omega_0 e^{-\frac{1}{\sqrt{2}}\omega_0 (t_i - t_j)} \cos\left(\frac{\omega_0 (t_i - t_j)}{\sqrt{2}} - \frac{\pi}{4}\right)$$

$\mathcal{L}$ compute time scales with N^3; limits usefulness to small data sets

Foreman-Mackey et al. (2017) introduces celerite:
Computation time scales with N, exploiting kernels composed of complex exponentials

Variability in the source star (Li et al. 2019)

- Asteroseismology to determine source size and distance; which breaks the distance-distance degeneracy in the microlensing model.
- GP models are consistent with χ^2 models to $\lesssim 3\sigma$.
- Quasi-periodic kernel

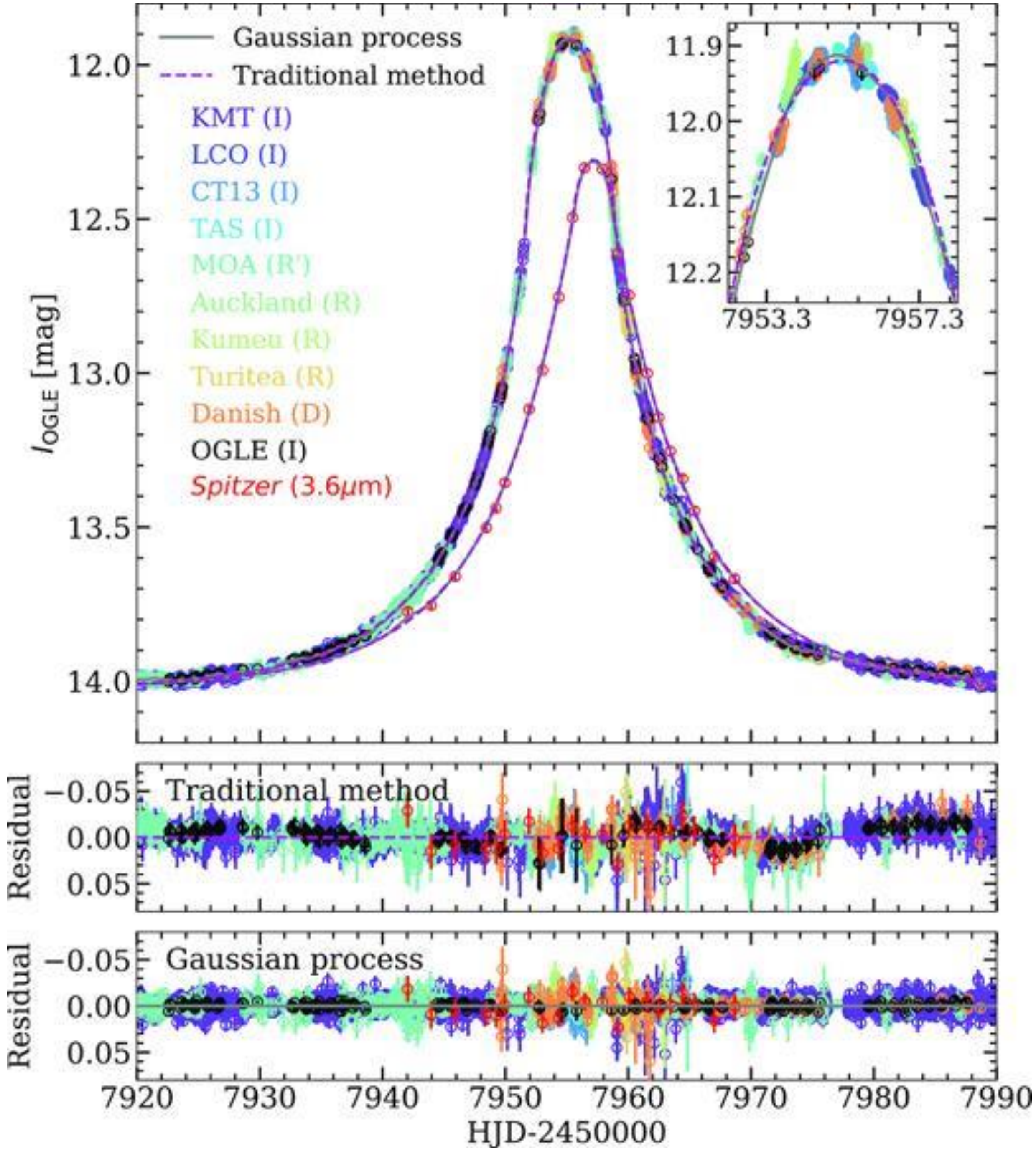
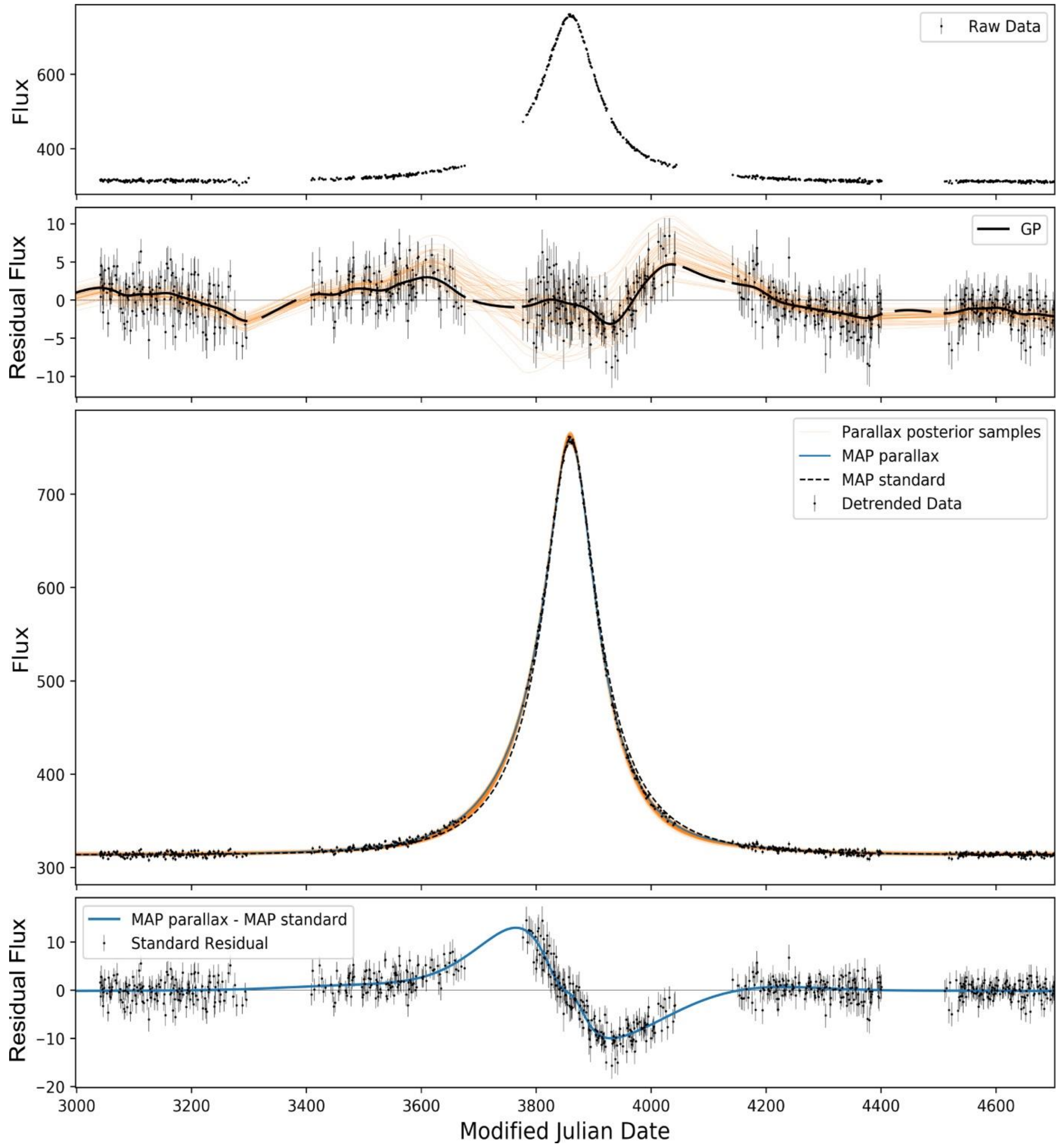


Figure 1 from Li et al. (2019)

Black Holes

- Events last multiple seasons
  - Blend star arrangements change due kinematics
- Physical sources of systematics includes, stellar variability, weather, sky conditions, etc.
- Golovich et al. (2022) used *celerite* for GP regression in the search for black holes in OGLE-III and -IV survey data
  - $\Sigma_{ij} = \kappa_{SHO}(t_i, t_j) + \kappa_{\frac{M3}{2}}(t_i, t_j) + \boxed{K^2\sigma^2\delta_{ij}.}$
- "[They] find that modeling the variability in the baseline removes a source of significant bias in individual events"



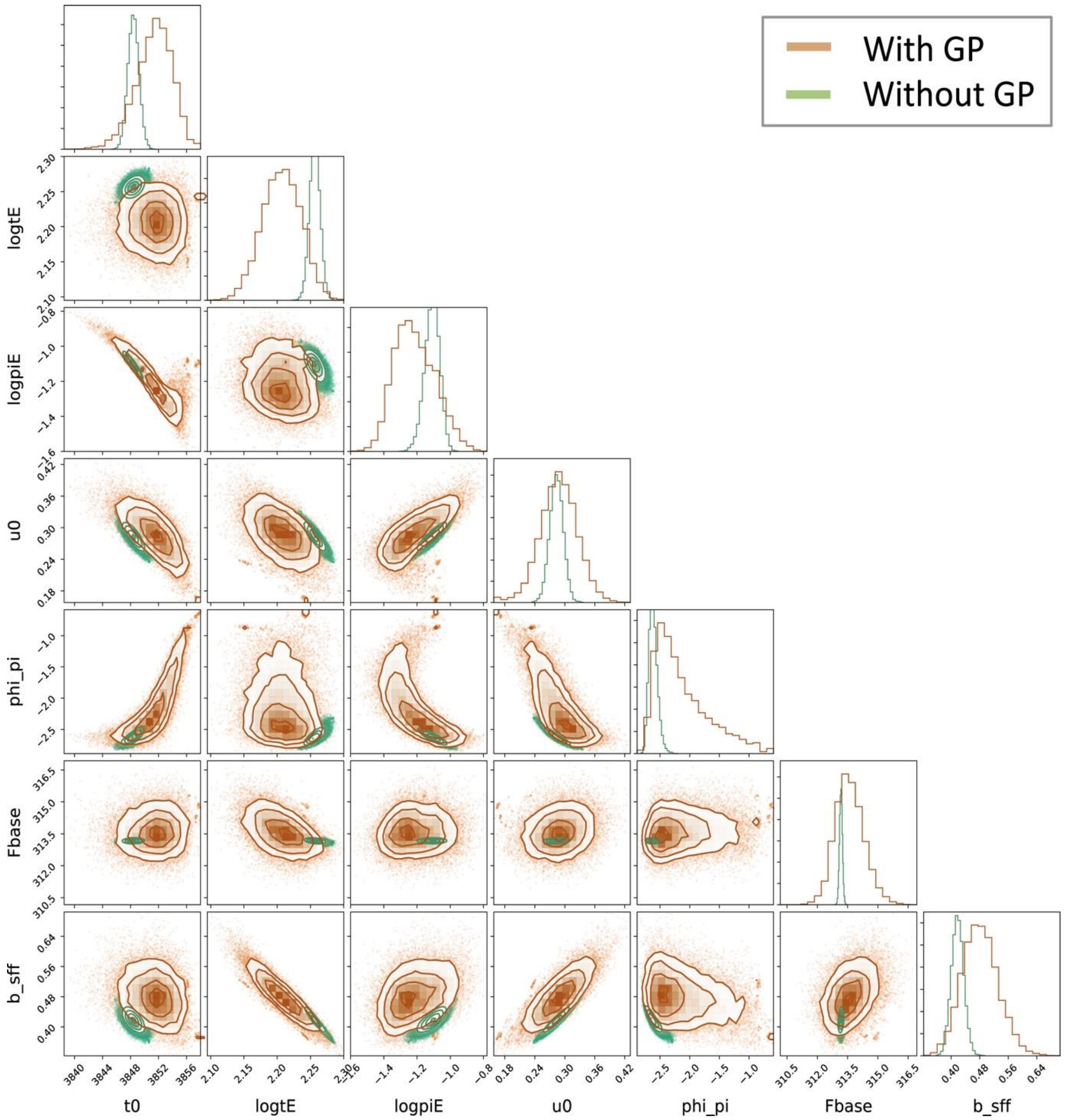


Figure 8 and Figure 16 from Golovich et al. (2022)

Data Systematics (Spitzer events)

- A few flux units, ~5 days
- Seasonal rotation of the camera and poorly defined neighboring star locations
- Are the strange kinematics solutions real? (Chung et al. 2019, Shvartzvald et al. 2019, 2017, Malpas et al. 2022).
- Are all the published mass and distances skewed towards not massive enough and too distant or are the Galactic-models and their inferences the problem?

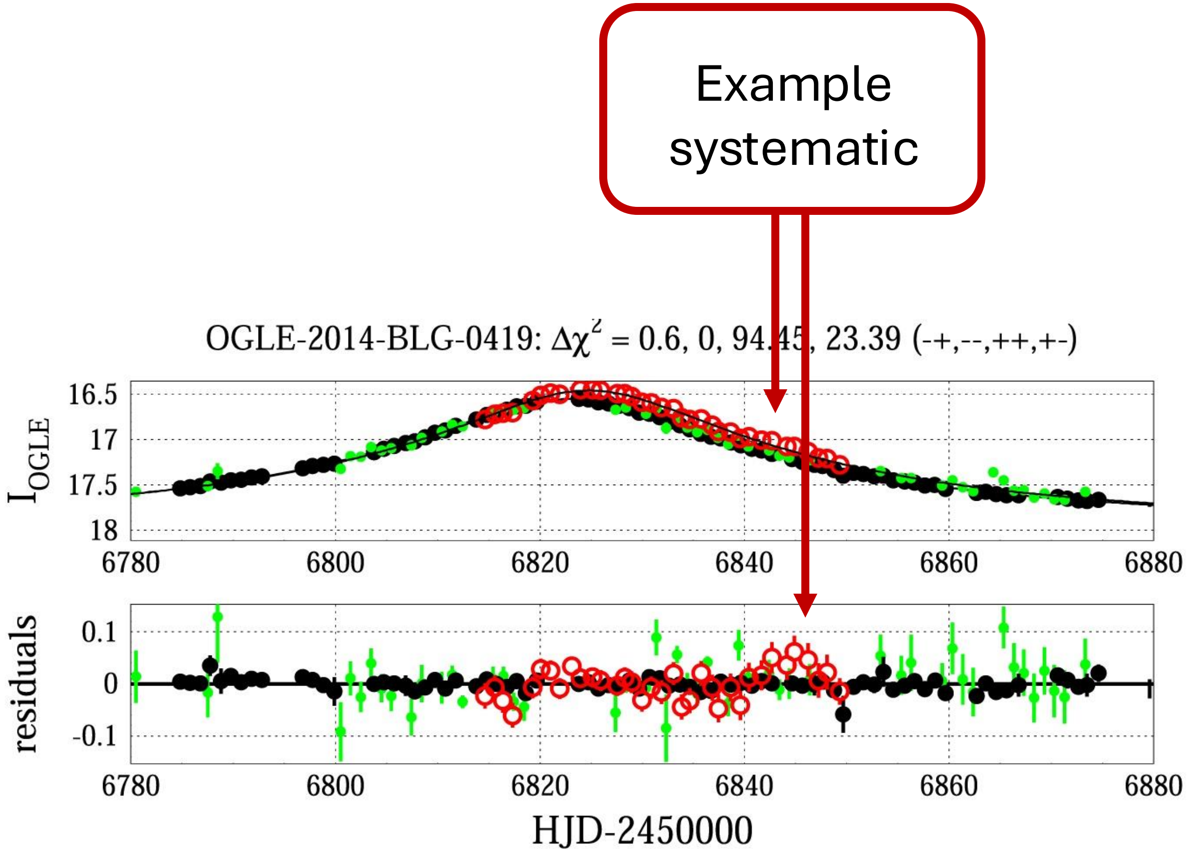
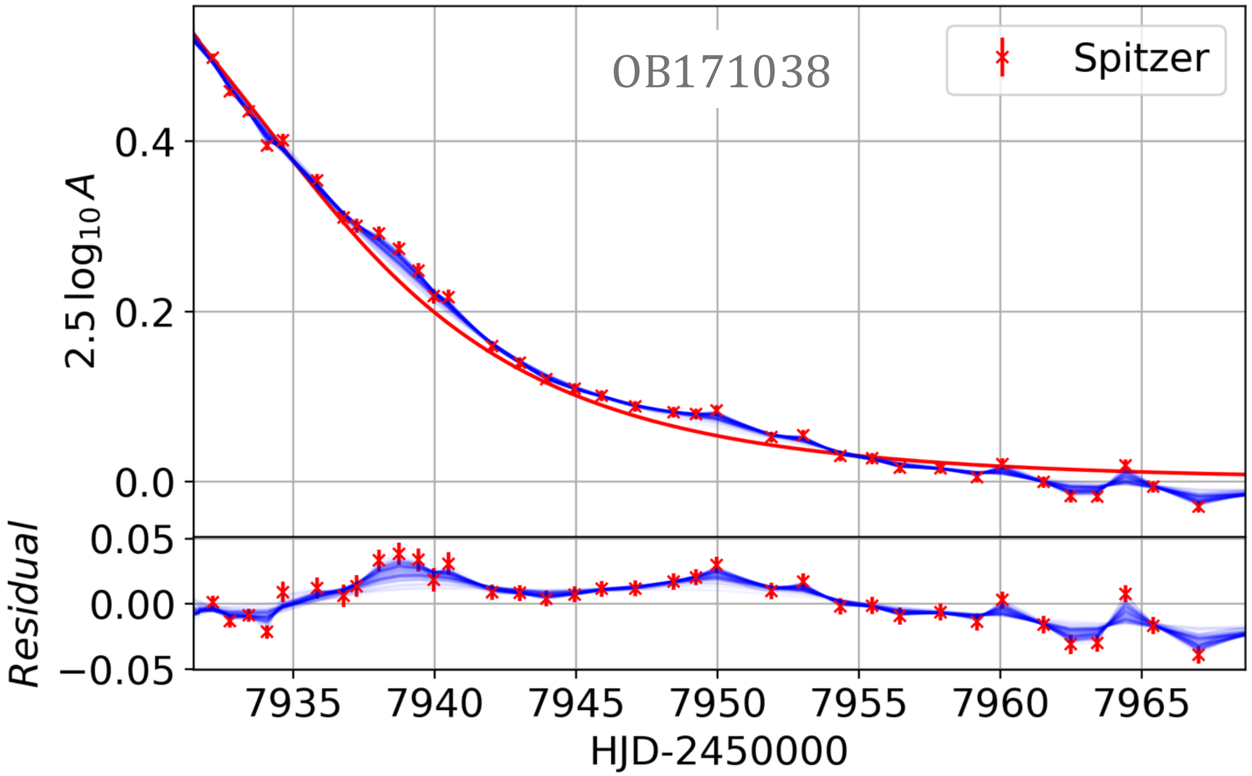
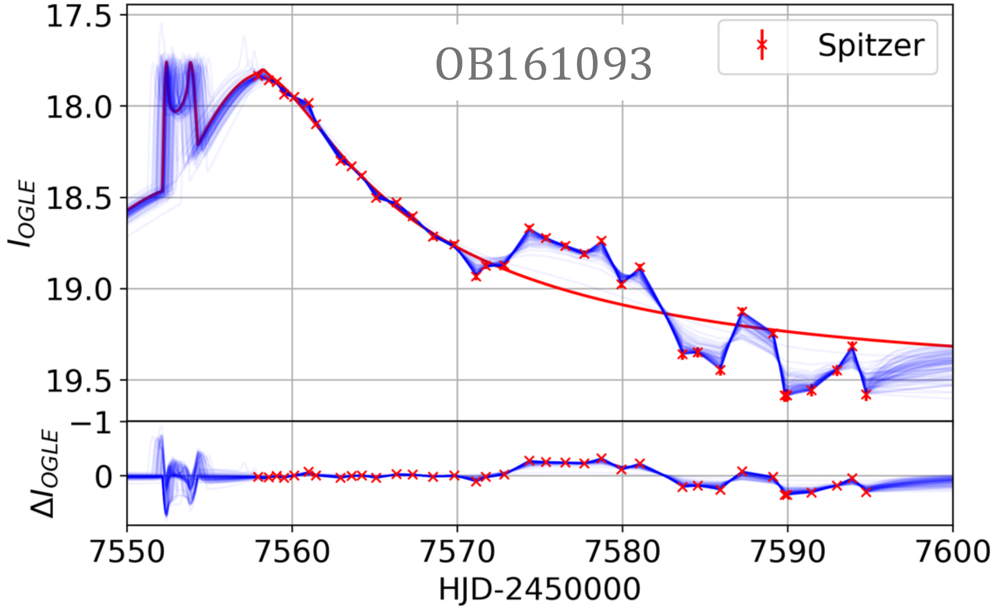
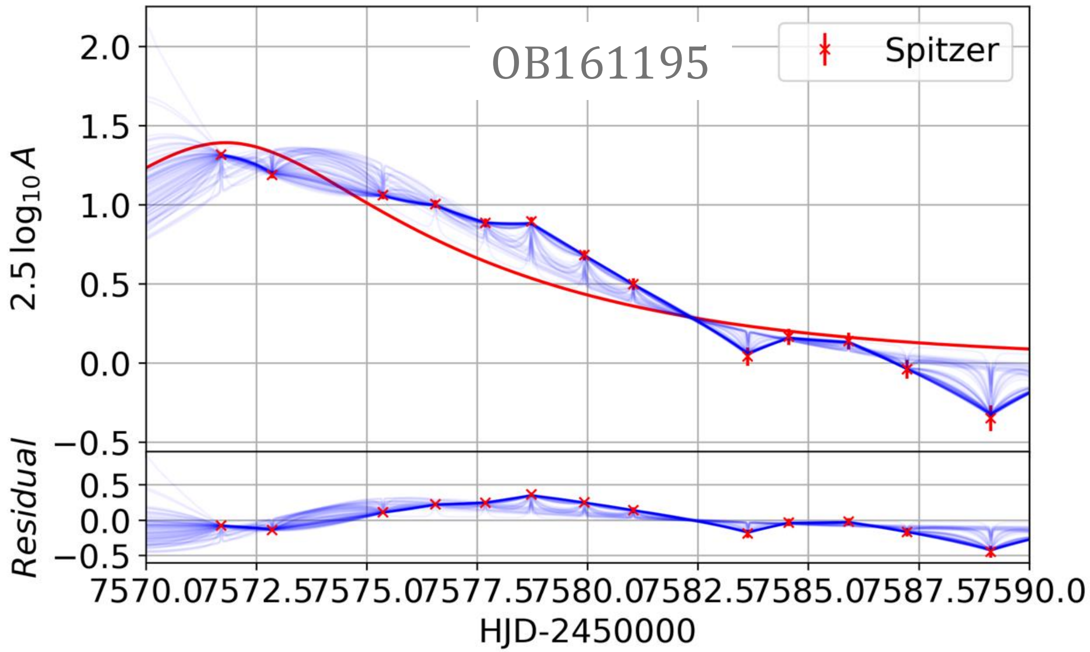


Figure from Novati et al. (2015)

Spitzer GP results

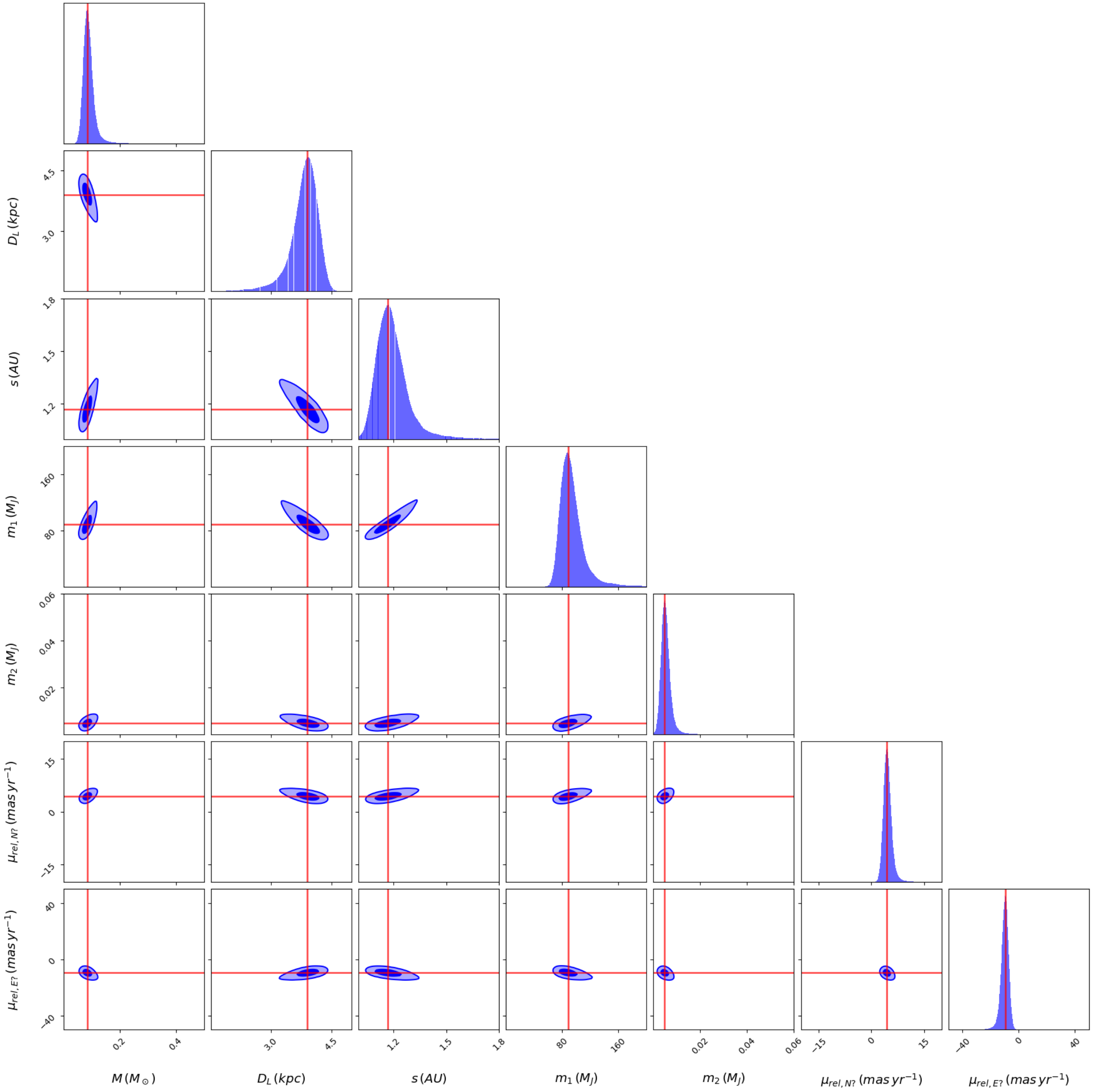






Physical Implications of using a GP with Binary-Lens Spitzer Data

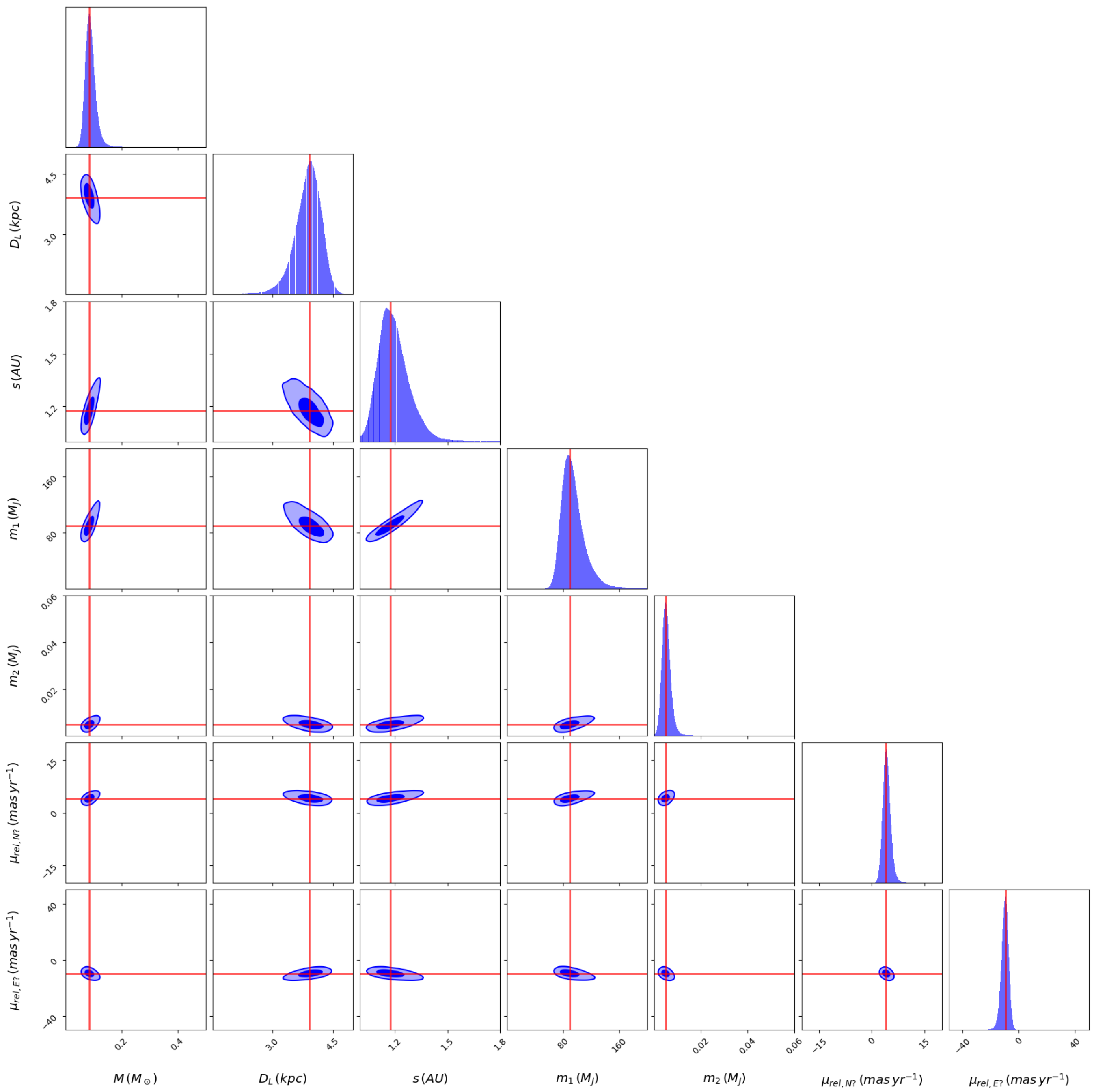
- The GP hyperparameters are not independent of the parallax measurement.
- In general, inclusion of a GP does not change the parallax measurement much, but it does widen the the posteriors on the parallax parameters.
- Greater agreement between the physical determinations from Spitzer parallax and the inferences from galactic models is predominantly due to wider posterior posterior.
- Reestablished degenerate solutions that might otherwise have been ruled out.



Fixed rescaling of errors

S=1

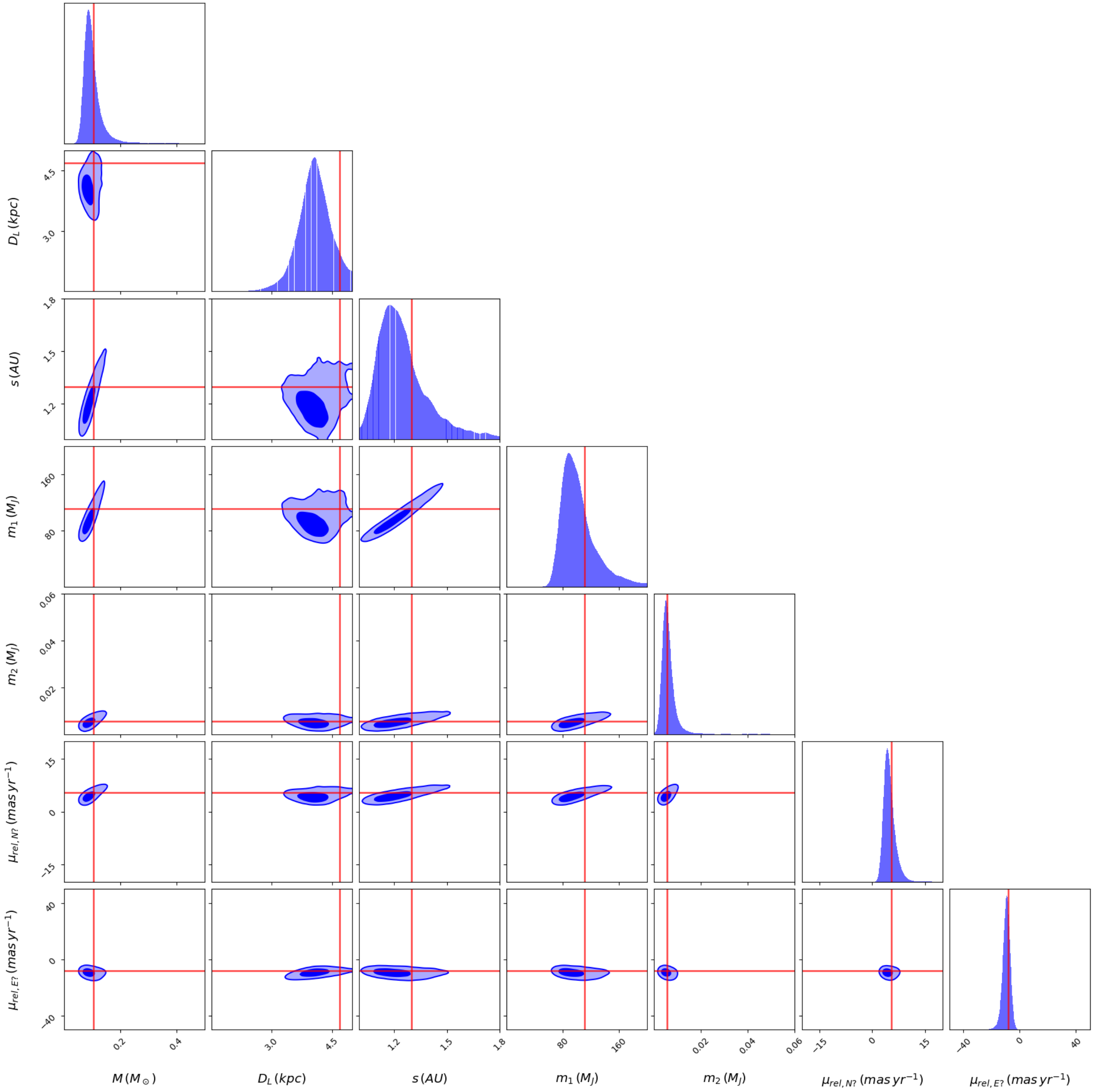
OB161195



Fixed rescaling of errors

S=1.95

OB161195



GP with free rescaling of errors

OB161195

GP weaknesses

- Kernel choice; it's complicated
- Fits take longer to run
- Start-up cost
- Potential for degeneracies with the physical model
- May not be supported by the data.
- Introduces complexities for the modeler when multiple data sources are involved

  e.g. Weakening of the baseline constraint between bands imposed by the expected color from color-color relations; The GP can "act" to undermine priors on the source color

Roman Era Microlensing and GP Usage

- GP provides a means for marginalizing the affects of data systematics; e.g. from variable blend stars, poorly estimated error bars, and blend compositions changes due to kinematics.
- The Roman GBTDS will runs for 6 years and will have events spanning observing seasons on a similar scale to those seen in the Golovich et al. (2022) sample; use of GP may be computational plausible with efficient likelihood computation.
- A potential tool for synergy between the fields of asteroseismology and microlensing. GBTDS has an expected yield of $\sim 10^6$ detections of oscillations in stars (Gould et al. 2015).