MNRAS **000**, 1–3 (2017) Preprint 22 June 2017 Compiled using MNRAS LATEX style file v3.0

Using Célérité to infer DRW parameters

Krzysztof Suberlak,[1]⋆ Željko Ivezić [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 Célérité. We compare the tools used to fitting for τ and SF_∞ to those of Kozłowski, 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, Kozlowski+2010,2016, 2017 Macleod+2010,2011,2012, Zu+2011,2013,2016, Kasliwal+2015,2017). We employ Célérité (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 Célérité 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 Kozłowski, 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 Célérité. We then compare Célérité to another well-tested tool (used by Kozłowski (2016), Zu et al. (2011), etc.) - JAVELIN. We then reproduce results of MacLeod et al. (2011) Fig.15, and Kozłowski, 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 Kozlowski'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 :

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

and

$$stdev^2 = 0.5\,\mathrm{SF}_\infty^2 \left(1 - e^{-2r}\right) \quad (2)$$

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 Kozlowski+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 Kozłowski, 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 $\rho = \tau / T$ (Kozlowski+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

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

(see Kozlowski+2010 eq. 1, Kozlowski+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:

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

. 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 $\texttt{log\_a} = 2\log\sigma$, and $\texttt{log\_c} = -\log\tau$. 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 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 :

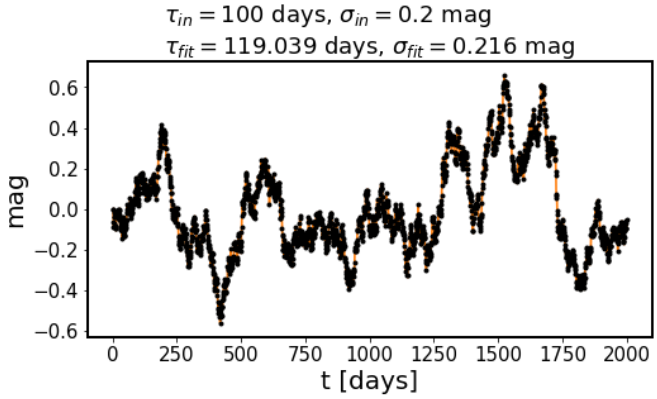

Figure 1. A Celerite fit to a simulated light curve using a flat prior.

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

so that in logarithmic space :

$$\log(prior) = \frac{1}{2}(\texttt{log\_c} - \texttt{log\_a} - \log 2) \quad (12)$$

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 parametes $\tau_{in} = 100$ 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

6 EXPERIMENTS WITH DAMPING TIMESCALE RETRIEVAL

7 CONCLUSIONS

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)

Jeffreys H., 1946, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 186, 453

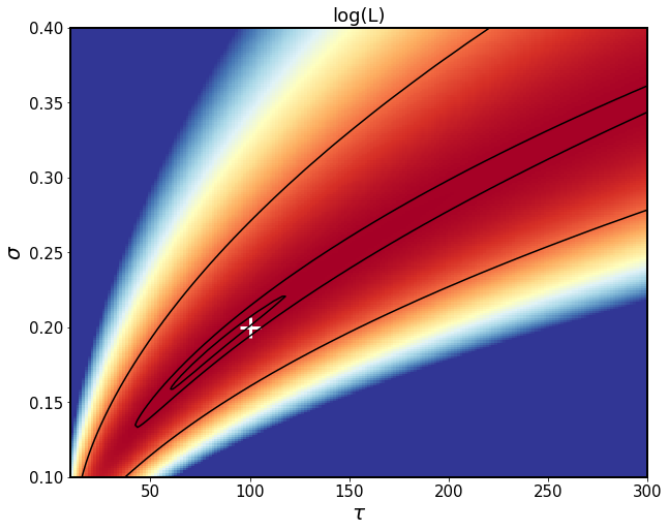

Figure 2. The log likelihood for the simulated light curve on Fig. 1. Black contours show 0.683, 0.955, 0.997 levels of the cumulative (integrated) posterior probability.

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 TEX/LATEX file prepared by the author.

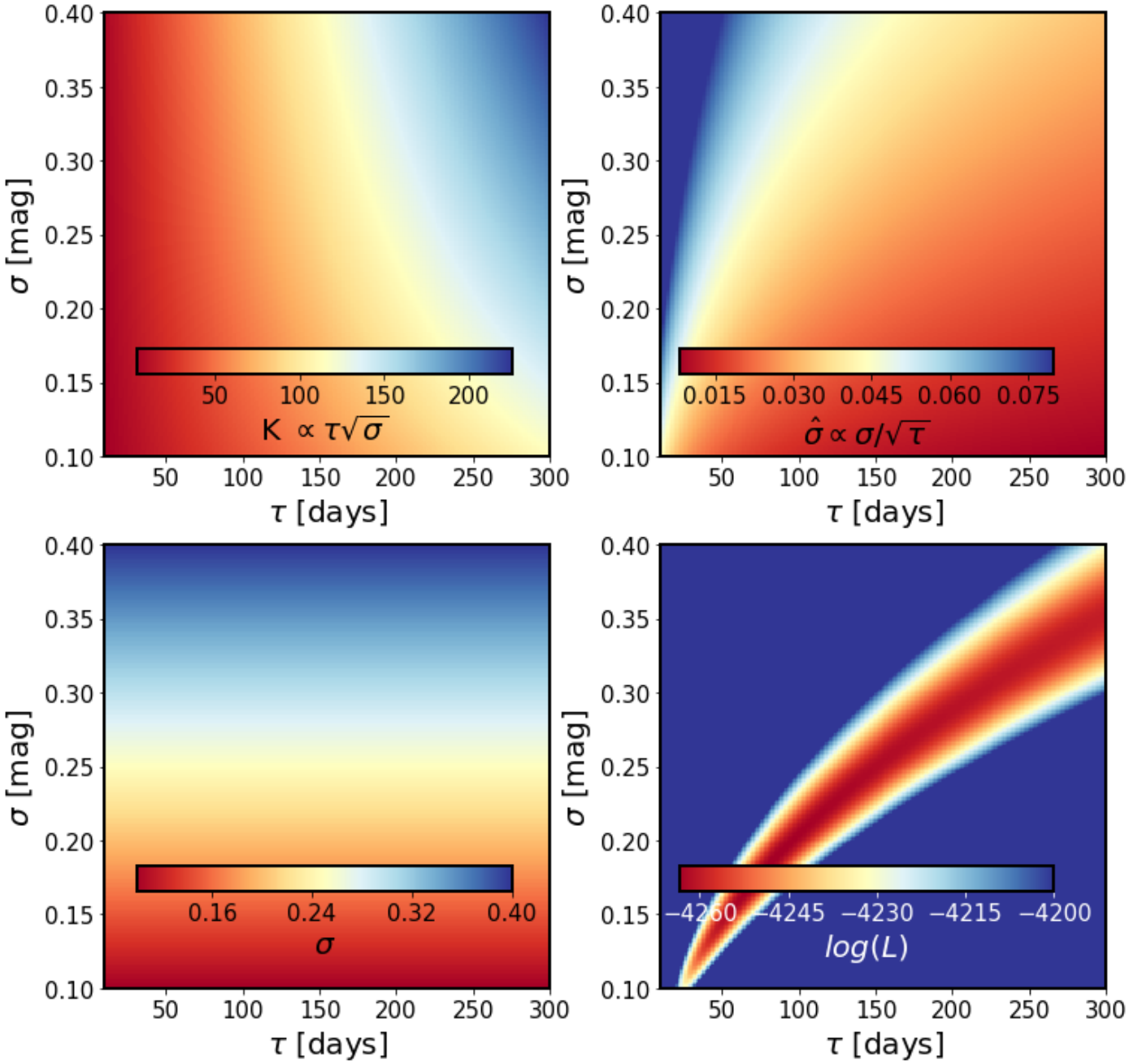

Figure 3. For each pixel on the σ - τ grid we evaluated the log-likelihood value, $\log L$, shown on the bottom-right panel (same as Fig. 2). In addition, given these σ and τ we also evaluated K and $\hat{\sigma}$, which enabled, given $\{\sigma, \tau, \hat{\sigma}, K, \log L\}$, plotting $\log L$ in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

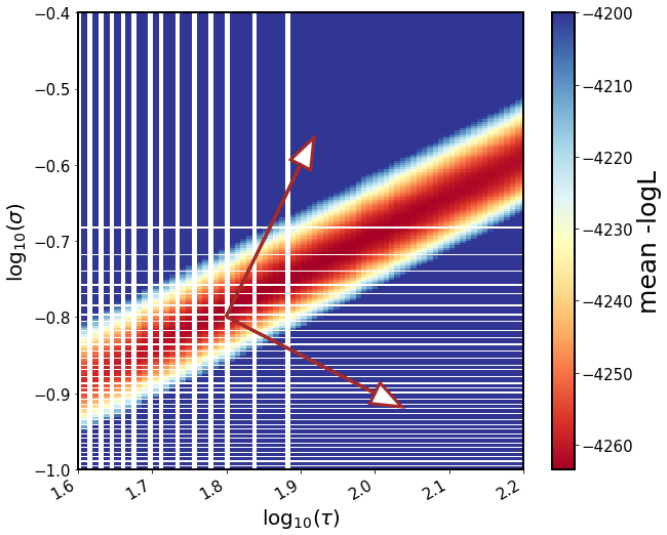

Figure 4. The log likelihood for the simulated light curve, plotted in $\log\sigma$-$\log\tau$ space. White gaps occur because originally the σ - τ grid on which we evaluated $\log L$ was linear, not logarithmic. Black contours show 0.683, 0.955, 0.997 levels of the cumulative (integrated) posterior 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.