

Using C  l  rit   to infer DRW parameters

Krzysztof Suberlak,^{1*}   eljko Ivezi  ¹

¹*Department of Astronomy, University of Washington, Seattle, WA, United States*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

A report to outline validation of C  l  rit  . We compare the tools used to fitting for τ and SF_∞ to those of Koz  owski, Szymon (2017) and MacLeod et al. (2011). To do so we reproduce some of experiments they conducted, and evaluate whether they can be mutually consistent.

1 INTRODUCTION

Quasars exhibit stochastic variability with characteristic timescales of hundreds of days (Kelly+2009, Koz  owski+2010,2016, 2017 MacLeod+2010,2011,2012, Zu+2011,2013,2016, Kasliwal+2015,2017). We employ C  l  rit   (Foreman-Mackey et al. 2017) , which allows to express any one -dimensional process as a Gaussian Process. Gaussian Process is defined by covariance and mean. Covariance parameters are often called hyperparameters. To find best-fit hyperparameters we optimize the marginal likelihood (eq. 5.4, 5.8, Rassmussen&Williams book). Since the marginal likelihood is the integral of the product of likelihood and prior, the logarithm of marginalized likelihood is the sum of the log-likelihood and log-prior (eq.2.28 Rassmussen&Williams).

In this report we summarize the tests that have been made to establish great usefulness of C  l  rit   in modelling the DRW light curves as Gaussian Processes. We first compare whether algorithms used to make mock DRW light curves are identical between MacLeod +2011 (which we use), and Koz  owski, Szymon (2017) (who challenges MacLeod+2011 results, and whose results we reproduce). Then we describe how a choice of boundaries and priors affects the results of best-fit hyperparameters with C  l  rit  . [We then compare C  l  rit   to another well-tested tool (used by Koz  owski (2016), Zu et al. (2011), etc.) - JAVELIN]. We then reproduce results of MacLeod et al. (2011) Fig.15, and Koz  owski, Szymon (2017) Fig.2 . We aim to answer the following questions:

- (i) are the tools for fitting tau and SFinf equivalent?
- (ii) can we reproduce Chelsea’s Fig 15?
- (iii) can we reproduce Koz  owski’s plot?
- (iv) are their plots mutually consistent, given our analysis?
- (v) can we reproduce best-fit tau and SFinf obtained using light curve fitting with the SF approach?

2 SIMULATING DRW

We simulate damped random walk light curves by drawing points from a Gaussian distribution, for which mean and

standard deviation are re-calculated at each timestep. Given an input of observation times t , SF_∞ - the asymptotic value of the structure function, mean magnitude $\langle y \rangle$, and the damping timescale τ , we start at time t_0 and signal at that time is equal to the mean $y_0 = \langle y \rangle$. The timestep is $\Delta t_i = t_{i+1} - t_i$. Given the signal at time t_i : y_i , and Δt_i , the signal at next time step y_{i+1} is drawn from $\mathcal{N}(loc, stdev)$, where :

$$loc = y_i e^{-r} + \langle y \rangle (1 - e^{-r}) \quad (1)$$

and

$$stdev^2 = 0.5 SF_\infty^2 (1 - e^{-2r}) \quad (2)$$

with $r = \Delta t_i / \tau$. Here we followed eq. A4 and A5 in Kelly et al. (2009), as well as Sec. 2.2 in MacLeod et al. (2010). To this ideal light curve signal we add photometric noise $\mathcal{N}(0, n_i)$, where n_i is the observational photometric noise. It is equivalent to Koz  owski+2017 formulation , who also starts with the signal s_i , drawing at each time step light curve points from a Gaussian distribution with dispersion $stdev$ and mean loc , subsequently adding the mean $\langle y \rangle$ and Gaussian noise (see Eq. (2) of Koz  owski, Szymon (2017)).

Apart from τ and SF_∞ , we choose N_{pts} - at how many points to sample the simulated DRW process, and the length of baseline $T = l \cdot \tau$. In this formalism the baseline multiplicity l is equivalent to $1/\rho$ where $\rho = \tau/T$ (Koz  owski+2017). We can sample the baseline either at regular intervals of Δt , or at random N_{pts} . One sets the other - given Δt , we find N_{pts} as the nearest integer to $t_{max} - t_{min} / \Delta t$.

3 DRW AS GAUSSIAN PROCESS

DRW is a stochastic process defined by the covariance matrix

$$S_{ij} = \sigma^2 \exp(-\Delta t_{ij} / \tau) \quad (3)$$

(see Koz  owski+2010 eq. 1, Koz  owski+2017 eq. 1, MacLeod+2011 eq.1, Zu+2013 eq. 3 , etc). A scatter of magnitude difference plotted as a function of time lag Δt_{ij}

is called the Structure Function (SF). SF for the Damped Random Walk is described by :

$$SF(\Delta t_{ij}) = SF_{\infty} \left(1 - e^{-|\Delta t_{ij}/\tau|}\right)^{1/2} \quad (4)$$

For large Δt_{ij} , we have

$$\lim_{\Delta t_{ij} \gg \tau} e^{-|\Delta t_{ij}/\tau|} = 1$$

so that:

$$SF(\Delta t_{ij} \rightarrow \infty) \rightarrow SF_{\infty}$$

. Following MacLeod+2011, we define the driving amplitude for shot-term variability as :

$$\hat{\sigma} = \sigma \sqrt{2/\tau} \quad (5)$$

We can relate SF_{∞} to σ and $\hat{\sigma}$:

$$SF_{\infty} = \hat{\sigma} \sqrt{\tau} = \sigma \sqrt{2} \quad (6)$$

thus SF_{∞} is just a scaled version of σ .

Another often used combination of hyperparameters is called K (as in MacLeod+2011) :

$$K = \tau \sqrt{SF_{\infty}} = \tau \sqrt{\sigma} 2^{1/4} \quad (7)$$

In the $\log \sigma$ - $\log \tau$ space, lines of constant K or $\hat{\sigma}$ are perpendicular to each other. This is because, if we take $\log \hat{\sigma}$, and rearrange, we have :

$$\log \sigma = \frac{1}{2} \log \tau + \log \hat{\sigma} - \frac{1}{2} \log 2 \quad (8)$$

and from $\log K$:

$$\log \sigma = -2 \log \tau + \log K - \frac{1}{2} \log 2 \quad (9)$$

These equations denote lines $y = ax + b$, and the slope of one is the inverse reciprocal of another, which proves that they are orthogonal in that space (see Fig. 4)

Covariance matrix, or kernel, is a function that defines similarity between two points. In general, a kernel is any function that maps x, x' onto \mathbb{R} . Thus a covariance function is a specific type of a kernel. A Gaussian process is defined by its covariance function and mean. To model light curves as DRW using Gaussian Process approach we use the Real Term kernel in Celerite :

$$S_{ij} = a_j e^{-c_j |t_j - t_i|} \quad (10)$$

with parameters \log_a and \log_c . It is clear that this is a DRW kernel if we substitute $a_j \equiv \sigma^2$, and $c_j \equiv \tau^{-1}$, so that $\log_a = 2 \log \sigma$, and $\log_c = -\log \tau$. By default there are no boundaries on parameter values, and there is no prior. We find that imposing very liberal boundaries does not affect the result of fit but helps ensure computational stability. Thus we choose to limit σ to between 0.01 and 1.0 mag, and τ to between 1 and 10000 days. Both MacLeod et al. (2011) and Kozłowski, Szymon (2017) use Jeffreys prior (Jeffreys

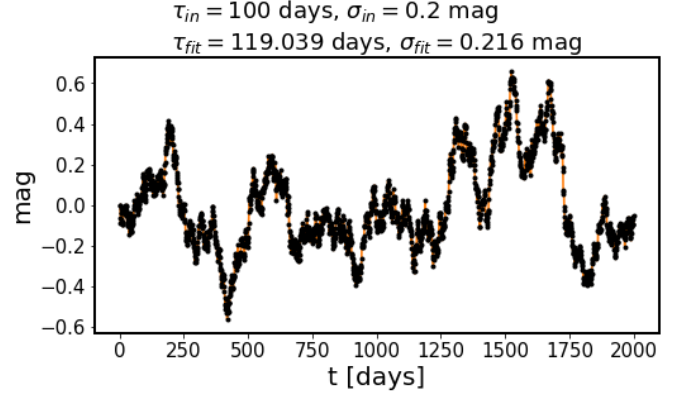


Figure 1. A Celerite fit to a simulated light curve using a flat prior.

1946) on τ and $\hat{\sigma}$: $prior(\tau) = 1/\tau$, and $prior(\hat{\sigma}) = 1/\hat{\sigma}$. Jeffreys prior can be effectively expressed in terms of a and c parameters :

$$prior(a, c) = c + \frac{1}{\sqrt{2ac}} \quad (11)$$

so that in logarithmic space :

$$\log(prior) = \frac{1}{2} (\log_c - \log_a - \log 2) \quad (12)$$

4 LIKELIHOOD FOR GP

Celerite efficiently evaluates the marginalized likelihood of the dataset under a Gaussian Process model with given kernel and hyperparameters. We optimize the log-likelihood for the best-fit hyperparameters with the stable L-BFGS-B (Byrd et al. 1995), (Zhu et al. 1997) algorithm using `scipy.optimize.minimize` (Jones et al. 2001) implementation. We illustrate the shape of log-likelihood for a simulated light curve with parameters $\tau_{in} = 100$ days, $\sigma_{in} = 0.2^{\text{mag}}$, Gaussian noise of 0.001^{mag} , with length 20τ , and regular sampling of $\Delta t = 1$ day, and flat prior. See Fig. 1 for the light curve and GP prediction, and Fig. 2 for the the shape of log-likelihood evaluated for this data on the grid of hyperparameters σ, τ .

5 EXPERIMENTING WITH NUMBER OF POINTS AND BASELINE

We simulate light curves as described in Sec. 2, using the same input parameters as MacLeod et al. (2011) : $\tau_{in} = 575$ days, $SF_{\infty} = 0.2$ mag, regular sampling interval of 10 days, $\sigma = SF_{\infty}/\sqrt{2} = 0.1414$. We start with 10 000 realizations of a very long (40 years) and well-sampled ($\Delta t = 10 \text{ days}$) light curve. We fit with Celerite, using bounds on σ : $[0.1 - 1.0]$ mag, and bounds on τ : $[1 - 10000]$ days. At each realization, we select 1-, 3-, 10- year, full sections of the light curve. We fit at each light curve section length with both flat prior or Jeffrey's prior.

We also performed two controlled experiments : how

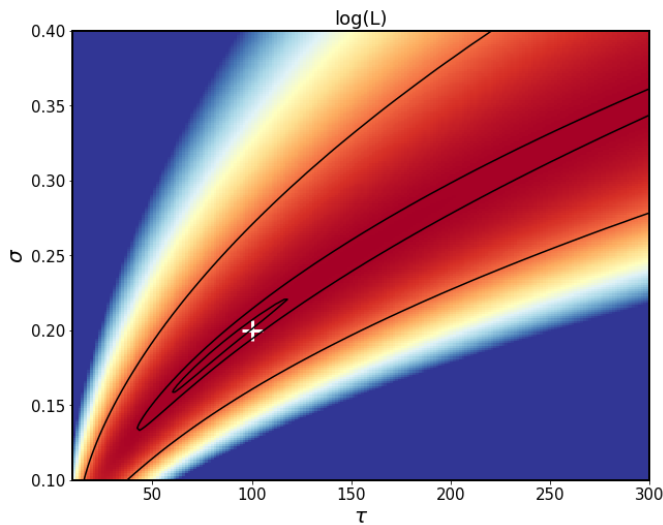


Figure 2. The log likelihood for the simulated light curve on Fig. 1. Black contours show 0.683, 0.955, 0.997 levels of the cumulative (integrated) posterior probability.

changing the number of points, or changing the baseline, affects the results.

With the former, we keep the light curve baseline fixed at 40 years, initially sampled by 1460 points (that corresponds to the regular interval of 10 days: $40 \times 365/10 = 1460$) - this is the same starting light curve as in the top panel of Fig. 5. We then increase the number of points by a factor $f \in 1, 2, 4, 8$. We illustrate that for flat prior on Fig. 7

We plot an equivalent measure as Fig. 6, detailing the results of this experiment - see Fig. 8

We also performed an experiment keeping the number of points per light curve fixed at $N = 1460$, but extending the baseline by factor $f \in 1, 2, 4, 8$. All other parameters are as before ($\tau = 575$ days, $SF_\infty = 0.2$ mag, $err = 0.001$ mag). We illustrate the light curves used on Fig. 9.

6 EXPERIMENTS WITH THE STRUCTURE FUNCTION SHAPE

We simulated 1000 light curves with $\tau = 100$ days, $\sigma = 0.2$ mag (so that $SF_\infty = 0.2828$ mag), baseline of 20τ , random sampling from a uniform distribution with sampling interval of $\Delta t = 5$ days, homoscedastic error 0.001 mag, $N = 400$ points. We fit them with flat or Johnson prior. Fig. 11

We perform a controlled experiments with this light curve setup: keeping τ , σ , baseline, error unchanged, we increase the sampling density Δt by a factor $f \in 1, 2, 4, 8$. Thus Δt , the sampling interval, is changed as set 5, 2.5, 1.25, 0.625 days, and correspondingly the number of points N increases as 400, 800, 1600, 3200. This experiment shows what effect on the Celerite fits has increasing the sampling density, with negligible photometric error.

To investigate the structure function, we used the base setup of $\tau = 100$ days, $\sigma = 0.2$, error = 0.001 mag, length = 20τ , sampling $\Delta t = 5$ days. For these light curves we considered raw pairs of $\Delta m_{i,j}, \Delta t_{i,j}$, as well as plotting the robust standard deviation in 200 bins of Δt . Note that here $\Delta t_{i,j}$

denotes the time difference between t_i and t_j . We find that the stochasticity of light curves prevents from using structure function for a single object, or even an ensemble with less than approximately 10 objects. We illustrate that on a series of Figs. 15, 16, 17 and 18.

7 EXPERIMENTS WITH DAMPING TIMESCALE RETRIEVAL

Here we repeat the basic tenets of the experiment performed by Kozłowski+2017. He simulated 100-year long light curves, with cadence $\Delta t = 2$ days, $\tau = 200$ days, $SF_\infty = 0.2$ mag, that were later degraded to the cadences of SDSS S82 or OGLE-III. Various sections of these light curves were used to test the bias caused by length of time series (baseline).

We simulated as the 'truth' 10 000 light curves with $\tau = 575$ days, $SF_\infty = 0.2$ mag, baseline of $100\tau = 57500$ days (157 years, but expressing the light curve length in terms of multiples of the underlying decorrelation timescale is most informative). We chose the cadence of $\Delta t = 1$ day, at regular intervals.

We illustrate the sampling procedure (in other words, the grid of ρ), on Fig. 19

8 CONCLUSIONS

We can confirm results of Kozłowski+2017 regarding an inherent biased introduced in fitting the DRW process. We find that even without any prior (flat prior), the results can be reproduced - the ρ_{out} is biased low for light curves shorter than $\approx 10\tau$. However, the bias is not large for light curves even of length only twice the input characteristic time scale. We argue that it is necessary to quantify the 'bias' and 'goodness' of theoretically possible performance of fitting the DRW process with available software. Our choice of software - Celerite - did not introduce any significant bias as compared to the tools used by Kozłowski+2017 (since the internal workings of Celerite are similar to Press-Rybicki-Hewitt method - the reason that Celerite is fast is the same that afforded the PRH method to be so quick). Indeed, the shape of likelihood used with Celerite, that better constrains $\hat{\sigma}$ than τ or σ individually, matches the shape of likelihood in Kozłowski (PRH) method. We propose the measure of percentage departure from the 'truth' as the measure for the goodness of fit, and we choose to consider results within 10% of the true input value of τ to be 'good'.

However, given that $\hat{\sigma}$ may be better constrained than τ purely due to the likelihood shape, we suggest that perhaps even for light curves that are too short to estimate 'well' the input timescale, we are able to estimate the asymptotic structure function value. This helps to select quasars from Stars, since even if the timescale cannot be well estimated, the amplitude of structure function (driven by σ), can help distinguish quasars from background noise, since their amplitude of variation is larger.

We argue that DRW fitting, and recovering the amplitude of variability within the damped random walk model, can help distinguish quasar light curves from background noise (or noisy stellar measurements) better than other sta-

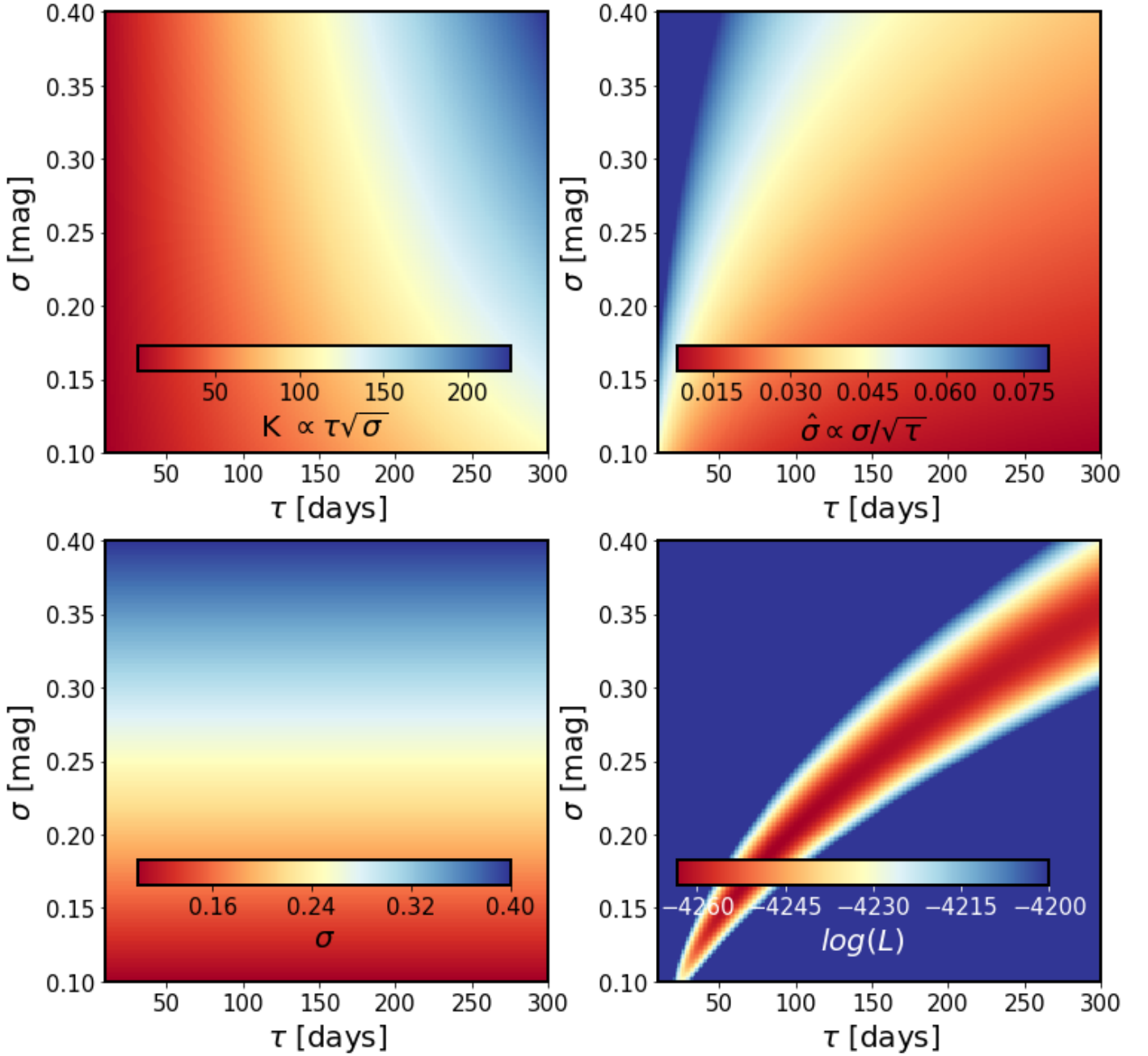


Figure 3. For each pixel on the σ - τ grid we evaluated the log-likelihood value, $\log L$, shown on the bottom-right panel (same as Fig. 2). In addition, given these σ and τ we also evaluated K and $\hat{\sigma}$, which enabled, given $\{\sigma, \tau, \hat{\sigma}, K, \log L\}$, plotting $\log L$ in space of K - $\hat{\sigma}$, or any other parameter as a function of the other two.

tistical measures (such as chi2 per degree of freedom, standard deviation, or rms of the light curve).

To show that, we simulated DRW light curves in range of τ , and range of σ . With the same sampling, we also simulated white noise, that would be reproducing the noisy measurements (is there any better model for random noise of measurement for SDSS or LSST ?). For each light curve, fitted with Celerite with DRW model, we recover τ and σ . We also calculate \mathcal{P} parameters: χ^2_{DOF} , χ^2_R , standard deviation, rms. For each set of input τ , σ , we have N

realizations, and for each i -th realization there are parameters \mathcal{P}_i . We plot a histogram of \mathcal{P} for each value of input τ , σ . We record the mean and median averaged over many realizations. We plot that as a two-dimensional histogram as $\log(\rho_{in})$ vs each parameter \mathcal{P} , overplotting the mean and median (along y-axis).

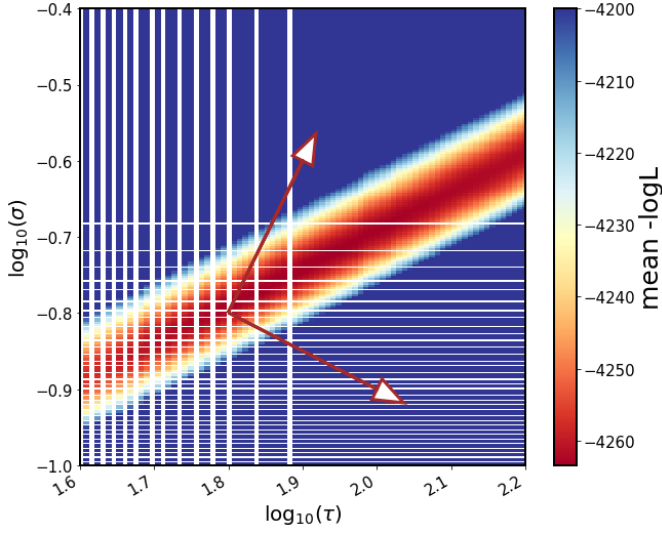


Figure 4. The log likelihood for the simulated light curve, plotted in $\log \sigma$ - $\log \tau$ space. White gaps occur because originally the σ - τ grid on which we evaluated $\log L$ was linear, not logarithmic. Black contours show 0.683, 0.955, 0.997 levels of the cumulative (integrated) posterior probability. Choosing the scale to be the same along both axes, arrows that point along direction of constant $\hat{\sigma}$ or constant K are perpendicular, as shown by Eqs.8 and 9.

REFERENCES

- Byrd R. H., Lu P., Nocedal J., Zhu C., 1995, *SIAM Journal on Scientific Computing*, 16, 1190
- Foreman-Mackey D., Agol E., Angus R., Ambikasaran S., 2017, preprint, ([arXiv:1703.09710](https://arxiv.org/abs/1703.09710))
- Jeffreys H., 1946, *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 186, 453
- Jones E., Oliphant T., Peterson P., et al., 2001, SciPy: Open source scientific tools for Python, <http://www.scipy.org/>
- Kelly B. C., Bechtold J., Siemiginowska A., 2009, *The Astrophysical Journal*, 698, 895
- Kozłowski S., 2016, *ApJ*, 826, 118
- Kozłowski, Szymon 2017, *A&A*, 597, A128
- MacLeod C. L., et al., 2010, *The Astrophysical Journal*, 721, 1014
- MacLeod C. L., et al., 2011, *The Astrophysical Journal*, 728, 26
- Zhu C., Byrd R. H., Lu P., Nocedal J., 1997, *ACM Trans. Math. Softw.*, 23, 550
- Zu Y., Kochanek C. S., Peterson B. M., 2011, *The Astrophysical Journal*, 735, 80

This paper has been typeset from a \LaTeX file prepared by the author.

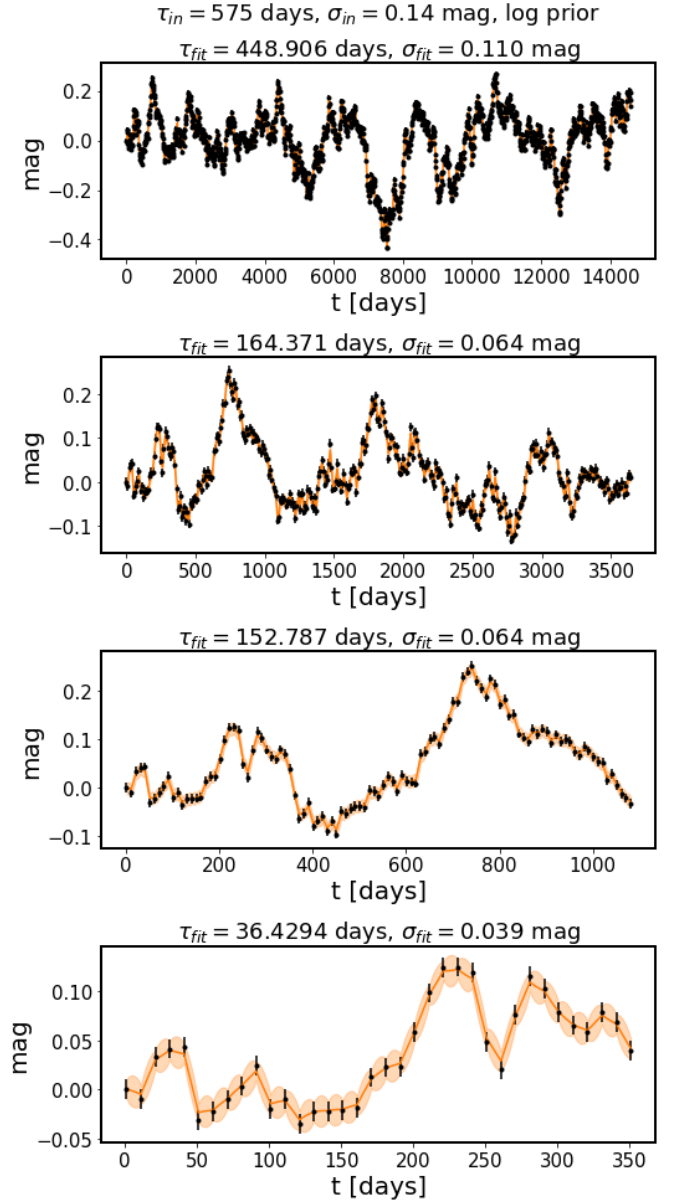


Figure 5. Sections of the 40-year light curve, fitted with the Jeffreys (log) prior. These sections are used to reproduce experiment from (MacLeod et al. 2011). From top to bottom: 40-year, 10-year, 3-year, 1-year sections. The input is $\tau = 575$ days, $SF_{\infty} = 0.2$, so that $\sigma = 0.14$, homoscedastic error of 0.001 mag.

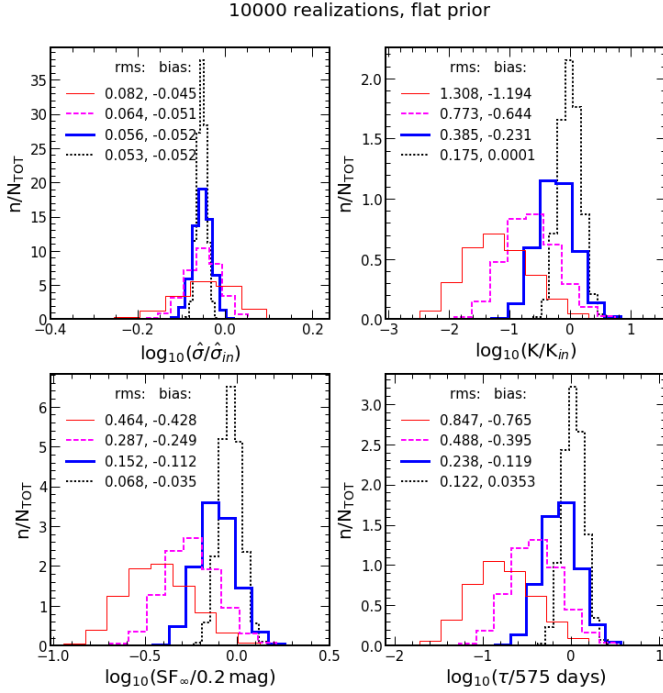


Figure 6. Distribution of results of 10000 iterations of DRW light curve, at each iteration fitting the full light curve, or its 1, 3, or 10 year sections, shown with dotted, thick solid, dashed, or thin solid lines, respectively. From top left panel, going clockwise, we display the ratio of each quantity derived from fitted τ , σ to the input values: $\hat{\sigma} = SF_{\infty}/\sqrt{\tau}$, $K = \tau\sqrt{SF_{\infty}}$, $SF_{\infty} = \sqrt{2}\sigma$, and τ . We display the rms and bias calculated for each distribution ($rms \equiv \sqrt{\langle x^2 \rangle}$, $bias \equiv \langle x \rangle$, with the latter being the distribution mean). Note that, especially as seen on upper right and bottom panels, the longer the section of the light curve that we use, the smaller the bias. It is surprising that even for a well-sampled 40-year light curve the bias in all four quantities is nonzero, but the overall conclusions: that the result of DRW fit asymptotically converge to true values only for light curves much longer than 10τ , are similar to those of (MacLeod et al. 2011).

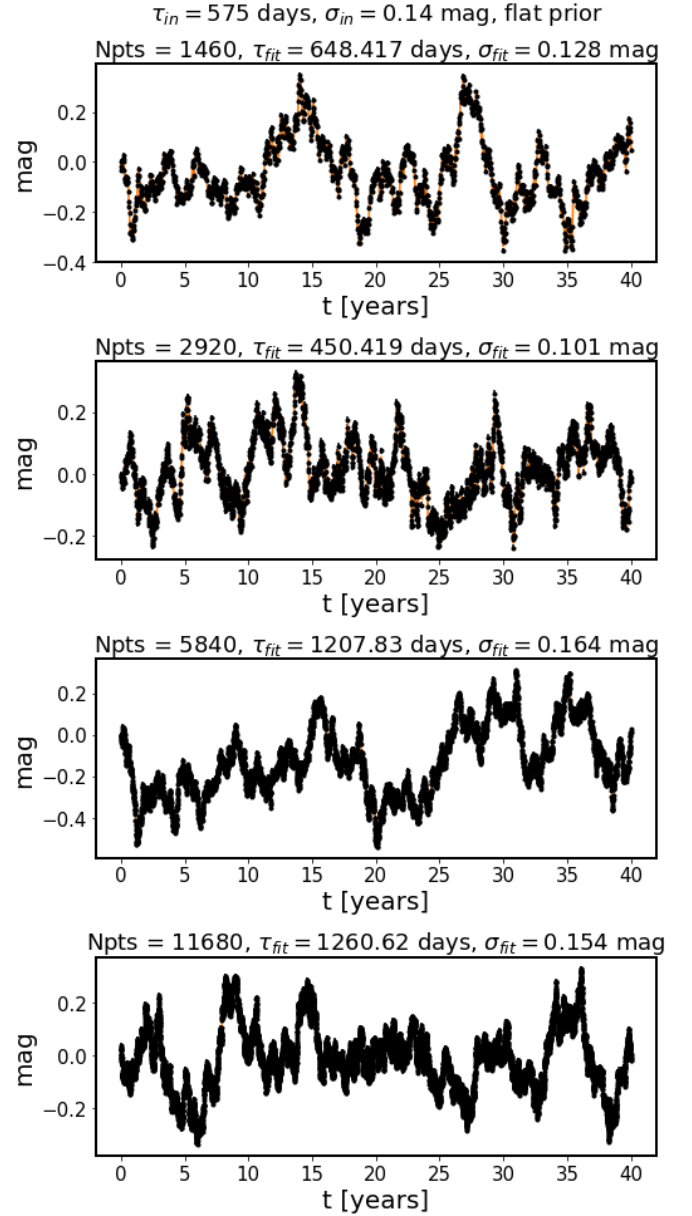


Figure 7. Experiment increasing the sampling density of the fixed baseline light curve. From top to bottom, we increase the initial number of points sampling the underlying process from 1460 to twice, four, and eight times more: second, third and fourth panels, respectively.

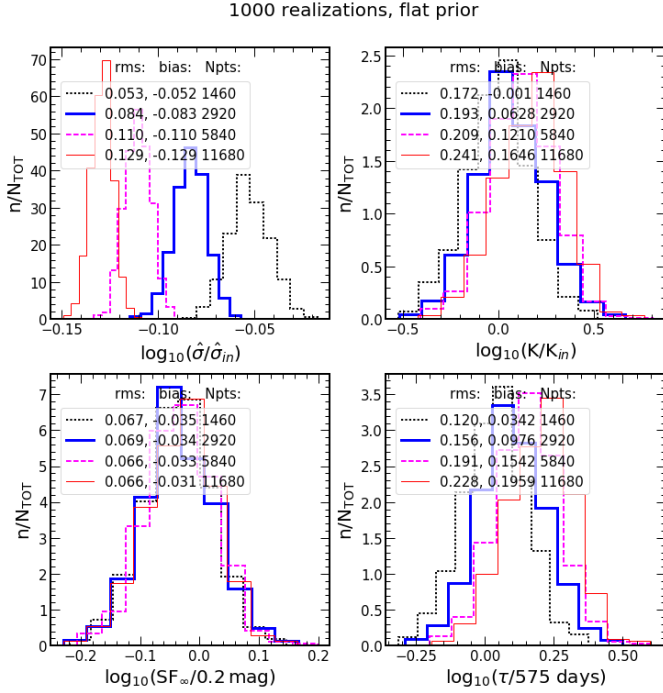


Figure 8. Distribution of Celerite fit results of 1000 iterations of DRW light curve simulations in controlled number of points experiment. At each iteration we make a realization of DRW light curve with input $\tau = 575$ days, $SF_{\infty} = 0.2$ mag, 40 -year baseline, sampled at regular intervals by $f \cdot 1460$ points, where $f \in 1, 2, 4, 8$.

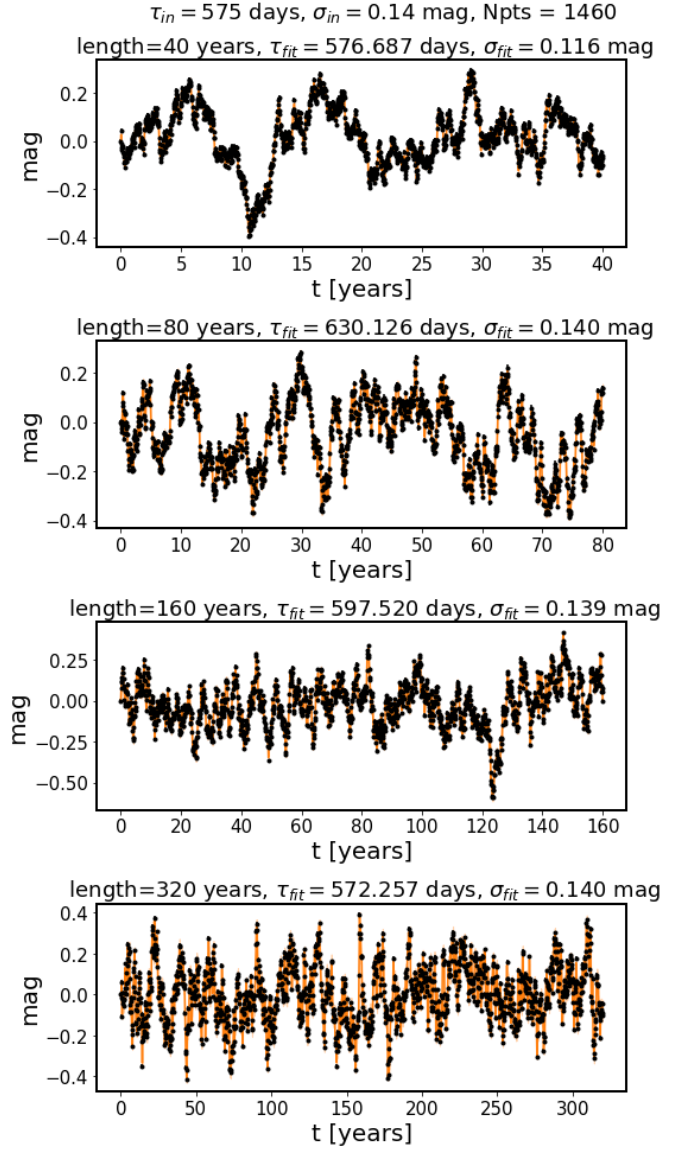


Figure 9. Light curves used in baseline experiment: we freeze parameters $\tau = 575$ days, $SF_{\infty} = 0.2$ mag, N (keep them fixed), and we extend the light curve baseline, starting from 40-years, and increasing it by a factor of $f \in 2, 4, 8$, from top to bottom. Note : at each iteration, longer light curves are not a mere shifted copy of the base 40-year length light curve, but new realizations of the DRW process with the same τ , σ , N , and different baseline.

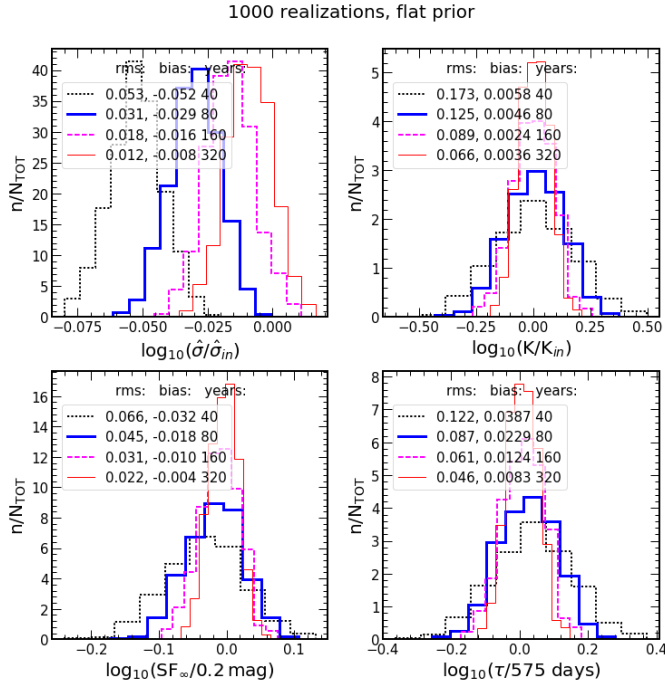


Figure 10. Distribution of Celerite fit results for 1000 iterations of DRW light curve simulations in controlled baseline experiment. We fix input parameters τ , σ , and N of points, but at each iteration we simulate four versions of the light curve : with 40-year baseline, and those longer by a factor $f \in 2, 4, 8$. We fit each baseline version with flat prior or Johnson prior. In each panel on top of histograms of ratios of best-fit results to input value, we display the *rms*, *bias*, and light curve length in years. The histograms are color-coded by light curve baseline : 40 year (black dotted line), 80 year (thick solid blue line), 160 year (dashed magenta line), 320 year (thin solid red line). As we would expect, the longer the baseline, the less is the bias. The rms also decreases with increasing baseline, because results of fit become more centered on a single value. It is worth noting that even for very long baseline (320 years), we do observe nonzero bias.

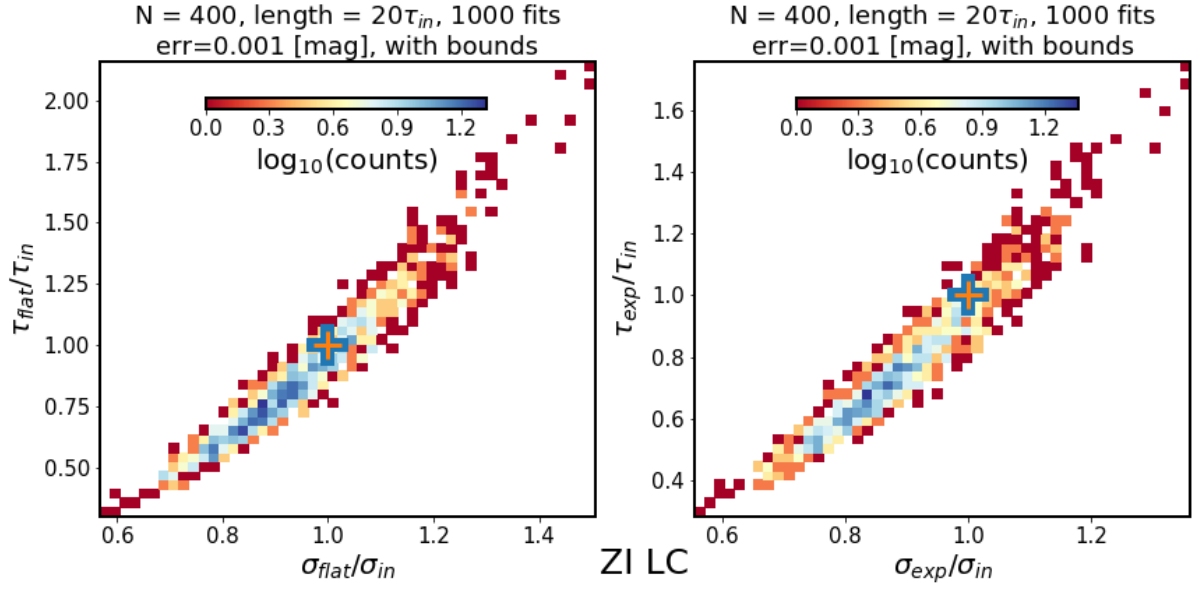


Figure 11. We plot results of fitting the simulated light curves with Celerite Real Term kernel, with flat (left), or Johnson prior (right). The cross shows the location of truth. Both are offset, and Fig. 12 shows the marginalized version of that plot. We repeat the same experiment, increasing the number of points tenfold to show the behavior of the bias.

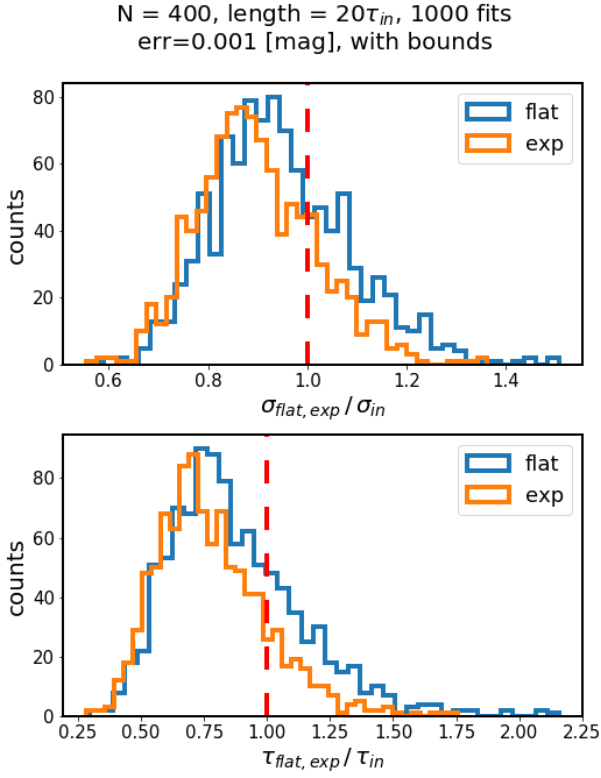


Figure 12. The marginalized version of Fig. 11. It shows that both with flat prior and Johnson prior, there is a bias in fitting well-sampled (400 points), long (20τ), with virtually no error ($\sigma_{phot} = 0.001$ mag). We investigate how this changes with increased number of points on Figs. 13 and 14

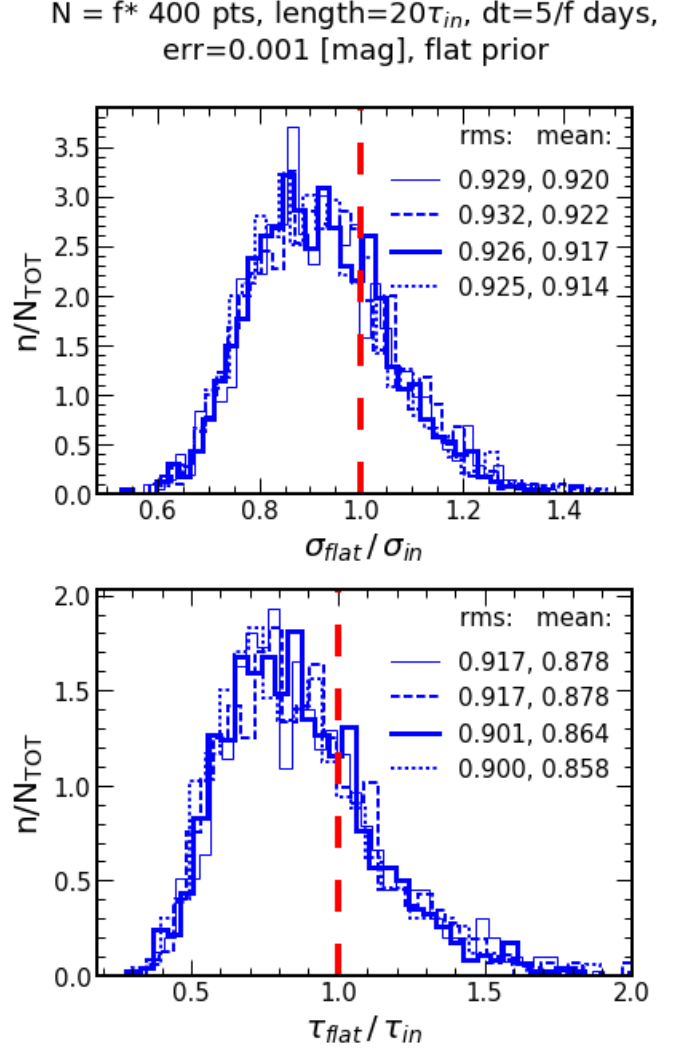


Figure 13. A controlled experiment to probe how the sampling density affects the fit results with a large number (1000) of light curves, with increasing number of points ($N = f \cdot 400$, where $f \in \{1, 2, 4, 8\}$), and thus increasing sampling density ($\Delta t = 5/f$, f as before), while keeping $\tau = 100$ days, $\sigma = 0.2$ mag, baseline of 20τ unchanged. As we expect, both bias and rms decrease as a function of sampling density, although even for very generous sampling of 12 hrs we still do not have a distribution that is centered on the true value for this baseline. On this plot we used flat prior.

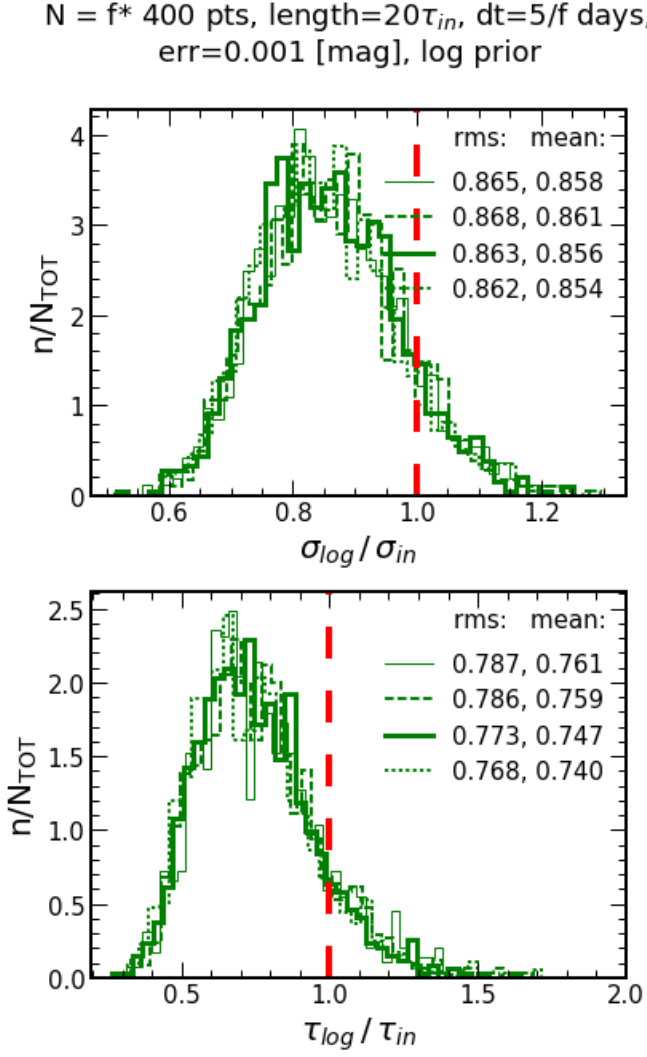


Figure 14. Same as Fig. 13, but with Johnson prior.

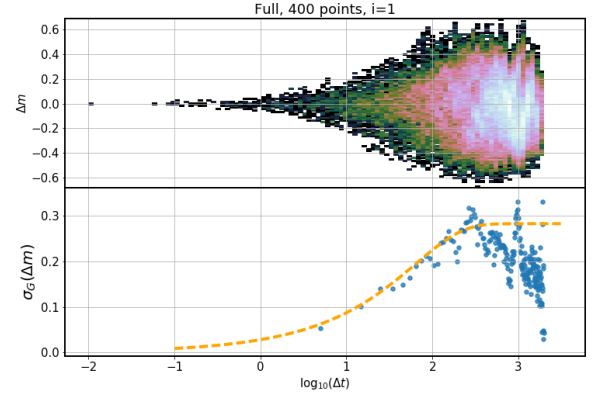


Figure 15. Raw time and magnitude pairwise differences on the top panel, and on the bottom panel with blue dots the robust standard deviation of Δm in 200 bins of Δt , with overplotted in orange dashed line the theoretical structure function (with $SF_\infty = 0.2\sqrt{2}$ mag, $\tau = 100$ days). Data for only one light curve.

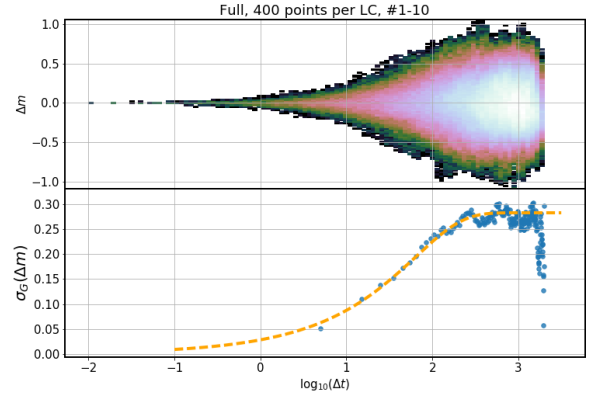


Figure 16. Same as Fig. 15, but combining data for 10 light curves.

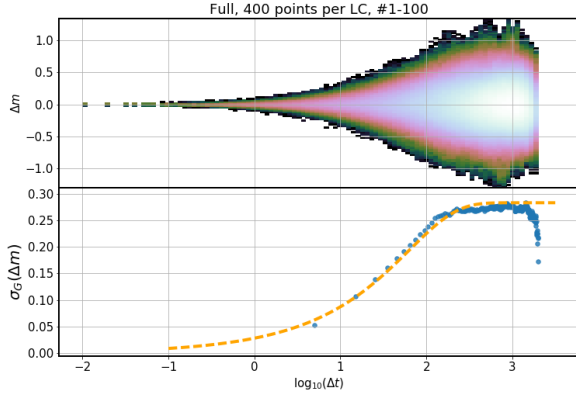


Figure 17. Same as Fig. 15, but combining data for 100 light curves.

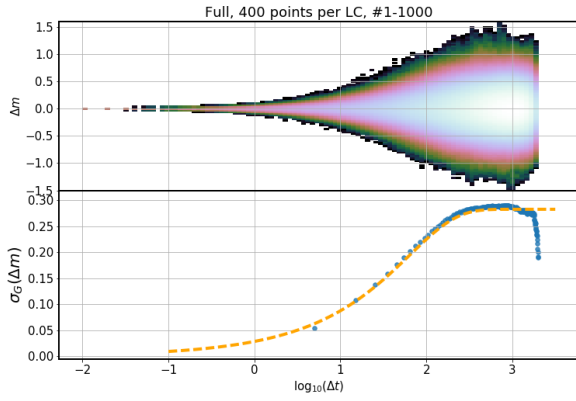


Figure 18. Same as Fig. 15, but combining data for 1000 light curves.

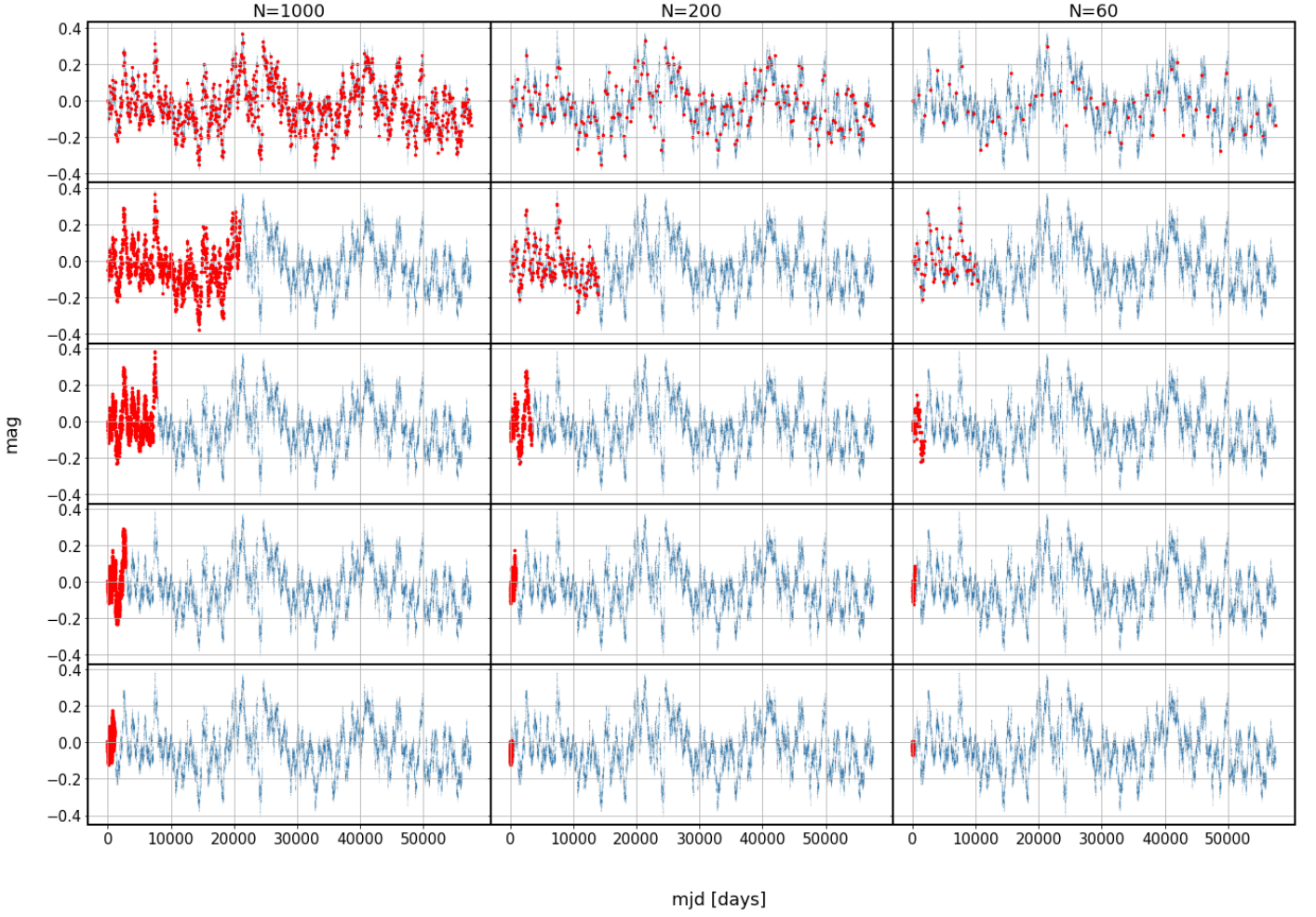


Figure 19. We illustrate the process of selecting different sections of a light curve, depending on the desired number of points N , and length of section : ρ . From left to right, we sample $N \in 1000, 200, 60$ points. Focusing on a single column, from top to bottom we sample on a logarithmic grid of ρ . The smallest ρ is set by the maximum attainable light curve section, which corresponds to the full length $l = 100\tau$, and since $\rho = 1/l$, $\rho_{min} = 0.01$. The largest ρ is related to the shortest possible light curve section conditional on the number of points chosen. Thus choosing $N \in 1000, 200, 60$ days, the shortest possible sections are $l \in 1000, 200, 60$ days given that we have adopted the $\Delta t = 1$ day in 'true' light curve.