Strong Lens Time Delay Estimation

Tak (Hyungsuk) Tak

SAMSI ASTRO / International CHASC Astrostatistics Collaboration

2 Nov 2016

Joint work with Kaisey Mandel (Center for Astrophysics; CfA),
David A. van Dyk (Imperial College London), Vinay L. Kashyap (CfA),
Xiao-Li Meng (Harvard), and Aneta Siemiginowska (CfA)

Outline

1. Introduction: Strong gravitational lensing
2. Data
3. Our model assumptions
   - State-space model
   - Distributions of the observed data
   - Distributions of the latent data
4. Bayesian: Prior distributions for unknown parameters
5. Bayesian: Full posterior and sampler
6. Frequentist: Profile likelihood
7. Our estimation strategy
8. Example: Time Delay Challenge
9. Microlensing (How to handle multimodality)
   - Problem
   - Model
   - Examples 1, 2 (Q0957+561), and 3 (J1029+2623)
10. Conclusion: Time Delay Challenge 2
11. (If time allows) A new MCMC method for multimodality

Introduction

Strong gravitational lensing

Credit: NASA's Goddard Space Flight Center

The strong gravitational field of a lensing galaxy splits light into two images.

- Light rays take different routes whose lengths can be different.
- Difference between their arrival times $\rightarrow$ **Time delay (Δ)**

Introduction

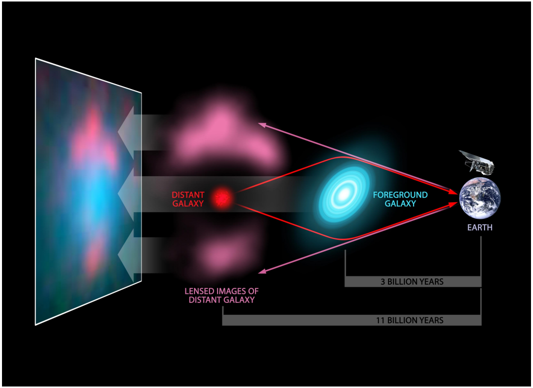
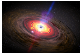


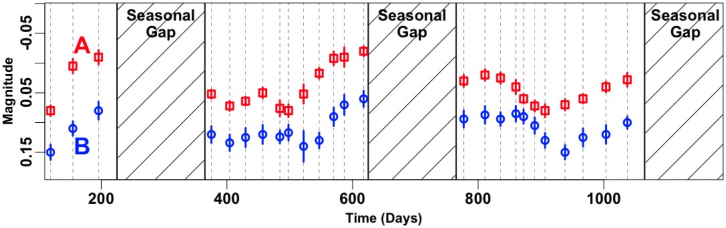
Image Credit: NASA/JPL-Caltech

Time delay is used to infer cosmological parameters, e.g.,

- Hubble constant H_0 (Refsdal, 1964)
- Equation of state of dark energy (Linder, 2011)

Data

Simulated data of a doubly-lensed quasar , an active black hole (Image Credit: NASA/ESA).

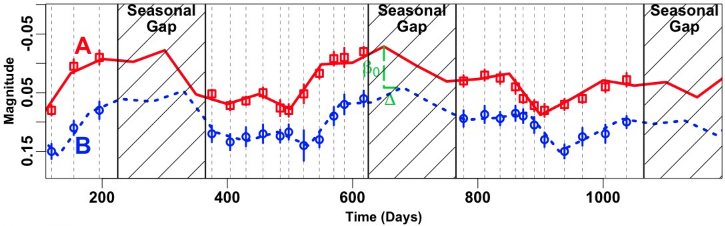


Data are composed of two time series with measurement errors.

- Observation times $\mathbf{t} \equiv \{t_1, t_2, \ldots, t_n\}^\top$
- Observed magnitudes $\mathbf{x} \equiv \{x_1, x_2, \ldots, x_n\}^\top$, and $\mathbf{y}$
- Measurement errors (SD) $\boldsymbol{\delta} \equiv \{\delta_1, \delta_2, \ldots, \delta_n\}^\top$ and $\boldsymbol{\eta}$

Our job is to estimate time delay (shift in the horizontal axis) between two time series.

State-space model



- $\exists$ latent light curves representing the unobserved true magnitudes in continuous time (red and blue dashed curves).

  $\mathbf{X(t)} = (X(t_1), X(t_2), \ldots, X(t_n))^\top$ and $\mathbf{Y(t)}$, values on curves at $\mathbf{t}$

- A curve-shifted model (Pelt et al., 1994):

$$\mathbf{Y(t)} = \mathbf{X}(\mathbf{t} - \Delta) + \beta_0,$$

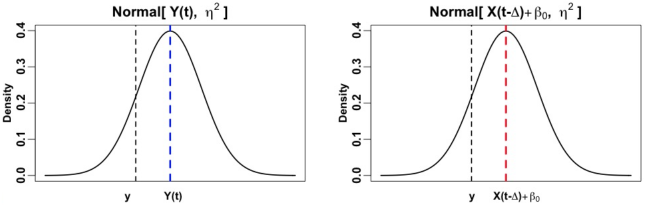
  where the time delay Δ and magnitude offset β_0 are unknown.

Distributions of the observed data

Observed data with independent Gaussian measurement errors

- $x_j \mid X(t_j) \overset{\text{indep.}}{\sim} \text{Normal}[X(t_j),\ \delta_j^2]$
- $y_j \mid Y(t_j) \overset{\text{indep.}}{\sim} \text{Normal}[Y(t_j),\ \eta_j^2]$

  $y_j \mid X(t_j - \Delta), \Delta, \beta_0 \overset{\text{indep.}}{\sim} \text{Normal}[X(t_j - \Delta) + \beta_0,\ \eta_j^2].$



- $p(\mathbf{x}, \mathbf{y} \mid \mathbf{X}(\mathbf{t}^{\Delta}), \Delta, \beta_0)$

  $= \prod_{j=1}^{n} p[x_j \mid X(t_j)] \times p[y_j \mid X(t_j - \Delta), \Delta, \beta_0],$

  where $\mathbf{t}^{\Delta}$ denotes the sorted $2n$ observation times of $\mathbf{t}$ and $\mathbf{t} - \Delta$.

Distributions of the latent data

Latent data with Ornstein-Uhlenbeck (O-U)/Damped RW process

- Many astrophysicists have supported the O-U process; Kelly+ (2009), Kozlowski+ (2010), MacLeod+ (2010), Zu+ (2013), Tewes+(2013), Hojjati+(2014), Bonvin+(2016), and more!
- $dX(t) = -\frac{1}{\tau}\big(X(t) - \mu\big)dt + \sigma dB(t)$, where τ is a mean-reversion time, μ is the overall mean, and σ is the short-term variability.
- O-U process is a Gaussian process with a Matérn(1/2) kernel.
- $\mathbf{X(t^{\Delta})} \sim \mathrm{Normal}_{2n}$ with some mean vector and covariance matrix
- $p(\mathbf{X(t^{\Delta})} \mid \mu, \sigma, \tau, \Delta) =$
  $p(X(t_1^{\Delta}) \mid \mu, \sigma, \tau, \Delta) \times \prod_{j=2}^{2n} p(X(t_j^{\Delta}) \mid X(t_{j-1}^{\Delta}), \mu, \sigma, \tau, \Delta)$

Bayesian: Prior Distributions for Parameters

From statistician's perspective,

- Prior distributions must guarantee posterior propriety, i.e.

$$\int p(\text{parameters} \mid \text{data})\, d(\text{parameters})$$
$$= \int Lik(\text{parameters}) \times p(\text{parameters})\, d(\text{parameters}) < \infty.$$

- Without knowing posterior propriety, no one can tell whether the resulting posterior sample is from the target posterior distribution or not (Hobert and Casella, 1994).
- When we are not sure about posterior propriety: Use proper priors!

From astrophysicist's perspective,

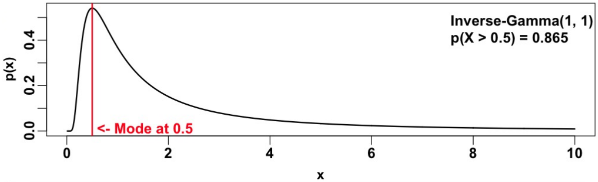
- Set up parameters in the proper prior distributions in a way to reflect on astrophysics and the dynamic of the O-U process.

Bayesian: Prior Distributions for Parameters

- $\Delta \sim \mathrm{Uniform}(u_1, u_2)$
  - $[u_1,\ u_2]$, if $\exists$ prior information to restrict the range of Δ, e.g., a physical model of the lens, redshift, and relative locations.
  - $[t_1 - t_n,\ t_n - t_1]$, otherwise.
- Mean of the O-U, $\mu \sim \mathrm{Uniform}(-30, 30)$, why 30?
- Magnitude offset, $\beta_0 \sim \mathrm{Uniform}(-60, 60)$, why 60?

Inv-Gamma distribution sets a soft lower bound of a random variable

If $X \sim \text{Inv-Gamma}(a, b)$, $p(x) \propto \frac{1}{x^{a+1}} \exp(-b/x)$ with a mode at $\frac{b}{a+1}$.



- Variance of the O-U, $\sigma^2 \sim \text{Inv-Gamma}(1, c)$, why 1&$c$?
- Timescale of the O-U, $\tau \sim \text{Inv-Gamma}(1, 1)$, why 1&1?
- Estimates for 9,275 SDSS quasar light curves (MacLeod+, 2010)

Bayesian: Full Posterior and Sampler

Notation: $\theta_{OU} \equiv (\mu, \sigma^2, \tau)$ and $D_{obs} \equiv \{\mathbf{x}, \mathbf{y}\}$

- Full Posterior: $p(\mathbf{X}(\mathbf{t}^{\boldsymbol{\Delta}}), \Delta, \beta_0, \theta_{OU} \mid D_{obs})$

$$\propto p(\mathbf{x}(\mathbf{t}), \mathbf{y}(\mathbf{t}) \mid \mathbf{X}(\mathbf{t}^{\boldsymbol{\Delta}}), \Delta, \beta_0) \quad \text{Observed data}$$
$$\times\ p(\mathbf{X}(\mathbf{t}^{\boldsymbol{\Delta}}) \mid \Delta, \theta_{OU}) \quad \text{Latent data}$$
$$\times\ p(\Delta, \beta_0, \theta_{OU}) \quad \text{Priors}$$

- Metropolis-Hastings within Gibbs sampler

  1. $p(\mathbf{X}(\mathbf{t}^{\boldsymbol{\Delta}}), \Delta \mid \beta_0, \theta_{OU}, D_{obs},)$
  2. $p(\beta_0 \mid \mathbf{X}(\mathbf{t}^{\boldsymbol{\Delta}}), \Delta, \theta_{OU}, D_{obs})$
  3. $p(\theta_{OU} \mid \beta_0, \mathbf{X}(\mathbf{t}^{\boldsymbol{\Delta}}), \Delta, D_{obs})$

- Pros: Complete investigation on all the model parameters
- Cons: Inefficient when $\exists$ multimodality

Frequentist: Profile likelihood

A profile likelihood function enables us to focus on the parameter of interest with nuisance parameters maximized out (Berger et al., 1996)

- $L(\Delta, \beta_0, \theta_{OU}) =$
  $\int_{\mathbf{R}^{2n}} p(\mathbf{x}, \mathbf{y} \mid \mathbf{X(t^{\Delta})}, \Delta, \beta_0) \times p(\mathbf{X(t^{\Delta})} \mid \Delta, \theta_{OU})\, d\mathbf{X(t^{\Delta})}$
- $L_{prof}(\Delta) \equiv \max_{\beta_0, \theta_{OU}} L(\Delta, \beta_0, \theta_{OU}) = L(\Delta, \hat{\beta}_0(\Delta), \hat{\theta}_{OU}(\Delta))$
- $L_{prof}(\Delta) \propto p(\Delta \mid D_{obs})$ asymptotically
- Provides approximate posterior mean, mode, standard deviation, and most importantly shape of the (approximate) distribution
- Pros: Simple to implement and easy to find multi-modes
- Cons: Computationally expensive for drawing a finer curve / no information about the relationship among parameters

Our time delay estimation strategy

Bayesian and frequentist methods complement each other!

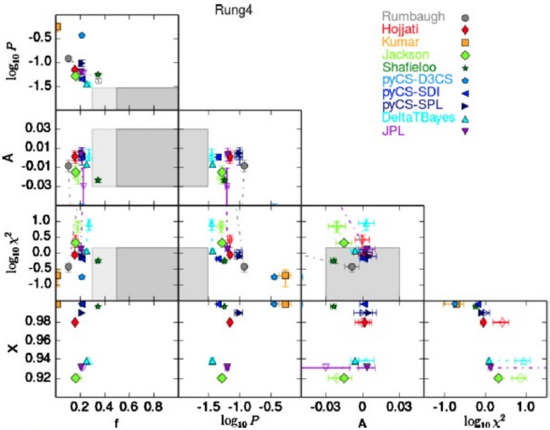
Our time delay estimation strategy:

1. Obtain the profile likelihood curve to check multimodality
2. Initialize Bayesian method near the modes identified by $L_{\mathrm{prof}}(\Delta)$

Example: Time Delay Challenge

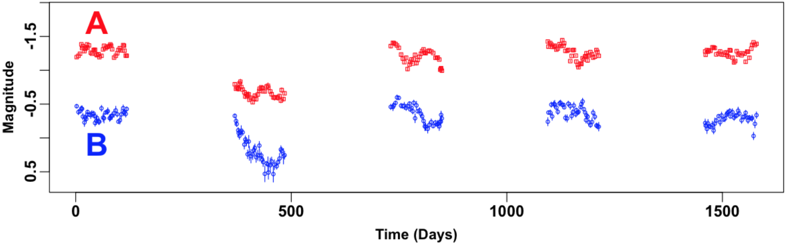
Time Delay Challenge (TDC, Dobler et al., 2015; Liao et al., 2015)

- A blind competition held by 8 astrophysicists from 2013 to 2014.
- Goals: (1) Providing an observation strategy for the LSST.
  (2) Improving current estimation methods.
- About 5,000 simulated data sets with some time delays (O-U).
- 13 teams blindly analyzed the simulated data.

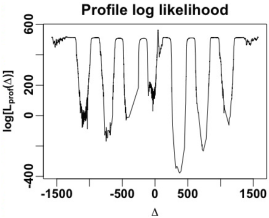


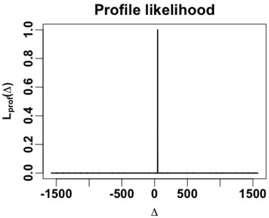
Example: Time Delay Challenge

Simulated data of a doubly-lensed quasar from TDC



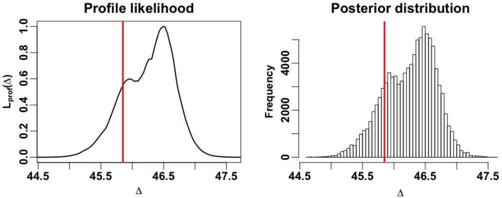
(1) The entire profile (log) likelihood curve





Example: Time Delay Challenge

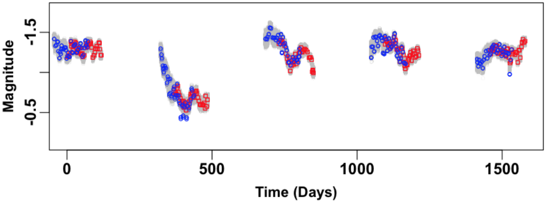
(2) Posterior distribution of Δ initialized near the dominant mode



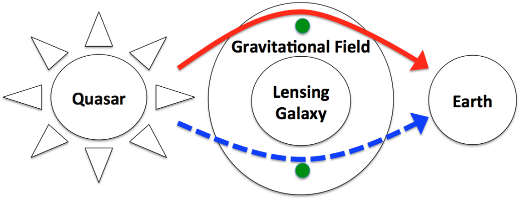
(3) Estimation summary for Δ

Method	Truth	Post. Mean	Post. SD
Bayesian	45.85	46.26	0.41
Profile likelihood		46.26	0.40

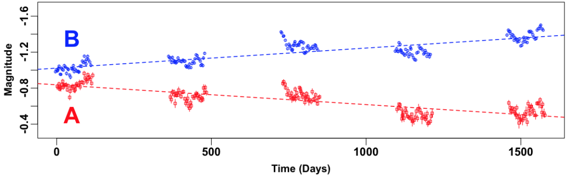
(4) Model Checking: Blue light curve is shifted by $\hat{\Delta}$, and $\hat{\beta}_0$.



Microlensing

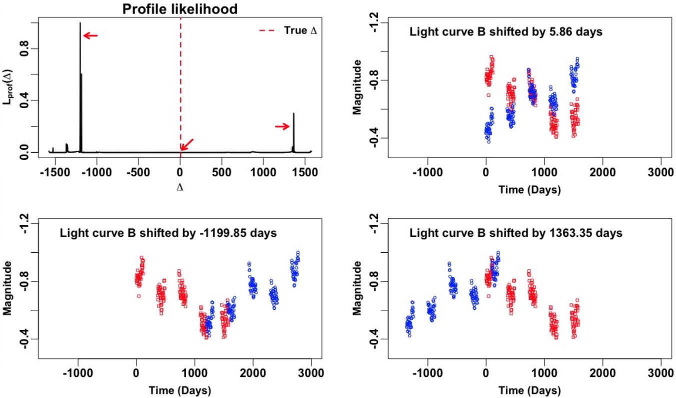


- **Microlensing** occurs when stars unusually close to the paths of light introduce independent noise into magnification of brightness light curves (Tewes et al., 2013).
- Two light curves may have different long-term trends, e.g., polynomial.



Microlensing: Problem

- A curve-shifted model does not work because one of the latent curves is no longer a shifted version of the other.



- A small overlap between two light curves (bottom plots) is the only similar fluctuation patterns detectable by shifting one of the light curves $\rightarrow$ several modes near margins of the range of Δ.

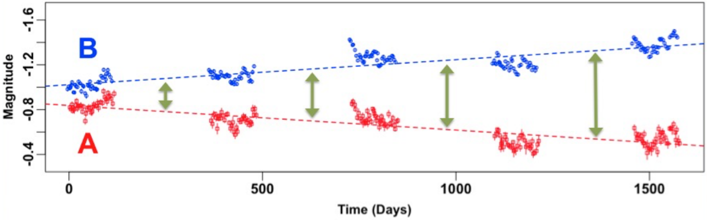
Microlensing: Model

Microlensing model

- Popular way is to model the long-term trend of *each* light curve via a polynomial regression.
- Our microlensing model accounts for the difference between long-term trends using an m^{th}-order polynomial regression.

$$Y(t) = X(t-\Delta) + \mathbf{w}_\mathbf{m}^\top(\mathbf{t}-\boldsymbol{\Delta})\boldsymbol{\beta},$$

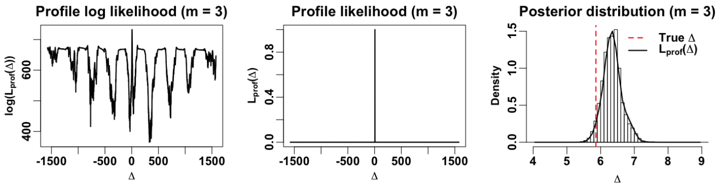
where $\mathbf{w}_m(t-\Delta) \equiv (1, t-\Delta, \ldots, (t-\Delta)^m)^\top$, and $\boldsymbol{\beta} \equiv (\beta_0, \beta_1, \ldots, \beta_m)^\top$ are regression coefficients.



- Our microlensing model reduces the number of regression coefficients by half!

Microlensing Example 1

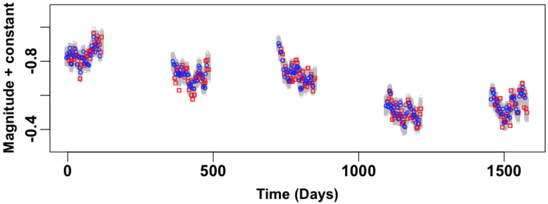
(1) We set $m = 3$ as a default (Kochanek+, Morgan+, Tewes+).



(2) Estimation summary (Error $\equiv |\Delta_{\mathrm{true}} - \hat{\Delta}|$ and $\chi \equiv \mathrm{Error/SD}$)

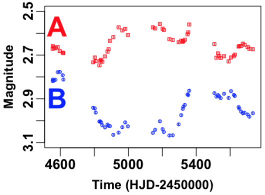
Method	Truth	E(Δ|Data)	SD(Δ|Data)	Error	χ
Bayesian	5.86	6.34	0.28	0.48	1.71
Profile Lik.		6.36	0.28	0.50	1.76

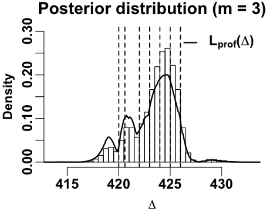
(3) Model Checking: Blue light curve is adjusted by $\hat{\Delta}$ and $\hat{\beta}$.

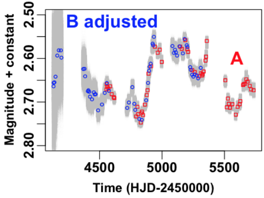


Microlensing Example 2: Q0957+561

r-band data collected at the US Naval Observatory (Hainline+, 2012).



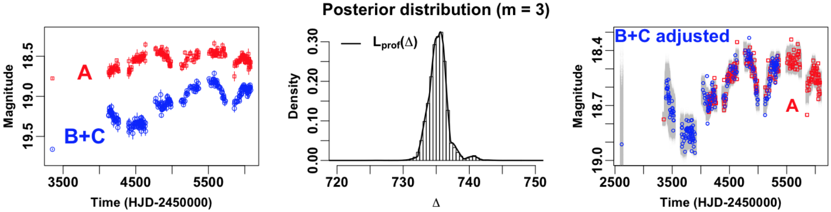




Researchers	Number of observations	Observation period	Measurement error (mag)	$\hat{\Delta}$	SE
Pelt et al. (1996)	831	1979–1994	0.0159	423	6
Oscoz et al. (1997)	86	1994–1996	0.01, 0.02	424	3
Serra-Ricart et al. (1999)	197	1996–1998	0.023, 0.025	425	4
Oscoz et al. (2001)	100	1994–1996	0.009, 0.01	423	2
Shalyapin et al. (2012)	371	2005–2010	0.012	420.6	1.9
This work	57	2008–2011	0.004	423.69 423.21	2.02 2.81

Microlensing Example 3: J1029+2623

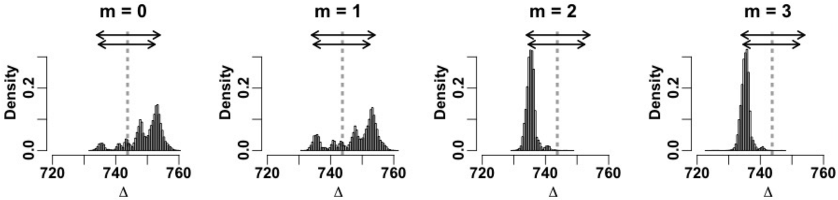
Data reported by Fohlmeister et al. (2013).



Researchers	Method	Estimate	90% Interval
Fohlmeister et al. (2013)	χ^2-minimization (AIC, BIC)	744	(734, 754)
Kumar, Stalin, and Prabhu (2014)	Difference-smoothing	743.5	(734.6, 752.4)
This work	Bayesian Profile likelihood	735.28 733.11	(733.08, 737.59) (732.94, 738.44)

Our point estimate is much smaller than theirs (by about 10 days) & our 90% intervals are much shorter than theirs. What's wrong here? Is it because our model is over-confident?

Microlensing Example 3: J1029+2623 (cont.)



Fohlmeister et al. (2013)

- A point estimator based on high-dimensional optimization
- Linear microlensing model (AIC) + a model w/o microlensing (BIC)

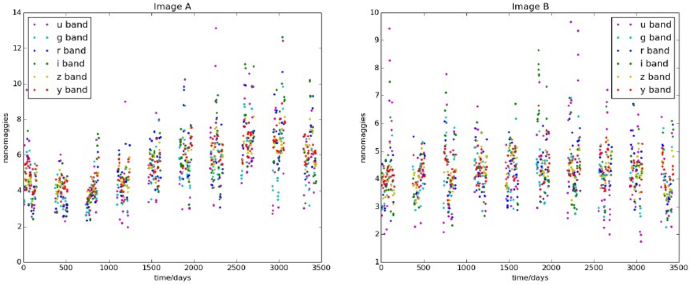
Kumar et al. (2014)

- A point estimator also based on high-dimensional optimization
- A spline with a Gaussian kernel to account for microlensing

It reveals that our model accounts for microlensing better than theirs.

Discussion: Time Delay Challenge II

Doubly-lensed multi-filter light curves (Marshall et al., 2016+)



- Six bands, u, g, r, i, z, y, lead to vector time series.
- 5,000+ systems, 6 bands for each system, 150 obs. for each band.
- Bluer filters are more sensitive to microlensing than redder filters.
- All the information (except Δ) needed to calculate H_0 will be given.
- Evaluation is based on two numbers, H_0 estimate and its uncertainty.
- http://timedelaychallenge.org for more information!

Reference

1. H. Tak, K. Mandel, D. A. van Dyk, V. Kashyap, X. Meng, and A. Siemiginowska (in progress) "Bayesian and Profile Likelihood Strategies for Time Delay Estimation from Stochastic Time Series of Gravitationally Lensed Quasars", tentatively accepted to *Annals of Applied Statistics*. (ArXiv 1602.01462)
2. M. Tewes, F. Courbin, and G. Meylan (2013) "COSMOGRAIL: the COSmological MOnitoring of GRAvitational Lenses" *Astronomy & Astrophysics*, 553, A120.
3. S. Kozlowski et al. (2010), "Quantifying quasar variability as part of a general approach to classifying continuously varying sources," *The Astrophysical Journal*, **708**, 927.
4. C.L. MacLeod et al. (2010), "Modeling the time variability of SDSS Stripe 82 quasars as a damped random walk, " *The Astrophysical Journal*, **721**, 1014.
5. Y. Zu et al. (2011), "An alternative approach to measuring reverberation lags in active galactic nuclei," *The Astrophysical Journal*, **735**, 80.
6. B. Kelly, J. Bechtold, and A. Siemiginowska (2009) "Are the variations in quasar optical flux driven by thermal fluctuation?" *The Aphysical Journal*, **698**, 895.
7. G. Dobler, C. Fassnacht, T. Treu, P. Marchall, K. Liao, A. Hojjati, E. Linder, and N. Rumbaugh (2015) "Strong lens time delay challenge: I. Experimental design." *The Astrophysical Journal*, **799**, 168.
8. K. Liao et al. (2015) "Strong lens time delay challenge: II. Results of TDC1." *The Astrophysical Journal*, **800**, 23.
9. C. S. Kochanek et al. (2006) "The time delays of gravitational lens HE0435+1223: An early-type galaxy with a rising rotation curve." *The Astrophysical Journal*, **640**, 47.
10. C. W. Morgan et al. (2012) "Further evidence that quasar X-ray emitting regions are compact: X-ray and optical microlensing in the lensed quasar Q J0158+4325." *The Astrophysical Journal*, **756**, 9.
11. S. Refsdal (1964) "The Gravitational Lens Effect." *Monthly Notices of the Royal Astronomical Society*, **128**, 295–306.
12. E. V. Linder (2011) "Lensing Time Delays and Cosmological Complementarity." *Phys. Rev. D*, **84**, 12.
13. A. Hojjati, A. G. Kim, and E. V. Linder (2014) "Robust strong lensing time delay estimation." *Phys. Rev. D*, **87**, 12.
14. Bonvin et al. (2016) "COSMOGRAIL: the COSmological MOnitoring of GRAvItational Lenses XV. Assessing the achievability and precision of time-delay measurements" *Astronomy & Astrophysics*, 585, A88.