# Master Thesis Notes: The Performance of Photometric Reverberation Mapping at High Redshift and the Reliability of Damped Random Walk Models

Shen et al. (2015a) construct 2 metrics in order to determine which combination of properties of their simulated light-curves yield the most accurate lag detections. They find that the ratio of the number of data points contributing to the calculation of the crosscorrelation function to the number of data points that contribute to resolving the true lag is typically ≈ 2 for detected lags. In the limit of Nepoch   1, this is equivalent to a requirement on the total observing run duration of 3 times the true observed lag, tspan/tlag = 3. We therefore imposed an additional criterion that the QSOs be observable for at least 3 times the length of their expected lag

Javelin supports a number of random walk covariance kernels which control the strength of the correlation between any two flux observations given the time between them. Zu et al. (2013) finds that the exponential covariance kernel is appropriate on time-scales, τ, between months and years, and we therefore adopt their recommendation for fitting with Javelin. Below a timescale of a few months, the correlation becomes stronger than can be accounted for by the exponential covariance kernel (Mushotzky et al. 2011; Zu et al. 2013) and the characteristics of stochastic behaviour at time-scales longer than a few years are not well known due to lack of data. There is further evidence that the DRW is not sufficient to explain high frequency light-curve variance as seen by Kepler and SDSS

Use a CARMA(2,1) model instead of a CAR(1) model?

The second order differential equation underlying the CARMA(2,1) process is familiar to many branches of physics and is therefore more easily interpretable than higher order processes. Indeed, the thermal motion with a fluid produces sound waves described by a PSD ∼ ν^2 (Mellen 1952), which suggests that CARMA(2,1) can be physically motivated by such distortions in the accretion disk.

| Parameter | Distribution | Source |
|---|---|---|
| Kepler Light curve | Choice[n=20] | Smith et al. (2018) |
| $\log(t_{lag})$ | $\mathscr{U}(0,300)$ | Set to cover |
| $\log\sigma$ | $\mathscr{N}(-2.2,1)$ | Prior from Target-10 |
| CARMA $\tau$ | $\sim P(\tau_{\text{Smith+18}})$ | Smith et al. (2018) |
| $w$ | $\mathscr{U}(0,13)$ | Set to cover |
| $s$ | $\mathscr{N}(1.70,1.21)$ | Measured from spectrum |
| $\alpha$ | $\mathscr{N}(1.20,0.53)$ | Measured from spectrum |
| $\sigma_z$ | $\mathscr{N}(0,0.3)$ | Set from zeropoint error |

**Table 2.** 50 000 draws were taken from these parameter distributions to create the simulated light curves for Target-10. Each draw created a different continuum lightcurve from the posterior distribution of CARMA(2, 1) fit to a randomly chosen Kepler light curve. The result was then propagated through a lagged smoothing window of width $w$ days, scaled by line scale $s$, and added onto the continuum at the position of the H$\alpha$ photometric filter $= \alpha * c(t)$ to create the narrow band light light curve.

We run Javelin with the default settings of a logarithmic prior which begins to penalize lag values larger than a third of the observational baseline (the time between the first observation and the last), and a hard limit on lags longer than the baseline itself. MCMC chains must have converged before any reliable parameter estimation can be performed. The model is run until convergence is achieved, whereby MCMC is halted when the autocorrelation time for all parameters changes less than 1 per cent and the number of iterations is larger than 50 times the largest autocorrelation time estimate

The mode within the highest-posterior-density (HPD) is found instead of using the median to account for the presence of multiple strong peaks in the lag posterior probability distribution. The HPD interval is the narrowest interval that is guaranteed to contain the mode of the distribution.

At all input lags and methods, we find artificial (i.e. incorrect) peaks at negative lags and so we can be justified in disregarding the peaks below -100 days. In particular, the Von Neumann estimator routinely places a large probability mass into a peak at -200 days. We find that there is always a large peak for all fitting methods at around 0-14 days, which coincides with the average cadence of observations (14 days). The KDE method allows us to assess the most likely peak without referring to the unstable maximum likelihood point, but it also implies a large uncertainty on the lag given that there are other regions of high probability which cannot be ruled out a priori. We can address the issue in four ways:

(i) Use the output lag distribution for our reliability simulations to mitigate the effect of non-linear artefacts that arise from the fitting process.

(ii) Apply a prior to the lag distribution based on previous lag and luminosity measurements, and established relations i.e. (Bentz et al. 2013).

(iii) Limit analysis to the range of lags bounded by the minima surrounding the tallest peak.

(iv) Combine estimations from each fitting method, thereby mitigating the biases which are not shared by both methods.

In conclusion, we find that if a Damped Random Walk (DRW) model is assumed by the fitting procedure when the light-curves are generated by a different Continuous Auto-Regressive Moving Average (CARMA) process, we can still recover accurate lags (despite a small loss in reliability). We find that by analysing the resulting probability distribution with more in-depth techniques, we can approach the precision demonstrated by spectroscopic reverberation mapping using photometric techniques.